



# Lexical and statistical evidence in the acquisition of second language phonemes

Rachel Hayes-Harb

## ► To cite this version:

Rachel Hayes-Harb. Lexical and statistical evidence in the acquisition of second language phonemes. Second Language Research, 2007, 23 (1), pp.65-94. 10.1177/0267658307071601 . hal-00570720

**HAL Id: hal-00570720**

**<https://hal.science/hal-00570720>**

Submitted on 1 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Lexical and statistical evidence in the acquisition of second language phonemes

Rachel Hayes-Harb *University of Utah*

Adults can often improve their perception of novel phoneme contrasts with exposure to a second language, but there is yet little understanding of how they accomplish this. The primary aim of this study was to compare two types of evidence that learners might use to learn the phonemes of second language: minimal pairs, or the lexicon, and statistical information. It was found that adults show evidence of perceptual learning based on statistical information alone, but that the availability of minimal pairs leads to more accurate perception of a novel contrast. In a follow-up experiment, it was found that adults can exhibit knowledge of a novel contrast without yet being able to use the contrast to distinguish words in a lexical task.

## I Introduction

Adults exhibit particular sensitivity to phonetic differences that correspond to the phonemic contrasts of their native language (Werker *et al.*, 1981; Werker and Tees, 1984; Best, 1994), but often exhibit difficulty distinguishing sounds that are not contrastive in their native language (Goto, 1971; MacKain *et al.*, 1981; Werker *et al.*, 1981; Best and Strange, 1992). This is taken as evidence that linguistic experience shapes sensitivity to phonetic distinctions; however, the issues of how linguistic experience affects this sensitivity and how it is applied to subsequent linguistic development – the development of phoneme inventories and contrastive lexical representations – are not yet well understood. The present research attempts to shed light on these issues

by comparing the effects of various types of linguistic experience on adult native English speakers' ability to discriminate two speech sounds that are not contrastive in English, and determining whether or not they can represent the sounds contrastively in their developing second language lexicon.

Many researchers have suggested that knowledge of word meanings can mediate a learner's acquisition of a language's phonemic contrasts (e.g. MacKain, 1982; Jusczyk, 1985; Werker and Pegg, 1992; Best, 1995; Lalonde and Werker, 1995).<sup>1</sup> For example, knowing that the words *goal* and *coal* have different meanings may provide the learner of English with information necessary to determine that their initial sounds, /g/ and /k/, are contrastive in English. However, recent research has provided evidence that listeners can learn to discriminate novel second language contrasts without any reference to word meaning, after simply listening to the target language during a training session (e.g. for adults, see Maye, 2000; for infants, see Maye *et al.*, 2002). These studies provide evidence that listeners are able to infer the phonemic inventory of a language from statistical patterns in the input speech, such that these patterns alone influence listeners' discrimination of phoneme contrasts even without the benefit of word meanings or minimal pairs.

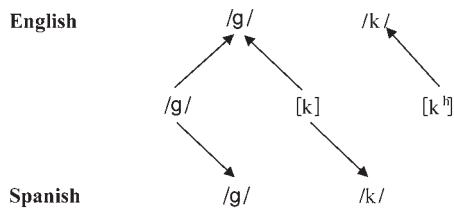
To date, no studies have directly compared the effects of statistical cues and the availability of minimal pairs on learners' ability to learn second language phonemes. The effects of these two types of evidence are compared directly in Experiment 1. Furthermore, although studies have shown that statistical patterns affect adults' discrimination of new phonetic distinctions, little is known about how changes in discrimination affect whether a learner treats a given phonetic distinction as lexically contrastive. Experiment 2 investigates whether these two types of information – statistical patterns and minimal pairs – contribute to the development of a target-like lexical-phonological structure, and addresses this issue by testing participants' ability to establish contrastive lexical representations for new second language words.

<sup>1</sup>However, in first language acquisition research, it has been shown that infants' ability to discriminate a contrast does not necessarily mean that they can use the contrast to discriminate words (Stager and Werker, 1997; Pater *et al.*, 2004; see also Swingley, 2003).

### 1 Native language and cross-language speech perception

Native language speech perception is characterized by sensitivity to phonetic differences that correspond to native phonemic contrasts and by a lack of sensitivity to non-contrastive phonetic variation. In English, for example, the phoneme /g/ is realized word-initially as some variant between voiceless unaspirated [k] and prevoiced [g], and native speakers of English perceive variants along the [k]–[g] continuum as the phoneme /g/ in word-initial position. They discriminate these variants of /g/ from variants of the phoneme /k/, which is produced with a relatively longer voice onset time or VOT ([k<sup>h</sup>]). In word-initial position, then, native English listeners ignore the phonetic distinction between [k] and [g] but not the distinction between [k<sup>h</sup>] and either [k] or [g].<sup>2</sup> On the other hand, listeners whose native language is Spanish break up the phonetic space differently, as the two phonemes are manifested differently in Spanish than in English. In Spanish, word-initial /g/ is realized as [g], while word-initial /k/ is realized as [k]. Figure 1 shows how Spanish and English assign the phones [g], [k] and [k<sup>h</sup>] to the phonemes /g/ and /k/.

It appears, however, that listeners do not always completely lose the ability to perceive non-contrastive phonetic variation. There is evidence that listeners are sensitive to sub-phonemic variation under certain task conditions (Pisoni and Tash, 1974; Werker and Logan, 1985; Polka, 1992; Werker, 1994; Pegg and Werker, 1997; Hayes, 2002; McMurray *et al.*, 2002). Pegg and Werker (1997) demonstrated that even though



**Figure 1** English and Spanish phonemes /g/ and /k/ and their phonetic variants in word-initial position

<sup>2</sup>The phonetic variants presented here are simplified for the purpose of illustration. In actuality, a range of voice onset time values for [g], [k] and [k<sup>h</sup>] are part of normal phonetic variation. A number of other phonetic characteristics also vary for these sounds. For the purpose of this illustration, the variation represented by the variants [g], [k] and [k<sup>h</sup>] will be discussed.

the difference between [d] and [t] does not correspond to a phonemic contrast word-initially in English, adult native speakers of English can discriminate the sounds with better than chance accuracy in certain tasks (though their discrimination of these sounds is less accurate than their discrimination of native phonemic contrasts). Thus the observed reduction in sensitivity to non-native contrasts in adults does not necessarily constitute an absolute loss of the ability to discriminate the sounds.

It should be noted that 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 many native speakers of Japanese with no English training discriminated English /l/ and /r/ with near-chance accuracy, and many studies of Japanese–English bilingual listeners' perception of English /l/ and /r/ have indicated that even experienced native Japanese learners of English can have difficulty discriminating these sounds (Goto, 1971; Miyawaki *et al.*, 1975; Mochizuki, 1981; Sheldon and Strange, 1982; Logan *et al.*, 1991; Yamada and Tohkura, 1992). A large literature addresses the issue of which non-native contrasts are most 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, including age of the learner and amount of second language input, are addressed in this literature on second language speech perception (e.g. Pisoni *et al.*, 1982; Werker and Tees, 1984; Werker and Logan, 1985; Polka, 1992; Best *et al.*, 2001; Flege, 1993; Flege *et al.*, 1995). 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 the adults' native language perceptual system, but also provides a window to the initial state for second language speech perception. In order to study

development of second language speech perception beyond this initial state, researchers often employ training paradigms, where listeners are trained on some aspect of a second language (an actual or artificial language) in a laboratory setting over the course of an experiment session (or 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.

## 2 Training studies

In speech perception training studies, participants are often presented pairs of auditory words that differ only in the contrast of interest; 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. In an example of this type of study, Strange and Dittman (1984) trained native speakers of Japanese to perceive the /l/-/r/ distinction in English, and the majority of participants showed improvement between pre- and post-training identification of the /l/ and /r/ sounds. This kind of feedback training, where participants are taught a second language contrast with evidence that the contrast distinguishes words in the target language, provides participants with lexicon-based evidence for a second language contrast, because they are taught that lexical items are distinguished by the phonetic differences (see Flege, 1989; 1995a; for more examples of this type of training, see Tajima *et al.*, 2002).

An alternative method for providing lexicon-based evidence for a phonemic contrast, one that does not involve explicit feedback, is to simply teach participants minimal pairs: sound strings differing only in the novel contrast that have different meanings (for examples of this type of training, see Curtin *et al.*, 1998; Pater, 2003). If a learner knows that the string [rak] has one meaning, while the string [lak] has a different meaning, he or she might infer the importance of the /l/-/r/ contrast even without feedback or having been explicitly told that the sounds distinguish the two words. In the experiments presented here, some participants were provided with this kind of evidence for a novel second language contrast.

In training studies that provide statistical evidence for a novel contrast, participants receive only acoustic information about the target language, and no information (explicit or implicit) about what distinguishes words in the language (for examples, see Maye, 2000; Maye *et al.*, 2002; Shea, 2005; Gilkerson, 2005). This kind of training exploits the statistical tendencies in speech with respect to phonemic distinctions: that separate phonemes have distinct distributions on some acoustic dimensions. Using statistical evidence, Maye (2000) trained native speakers of English either to discriminate or not to discriminate voiceless unaspirated stops from prevoiced stops (e.g. [k] vs. [g]). 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]; Pegg and Werker, 1997). Maye created a series of eight tokens on a continuum between prevoiced [g] and voiceless unaspirated [k] in three different vowel environments. The bimodal training group heard tokens near the endpoints of the continuum four times as often as the tokens in the centre of the continuum, while the unimodal training group heard tokens from the centre four times as often as the tokens at the endpoints. The distributions of tokens along the continuum reflect the intended phonemic status of the continuum, such that bimodal training provides listeners with two distinct clusters of tokens (two phonemes) along the continuum, while unimodal training provides them with only one cluster (a single phoneme). After a training session where participants listened to one of the two distributions of the various [g]–[k] tokens plus a 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. Participants trained with the bimodal distribution of tokens more accurately discriminated the endpoint tokens (tokens 1 and 8) than participants trained on the unimodal distribution of the same stimuli, while performance by both groups on all filler pairs was near ceiling. These results suggest that participants in the bimodal group were able to infer the phonemic status of the difference between [g] versus [k] based only on the statistical evidence they received during training, and that the availability of minimal pairs (or lexical evidence) is not a necessary condition for this type of phoneme learning.

Experiment 1 addresses the question of the relative impact of lexical and statistical evidence on the acquisition of novel phonemic contrasts

by providing participants with various kinds of lexical and statistical evidence about the presence or absence of a novel distinction in an artificial language. The training that participants received was manipulated such that some received statistical evidence in support of or against a novel contrast, some received lexical evidence in support of or against a novel contrast, and some received no training. Following the training session, participants were tested on their ability to discriminate the novel contrast in a sound discrimination test.

## **II Experiment 1**

This experiment examines how the availability of two different types of evidence during a training session – lexical and statistical evidence – affects adult monolingual English speakers' ability to discriminate a novel phonemic contrast in a subsequent perception task.

The first training groups received statistical evidence, and were exposed to the same training as 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 statistical evidence were presented stimuli in two different distributions along an acoustic continuum between [g] and [k]. The question addressed with these training groups is whether participants can learn to discriminate or not to discriminate a novel sound contrast based on this kind of statistical evidence alone.

Participants who received lexical evidence were taught the meanings of the auditory stimuli that they heard via the simultaneous presentation of pictures. 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.

### *1 Participants*

Participants were 132 native English speakers enrolled in undergraduate courses at the University of Arizona who received either course credit or financial compensation for their voluntary participation. All participants were monolingual English speakers whose parents and/or primary childhood caretakers were also monolingual English speakers,



and spoke English exclusively on a daily basis. Participants had between one and 13 years of foreign language instruction (mean = 4.3 years), but none considered themselves fluent in any language other than English. Participants were randomly assigned to one of six training conditions (described below), with a total of 22 participants per condition.

## *2 Training stimuli and methods*

All training stimuli were the same as those used in Maye (2000; Experiment 2). The target training stimuli were three continua between [g] and [k]: [ga] to [ka], [gæ] to [kæ] and [gɤ] to [kɤ]. First, a female native speaker of American English produced each of the syllables, and from each a series of four digitally-synthesized items was created to make a continuum of eight tokens, such that tokens 1–4 were created from an original [g]-syllable production, and tokens 5–8 were created from an original [sk]-syllable production. The acoustic dimensions that were manipulated were duration of prevoicing and formant onset frequencies (and the [s] was removed from the [sk]-syllables). The training stimuli ranged from 9 milliseconds of prevoicing (the most [g]-like) to 0 milliseconds of prevoicing (the most [k]-like).

The six training conditions in this experiment differed in 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 16 target training items for each of the three vowel environments, totalling 48 training stimuli per block. Participants also heard filler stimuli during training. Fillers began with /m/ and /l/ in the same vowel environments as the target training items: /ma/, la/, /mæ/, /læ/, /mɤ/ and /lɤ/. Four different tokens of each filler item were presented twice each (six filler words  $\times$  four tokens  $\times$  two presentations each) for a total of 48 fillers per block. The total for target plus filler training stimuli was 96 per block, and the entire training session consisted of four repetitions of the block, for a total of 384 auditory stimuli 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 giving a response.

*a Statistical training conditions:* Participants in the unimodal training condition heard target training stimuli in a unimodal distribution between [g] and [k]: tokens from the centre 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; see Figure 2 below). Participants in the bimodal training condition heard target 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 centre of the continuum. Participants in both of these training conditions heard the same set of target training stimuli; the only difference between the bimodal and unimodal training conditions is the number of times they heard each individual token. 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).

To summarize, participants in the bimodal and unimodal groups were trained by presenting auditory stimuli in particular distributions, without any indication of what the words meant; that is, without lexical evidence for the phonemic status of the sounds [g] and [k].

*b Lexical training conditions:* In the lexical 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. 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). Therefore, the acoustic distance between the peaks in the bimodal distribution and the tokens heard by participants in the lexical training conditions was identical (this was done in an attempt to make the various types of training maximally similar).

There were two kinds of lexical training: contrastive and non-contrastive. In the contrastive 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 pot while

token 7 was paired with a picture of a mouse. 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.

In the non-contrastive training condition, participants were presented pictures paired with auditory stimuli consistent with there not being a 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 pot. Despite the phonetic difference between the two tokens, they were both paired with the same meaning, suggesting to the listener that the initial consonant sounds in tokens 2 and 7 belong to the same phoneme. Table 1 provides example contrastive and non-contrastive training stimuli.

Filler training stimuli for the lexical training groups were also presented with accompanying pictures. All four tokens of each filler item were accompanied by the same picture, giving participants some phonetic variation during the training phase. Example fillers include four different tokens of [ma] paired with a picture of a bicycle and four different tokens of [læ] paired with a picture of a raccoon.

*c Additional training conditions:* The statistical training conditions were intended to provide statistical evidence without lexical evidence for a distinction (or lack thereof) between [g] and [k], while the lexical training conditions were intended to provide lexical evidence without statistical evidence for a distinction (or lack thereof). While it was possible to accomplish the former, by not including pictures to indicate what the words meant in the statistical training conditions, it was not entirely possible to accomplish the latter. In fact, the lexical training

**Table 1** Example training stimuli for the contrastive and non-contrastive training conditions

Participant heard...	Participant saw a picture of a...	
	Contrastive training	Non-contrastive training
[ga] (token 2)	pot	pot
[ka] (token 7)	mouse	pot

conditions did provide statistical evidence for a contrast between [g] and [k]. The distribution of tokens in the contrastive and non-contrastive training conditions was in fact bimodal, as participants heard only tokens 2 and 7. This leads to two potential problems: first, it appears that the contrastive and non-contrastive training groups receive both lexical and statistical information about the language during training. Thus performance by participants in both of these groups was potentially influenced by both types of information, making it difficult to interpret their discrimination test performance. The second potential problem is that participants in the non-contrastive 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. Because of this confound between lexical and statistical evidence in the contrastive and non-contrastive training conditions, it is not possible to know whether discrimination success by participants in either of these groups results from the distributional evidence or the lexical evidence they received during training. To address this problem, a fifth training condition was added. Participants in the two–seven training condition heard only tokens 2 and 7 (identical to the distribution heard by the contrastive and non-contrastive training conditions), but they were not provided with pictures to indicate word meanings. If participants in the contrastive training condition exhibit discrimination of [g] and [k] that is superior to participants' discrimination in the two–seven training condition, then the addition of lexical information is beneficial to learning to discriminate the new contrast. And if performance by participants in the non-contrastive training group is worse than that by participants in the two–seven group, then we also learn that the impact of lexical information overrules the effect of the distributional evidence that is inherent to the lexical training conditions.<sup>3</sup>

Finally, in order to fully interpret the effects of the different training conditions on participants' test performance, it was necessary to include

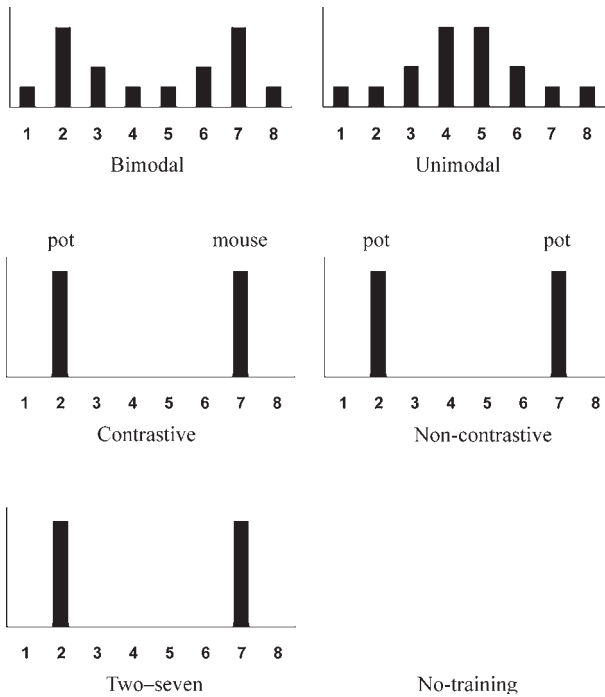
<sup>3</sup>Please see Hayes (2003) for a discussion of comparisons between performance by participants in the two–seven and bimodal training groups in light of the literature on the role of phonetic variability in novel phonetic category learning (see, e.g., Pisoni and Lively, 1995).

a group of participants who received no training at all. Maye (2000) also included a control group of untrained participants, and, although no direct comparisons between the untrained group and the monomodal and bimodal groups were reported, Maye (2001: 61) suggested that there is 'reason to hypothesize that the training phase had an effect on both monomodal and bimodal training groups'. On the one hand, training that provides evidence for a distinction between [g] and [k] may have the effect of enhancing participants' ability to discriminate the sounds relative to untrained participants (as typically reported in perceptual training studies). On the other hand, given that voiced–voiceless unaspirated consonant contrasts are discriminable by native speakers of English to some extent (Pegg and Werker, 1997), training that provides evidence against a distinction between [g] and [k] may have the effect of suppressing participants' pre-existing discrimination ability. A third possibility, that suggested by Maye (2000), is that the no-training participant group's performance will be between that of the groups who receive evidence for and evidence against a contrast. The six training conditions are summarized in Figure 2.

### *3 Discrimination test 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 and that their task was to indicate whether they thought the two stimuli in each pair were the same word repeated twice (by pressing a button labelled 'same') or two different words in the new language (by pressing a button labelled 'different'). They were told that they should use what they 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.

*a Target test pairs:* There were five types of pairs presented in the discrimination test. Tokens in all pairs were taken from the token sets presented during training. For each target training stimulus continuum ([gɑ] to [kɑ], [gæ] to [kæ], and [gɤ] to [kɤ]), tokens 1 and 8 were paired to create target–different pairs. In half of these pairs, token 1 was



**Figure 2** Summary of the six training conditions; token number is on the horizontal axis, frequency of presentation is on the vertical axis

presented first, and in the other half, token 8 was presented first (token 1–token 8; token 8–token 1). These are the critical test pairs, as ‘different’ responses to these pairs indicates an ability to discriminate [g] and [k].

For each continuum, two target–same pairs were also created (token 1–token 1; token 8–token 8). 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 target–different pairs can be interpreted. These two types make up the target pairs. Each of the six target–different and the six target–same pairs was presented twice, for a total of 24 target pairs in the discrimination test.

*b Filler test pairs:* There were also 24 filler test pairs, half of which required the response ‘same’ and half ‘different’. Filler pairs requiring

'same' responses were of two types. 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 [mæ] presented twice). The filler-same pairs consisted of non-identical tokens of the same word (e.g. two different tokens of [mæ]). The filler-different pairs consist of filler words paired with different filler words (e.g. [mæ]–[læ]; the vowels were always the same). Decisions about which tokens of each filler word to use in the test were made randomly. Each filler pair was presented once in the discrimination test (12 filler-different + six filler-identical + six filler-same = 24 filler pairs), plus the 24 target pairs, making a total of 48 pairs in the discrimination test.

*c Practice discrimination task:* At the very beginning of the experiment, before the training session, participants performed a practice discrimination task, 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 contained two 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 task was to get participants accustomed to the task and to pressing the 'same' and 'different' response buttons. The practice task was also intended to show participants that they should pay attention to phonemic differences between the words, and not to sub-phonemic phonetic differences in making their 'same' or 'different' decisions. One half of the 10 practice pairs were different words (e.g. *need* and *lead*) and one half were two different tokens of the same word (e.g. *need*<sub>1</sub> and *need*<sub>2</sub>). All practice test words were unmodified productions by a male native speaker of American English, and were presented once each. After the practice test, participants continued directly to the training phase.

All parts of the experiment took place in a sound-attenuated testing booth. Visual and auditory stimuli were presented using the DMDX experiment presentation software (Forster and Forster, 2003), and presentation of training stimuli and test items was randomized separately for each participant. In the discrimination test, items were presented with an inter-stimulus interval of 3000 milliseconds. All auditory

stimuli were presented over headphones, and visual stimuli were presented on a computer screen that participants viewed through a window in the testing booth. Participants responded to test items by pressing labelled buttons on a keyboard.

#### 4 Results

Participants' responses in the discrimination test were scored as either correct ('same' responses to target–same, filler–same and filler–identical pairs, and 'different' responses to target–different and filler–different pairs) or incorrect ('different' responses to target–same, filler–same, and filler–identical pairs, and 'same' responses to target–different and filler–different pairs). Overall, participants' performance on the control test pairs (target–same, filler–same and filler–different) was near ceiling, while discrimination accuracy for the target–different pairs was below chance (50%) for all training groups. The percent correct data is provided in Table 2.

An ANOVA with training condition (6 levels: bimodal, unimodal, contrastive, non-contrastive, two–seven, and no-training) as a between-subjects variable and test pair condition as a within-subjects variable (5 levels: target–same, target–different, filler–same, filler–different, filler–identical) was performed on the percent correct data. There was a significant effect of test pair condition ( $F(4,504) = 828.1$ ,  $p = .000$ ; partial eta squared = .868), a significant effect of training condition ( $F(5,126) = 2.882$ ,  $p = .017$ ; partial eta squared = .103), and a significant interaction of the two ( $F(20,504) = 4.321$ ,  $p = .000$ ; partial eta squared = .146).<sup>4</sup>

The item condition of primary interest is target–different; the other four serve as controls to determine whether the target–different results are interpretable. Recall that performance on every test pair condition other than target–different was expected to be highly accurate: native speakers of English were expected to respond 'different' to filler–different pairs (e.g. [mæ]–[læ]) and 'same' to filler–identical (e.g. [mæ]<sub>1</sub>–[mæ]<sub>1</sub>),

<sup>4</sup>The partial eta squared values are interpreted according to Cohen (1988): 0.01 represents a small effect, 0.06 a moderate effect, and 0.14 a large effect. In general, the statistical tests reported here were significant for moderate-to-large effects.



**Table 2** Discrimination test percent correct data by test pair and training condition (standard error)

Training condition	Target-different	Target-same	Filler-same	Filler-identical	Filler-different
Bimodal	22.7 (4.5)	97.0 (1.2)	94.7 (2.3)	97.7 (1.2)	98.9 (6.2)
Contrastive	43.6 (6.6)	96.2 (1.5)	89.4 (2.6)	97.7 (1.7)	98.9 (8.3)
Unimodal	11.0 (3.3)	97.7 (9.9)	93.2 (2.1)	97.7 (1.2)	98.1 (7.6)
Non-contrastive	18.2 (4.7)	98.1 (9.4)	96.2 (1.5)	95.5 (2.5)	98.1 (9.4)
No-training	32.6 (6.8)	96.6 (1.3)	93.2 (2.6)	97.0 (1.4)	97.3 (8.5)
Two-seven	17.4 (4.8)	98.9 (6.2)	96.2 (1.9)	96.2 (1.5)	100.0 (0.0)

filler-same (e.g. [mæ]<sub>1</sub>–[mæ]<sub>2</sub>) and target-same pairs (e.g. [gæ] token 1–[gæ] token 1). Planned comparisons revealed that the only significant effect of group was for the target-different test pairs ( $F(5, 126) = 5.128$ ,  $p = .000$ , partial eta squared = .169).

Following up on the effect of group for the target-different test pairs, planned comparisons indicated that the bimodal training group (22.7% correct) performed significantly better than the unimodal training group (11.0% correct;  $F(1,42) = 4.516$ ,  $p = .04$ ; partial eta squared = .097), replicating the finding first reported in Maye (2000) that statistical training of this type can affect discrimination of a novel phonemic contrast. The contrastive training group (43.6% correct) performed significantly better than the non-contrastive training group (18.2% correct;  $F(1,42) = 10.032$ ,  $p = .003$ ; partial eta squared = .193), providing evidence that the availability of lexical information can affect discrimination of a novel phonemic contrast.

Given that there is an effect of both lexical and statistical evidence on performance in this discrimination task, the next question asked is which type of evidence has a greater influence on learning the contrast. The contrastive training group performed significantly better than the bimodal training group ( $F(1,42) = 7.039$ ,  $p = .011$ ; partial eta squared = .144), providing preliminary evidence that lexical evidence has a greater influence than that of statistical evidence on learning to discriminate a novel phoneme contrast. However, recall that there was a confound in the contrastive and non-contrastive training conditions between lexical and statistical evidence, and that the two-seven training conditions was included to address this confound. The contrastive

training group performed significantly better than the two–seven training group ( $F(1,42) = 10.528, p = .002$ ; partial eta squared = .200); the addition of lexical information in the contrastive training condition appears to enhance discrimination performance by these participants over the two–seven 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 contrastive versus bimodal conditions, suggests an important influence of lexical information in learning novel phoneme contrasts.

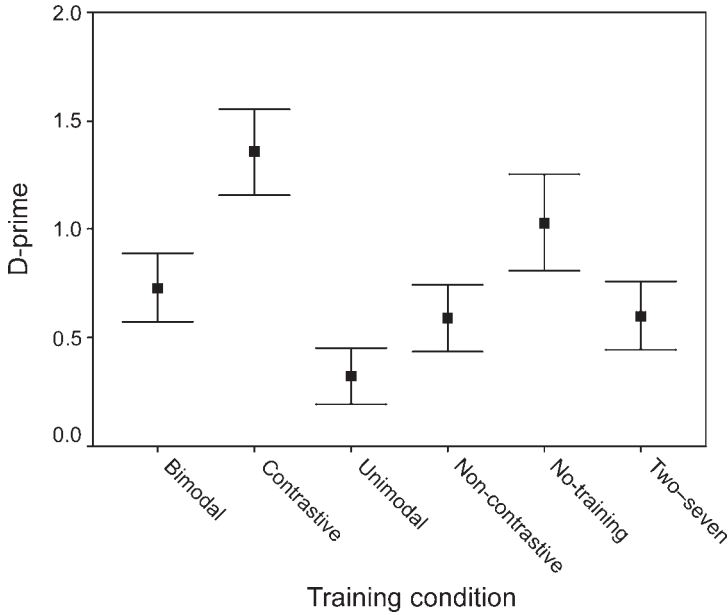
Next we consider the more general issue of the impact of training of any kind on discrimination performance, by comparing the bimodal, contrastive and two–seven conditions to the no-training condition. The purpose of these comparisons is to determine whether participants with training that supports a contrast between [g] and [k] will 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 (32.6% correct) and performance by participants in the bimodal training group ( $F(1,42) = 1.453, p = .235$ ; partial eta squared = .033), the contrastive training group ( $F(1,42) = 1.361, p = .250$ ; partial eta squared = .031), or the two–seven training group ( $F(1,42) = 3.28, p = .077$ ; partial eta squared = .072), the training groups that received evidence in support of a contrast between [g] and [k]. This null result suggests at least that the no-training participants were able to discriminate the sounds to some extent (which is not unexpected in light of previous studies; e.g. Pegg and Werker, 1997), and that the impact of training indicative of the phonemic status of the distinction did not serve to significantly enhance the discrimination ability that participants had prior to training. Another way of considering this issue is to determine whether the unimodal and non-contrastive 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 unimodal training group ( $F(1,42) = 8.106, p = .007$ ; partial eta

squared = .162), suggesting that statistical evidence against the contrast between [g] and [k] in fact had the effect of suppressing participants' previously existing ability to discriminate the sounds. The difference between discrimination by the no-training group and the non-contrastive training group, although tending in the same direction as the previous comparison, was not significant ( $F(1,42) = 2.989$ ,  $p = .091$ ; partial eta squared = .066). A liberal interpretation of these comparisons may be that participants started with a baseline ability to discriminate [g] and [k] that was suppressed by training of either kind against the distinction. However, the contrastive group's more accurate performance relative to the bimodal group suggests that there in fact was an effect of training for the distinction; it is possible that the contrastive group's performance was enhanced relative to the baseline performance of the no-training group, but that the comparison did not have enough power (given the small effect size) to be statistically significant.

Because the task was forced-choice between two options—'same' or 'different'—it is possible that participants were biased in general toward 'same' or 'different' responses, and that the different groups of participants had different biases (despite random assignment of subjects to groups). In order to fully interpret responses to the target–different items, it is necessary to separate these possible biases from detectability, or participants' ability to distinguish between target–same and target–different pairs. To accomplish this, a signal detection analysis was performed on each subject's scores separately, where 'different' responses to different items is the hit rate and 'different' responses to same items is the false positive rate. Separate d-prime scores were calculated for the target pairs (target–same and target–different pairs) and the filler pairs (filler–same and filler–different pairs) for each participant. An ANOVA with training condition as a between-subjects factor revealed that there was no effect of training condition on the d-prime scores for filler pairs ( $F(5,126) = 1.572$ ;  $p = .173$ ; partial eta squared = .059), but there was a significant effect for target pairs ( $F(5,126) = 4.526$ ,  $p = .001$ ; partial eta squared = .152). The d-prime scores for target pairs are presented in Figure 3.

Planned comparisons reveal a nearly-significant difference in detectability between the unimodal (.32) and bimodal groups (.73;  $F(1,42) = 3.964$ ,  $p = .053$ ; partial eta squared = .086). There are also



**Figure 3** D-prime scores for target pairs for all training conditions; bars represent  $\pm 1$  standard error

significant differences between the contrastive (1.36) and non-contrastive groups ( $.59$ ;  $F(1,42) = 9.331$ ,  $p = .004$ ; partial eta squared = .182), the contrastive and bimodal groups ( $F(1,42) = 6.016$ ,  $p = .018$ ; partial eta squared = .125), the two-seven (.60) and contrastive groups ( $F(1,42) = 8.882$ ,  $p = .005$ ; partial eta squared = .175), and the no-training (1.03) and unimodal groups ( $F(1,42) = 7.632$ ,  $p = .008$ ; partial eta squared = .154). Differences between the bimodal and no-training groups ( $F(1,42) = 1.211$ ,  $p = .277$ ; partial eta squared = .028), the contrastive and no-training groups ( $F(1,42) = 1.182$ ,  $p = .283$ ; partial eta squared = .027), the two-seven and no-training groups ( $F(1,42) = 2.504$ ,  $p = .121$ ; partial eta squared = .056), and the no-training and non-contrastive groups ( $F(1,42) = 2.672$ ,  $p = .110$ ; partial eta squared = .060) were not significant. Note that the results of the statistical tests using d-prime as the dependent variable mirror the results of the tests using percent correct on target-different pairs, as reported above, in all cases. Thus the training condition effects reported above remain even once possible biases are taken into consideration.

### III Experiment 2

Experiment 2 was performed in order to determine whether participants were able to establish lexical representations of new words in the new language that encoded the [g]–[k] distinction. In other words, Experiment 2 tests whether participants' demonstrated sensitivity to language-specific sound contrasts (in Experiment 1) can be extended to learning new second language words, that is, the experiment tests the linguistic relevance of the different types of training.

Discrimination data can provide information about participants' knowledge of a contrast, but does not directly indicate whether or not they can exploit this knowledge for subsequent linguistic development, e.g. the development of contrastive lexical representations. Following traditional generative phonology assumptions about the phonological content of lexical representations (Chomsky and Halle, 1968), lexical representations contain only contrastive information. We might therefore expect second language learners to initially neutralize novel contrasts in their lexical representations of second language words (but for a discussion of this issue within the framework of Optimality Theory, see Pater, 2003). In order to test whether participants who are trained on a novel phonetic difference can establish contrastive lexical representations of words containing the new sounds, a randomly-selected subset of participants from experiment one 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 and statistical) 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 discriminate a new contrast perceptually and learning to represent the contrast lexically.

If participants' ability to discriminate the novel contrast precedes their ability to represent the contrast lexically, this provides evidence for an intermediate stage between having a discrimination profile indicative of sensitivity to a phonemic contrast and being able to exploit the contrast lexically, as is often found in studies of first language acquisition (e.g. Stager and Werker, 1997; Pater *et al.*, 2004).

There are two parts to this experiment. In the first part, participants were taught words in the new language associated with pictures indicating the

words' meanings: this is the word learning phase. In the second part, participants were given a matching test to evaluate their lexical representations for the new words. Eight randomly-selected participants from each of the training groups in Experiment 1 also took part in Experiment 2, which immediately followed Experiment 1 during the same experiment session.

### *1 Word learning phase methods and stimuli*

During the word learning phase, participants were presented auditory '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' meanings in a matching test. It was expected that participants would establish representations for the words in memory that reflected their phonological knowledge of the new language. Thus it was predicted that if they had learned during experiment one that the [g] and [k] sounds were distinct, they might exploit that knowledge further in learning words in the new language that contain these sounds.

Target words were constructed using tokens from the [g]–[k] continua from Experiment 1, because it was important that the onsets and nuclei of the words that participants learned in this experiment had identical acoustic properties to the stimuli that they had heard earlier. Tokens 1 and 8 were used because they were the same tokens used in the discrimination test from Experiment 1. 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 original training stimuli, which were all single open syllables. For each of the three vowel continua, four words were created: two from token 1 and 2 from token 8, for a total of 12 target words. The filler words were created from one token of each of the filler words from Experiment 1, for a total of six filler words.

To create these words, the training stimuli were prepared for the addition of a coda by deleting the final 50 milliseconds of the vowel. Next, a female native speaker of English produced several tokens of the 'words' containing the desired codas, in the appropriate vowel environments. From these recordings, the final 50 milliseconds of the vowel, plus the coda, were copied from the whole word and pasted to the end

of the shortened training stimuli. Although the onset–vowel and vowel–coda sequences came from different utterances, all of the resultant word stimuli were judged to be natural-sounding by a native speaker of English. Table 3 provides an exhaustive list of target and filler words presented during the word learning phase.

Each word was presented with a picture to indicate its meaning. In order to prevent confounding word learning with lexical training, the set of words taught in Experiment 2 did not contain minimal pairs. 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 on the basis of lexical evidence, which might neutralize any effect of statistical versus lexical training from Experiment 1.

As mentioned earlier, it was assumed that participants in this study would store only contrastive information in their lexical representations of the words that they learned. Therefore, they were expected to be able to discriminate, for example, a [g]-initial word (e.g. [gant]) from a [k]-initial ‘non-word’ (e.g. [kant]) at test if they had knowledge that the sounds [g] and [k] were phonemically distinct and were able to exploit that knowledge lexically.

**Table 3** Target and filler words presented during the word learning phase

Training stimulus (from Experiment 1)	Coda added	Word created	Meaning (picture)
‘Word’ type: Target:			
[ga] (token 1)	[-nt]	[gant]	boot
	[-ŋ]	[gang]	apple
[ka] (token 8)	[-v]	[kav]	fork
	[-pt]	[kapt]	foot
[gæ] (token 1)	[-vs]	[gævz]	carrot
	[-t]	[gæt]	chair
[kæ] (token 8)	[-ŋ]	[kæŋ]	grapes
	[-sp]	[kæsp]	hand
[gə] (token 1)	[-st]	[gəst]	elephant
	[-k]	[gə:k]	duck
[kə] (token 8)	[-z]	[kəz]	helicopter
	[-m]	[kəm]	lamp
[ma]	[-mp]	[mamp]	motorcycle
[la]	[-n]	[lan]	onion
[mæ]	[-ŋk]	[mæŋk]	pig
[læ]	[-z]	[læz]	shirt
[mə]	[-fs]	[məfs]	spider
[lə]	[-p]	[ləp]	trumpet

## 2 Word learning test methods and stimuli

In the word learning test, 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 that of the practice test; the difference was in the incorrectly-matched test pairs. In this test, the target–unmatched test pairs were incorrect only in whether the word began with [g] or [k]. For this part of the experiment, a counterpart with the wrong initial consonant ([g] or [k]) was made for each of the experimental words that were used in training. These ‘non-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 matched word (filler–matched test pairs) and once with another filler word (filler–unmatched test pairs). In the final word learning test, participants saw each picture two times: once with the matched word and once with an unmatched ‘wrong’ word; again, their task was to decide whether the word they heard and the picture they saw were correctly matched.

## 3 Results

Accurate responses to target–unmatched items would indicate that the participant noticed that the word presented with the picture began with the wrong consonant (e.g. [g] instead of [k]). In order to respond accurately to these items, participants would need to have created representations for the words in memory that encoded the [g]–[k] distinction. Despite their highly accurate performance on the filler–matched, filler–unmatched, and target–matched items, average accuracy for each of the six training groups was below 20% on the target–unmatched pairs, and no individual participant in any training condition performed above chance (50%; see Table 4).

There were no training group differences or correlations between performance on this task and discrimination performance in experiment one. These results suggest that it is possible for adults to exhibit differential discrimination ability depending on the type of second language input they receive that does not translate immediately into differential encoding of the contrast in the lexicon, or, at least, that does not translate into a differential ability to demonstrate the lexical contrast in the



**Table 4** Word learning test percent correct data by test pair and training condition (standard error)

Training condition	Target–unmatched	Target–matched	Filler–unmatched	Filler–matched
Bimodal	14.6 (4.9)	85.4 (5.2)	93.8 (4.4)	87.5 (5.2)
Contrastive	15.6 (4.8)	88.5 (2.7)	95.8 (2.7)	95.8 (2.7)
Unimodal	16.7 (3.9)	90.6 (2.9)	89.6 (4.4)	97.9 (2.1)
Non-contrastive	7.3 (2.9)	89.6 (3.1)	97.9 (2.1)	97.9 (2.1)
No-training	5.2 (2.7)	88.5 (4.1)	95.8 (2.7)	95.8 (2.7)
Two–seven	7.3 (3.3)	95.8 (6.3)	95.8 (2.7)	97.9 (2.1)

on-line auditory word–picture matching task. It should be noted that the task demands associated with discrimination and with matching auditory words and meanings are different. It is possible that the word learning experiment was more difficult in general for participants than the discrimination experiment, as it required, for the target–unmatched test pairs, that participants have the contrast encoded in the lexicon; and that they perceive the contrast in the on-line auditory task in a way that allowed them to make use of the stored lexical contrast.<sup>5</sup>

#### IV General discussion

The results of Experiment 1 provide evidence that both lexical and statistical evidence can contribute to learning novel second language phoneme contrasts. However, it seems that training in support of a contrast did not significantly enhance participants' ability to perceive the contrast; instead, training that did not support the existence of a contrast significantly reduced participants' ability to discriminate the sounds. This finding is reminiscent of a typical pattern seen in infant speech perception over the course of the first year of life; for most contrasts tested, infants' discrimination is initially good, but sensitivity declines by the age of 12 months for those contrasts that are not utilized in the native language (for a review of this extensive literature, see Aslin *et al.*, 1998). That untrained participants in this study demonstrated an ability to discriminate the contrast to some extent is consistent with previous research on native English speakers' perception of stop consonant voicing (Pegg and Werker, 1997).

<sup>5</sup> Recent studies by Pater (2003) and Weber and Cutler (2004) have attempted to experimentally tease apart effects of lexical representations from effects of on-line task demands.

Despite the lack of statistical difference between discrimination by the bimodal/contrastive groups and the no-training group, it was found that participants in the contrastive group exhibited more accurate discrimination than the bimodal group. This suggests that the availability of word meanings and minimal pairs in the linguistic input leads to more accurate discrimination of a novel contrast than statistical evidence alone.

However, it was argued that a discrimination task does not necessarily indicate that learners have acquired a novel contrast in a linguistically useful way. Experiment 2 was designed to determine whether discrimination ability demonstrated in Experiment 1 had lexical consequences; that is, whether participants who were able to discriminate the contrast in a perception task were further able to encode the contrast lexically. The word learning test revealed that despite near-ceiling memory for filler words, no participants, regardless of training condition, showed evidence of establishing lexical representations for new words that encoded the [g]–[k] contrast. The duration of training may not have been sufficient for participants to generalize the newly-learned contrast to the lexicon, providing evidence for an intermediate stage in the acquisition of second language phonology where learners have perceptual sensitivity to a novel contrast that they may not yet be able to represent lexically. As discussed earlier, a parallel finding is reported in the first language acquisition literature, where infants demonstrate an ability to discriminate sounds without being able to contrastively match the sounds to meanings in lexical tasks. This finding is also consistent with the common observation among language teachers that even when adult learners exhibit target-like pronunciation in controlled classroom tasks, they often revert to accented speech in less formal settings. It is possible that controlled classroom tasks make use of short-term memory of target-language sounds, but it may be that in less formal settings, learners rely on non-target-like lexical representations. In other words, learners may have developed the ability to perceive and produce novel sound contrasts under certain circumstances, but be yet unable to represent the contrasts lexically and/or to implement the lexically stored contrasts.

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 speech alone (statistical evidence) and from knowledge of words and meanings in the language (lexical 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. However, although participants who were exposed to minimal pairs during the training phase discriminated the novel sound contrast with greater accuracy than those who received only statistical evidence, this research additionally provides evidence supporting learners' ability to learn to discriminate a novel sound contrast in the absence of a lexicon, on the basis of the statistical properties of the auditory second language input alone.

### *Acknowledgements*

This research was supported by a grant from the National Science Foundation (Dissertation Improvement Grant Bcs-0236641). The author thanks Mike Hammond, Natasha Warner and Janet Nicol for their contributions to this project, Jessica Maye for her comments and for generously sharing her experimental stimuli, Anne Cook for statistics advice, Mara Haslam for comments on the manuscript, and the anonymous *Second Language Research* reviewers for their helpful suggestions.

### **V References**

- Aslin, R.N., Jusczyk, P.W. and Pisoni, D.B.** 1998: Speech and auditory processing during infancy: constraints on and precursors to language. In Kuhn, D. and Siegler, R., editors, *Handbook of child psychology volume 2: cognition, perception and language*. Wiley, 147–98.
- Best, C.T.** 1994: The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Goodman, J.C. and Nusbaum, H.C., editors, *The development of speech perception: the transition from speech sounds to spoken words*. MIT Press, 167–224.

- 1995: Learning to perceive the sound pattern of English. In Rovee-Collier, C. and Lipsitt, L., editors, *Advances in infancy research*. Volume 9. Ablex, 217–304.
- Best, C.T. and Strange, W.** 1992: Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–30.
- Best, C.T., McRoberts, G.W. and Goodell, E.** 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, 775–94.
- 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, 84–91.
- Chomsky, N. and Halle, M.** 1968: *The Sound Pattern of English*. Harper and Row.
- Cohen, J.** 1988: *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Curtin, S., Goad, H. and Pater, J.V.** 1998: Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14, 389–405.
- 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, 1684–97.
- 1993: Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America* 93, 1589–608.
- 1995a: Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics* 16, 425–42.
- 1995b: Second language speech learning: theory, findings, and problems. In Strange, W., editor, *Speech perception and linguistic experience: issues in cross-language research*. New York Press, 233–77.
- Flege, J.E., Munro, M.J. and MacKay, I.R.A.** 1995: Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16, 1–26.
- Forster, K.I. and Forster, J.C.** 2003: DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods: Instruments and Computers* 35, 116–24.
- Gilkerson, J.** 2005: Categorical perception of natural and unnatural categories: evidence for innate category boundaries. In Okabe, R. and Nielsen, K., editors, *UCLA working papers in linguistics, volume 2: papers in psycholinguistics*, 34–58.
- Goto, H.** 1971: Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia* 9, 317–23.
- Hayes, R.L.** 2002: The perception of Japanese consonant length by non-native listeners. *Laboratory Phonology* 8, New Haven, 2002. Poster.

- 2003: How are second language phoneme contrasts learned? Unpublished doctoral dissertation, University of Arizona.
- Iverson, P., Kuhl, P.K. and Akahane-Yamada, R.** 2003: A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.
- Jusczyk, P.W.** 1985: On characterizing the development of speech perception. In Mehler, J. and Fox, R., editors, *Neonate cognition: beyond the blooming, buzzing confusion*. Lawrence Erlbaum, 199–229.
- Kuhl, P.K.** 1991: Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50, 93–107.
- 1993: Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics* 21, 1–2, 125–39.
- Lalonde, C.E. and Werker, J.F.** 1995: Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development* 18, 459–75.
- Logan, J.S., Lively, S.E. and Pisoni, D.B.** 1991: Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89, 874–86.
- MacKain, K.S.** 1982: Assessing the role of experience on infants’ speech discrimination. *Journal of Child Language* 9, 527–42.
- MacKain, K.S., Best, C.T. and Strange, W.** 1981: Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 2, 369–90.
- Maye, J.** 2000: The acquisition of speech sound categories on the basis of distributional information. Unpublished doctoral dissertation, University of Arizona.
- Maye, J., Werker, J.F. and Gerken, L.A.** 2002: Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B10–B111.
- McMurray, B., Tanenhaus, M.K. and Aslin, R.N.** 2002: Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86, B33–B42.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J. and Fujimura, O.** 1975: An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics* 18, 331–40.
- Mochizuki, M.** 1981: The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics* 9, 283–303.
- Pater, J.** 2003: The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research* 19, 209–23.

- Pater, J., Stager, C. and Werker, J.** 2004: The perceptual acquisition of phonological contrasts. *Language* 80, 384–402.
- Pegg, J.E. and Werker, J.F.** 1997: Adult and infant perception of two English phones. *Journal of the Acoustical Society of America* 102, 3742–53.
- Pisoni, D.B. and Lively, S.** 1995: Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In Strange, W., editor, *Speech perception and linguistic experience: issues in cross-language research*. York Press, 433–59.
- Pisoni, D.B. and Tash, J.** 1974: Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15, 285–90.
- Pisoni, D.B., Aslin, R.N., Perey, A.J. and Hennesy, B.L.** 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, 297–314.
- Polka, L.** 1992: Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52, 37–52.
- Shea, C.** 2005: Learning allophones from the input. *Acoustical society of America annual meeting*, Vancouver, 2005. Poster.
- Sheldon, A. and Strange, W.** 1982: The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243–61.
- Stager, C.L. and Werker, J.F.** 1997: Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 6640, 381–382.
- Strange, W. and Dittman, S.** 1984: Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics* 36, 131–45.
- Swingle, D.** 2003: Phonetic detail in the developing lexicon. *Language and Speech* 46, 265–94.
- Tajima, K., Rothwell, A. and Munhall, K.G.** 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, 2387.
- Weber, A. and Cutler, A.** 2004: Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50, 1–25.
- Werker, J.F.** 1994: Cross-language speech perception: developmental change does not involve loss. In Goodman, J.C. and Nusbaum, H.C., editors, *Development of speech perception: the transition from speech sounds to spoken words*. MIT Press, 93–120.
- Werker, J.F. and Logan, J.S.** 1985: Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37, 35–44.
- Werker, J.F. and Pegg, J.E.** 1992: Infant speech perception and phonological acquisition. In Ferguson, C.A., Menn, L. and Stoel-Gammon, C., editors,

*Phonological development: models, research, implications*. York Press, 285–311.

**Werker, J.F. and Tees, R.C.** 1984: Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75, 1866–78.

**Werker, J.F., Gilbert, J.H., Humphrey, K. and Tees, R.C.** 1981: Developmental aspects of cross-language speech perception. *Child Development* 52, 349–55.

**Yamada, R.A. and Tohkura, Y.** 1992: The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception and Psychophysics* 52, 376–92.