SHALLOW VS. DEEP: BILINGUAL CONTRAST PROCESSING

by

KATHERINE AYA YOSHIDA

B. Sc., McGill University, 2002

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Department of Psychology, Cognitive Science Programme)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2004

THE UNIVERSITY OF BRITISH COLUMBIA FACULTY OF GRADUATE STUDIES

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Katherine Yoshida
Name of Author *(please print)*

06/08/2004
Date (dd/mm/yyyy)

Title of Thesis: Shallow vs. Deep: Bilingual Contrast Processing

Degree: M.A.

Year: 2004

Department of Psychology
The University of British Columbia
Vancouver, BC Canada

Abstract

Languages differ in the range of phones that cluster together to carry the same meaning, forming phonemic categories (e.g. all the possible manifestations of /l/ in English). These phonemes then contrast with one another to denote differences of meaning (e.g. /l/ vs. /r/, a valid contrast in English but not in Japanese). Infants are born with the ability to discriminate contrasts belonging to their native language, as well as those that are non-native contrasts. However, by 10-12 months of age they refine their repertoire, leaving only those utilized in the language input (Werker & Tees, 1984). The subsequent formation of non-native phonemes belonging to an ensuing second language (L2) is less clearly understood. This work uses Cantonese-native ESL speakers to investigate the distinction between "shallow" knowledge of non-native (English-specific) isolated phonemes and "deep" knowledge of the phonemes embedded within a linguistic context (Werker et al, 2001). Deep knowledge of the non-native contrasts is tested using a lexical decision task (Pallier, Colomé & Sebastián-Gallés, 2001) and compared to shallow knowledge, operationalized as ability to perceive differences between pairs of isolated phones. Results indicated that the bilinguals had difficulty perceiving the English-specific contrasts within the deep context, but not within the shallow context. Thus, support is found for a shallow-deep distinction within L2 speakers of a language. This helps explain previous divergent findings of L2 perception, and may potentially have implications for hypotheses of interference and a critical period.

Table of Contents

List of Tables

List of Figures

Dedication

Firstly and mostly I have to thank Janet Werker, for her extreme wonderfulness and brilliance.

Thanks to Ramesh Swamy for programming it all. Ralph Hakstian for all of his statistical help, along with Darcy Hallett and Mike Papsdorf. Geoff Hall and Todd Handy for their insightful comments. Christiane Dietrich for introducing me to and answering all of my endless questions about Praat. Carrie Cuttler and Jess Gao for their help in recruiting subjects. Jen Corbett for help getting this document ready. Amanda LaMarre for providing the vocal stimuli. Shahab Ghorashi for the graphs. Laurie Fais for her editing and comments, and linguistical advice. Chris Fennell, Marie Jetté, Nenagh Kemp, Erin Moon, Laura Sabourin, Athena Vouloumanos and Whitney Weikum for being such great labmates. Marisa Cruickshank for coming back. My parents for all of their love and support. My brother, Greg, for cleaning my bathroom. And, finally, thanks to Frédérique Delaprée, Shebob Ghorashi, Anjula Joshi, Eli Puterman and Misty Russell for keeping me alive in the department. You are all amazing.

Introduction

The most obviously distinguishing characteristics of languages are their vocabularies and syntactic systems, yet even more fundamentally there also exist important cross-linguistic differences in phonological systems. These govern at the level of auditory sound, depicting a number of linguistic aspects, including the range of phones[1] that cluster together in specific languages to form categories called phonemes (e.g., all the possible manifestations of /r/ in English). The language-specific phonemes then contrast with one another to denote differences of meaning. For example, the "liquid" phonemes /r/ and /l/ unmistakably contrast in English, as substituting one for the other would change the meaning of a word, for example, from "rake" to "lake". However, there exists only one liquid in Japanese, preventing any meaningful /r/ - /l/ contrast, and a substitution of one phone for the other here would not result in any change of concept. The ability to differentially discriminate those contrasts denoting meaning from those that represent irrelevant phonetic variation within a phoneme is imperative for accurate comprehension of a language.

The question under investigation here is broadly concerned with the manner in which adult second language (L2) speakers discriminate contrasts specific to that L2 that do not denote meaning in their native language (L1). For example, how does a native speaker of Japanese who has learned English as an L2 perceive /r/ - /l/? This generally addresses the impact of early linguistic experience upon later perceptual abilities. Specifically, the hypothesis presently put forth is that L2 speakers are more easily able to discriminate L2-specific contrasts when they are embedded within meaningless syllables than within full meaningful words. If such a difference in L2 perceptual ability did in fact exist, it would mirror that found in infants learning their L1.

---

[1] A phone is taken here as a linguistic segment.

Infants: First Language Acquisition

It is well known that native speakers of a language often experience difficulty in discriminating contrasts that are not meaningful in their language (non-native). In one demonstration of this, native speakers of English and Hindi were tested on two contrasts specific to Hindi (a place-of-articulation retroflex versus dental voiceless stop /Ta/ - /ta/ contrast and a breathy voiced vs. unvoiced unaspirated dental stop /d^ha/ - /t^ha/ contrast; Werker, Gilbert, Humphrey & Tees, 1981). All of the Hindi-speaking adults were able to discriminate the contrasts, and 6-8 month-old English-learning infants were also successful. However, the English speaking adults displayed significant difficulty with these non-native contrasts (also see Goto, 1971; Lisker & Abramson, 1970; MacKain, Best & Strange, 1981; Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975; Sheldon & Strange, 1982; Singh & Black, 1966; Snow & Hoefnagel-Hohle, 1978; Trehub, 1976). It appears that from birth infants discriminate both native and non-native contrasts (e.g., Aslin, Pisoni, Hennessy & Perey, 1981; Lasky, Syrdal-Lasky & Klein, 1975; Streeter, 1976; Trehub, 1976). However, by the end of the first year of life this broad ability is narrowed as the perceptual repertoire is refined to leave only native contrasts, consistent with the language input.

Werker and Tees (1984a) expanded the sample of 6-8 month-olds tested in Werker et al. (1981) on the Hindi-specific contrasts by adding 8-12 month old infants and an additional non-native contrast, a glottalized voiceless place-of-articulation uvular versus velar Thompson distinction (Interior Salish spoken in south central British Columbia), denoted /k'/-/q'/. Thus their final sample of 6-12 month-olds was tested on two contrasts. Infants of 6-8 and 8-10 months of age were able to discriminate both non-native contrasts. However, infants of 10-12 months of age performed more poorly, displaying decreased detection of the non-native changes. Thus, by one year of age, infants appear able to disregard nonmeaningful phonetic variation, indicating that they have "tuned into" their L1. These results suggest a perceptual reorganization that would ostensibly help the young learner acquire language, and have been replicated using a number of different contrasts (e.g., Tsao, Liu, Kuhl & Tseng, 2000; Tsushima et al., 1996).

Werker and Lalonde (1988) exposed adults to artificially created tokens spaced along a /ba/-/da/ continuum that the native speakers of English grouped into the two categories. The native Hindi speakers, however, grouped the tokens into three categories, assigning tokens from the single English /da/ phoneme into the Hindi retroflex versus dental /Da/ and /da/ categories. Infants of 6-8-months discriminated all three categories, regardless of their language input, but at 11-13-months only those learning Hindi continued to distinguish between /Da/ and /da/. Everyday exposure to the lone intermediate English /da/ category appears to lead English infants to collapse the previously discriminated Hindi distinction into a single phoneme.

Mechanism of perceptual reorganization.

The necessity of early language experience for the maintenance (Gottlieb, 1976; Tees, 1976) of native perceptual abilities is generally accepted (e.g., Walley, Pisoni & Aslin, 1981). In fact, Polka, Colantonio and Sundara (2001) even find evidence for the sharpening of perceptual abilities with experience. English speaking adults and English learning infants of 6-8 and 10-12 months of age were tested on their abilities to perceive a native voiced dental stop vs. fricative contrast /d/ - /D/. The two groups of infants showed comparably successful discrimination of the contrast, but the adults showed increased discrimination compared to both infant groups, suggesting a facilitation of discriminatory abilities that occurs somewhere between 12 months of age and adulthood. Similar results of improvement have also been reported by Kuhl, Tsao and Liu (2003).

The Perceptual Magnet Effect provides one theoretical formalization of the effects of language exposure upon perceptual abilities (e.g., Kuhl, 1995). This model conceptualizes the full range of speech sounds as if displayed in space, with the distance between tokens corresponding to their degree of perceptual difference from one another. Exposure to the native language input effectively alters the distances, pulling nonmeaningfully contrasting phones closer together to result in less easily discriminated contrasts. Additionally, phonemes denoting meaningful distinctions are thought to be pushed farther apart, leading to increased prominence of these contrasts. This recalibration would lessen the perceptual load on the listener, allowing for eased disregard of irrelevant differences and increased saliency of meaningful contrasts between phonemes.

In evidence of this effect, adults presented with a category prototype and asked to rate the goodness of its surrounding tokens judge those that belong to the category to be disproportionately closer to the prototype than nonmember tokens (Kuhl, 1991). Six-month-old infants were less likely to notice a perceptually equidistant stimulus change when initially exposed to a category exemplar than a poor exemplar token (Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). Importantly, English-learning infants showed the effect only in correspondence with English and not Swedish language categories, whereas Swedish-learning infants displayed the reverse effect (however, for contradictory evidence and an alternative explanation, see: Polka & Bohn, 2003).

An alternative to the Perceptual Magnet Effect is a model focusing on statistical learning of the frequency distribution of the input. In reminder, the difficulty that 11-13 month-old, but not 6-8-month-old English learning infants experience with the Hindi retroflex-dental distinction was suggested to result from exposure to the single intermediate English category, leading the older, more experienced infants to collapse the earlier discriminated categories into one (Werker & Lalonde, 1988). If the frequencies of the input tokens were graphed, the pattern of the input categories would resemble a monomodal distribution for the English-learning infants and a bimodal distribution for the Hindi-learning infants. Maye, Werker and Gerken (2002) used 6-8 month-olds to test the hypothesis that infants can keep track of and use the statistical distribution of the input. They familiarized two groups of infants to 8 artificial tokens from a /da/ - /ta/

continuum, presenting each token with frequencies that formed either a monomodal or bimodal distribution (see Figure 1). The monomodal distribution contained more instances of tokens 4 and 5 while the bimodal distribution most often presented 2 and 7. The infants were then tested on their ability to discriminate tokens 1 and 8. If distributional learning is the mechanism of perceptual change, then familiarization with the monomodal distribution would indicate the existence of a single category and lead to decreased discrimination of its endpoints. However, the bimodal distribution would suggest a useful distinction between two categories, enhancing discrimination of the continuum endpoints. Indeed, infants familiarized with the bimodal distribution treated the two endpoint tokens as different, suggesting that they had inferred the existence of two categories. Conversely, those familiarized with the unimodal distribution considered the endpoints instances of the same category, showing poorer discrimination. These results strongly suggest that infants are able to keep track of and use the statistical distribution of the input.

Anderson, Morgan and White (2003) provided further evidence in support of a statistical frequency basis for perceptual reorganization by showing that discriminatory decline of non-native contrasts commences earlier for those comprised of phones that appear more frequently in language usage than less frequent phones. If statistics tracking did in fact influence the development of the phonological system, infants should be expected to more quickly learn the nonmeaningful nature of more frequently presented contrasts than those with which they have had less experience. When tested on the previously described Hindi- and Thompson-specific contrasts that are equally poorly discriminated by English speaking adults, 6.5 month-old English learning infants were found to equally discriminate the less frequently appearing phones comprising the Salish velar vs. uvular contrast and the more frequently appearing phones in the retroflex vs. dental Hindi contrast. However, English learning 8.5-month-olds discriminated the former less frequent contrast more easily than the latter, demonstrating that impairment of the more frequent contrast phones was greater than that for those less frequent. These results are consistent with the hypothesis that the statistical learning of frequencies is the mechanism of reorganization.

Phonetic detail in words.

To demonstrate infant use of phonetic detail in word form, Jusczyk and Aslin (1995) familiarized 7.5-month-old infants to either *cup* and *dog* or *feet* and *bike*, and measured the respective looking times of the two groups to four sentences, each containing one of the four initial stimuli. They found that the infants listened significantly longer to the two sentences containing the tokens they had been familiarized to. However, if the familiarization stimuli were replaced with the similar word forms *tup* and *zeet* or *gike* and *bawg* (all of which differed by a meaningful phonemic contrast) and followed by the same four sentences (containing the original word forms), this did not lead to any longer listening times for the sentences containing the similar words. This demonstrates the ability of infants to perceive fine phonetic detail within word forms, none of which were associated with any meaning by the infants.

As detailed, this ability to distinguish meaningful variation is fundamental to learning the words of a language. However, there appears to be a time lag from the acquisition of native phonemes and the resulting contrasts, to their application in the service of learning meaningful words.

Stager and Werker (1997) showed that at 14-months, infants succeed in a "Switch" paradigm requiring them to map the dissimilar sounds "lif" and "neem" to separate objects (word learning) and notice when the labels are switched. Yet if the labels used are similar to one another ("bih" and "dih"), they fail to notice the swap – even though those exact sounds are perceptually discriminated when presented with a visual checkerboard pattern (i.e., a simple speech discrimination task as opposed to word learning). Furthermore, infants of 8-months can succeed in the Switch task with the similar labels, indicating that for them, the task is not one of word learning, but of simple discrimination.

Thus, although infants may be in the process of learning their native categories and able to perceptually discriminate the L1 phonemes in word forms, they may not be subsequently successful with their use in word learning tasks. It appears that the language-specific perceptual knowledge used in simple discrimination tasks may differ from that of more involved tasks such as word learning.

To wit, a theoretical distinction has been proposed between "deep" and "shallow" contrast knowledge (Werker, Marcus, Mehler, Neville & Sebastián-Gallés, 2001)[2]. Deep knowledge encompasses perceptual abilities within a lexical (meaningful word) context, whereas shallow knowledge refers to simple surface perception. The different types of knowledge are applied in different contexts, which could account for the application of knowledge to one type of task but not another. This distinction is consistent with PRIMIR, a speech processing model that considers infant attention to focus upon different aspects of the input speech signal depending upon the stage of development and type of task (Werker & Curtin, under revision). Here, the attentional distinction appears to be between the word or stimulus form (shallow), and the functional meaning of the word form (deep).

Whereas infants appear to be able to perceive contrasts in a shallow context within the first year of life, only one experiment reviewed here tested deep knowledge, that by Stager and Werker (1997), which ascribed meaning to the sounds (by labeling an object). This study explicitly contrasted deep and shallow and suggested that simple phonetic discriminatory ability (shallow knowledge) appears to consolidate within the first year of life, as demonstrated by the ability of the infants to discriminate "bih" from "dih" (consistent with numerous studies showing the sensitivities of young infants to most phonetic distinctions). However, deep knowledge may represent a separate realm, as the infants were unable to discriminate the same sounds when they were used as labels for objects.

To summarize, the critical difference between deep and shallow knowledge is use in a linguistic task, in this case attaching meaning to an object. Shallow abilities are isolated. They can be used in simple perceptual situations that lack meaning, while deep abilities are integrated into the language system and are applied in meaningful contexts. The Stager and Werker article was one of the first suggestions of this perceptual distinction, and is the only published explicit comparison of deep and shallow to date. Nevertheless, there have also been indications that this knowledge type distinction found in infants learning their L1 may be paralleled in adults learning their L2.

Adults: Second Language Acquisition

As earlier noted, the perceptual attunement to the contrasts of the native language is thought to be extremely valuable in bootstrapping the infant into language. But because different languages contain different phonemic organizations, these exquisite L1 sensitivities necessarily come at the expense of some non-native contrasts, for the perception of which the system is now nonoptimally calibrated. Thus the question is posed of how this reorganization affects perception of a later-learned L2.

According to the Perceptual Assimilation Model (PAM; Best, 1993, 1995; Best, McRoberts & Sithole, 1988), perception of non-native phonemes is dependent upon the particular inventory of native phonemic categories. The unfamiliar phones will be perceptually assimilated to existing categories in one of three ways, determined by their degree of articulatory difference from the closest existing L1 categories. They could be mapped onto a native category, considered an uncategorizable speech sound, or not assimilated to speech at all. The manner of category assimilation determines the ease of accessibility of resulting contrasts, which

---

[2] The shallow vs. deep terminology was suggested by Marcus.

varies widely. For example, excellent discrimination can result if each new phoneme is assimilated to a different L1 category, or even if they are mapped to the same one where one is a markedly better exemplar in the L1 than the other. However, if the two phonemes are mapped onto the same previously existing category with a slight or no goodness difference between them, ability of discrimination will likely be quite poor.

The Speech Learning Model (SLM; Flege, 1995) allows for the possibility of novel L2 category formation, contingent upon the degree of perceived dissimilarity between the L1 and L2 phonemes, and the age of learning the L2 (AOL). Greater dissimilarity between the new L2 and already-formed L1 categories, coupled with an earlier AOL are theorized to lead to a higher probability of new category formation.

For example, the single liquid manifested in Japanese is perceptually more similar to the English /l/ than /r/ for native speakers (Sekiyama & Tohkura, 1993; Takagi, 1993). Here, the SLM predicts that native Japanese speakers will more easily form a category for the more distant English /r/ as opposed to the closer, more similar /l/. In fact, Flege, Takagi and Mann (1995b) found that when auditorally exposed to English /r/-/l/ contrastive minimal pairs (words that differ by only a contrast, e.g., /rock/ - /lock/), native speakers of Japanese were able to correctly identify more tokens containing /r/ than those with /l/.

Therefore, both the PAM and SLM fundamentally agree that although adults have the ability to perceive non-native or L2 contrasts, actual performance will vary, largely dependent upon the specific L2 contrasts being learned, and how they compare to the previously formed L1 phonemes. Indeed, certain experiments have shown that adults can in fact be sensitive to non-native contrasts, and the following review of the literature reveals that it is more likely that adults will be sensitive to non-native contrasts when tested in simple perceptual tasks (shallow) than when tested on their ability to use the L2 categories to distinguish meaningful words (deep).

Shallow.

The AX (same-different) task involves presenting the participant with pairs of stimuli on which they make a "same" or "different" judgment. The stimuli pairings consist of both within-category pairings (e.g., a /ba/ vs. a different /ba/) and those between-category (e.g., a /ba/ vs. a /pa/) that a native speaker of a language that makes such distinctions could be expected to correctly judge to be the same and different, respectively. However, this discrimination is potentially more difficult for native speakers of a language with no such distinctions, where the contrasts are non-native.

Carney, Widin and Viemeister (1977) used a /ba/ - /pa/ continuum to demonstrate non-categorical perception (potentially non-native contrast discrimination) using a number of tasks, including the AX task. Their work illustrated that even within the group of shallow tasks where meaning is absent, some allow for greater non-categorical discrimination than others. These comparisons demonstrate the effect of the varying cognitive demand within different contexts and indicate the existence of a range of shallowness.

In a series of AX experiments, English speaking adults were found capable of discriminating the Thompson /k'/ - /q'/ contrast and the Hindi /Ta/ - /ta/ distinction (Werker & Tees, 1984b). When the inter-stimulus interval (ISI) between the stimulus pairs was 1500ms, the subjects failed to discriminate between the two, ostensibly because the auditory memory trace of the first stimulus had decayed by the time the second was presented. In its place was only a summary representation of an English category label that also described the second. However, when the ISI was lowered to 500ms, reducing the cognitive demand of the task, individuals were able to successfully discriminate the non-native contrast pairs. The shorter ISI encouraged perception of nonmeaningful variation in the native language, or phonetic (non-categorical) response, whereas the longer ISI with higher cognitive demand masked this variation, encouraging subjects to respond phonemically or categorically. Thus, the AX task with a shorter ISI can be seen as tapping shallower knowledge than one with a longer ISI.

In an AXB discrimination task, participants are exposed to three stimuli and are asked to decide whether the middle stimulus is from the same category as the first stimulus or the second. Best, McRoberts and Sithole (1988) used this paradigm to show that English speaking adults are capable of discriminating a number of isiZulu nonnasal voicing and place click contrasts. The total inexperience with these click phonemes afforded by the native language did not impair discrimination. This task is likely generally more demanding (deeper) than the AX task, as it requires comparison of three stimuli as opposed to only two. However, the isiZulu stimuli are not similar to the sounds of English and may remain discriminable for other reasons (Best, McRoberts, LaFleur & Silver-Isenstadt, 1995).

Training has also been found to improve the non-native contrast perception abilities of English speakers (Tees & Werker, 1984). Sensitivity to a change of stimulus was investigated for the Hindi-specific /Ta/ - /ta/ and /t^ha/-/d^ha/ contrasts. Individuals who were initially unable to discriminate the contrasts were further exposed to the sounds, and given feedback on their decisions before retesting. This training had less of an impact upon the voicing /t^ha/-/d^ha/ contrast, but significantly improved performance on the place /Ta/ - /ta/ contrast, with improvements generally maintained after 30-40 days.

All of these non-native contrast perception abilities would be considered shallow, as the phonemes tested were outside of a lexical context and tested within syllables devoid of meaning. However, it is also important to keep in mind the impact of the degree of cognitive demand upon the shallowness of the task, as it seems to determine the degree of shallowness and may result in a less discriminable contrast. In general it appears that, similar to infants learning their L1, in certain shallow situations individuals are able to demonstrate non-native or L2-specific contrast knowledge. However, non-native contrast perception work within a deep context reveals a different picture.

Deep.

Much of the L2 speech processing research has been conducted in Catalonia, a region in Spain, whose inhabitants are typically bilingual speakers of Spanish and Catalan. Usually only one of the languages is spoken at home and exposure to the other does not begin until formal schooling commences at about six years of age. Complete fluency is eventually attained in both languages but the first is considered dominant.

In an investigation into the vowel representations possessed by these bilinguals[3], Pallier, Bosch & Sebastián-Gallés (1997) contrasted the two mid front Catalan vowels: high /e/ and low /ɛ/ with the single intermediate /e/ manifested in Spanish. Seven vowels were synthesized at equal intervals along the Catalan /e/ - /ɛ/ continuum, and participants were asked to decide whether a given token was a better example of the first vowel in the Catalan minimal pairs */Pere/*, meaning *Peter*, or */pɛre/,* meaning *father*. Results indicated that the Catalan-dominant subjects alone were able to map the stimuli differentially onto the two different words, assigning significantly more of the /e/-end tokens to */Pere/* and /ɛ/-end tokens to */pɛre/.* The Spanish-dominant subjects produced an essentially flat curve, indicating that they failed to differentiate along the continuum and could not discriminate between the Catalan L2 vowels in a lexical, meaningful context.

The subjects further performed typicality judgments, rating the goodness of the tokens as compared to an imagined vowel in a given word that utilized either the Spanish /e/, the Catalan /e/, or the Catalan /ɛ/. When asked to rate for goodness of the Spanish /e/, Catalan-dominants more highly rated tokens that were similar to the Catalan /e/, suggesting that they had conflated

[3] *L2 speaker* and *bilingual* will be used interchangeably.

the two, using the L1 category to additionally serve the L2 vowel. Thus, even when exposed to the L2 by 6 years of age, neither group of bilinguals appeared to form categories for all three separate vowels.

In work with the same Catalonian population, Sebastián-Gallés and Soto-Faraco (1999) used a gating paradigm to investigate discrimination of four contrasts that exist in Catalan, but not Spanish. Minimally paired nonwords utilizing the Catalan-specific contrast were divided into nine gates or segments. The third gate was placed in each nonword at the divergence point of the minimal pairing, and the others in 10 ms increments about it (i.e. the first gate was 20 ms before the third gate). The tenth gate was the entire nonword, regardless of how much of the word had been presented up to the ninth gate.

The experiment began with visual presentation of the two nonword Catalan-specific contrast minimal pairs which were auditorily articulated in full upon key-press request to a maximum of twelve times. Once subjects had ceased to request further presentation or had exhausted the allowable number, gated presentation commenced. One of the onscreen minimally-paired nonwords was presented up to the first gate and subjects performed a forced-choice identification decision between the two visible nonwords, along with a confidence rating of that decision. This same nonword was then presented up to the second gate with the same decisions elicited; this was repeated through all ten gates.

Results indicated that although both groups generally were able to discriminate the contrasts, the Spanish-dominant and Catalan-dominant subjects did not perform at the same level. Firstly, the Spanish-dominants initially requested to listen to the stimuli significantly more times than did the Catalan-dominants, indicating their initial uncertainty with the contrasts. And more tellingly, they required a greater number of gates before both correct and confident identification of the nonword for three of the four contrasts. The one contrast for which this did not hold likely formed a category-goodness difference in Spanish, easing discrimination for the Spanish-dominants.

Not only does an L2 speaker require more information to process non-native contrasts, but it appears that for some tasks two phonemes may actually be represented identically, as homophones (Pallier, Colomé & Sebastián-Gallés, 2001). In a repetition priming task, subjects perform lexical decisions (word/nonword judgments) on a list of stimuli in which the same word might occur more than once. A priming effect is observed in response to repeated words whereby a faster decision time is recorded to the repeat than to the earlier presentation of the word. The Spanish-dominant bilinguals were vulnerable to this effect not only when a subsequent presentation was a repeat of a previously presented word, but also when its minimal pair was presented, where the tested contrast was Catalan-specific.

Therefore, based on a body of work conducted in a population where the L2 is learned at a very early age, it appears that the L2 speakers do not possess deep knowledge of the tested contrasts. Interestingly, the results of the gating task, which placed the lightest perceptual load upon the subjects, were most favourable to the discriminatory abilities of the L2 speakers, suggesting that this task was less deep than the others. These results are at odds with those found with in certain sensitive shallow non-native contrast discrimination tasks, in which subjects succeeded, often despite a lack of experience with the language from which the non-native contrast was selected.

Shallow vs. deep.

The adult literature supports the hypothesis that there exist important differences between the inferences drawn from shallow tasks (where L2 speakers have been shown to be quite successful), and deep tasks (where clear differences between natives and L2 speakers have been found, even when the L2 was learned early in life). Furthermore, results indicate that there is variation within each of shallow and deep, with some tasks tapping knowledge that is shallower or deeper than others. Therefore, within L2 contrast learning, a shallow/deep distinction of some type appears likely, similar to that of infants learning their L1. However, this presumption is extrapolated from a body of research that is not all directly comparable as it was conducted on a variety of populations and tested a range of contrasts chosen from different languages.

In the present work, explicit evidence is sought for this shallow/deep knowledge distinction in a group of English L2 speakers. Two experiments were designed in an attempt to tap into the two hypothesized separate types of contrast knowledge and sought to determine whether bilinguals would more easily discriminate L2 contrasts when embedded in meaningless syllables than within meaningful words.

Additional Theoretical Issues

The present search for a shallow/deep distinction within L2 learners could also potentially speak to two overarching theoretical questions, due to the unique demographics of the chosen population, as will be later detailed. Therefore, the data will be analyzed in such a way as to speak to each of these.

Critical period.

One of the most compelling questions in the psycholinguistic literature is that of the critical period hypothesis. This is the notion of a biologically based window within which a learner must be exposed to language in order to become a "native-like" speaker (Lenneberg, 1967; Penfield & Roberts, 1959).

That increasing age is related to greater difficulty in learning a language is uncontroversial. However, whether or not this constitutes evidence of a true critical period remains an open question; it is unclear whether there is a point after which level of attainment significantly diminishes, or whether linguistic potential simply diminishes gradually with greater age (Hakuta, Bialystok & Wiley, 2003). Clouding the picture, there are a myriad of other factors that influence second language acquisition and as a result the literature has so far been unable to draw a clear conclusion, with some finding evidence of such a critical period (e.g., Johnson & Newport, 1989; 1991) and others failing (e.g., Bialystok & Hakuta, 1999; Birdsong & Flege, 2000).

If such a critical period does exist, it is further possible that different aspects of contrast perception (i.e., shallow vs. deep), as different aspects of language, may have different critical periods. Alternatively, the critical period may apply to deep but not shallow knowledge.

Interference.

Another overarching issue is that of interference. Ventureyra, Pallier and Yoo (2004) studied French adults who had been adopted as children from Korea between the ages of 3 to 9. At the time of adoption, the children ceased to be exposed to their native Korean and began learning French without interference from the L1. These adoptees, with French and Korean monolinguals, were tested on Korean-specific VOT contrasts within a full nonword (consonant-vowel-consonant-vowel; CVCV). These bisyllabic nonwords were presented paired in an AX task with an ISI of 500 ms, encouraging detection of the differences. Essentially, the adoptees were found to be just as poor as the French natives in discrimination of the Korean-specific contrasts, indicating that although Korean was their first-learned language, the blanketing interference from French had overridden the Korean-tuned perceptual abilities.

Another study tested individuals from the same population who had been adopted between the ages of 3-8 years (Pallier, Dehaene, Poline, LeBihan, Argenti, Dupoux & Mehler, 2003). These adoptees did not differ from French natives in conscious behavioural tasks, nor did fMRI data indicate any unconscious recognition of the native Korean language. In fact, the imaging results did not show any greater activation in the adoptees to Korean than to other unknown languages (Polish and Japanese), and their responses to French were similar to those of native monolinguals, although the magnitude was greater in the native speakers.

Conversely, other work has found that early experience does affect later abilities, although in these "overhearing" populations the early linguistic experience is aided by continued maintenance of the overhearing into adulthood; the exposure is not truncated as in the case of the adoptees. Spanish-learning adults who had "overheard" Spanish as children (but never learned to speak it) were judged by native speakers to have better Spanish accents than non-overhearers (Au, Knightly, Jun & Oh, 2002). Mere exposure was apparently sufficient as neither the duration or amount of overhearing were related to performance.

From a group of adults enrolled in an introductory Korean course, those who were native speakers and those who were childhood overhearers of Korean (Oh, Jun, Knightly & Au, 2003) were tested and compared on ability to discriminate Korean-specific VOT contrasts. These three-way alveolar stop contrasts were used to create minimal triplets (three words differing only by the three-way contrast) of which one was presented and an identification judgment elicited. Both the childhood speakers and overhearers were found to perform at the level of native speakers, significantly above the level of individuals with no early exposure to Korean, underscoring the effect of simple exposure to a language in childhood.

Method

The present research explicitly compares deep and shallow L2 contrast knowledge in a group of Cantonese-English bilinguals. Deep knowledge was tested using the repetition priming paradigm of Pallier, Colomé & Sebastián-Gallés (2001), and investigated whether or not L2 English speakers would be able to perceive the differences between English words that differ only by a contrast not utilized in their L1. Shallow knowledge was tested in an AX task by assessing ability to discriminate between pairs of nonmeaningful syllables differing only by the same L2-specific contrasts. Three different English-specific contrasts were selected for use in the two tasks to assess the generalizability of the L2-specific findings. It was hypothesized that

L2 speakers would more easily discriminate the L2-specific contrasts when they were embedded within the shallow task syllables devoid of meaning than within the deep task meaningful words.

Subjects

Native speakers of Cantonese with fluent L2 knowledge of English along with English monolingual controls were recruited to take part in these studies from personal contacts and the University of British Columbia (UBC) undergraduate subject pool, via postings around campus. A relatively high level of English proficiency could be anticipated in these Cantonese-English bilinguals as UBC's language of instruction is English, and L2 students are required to demonstrate English proficiency by attaining a pre-set standard on one of a range of offered tests, including the BC English 12 Provincial Exam (70%) and the Test of English as a Foreign Language (TOEFL).

For an hour of their time, participants were offered their choice of $10 or, if they were enrolled in a designated Psychology class, 1 course credit.

There were 84 females and 39 males for a total of 123 subjects. 52 were native English monolinguals and 71 were bilinguals who had acquired English at a range of ages.

Stimuli

Three <u>English-specific contrasts</u> were tested. All three denote meaningful differences in English, but not in the Cantonese L1 (Chan & Li, 2000). They differ in the potential difficulty of discrimination they pose for the bilinguals.

1. /v/-/w/

/v/ is a voiced labiodental fricative and /w/ is a voiced bilabial approximant. They thus differ in both place and manner of articulation. Both of these phones appear in Cantonese production, yet never contrast with one another. They are considered allophones of the same phoneme, or different manifestations of the same category, meaning that if two words differed by only those phones, they would be considered identical. However this never happens as their distribution is complementary, meaning that /v/ appears in certain linguistic environments and /w/ in the others; they never contrast. Given their allophonic status, both can be expected to be assimilated to the same L1 phoneme category, making discrimination difficult. However, it is also possible that the experience with both of the L2 phonemes as allophones in the L1 may aid in the maintainance of their perceptual distinctiveness (MacKain & Stern, 1982; Tees & Werker, 1984). Moreover, the fact that the two phones differ by two features (place and manner) may make them more discriminable.

2. /s/-/T/

/s/ denotes a voiceless alveolar fricative while /T/ is a voiceless interdental fricative. Whereas the former also exists within Cantonese, /T/ does not. Its manner of assimilation is unpredictable as it falls between the common phonemes /s/ and /f/, the latter of which is a voiceless labiodental fricative. The /T/ could be assimilated to either of the Cantonese fricatives, but the choice of /s/ was guided by the assimilation pattern of native French speakers, who lack the unvoiced interdental fricative /D/ and assimilate it to the unvoiced alveolar /z/ (Werker, Frost & McGurk, 1992).

3. /ε/-/Q/

/ε/ is a lax mid front unrounded vowel; /Q/ is a short low front unrounded vowel. /ε/ alone exists in Cantonese, /Q/ exists only in English, and the two differ by only a single feature. Thus this pairing may represent the most difficult contrast for the bilinguals.

As control, three "common" contrasts that are meaningful in both Cantonese and English were also chosen. They were matched as closely as possible to the English-specific contrasts, and two of the three pairings involve phonemes articulatorily closer to one another than the English-specific contrasts that they control for.

It is important to note that although these phonemes exist in both languages, their acoustic-phonetic features likely differ cross-linguistically. That is, the phonemes may manifest themselves slightly differently in each language. It is nevertheless expected that the L2 phonemes will map to their corresponding but possibly slightly shifted L1 partners, which, crucially, contrast in the L1, making the L2 contrast common to both languages.

Phonemic contrasts that are meaningful in both English and Cantonese:

1. /m/-/n/

/m/ is a voiced bilabial nasal and /n/ is a voiced alveolar nasal. These two phonemes both exist and are used to contrast meaning in both English and Cantonese. They differ by only a single phonetic feature: place of articulation.

2. /f/-/v/

/f/ is a voiceless labiodental fricative and /v/ is a voiced labiodental fricative. Both of these phonemes exist and contrast meaning in both languages. They differ only in voicing, a single feature.

3. /u/-/ç/

/u/ is high back rounded vowel and /ç/ is a mid back rounded vowel. They exist and distinguish meaning in both languages, with the single feature difference of height.

Deep task stimuli.

From these six contrasts (three English-specific and three common), lists of English minimal pairs were created. For each contrast eight pairs of minimally different words were recorded (16 words, 8 of each phoneme). Filler items were also included. See Table 1 for a list of tested stimuli by contrast.

There were a total of 28 instances of each phoneme in its tested position (word-initial for consonants and syllable-medially for vowels) over the entire list.

This yielded a total of 290 English words and nonwords[4].

a) 48 words (24 minimal pairs) utilizing common contrasts (meaningful in both Cantonese and English

b) 48 words (24 minimal pairs) utilizing the L2-specific contrasts

• a) and b) allow for a comparison of how L2 speakers discriminate contrasts that are meaningful in both languages versus those that denote meaning only in the L2

c) 48 nonwords (24 minimal pairs) utilizing the L2-specific contrasts

• the inclusion of nonwords will reveal whether or not any effects found are specific to words

d) 146 filler word and nonword tokens.

• these filler tokens were carefully chosen to allow for balanced exposure to phoneme types

The comparisons of interest (repeats and minimal pairs) were embedded in the list of stimuli a minimum of eight slots, and a maximum of eighteen slots apart (see Table 2 for a full list of stimuli by order).

These English words were chosen so as to not violate the phonotactics of Cantonese, which generally places consonants in the syllable-initial position and vowels syllable-medially or syllable-finally. Nasals are the consonantal exceptions that can be found in syllable-final position in Cantonese as well as syllable-initial, but to maintain consistency consonants were all placed syllable-initially here. Thus, consonantally-contrasting minimal pairs that diverged in a syllable-final position were rejected (e.g., /pass/ - /path/) in favour of those that contrasted word-initially (e.g., /sick/ - /thick/). Similarly, only vocalicly contrasting minimal pairs that differed syllable-medially were included (e.g., /bad/ - /bed/). Furthermore, vocalic minimal pairs that

[4] All stimuli tokens were recorded in a soundproof booth by a native Western Canadian English monolingual speaker who was raised in Maple Ridge, BC.

placed nasals such as /m/ or /n/ immediately following the vocalic contrast (e.g., /command/ - /commend/) were discarded due to the significant "colouring" of the vowel imparted by the nasal, potentially tainting the contrast.

Note that the /v/ phoneme is utilized in this design both as a member of a common contrast (/f/ - /v/) and one that is English-specific (/v/ - /w/). In order to avoid any differential effects of adaptation to /v/, the filler tokens (words and nonwords) were carefully chosen to balance the overall exposure to the phoneme used in each condition, resulting in an equal number of all phonemes within a tested contrast.

Shallow task stimuli.

To ensure that the respective sets of stimuli used in the two tasks were as comparable as possible to one another, the initial CV syllables from deep task words were excised to create stimuli for the shallow task. Four deep words were chosen per contrast to create the shallow stimuli. This yielded two instances of two CV syllables each. For example, the deep tokens *bad, bat, bed, bet* yielded two /ba/ syllables (from *bad* and *bat*) and two /be/ syllables (from *bed* and *bet*). These four syllabic tokens then generated 16 different pairing combinations, for example:

a) 4 physically identical pairs: ba1-ba1, ba2-ba2, be1-be1, be2-be2

b) 4 same-category pairs: ba1-ba2, ba2-ba1, be1-be2, be2-be1

c) 8 different-category pairs: ba1-be1, ba1-be2, ba2-be1, ba2-be2; then in reverse order

These 16 pairings were reproduced across all six tested contrasts yielding a total of 96 pairs (see Table 3). In addition, to again balance exposure to all tested phones as a result of the repeated /v/ phone, ten filler syllables (each including one of the tested phonemes other than /v/) were each randomly paired sixteen times with each other. This resulted in a total of 186 pairings (tested and filler).

Procedure

Participants completed both the deep and the shallow experiments[5] within an hour. Subjects first completed either the shallow or the deep task, the ordering of which alternated from subject to subject. Next, they filled out a questionnaire for language dominance (Desrocher, 2003), then completed the remaining shallow or deep task, and finally, answered the Vancouver Index of Acculturation (bilinguals only; Ryder, Alden & Paulhus, 2000) and a vocabulary check questionnaire.

Deep task.

Deep contrast knowledge was tested using a repetition priming task. Reaction times (RTs) were recorded to lexical decisions on a list of auditorily presented words and nonwords in which the minimal pairs were embedded.

Participants were tested individually in a sound-attenuated room where they were seated in front of a keyboard and monitor and fitted with Peltor HTB79A-02 headphones. They were given verbal instructions to indicate whether each stimulus was an English word or nonword by pressing either the *b* (= word) or *n* (= nonword) keys on the keyboard.[6] Subjects were informed that their reactions times would be measured, and asked to leave their index fingers on the respective keys to allow for the quickest response possible. When the participant indicated that he or she was comfortable with the procedure and was wearing the headphones, the experimenter left the room, and the testing session commenced. The software DMDX (Forster & Forster, 2003) was programmed on an Intel-based personal computer to present stimuli and record responses[7].

---

[5] One individual completed only the shallow task due to experimenter error on the deep task.
[6] In both the deep and shallow task, key labels were clearly identified on the monitor and were counterbalanced across subjects.
[7] Thanks are extended to Ramesh Thiruvengadaswamy for programming the experiments.

Each subject first completed a training phase containing eight practice stimuli. No test phonemes were included. The purpose was simply to familiarize participants with the procedure, and to give them practice making a lexical decision and pressing the *b* and *n* keys. Regardless of performance every subject continued on to the test phase at conclusion of this training. No feedback was given.

In both training and test, each token was presented 2000 ms after the response to the previous stimulus. If there was no response recorded after 3000 ms, presentation of the next stimulus commenced.

In the test phase, participants received one of four different counterbalanced word orders (see Table 2). Thus, any given ordering consisted of half of the complete minimal pairings (36 of the 72 pairs). The remaining half of the pairs appeared in the ordering as a repeat; only one member of the pair was presented across each order (for each subject), and it occurred twice while its minimal pair was not presented at all. For example, the minimal pair *bad* and *bed*: half of the subjects received one token of *bad* and one token of *bed*, while the other half of the subjects were exposed to two tokens of either *bad* or *bed* (and no tokens of the other). Two trials of any given minimal pair appeared in every order; one order contained a repetition of the same token (true repetition) while the other contained one presentation of each member of the minimal pair (minimal pair).

Deep task: Data set-up.

To assess repetition priming, the RT to the second member of the pair (minimal pair or repetition) was subtracted from the RT to the first member of the pair. A positive difference RT represents a repetition priming effect (speeded response to the second stimulus), while a zero difference RT represents no repetition priming.

Deep task: Predicted results.

Both the Cantonese-English bilinguals and the English monolinguals were expected to show the repetition priming response to true repetitions of words (irrespective of whether the single repeated word was a part of a common or English-specific contrast minimal pair). The first presentation should prime the second, resulting in a faster second RT, and a positive difference RT.

Similarly, both groups were expected to show equivalent reaction times to each member of the common contrast minimal pairs (difference RT of 0). The contrasts are available to both groups, allowing all subjects to discriminate the minimally-paired words, and preventing any repetition priming effect.

The predicted difference between the groups lay in the RTs to the English-specific minimal pairs. The monolinguals were expected to treat these minimal pairs identically to those of the common contrast, as both groups of contrasts are native and easily perceived. They should easily distinguish between the minimal pairs, avoiding any priming effect, and recording equivalent RTs to each member (difference RT of 0). The bilinguals, however, were anticipated to show more confusion between the two. If, at some level of perceptual or lexical representation, they did not distinguish between the L2-contrast minimal pairs, the tokens would be treated as repeats, resulting in significant repetition priming and a positive difference RT between the first and second presentation.

No repetition priming effect was expected for either the nonword repeats or minimal pairs. Repetition priming effects do not typically hold for nonwords (Pallier, Colomé & Sebastián-Gallés, 2001), possibly due to a processing lag masking RT differences.

Shallow task.

Discrimination of L2 contrasts within a shallow context was tested using an AX procedure to investigate the same contrasts taken out of a lexical, meaningful context. Subjects were presented with a series of syllabic pairings, and requested to make judgments of either "same" or "different" for each. Correct performance would be "same" judgments for both the physically identical and same category pairs, and "different" judgments for any pair of syllables containing two syllables differing in one of the contrasts of interest.

In previous work an ISI of 500 ms has been shown to facilitate discrimination of non-native and even within-category differences between stimuli (Werker & Logan, 1985; Werker & Tees, 1984b). Thus, this brief ISI was used to ensure the task could be performed at a shallow level. Inter-pair intervals were identical to that of the deep task; 2000 ms from response, or 3000 ms if no response was recorded.

Participants were given oral instructions to press *b* if the two syllables were the same, and *n* if they heard them as different. Again, subjects were told that their RTs would be measured, and asked to leave their index fingers on the respective keys.

As there were multiple instances of each syllable type, subjects were instructed to ignore non-phonetic differences. Therefore, although the ISI should allow for individuals to perceive the within-category stimuli differences (the "same-category" pairs), they were instructed to consciously consider whether or not the differences were irrelevant in English, or whether they indicated a change of category.

Once the subject was comfortable with the paradigm and the headphones, the experimenter started the program and left the room. The training phase contained eight pairs for procedural familiarization, and was followed by the test phase, which randomized all 186 pairings individually for each subject.

Shallow task: Data set-up.

One decision was made in response to each pair (same or different), and both accuracy rates and RTs to the decision were recorded.

Shallow task: Predicted results.

In some cases, listeners can discriminate non-native contrasts in shallow perceptual tasks. This sample tested fluent bilinguals with English as an L2. Given their expertise with English and the ease of the task, it is predicted that the performance of the bilinguals will approach that of the monolinguals.

Questionnaires

Language dominance.

The Cantonese-English bilinguals completed all four pages of a questionnaire for language dominance (Desrochers, 2003) to ascertain, among other things, their AOL and percentage of current language usage in both languages. Monolingual participants filled out only the first page, which contained information about handedness, first languages spoken by parents, and any foreign language courses enrolled in.

Vancouver Index of Acculturation.

All of the bilinguals tested were of Southeast Asian descent, many of whom had been born in Asia, and the remainder of whom were born in Canada to immigrant families. To capture the impact of degree of acculturation to mainstream North American society, the bilinguals alone filled out the Vancouver Index of Acculturation. In it, individuals rate 20 statements on a scale of 1 – 9 ("strongly disagree" to "strongly agree"), half of which pertain to the heritage culture, the other half to North American culture. For example, "I believe in the values of my heritage culture" and its counterpart "I believe in mainstream North American values."

This questionnaire was filled out online, as it was added to the procedure after testing had commenced. Previously tested subjects were contacted by electronic mail and directed to a website where they could submit their responses.[8] Later tested subjects used a lab computer to respond during the experimental session.

Vocabulary check.

Given that the vocabularies of subjects may reflect a variation in language proficiencies, a post-test was administered to ensure that all subjects knew the meaning of all words included in the deep task. Monolinguals also completed the form. Participants were given a printed list of the tested words (English-specific and common contrast) and asked to rate tokens each on a scale of 1-7, where 1 corresponded to "do not know", 4 was "know word", and 7 for "know and understand word". If a participant did not know one or both members of a tested minimal pair, the data from that pair was thrown out.

Summary of Predictions

The monolinguals were expected to perceive the common contrast and English-specific stimuli with equal ease in both the shallow and deep tasks. The bilinguals were expected to perceive the common contrasts equally across the two tasks, but were anticipated to show more difficulty with the English-specific contrasts in the deep context than in the shallow.

Results

In reminder, the two tasks were chosen to test the prediction that they would elicit different results; that the bilinguals would be able to perceive the L2-specific contrasts more easily in the shallow but not the deep task.

Deep Task

Overall differences between linguistic groups.

Two initial analyses were conducted to assess general differences between the bilinguals and monolinguals. Independent samples t-tests were performed on the RTs to both words and nonwords (stimuli that were either repeats or a minimal pair of a previously presented word were

[8] Three subjects did not submit data. One could not be contacted, and two more declined.

excluded to avoid any repetition effects) in order to compare overall RTs. Both were found to be significant [for words: $t(121) = -2.75, p < .01$, for nonwords: $t(121) = -2.59, p < .01$], indicating that the bilinguals were significantly slower than the monolinguals in their responses to both the words and nonwords.

An independent sample t-test was also performed on the overall accuracy rates of the groups, and the monolinguals were found to be significantly more accurate than the bilinguals [$t(121) = 5.26, p < .001$].

Repetition priming of common contrasts.

To assess the repetition priming effect, two 2x2 (linguistic group by condition) between-within ANOVAs were conducted on the overall difference RTs for both the common contrast and the English-specific stimuli.

For the common contrasts, the linguistic group comparison was not significant, indicating that there were no overall differences between the difference scores of the two groups [$F(1, 120) = .034, p > .50$][9]. The condition effect (repetition vs. minimal pair) was highly significant, denoting differences in repetition priming found in response to the true repetitions and the minimal pairs [$F(1, 120) = 14.80, p < .001$], and there was no interaction [$F(1, 120) = .326, p > .50$].

Repetition priming of English-specific contrasts.

For the English-specific contrasts, the linguistic group comparison was not significant, again indicating the absence of overall group differences in repetition priming [$F(1, 121) = 2.36, p > .10$]. The condition effect (repetition vs. minimal pair) was highly significant [$F(1, 121) = 15.80, p < .001$], but more importantly the interaction effect was also significant in precisely the way predicted [$F(1, 121) = 3.73, p < .05$, directional]. This indicates that the bilinguals alone

[9] One subject did not submit data on this measure.

showed significant repetition priming to the English-specific contrast minimal pairs. This analysis was therefore followed by analyses of simple main effects, with the following results: (a) condition for monolinguals, [$F(1, 121) = 16.44, p < .001$], and (b) condition for bilinguals, [$F(1, 121) = 3.13, p > .05$]. This tells us that although the repetition priming effect of monolinguals was significantly greater in response to the true repetitions than to the minimal pairs, the bilinguals showed an equivalently strong repetition priming effect to both repetitions and minimal pairs.

English-specific individual contrasts.

An analysis of each L2 contrast was performed to determine if the predicted differences in difficulty of the L2 contrasts were seen. It appears that the English-specific interaction found is driven by the /æ/ - /ɛ/ contrast [$F(1, 121) = 15.71, p < .001$]. Neither the /s/ - /θ/ or /v/ - /w/ interactions were alone significant [$F(1, 120) = 2.123, p > .10$ and $F(1, 108) = 1.317, p > .10$, respectively].

English-specific nonword contrasts.

In order to confirm that the English-specific repetition priming response was specific to meaningful words, a 2x2 (linguistic group by condition) between-within ANOVA (identical to those performed on the common contrasts and English-specific word contrasts) was performed. It failed to reveal any effects of linguistic group, condition, or an interaction [$F(1, 120) = .044, p > .50$, $F(1, 120) = 3.17, p > .05$, and $F(1, 120) = .2.02, p > .05$, respectively], confirming that the repetition priming effect of English-specific contrast minimal pairs were restricted to meaningful word stimuli.

AOL, interference and culture.

The English-specific minimal pair RTs were regressed onto reported AOL, average percentage of English spoken in everyday life, and the score on heritage culture from the Vancouver Index of Acculturation. It was found to be a very poor fit ($R^2_{adj} = -.054$) and the overall relationship was nonsignificant [$F(3, 54) = .025, p > .50$]. None of the effects were individually significant.

Shallow Task

Two 2x2 (linguistic group by condition) between-within ANOVAs were performed on both accuracy and reaction time data.

Accuracy.

Across groups, accuracy rates were significantly higher for the English-specific contrasts than for the common contrasts [$F(1, 122) = 44.6, p < .001$]. Across conditions the monolinguals performed significantly better than the bilinguals [$F(1, 122) = 10.3, p < .005$], however most importantly, the interaction was highly nonsignificant [$F(1, 122) = .031, p > .50$], indicating that from one condition to the next, performance of the bilinguals was equal to that of the monolinguals.

Reaction time.

In terms of RT, there was a significant interaction between linguistic group and condition [$F(1, 122) = 6.76, p < .05$]. Follow-up simple main effects showed that although the monolinguals performed equally across the conditions [$t(51) = -.517, p > .50$], the RTs of the bilinguals were significantly slower to the English-specific than to the common contrasts [$t(71) = 3.248, p < .01$].

AOL, interference, and culture.

The English-specific pair discrimination accuracy scores were regressed onto AOL, average percentage of English spoken in everyday life, and the score on heritage culture from the Vancouver Index of Acculturation. The regression fit poorly ($R^2_{adj} = .045$), and the overall relationship was nonsignificant [$F(3, 49) = 1.812, p > .10$]. However, average percentage of English spoken was a significant effect [$t(49) = 2.32, p < .05$].

Discussion

Interpretation of Results

Overall speed and accuracy.

The overall less accurate performance of the bilinguals is to be expected as the stimuli were all English language. Even the phonemes that also existed in Cantonese likely differed in their acoustic-phonetic manifestations here. In illustration, Ju and Luce (2004) used Spanish-English bilinguals to show the confusing effects of the language-specific VOTs upon phonemes common to both languages. When exposed to Spanish words (e.g., *playa*, meaning "beach") in which the initial phoneme had been altered to the English-appropriate length, the bilinguals looked longer at the interlingual distracter whose English label began with the same phoneme (e.g., *pliers*). This did not occur when the Spanish word contained the Spanish-appropriate VOT, indicating that bilinguals are sensitive to fine acoustic-phonetic information, making them vulnerable to confusion even when language processing is ostensibly in one language only. Thus, the generally slower and less accurate responses of the bilinguals are consistent with expectations.

Deep task.

Both groups showed significant repetition priming in response to the true repetitions (those from both the common contrast and English-specific lists) and neither group showed significant repetition priming in response to the common contrast minimal pairs. The latter result confirms that both monolinguals and bilinguals were able to discriminate contrasts common to both languages with comparable ease (even when the contrasts were presented in a lexical decision task that requires listening for meaning). The interaction indicates that there was a difference in how the groups performed on the English-specific contrast in the lexical context. The bilinguals showed significant repetition priming, suggesting that the minimally-paired words were in fact treated as repetition of the same word and that significant difficulty was experienced in discriminating the English-specific contrasts.

Shallow task.

Control analyses indicated that the common contrasts were significantly less well discriminated than the English-specific contrasts. This can be explained by the closeness of their constituent phonemes, as the common contrast phonemes were chosen to be articulitorily more distant than those of the English-specific contrasts. This was done in order to rule out the possibility that simple acoustic-phonetic distance could account for differential results between contrasts in the deep task. Thus, the closeness of the common contrasts allowed for a stronger result; that even though the common contrasts were more similar, the English-specific contrasts were more confusing to the bilinguals in the deep task. However, the greater similarity of the common contrasts leads to greater confusion during shallow discrimination, explaining the poorer performance of both monolinguals and bilinguals here.

The accuracy results show that the bilinguals perform equally as accurately as the monolinguals across the English-specific contrasts and the common contrasts. This confirms the hypothesis that within a shallow context bilingual discrimination of L2 contrast perception will be eased.

However in terms of RT, the bilinguals respond significantly more slowly to the English-specific contrasts than to the common contrasts. Therefore, bilinguals discriminate the English-specific contrasts just as accurately as they do common contrasts, but they take significantly longer to do so.

Comparison with predictions.

As anticipated, in the deep task the bilinguals showed impaired discrimination of the English-specific contrasts as indicated by the significant repetition priming effect in response to the L2-specific minimal pairs. In the shallow task, also as predicted, the bilinguals were correctly able to discriminate the L2 contrasts with as much accuracy as the common contrasts within the syllabic pairs. However, an unanticipated shallow result was that the RTs to making the respective decisions were not equivalent. The bilinguals took significantly longer to respond to the English-specific contrast pairs than the common contrast pairs – even though their decisions were ultimately just as accurate.

Critical period and interference.

It had been hoped that the results of the experiments conducted here would be relevant to these two theoretical issues. Yet, only the percentage of English spoken in everyday life was found to relate to shallow performance. AOL was not found to affect either task.

With respects to interference even a small amount of overhearing has been shown effective in maintenance of perceptual abilities (Au et al., 2002; Oh et al., 2003). The amount of Cantonese exposure experienced by most of the present participants is likely sufficient to interfere with the L2, unlike the absolute lack of L1 input experienced by the adoptee populations (Pallier et al., 2003; Ventureyra et al., 2004). Even those of our subjects who socialize entirely with English speakers likely have a minimal amount of exposure to their native Cantonese in Vancouver, or in communication with family. Furthermore, data that was not collected was who this English was being spoken with. It is possible that the English being spoken was a "Cantonese dialect" of English, that did not allow for exposure to English-specific contrasts. Thus, it is interesting that amount of English was able to affect shallow performance.

In the case of AOL, the lack of effect is likely a result of poor data. Subjects were asked to note the age at which they began learning English. This is not as transparent a measure as first assumed, as English is frequently taught in grade schools in Hong Kong. However, their manner and quality of instruction is varied, with some learning to read in English yet being verbally instructed in Cantonese, and some being taught English since commencing school, yet not really learning the language until moving to Canada. Furthermore, the dialect of English being taught could be questioned. This great variation and somewhat subjective measure likely resulted in inaccurate data, and in the future more care should be taken to gather a more complete language history.

General results.

A comparison of the results from the respective tasks supports the notion of a shallow/deep distinction. In the deep repetition priming task, the bilinguals had difficulty discriminating the English-specific contrasts, significantly more than with the contrasts that were meaningful in both English and Cantonese. Yet, they displayed no such difficulties in the shallow AX task, discriminating the English-specific contrasts just as easily as those common to both languages. Therefore, there appears to be a distinction between the types of knowledge applied within the two different contexts.

Such a shallow/deep distinction can account for many of the otherwise difficult to reconcile, disparate findings of research in speech perception. In the infant literature, it can explain the findings of Stager and Werker (1997) with 14-month-olds, who can discriminate “bih” and “dih” in a shallow context, but not in a deep word-learning task. And in the adult work, such a discrimination could potentially separate the deep context failures of Catalan L2 speakers (e.g., Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999) from the shallow successes of non-native listeners (e.g., Carney et al., 1977; Werker & Tees, 1984b). A shallow/deep distinction would be of practical use in explaining past results, but an evaluation of its compatibility with theoretical models of speech processing is necessary.

Three issues of shallow and deep constructs will be further discussed: their compatibility with present theoretical models of speech perception (including their potential in the L1), whether they are continuous or discrete, and potential mechanisms to help explain them.

Models of Speech Perception

PRIMIR.

As earlier noted, a shallow/deep distinction would be consistent with the speech processing model PRIMIR, which allows for attention of the language learner to focus upon different aspects of the speech signal, dependent on the nature of the task (Werker & Curtin, under revision). Here, the emergence of language-specific phonemes is considered to ease perception by collapsing differences between the phones of the input and allowing attention to focus upon those parts of the word form that denote meaning. PRIMIR would explain the difficulty experienced by L2 listeners on the deep task as an indication that they had not yet formed representations of the L2-specific phonemes. The bilinguals are still able to discriminate the L2 contrasts in a shallow task when attention is focused upon the phonetic detail in the speech signal, but in a deep task where attention is directed elsewhere (towards the word meaning), the lack of L2 phonemes to ease perceptual demand results in failure to discriminate relevant differences.

PAM.

Breaking the overall English-specific contrast repetition priming results into the three individual contrasts reveals differences between them that are explained by PAM. The contrast on which bilinguals experienced the most difficulty was the vocalic contrast /ε/-/Q/. PAM would predict that the non-native /Q/ would assimilate to the existing /ε/, resulting in the two phones belonging to the same Cantonese category, correctly anticipating the difficulties. The same type of assimilation occurred with the /v/ - /w/ contrast; whereas in English these represent two phonemes, they are members of a single phoneme in Cantonese. However, unlike the vocalic contrast, both phones are found in Cantonese production, and this continuing exposure likely aided in the maintenance of their perception (e.g., MacKain & Stern, 1982). Thus, discrimination was eased compared to that of /ɛ/-/Q/. Finally, PAM can easily explain the success of the bilinguals in discriminating the non-native /θ/ from the native /s/ by the assimilation of /θ/ to /f/ instead of /s/, resulting in what is, in essence, a comparison of /s/ and /f/. This would be a native contrast to the Cantonese speakers, and as such, would be expected to be easily discriminated. Further work will need to be conducted with Cantonese-English bilinguals to confirm that they, in fact, do assimilate the /θ/ to /f/.

SLM.

The apparent failure of the bilinguals to form representations of the L2 categories is not inconsistent with the SLM. In reminder, this model, unlike PAM, allows for the possibility of new category formation, if the AOL is early enough, and there is enough dissimilarity between the new and native phonemes. The fact that /v/ and /w/ are allophones of the same phoneme in Cantonese would not provide sufficient distance to allow formation of new representations, and similarly, the /Q/ phone would likely be perceived as too close to the Cantonese /ɛ/ to form its own representation. There is a possibility that the success of the bilinguals in discriminating /θ/ from /s/ could be due to formation of a /θ/ category. If this was the case, it could be inferred from the success of the bilinguals in discriminating from /f/, indicating that /θ/ had not been assimilated to the /f/ category and instead had formed its own. The SLM would also predict AOL-related differences within the bilingual group.

Shallow and deep in the L1.

Whereas deep represents greater difficulty of discrimination in L2 speakers, the opposite seems to be true within L1 speakers. Rubin, Turvey and Van Gelder (1976) showed differential performance on L1 word and nonword stimuli in a way that would not be expected in the L2 speaker. They found that RTs to label the initial phoneme in a lexical-like stimulus was faster for meaningful words than nonwords, with subjects more quickly identifying the /b/ in "bit" than in "bip". The authors interpreted this word/nonword difference as a result of a privileged status of meaning, suggesting that it is the presence or absence of meaning in stimuli that determines speed of availability to conscious awareness. However, the perspective of shallow/deep could suggest that the differences actually result from an eased perceptual load of more familiar meaningful words, allowing for faster phonetic consideration.

The Ganong Effect may also support a shallow/deep distinction within adult L1 speakers (Ganong, 1980). This effect demonstrates the differential phonetic categorization of stimuli dependent upon presence of meaning. Subjects were more likely to label a phone as belonging to the category that would bestow meaning upon the token containing the phone. For example, if given a continuum from /dash/ to /tash/, subjects were more likely to categorize more of the stimuli as beginning with a /d/ rather than a /t/. If the continuum ranged from /dask/ to /task/, the opposite effect was found (preference for /t/ over /d/), revealing a bias toward meaning.

This work in the L1 seems to be consistent with a distinction between shallow and deep. However, whereas meaningful words constitute a heavier load for the language learner, the greater familiarity with such stimuli of adult L1 speakers may actually ease perception. Meaningless stimuli may in fact represent a greater processing load, requiring more consideration than the familiar meaningful words. Indeed, there is evidence to suggest that if nonwords are presented as few as eight times they can elicit word-like consideration in a Ganong-like paradigm (Remez, Rubin, Katz and Dodelson, 1985), indicating that familiarity is more important here than meaning.

Continuous or Discrete?

It is as yet unclear whether the proposed distinction between shallow and deep is one of two independent constructs, or whether shallow and deep represent two endpoints of a continuum. The repetition priming and AX tasks could have tapped either two separate (discrete) constructs of L2 contrast knowledge, or different points along a shallow/deep knowledge continuum, far enough apart to reveal distinct differences.

The previous examination of the literature is taken here to support the continuum hypothesis. Tasks were reviewed that reveal a gradation of both shallow and deep knowledge, such as the variation of shallow within-category discrimination (Carney, Widin & Viemeister, 1977), the deepening effect of a longer ISI in the shallow AX task (Werker & Tees, 1984b), and the deep gating task which allowed the bilinguals to direct attention towards phonetic detail of L2 contrast minimal pairs (Sebastián-Gallés & Soto-Faraco, 1999). These tasks fit nicely on a shallow/deep continuum that would be the shallowest at one end and the deepest at the other with a continuous gradient in between.

It is alternatively possible that shallow and deep may turn out to be separate (discrete) constructs. However, considering the variation that has been shown within each hypothetically discrete category ("deeper" shallow tasks and "shallower" deep tasks), this would still represent, in essence, a shallow/deep continuum with a gap in the middle. In order for any firm conclusion to be drawn, work must be conducted to explicitly compare these varyingly shallow and deep tasks with comparable stimuli to determine whether they have a continuously gradient impact upon L2 contrast processing.

The two options of continuous and discrete support different hypotheses for the mechanism behind shallow and deep. A continuous distinction would more likely result from a single process, while two discrete constructs would imply separate processes underlying each. As the continuous hypothesis is endorsed here, one feasible mechanism that could explain how a single process could yield these different results is outlined below.

Simultaneity mechanism.

A theory that could help explain how a single process explanation could result in functional differences between deep and shallow is the Simultaneity Mechanism (SM), which was conceived to explain age-related differences in cognition (Salthouse, 1996). The slowing of cognitive function is thought to be a major contributor to age-related differences in certain tasks (e.g., Salthouse, 1980, 1985, 1992, 1994). The SM is relevant here in that bilingual status is associated with slower processing of the L2 (e.g., the slower overall RTs found here in the AX and repetition priming task; Sebastián-Gallés & Soto-Faraco, 1999). SM hypothesizes that it is this slower processing that leads to performance differences. It is concerned with tasks that are reliant upon multiple processes that individually slow, potentially allowing the degradation of the earlier processing results by the time later processing is complete. In bilinguals, this would apply to deep contexts where the greater processing load undoubtedly draws upon a greater number of processes than shallow contexts. For example, SM could explain the distinction between shallow and deep found here as the loss of fine acoustic-phonetic detail by the time the word is retrieved from the lexicon, a step that is unnecessary in shallow contexts.

Change Over Time

A theory of shallow and deep should be able to explain perceptual change over time, as numerous studies have detailed the impact of training upon non-native contrast discrimination. After repeated exposure to non-native contrasts, often with feedback, individuals typically display much improved discrimination, in both shallow (Tees & Werker, 1984) and deep contexts (Logan, Lively & Pisoni, 1991; Pisoni, Aslin, Perey & Hennessy, 1982). Additionally, shallow results here suggest that amount of English usage in everyday life affects perceptual abilities. However, the mechanism of this change is unknown.

One possibility is that of statistical learning. As shown by Maye, Werker and Gerken (2002), the frequency distribution of phones in the input speech stream impacts infants' categorical perception. Use of statistics has similarly been shown in and to affect the processing of adults (Maye, 2000; Maye & Gerken, 2000), and could potentially be a mechanism aiding in the acquisition of categories by L2 learners. What is unknown is whether statistical learning could influence acquisition of both shallow and deep, or only shallow.

Another possibility is that training could affect perception by shifting the cue-weighting of the L2 listeners. Cue-weighting is thought to factor into developmental shifts in listening strategies by children learning their native language (e.g., Nittrouer, 2002). For example, English adults are known to distinguish /s/ from /ʃ/ based largely upon the noise spectrum (e.g., Harris, 1958), but young children have been shown to base their discrimination of these phonemes more on the basis of formant transitions. A developmental shift has been documented revealing that the extent to which children rely upon the cue of the noise spectrum increases between the ages of 3.5 and 7.5 years (Nittrouer, 1992; Nittrouer & Miller, 1997a, 1997b; Nittrouer & Studdert-Kennedy, 1987), and it is suggested here that such a shift may also occur when training L2 speakers. The greater abilities of L2 discrimination following training could potentially be due to a shift in cue-weighting that allows for more accurate discrimination. Again, the question would remain of whether cue-weighting could affect both shallow and deep, or only shallow.

Limitations and Future Research

In the presently discussed work, the stimuli were chosen to allow for as much comparison as possible between the shallow and deep tasks. However, the tasks themselves could be more directly comparable. The inference as to whether or not the L2-specific contrasts are perceived by the bilinguals is drawn differently between tasks. The deep task measure is wholly implicit, using differential RTs, whereas the fundamental measure of the shallow task tests is explicit (accuracy rates). There is an additional implicit shallow measure of RTs, but it speaks more towards the speed of successful discrimination decisions than towards actual success of discrimination. However, no shallow/deep distinction can be concluded if one compares only the implicit RT results from the two tasks as they lead to the same conclusion (impaired perception of the English-specific contrasts). In future work, a deep task that allows for an explicit measure is desirable. Having noted that, it is extremely unlikely that an explicit deep task measure would return results divergent from those found in the repetition priming task. It is highly doubtful that a bilingual subject would explicitly discriminate, for example, *bad* and *bed*, perceiving them as two separate words, yet show repetition priming in response to them.

The theoretical issue considered here to be the most important for future research is the issue of whether or not shallow and deep are continuous or discrete constructs. The search for underlying processes is also of great interest, but will be simplified if it is known beforehand whether one or two processes is sought.

Conclusions

It appears that there is a distinction between shallow and deep bilingual contrast knowledge. This distinction is potentially very useful as it may help to explain why, in some testing situations, bilinguals are able to perceive L2-specific differences, while they are unsuccessful in others.

References

Anderson, J. L., Morgan, J. L. & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46(2-3), 155-182.

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

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies & S. de Schonen et al. (Eds.). *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). New York: Kluwer.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171-206). Baltimore: York Press.

Best, C. T., McRoberts, G. W., LaFleur, R. & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, 18, 339-350.

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360.

Bialystok, E. & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161-181). Mahwah, NJ:Erlbaum.

Birdsong, D. & Flege, J. E. (2000). Regular-irregular dissociations in the acquisition of

English as a second language. *Paper presented at the 25th Boston University Conference on Language Development*, Boston, MA.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B. & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977-985.

Carney, A. E., Widin, G. P. & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62(4), 961-970.

Chan, Alice Y. W. & Li, David C. S. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum*, 13(1), 67-85.

Crosson, B. (1992). *Subcortical functions in language and memory*. New York: Guilford Press.

Crowder, R. G. (1982). A common basis for auditory sensory storage in perception and immediate memory. *Perception & Psychophysics*, 29, 26-36.

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P-F., Lehericy, S. & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*, 8, 3809-3815.

Desrochers, A. (2003). Fluency assessment questionnaire for English-French bilinguals. *Unpublished instrument, Cognitive Psychology Laboratory*, *University of Ottawa*, Ottawa, Canada.

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

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89: 395-411.

Flege, J. E., Takagi, N. & Mann, V. (1995a). Japanese adults can learn to produce /r/ and /l/ accurately. *Language and Speech* 38, Part I. *In press.*

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99(2), 1161-1173.

Flege, J. E., Takagi, N., & Mann, V. (1995b). Experienced Japanese speakers of English can identify English [r] accurately, but not [l]. *Journal of the Acoustical Society of America.*

Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116-124.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110-125.

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

Hakuta, K., Bialystok, E. & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, 14(1), 31-38.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1-7.

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

Johnson, J. S. & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, 39, 215-258.

Ju, M. & Luce, P. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15(5), 314-318.

Jusczyk, P. W. & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1-23.

Klein, D., Zatorre, R. J., Milner, B., Meyer, E. & Evans, A. C. (1994). Left putaminal activation when speaking a second language: evidence from PET. *NeuroReport*, 5, 2295-2297.

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

Kuhl, P. & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 121-154). Baltimore: York Press.

Kuhl, P. K., Tsao, F. M. & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100, 9096-0101.

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

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 215-225.

Lenneberg, E. H. (1967). *Biological foundations of Language.* New York: Wiley.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences, Prague, 1967* (pp. 563-567). Prague: Academia.

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

MacKain, K. S., Best, C. T. & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 368-390.

MacKain, K. S. & Stern, D. N. (1982). The concept of experience in speech development. In K. E. Nelson (Ed.), *Children's Language*. New York: Gardiner.

Maye, J. (2000). Learning speech sound categories on the basis of distributional information. *Unpublished doctoral dissertation, University of Arizona*, Tucson.

Maye, J. & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish & T. Keith-Lucas, *Proceedings of the 24th Boston University Conference on Language Development* (pp. 522-533). Somerville, MA: Cascadilla Press.

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

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

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 1-32.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, 112(2), 711-719.

Nittrouer, S. & Miller, M. E. (1997a). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America*, 102, 572-580.

Nittrouer, S. & Miller, M. E. (1997b). Perdicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, 101, 2253-2266.

Nittrouer, S. & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, 30, 319-329.

Oh, J. S., Jun, S.-A., Knightly, L. M. & Au, T. K.-f. (2003). Holding on to childhood language memory. *Cognition*, 86, B53-B64.

Pallier, C., Bosch, L. & Sebastián-Gallés, N. (1997). A limit on behavioural plasticity in speech perception. *Cognition*, 64, B9 – B17.

Pallier, C., Colomé, A. & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445-448.

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M. Dupoux, E. & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex*, 13, 155-161.

Penfield, W. & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., Fazio, F. & Mehler,

J. (1996). Brain processing of native and foreign languages. *NeuroReport*, 7, 2439-2444.

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

Polka, L. & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221-231.

Polka, L., Colantonio, C. & Sundara, M. (2001). A cross-language comparison of /d/-/D/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, 109(5), 2190-2201.

Remez, R., Rubin, P. E., Katz, M. & Dodelson, S. (1985). On the influence of lexical status in phonetic perception. In *Twenty-sixth annual meeting of the Psychonomic Society, pp. 29-30. Boston, MA*: Psychonomic Society.

Rubin, P., Turvey, M. T. & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19, 394-398.

Salthouse, T. A. (1980). Age and memory: Strategies for localizing the loss. In L. W. Poon, J. L. Fozard, L. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New directions in memory and aging*. (pp. 47-65). Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam: North-Holland.

Salthouse, T. A. (1992). *Mechanisms of age-cognition relations in adulthood.* Hillsdale, NJ: Erlbaum.

Salthouse, T. A. (1994). The nature of the influence of speed on adult age differences in cognition. *Developmental Psychology*, 30, 240-259.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition.

Psychological Review, 103(3), 403- 428.

Sebastián-Gallés, N. & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72, 111-123.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21: 427-444.

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

Singh, S. & Black, J. W. (1977). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America*, 39, 371-387.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition*, 14, 230-237.

Snow, C. E. & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49, 1114-1128.

Stager, C. L. & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, 388, 381-382.

Streeter, L.A. (1976). Language perception of two-month old infants shows effects of both innate mechanisms and experience. *Nature*, 250, 39-41.

Takagi, N. (1993). Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view. Unpub*lished Ph.D dissertation, University of California – Irvine.*

Tees, R. C. & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, 38(4), 579-590.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 47, 466-472.

Tsao, F., Liu, H., Kuhl, P. K. & Tseng, C. (2000, July). Perceptual discrimination of a Mandarin fricative-affricate contrast by English-learning and Mandarin-learning infants. *Poster presented at the meeting of the International Society on Infant Studies*, Brighton, U.K.

Tsushima, T., Takizawa, O., Sasaki, M, Shiraki, S., Nishi, K., Kohno, M., Menyuk, P. & Best, C. (1996). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *The Emergence of Human Cognition and Language*, 3, 57-61.

Ventureyra, V., Pallier, C. & Yoo, H. (2004). The loss of first language phonetic perception in adopted Koreans. Journal of Neurolinguistics, 17, 79-91.

Walley, A. C., Pisoni, D. B. & Aslin, R. N. (1981). The role of early experience in the development of speech perception. In: R. N. Aslin, J. R. Alberts & M. R. Petersen (Eds.), *Development of Perception: Psychobiological Perspectives* (pp. 219 – 255). New York: Academic Press.

Werker, J. F. & Curtin, S. (under revision). PRIMIR: A developmental framework of infant speech processing.

Werker, J. F., Frost, P. E. & McGurk, H. La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, 46(4), 551-568.

Werker, J. F., Gilbert, J. H. V., Humphrey, K. & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-353.

Werker, J. F. & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24(5), 672-683.

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

Werker, J. F., Marcus, G. F., Mehler, J., Neville, H. J., & Sebastián-Gallés, N. (2001). Flexibility and specificity in language systems of the brain. *Human Frontiers Science Project Grant.*

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

Whalen, D. H., Best, C. T. & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25, 501-528.

Table 1

Deep Task Stimuli by Contrast

English-Specific						Common					
/æ/ - /ɛ/		/s/ - /θ/		/v/ - /w/		/f/ - /v/		/m/ - /n/		/ɔ/ - /u/	
bad	bed	sank	thank	veil	wail	fairy	very	might	night	bought	boot
bat	bet	saw	thaw	vein	wane	fast	vast	mob	knob	flaw	flew
blast	blessed	seem	theme	vent	went	fat	vat	mock	knock	prawn	prune
last	lest	sick	thick	vest	west	fault	vault	moan	known	spawn	spoon
lather	leather	sin	thin	vet	wet	fear	veer	mail	nail	lop	loop
laughed	left	song	thong	vie	why	fender	vendor	mice	nice	lot	loot
mat	met	sum	thumb	vine	wine	few	view	mine	nine	dawdle	doodle
past	pest	symbol	thimble	vow	wow	file	vile	me	knee	drawl	drool

Table 2

Deep Task Stimuli by Ordering

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
1	milk					
2	fine					
3		bad	bed	bad	bed	1
4	now					
5		saw	thaw	saw	thaw	2
6	sland					
7		vab	wab	vab	wab	3
8	great					
9		thong	song	thong	song	4
10	flane					
11		bet	bat	bet	bat	5
12	marching					
13		bed	bad	bad	bed	1
14	face					
15		wab	vab	vab	wab	3
16	crive					
17		thaw	saw	saw	thaw	2
18	nush					
19		laughed	left	laughed	left	6
20	third					
21		bat	bet	bet	bat	5
22	route					
23		war	var	war	var	7
24	nake					
25		song	thong	thong	song	4
26	number					
27		sin	thin	sin	thin	8
28	rop					
29		veeg	weeg	veeg	weeg	9
30	thousand					
31		laughed	left	left	laughed	6
32	shrabes					
33		theme	seem	theme	seem	10
34	maybe					
35		var	war	war	var	7
36	white					
37		lest	last	lest	last	11
38	faded					
39		vast	fast	vast	fast	12
40	twade					
41		sin	thin	thin	sin	8
42	eept					
43		veeg	weeg	weeg	veeg	9
44	marine					

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
45		very	fairy	very	fairy	13
46	slont					
47		theme	seem	seem	theme	10
48	glave					
49		loot	lot	loot	lot	14
50	never					
51		lest	last	last	lest	11
52	woone					
53		fast	vast	vast	fast	12
54	gwot					
55		vile	file	vile	file	15
56	sad					
57		lot	loot	loot	lot	14
58	hang					
59		knob	mob	knob	mob	16
60	falk					
61		fat	vat	vat	fat	17
62	kinose					
63		fairy	very	very	fairy	13
64	thousand					
65		weej	veej	weej	veej	18
66	moust					
67		blast	blessed	blast	blessed	19
68	nod					
69		mob	knob	knob	mob	16
70	gia					
71		boot	bought	boot	bought	20
72	near					
73		vile	file	file	vile	15
74	fiddle					
75		blessed	blast	blast	blessed	19
76	festival					
77		vat	fat	vat	fat	17
78	milk					
79		pab	peb	pab	peb	21
80	fiddle					
81		weej	veej	veej	weej	18
82	gorbs					
83		mock	knock	mock	knock	22
84	kruntz					
85		bought	boot	boot	bought	20
86	fire					
87		dawdle	doodle	dawdle	doodle	23
88	fixed					
89		wane	vain	wane	vain	24
90	troot					
91		might	night	might	night	25

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
92	gwals					
93		peb	pab	pab	peb	21
94	lurps					
95		symbol	thimble	symbol	thimble	26
96	vaps					
97		knock	mock	mock	knock	22
98	feast					
99		doodle	dawdle	dawdle	doodle	23
100	tarb					
101		moan	known	moan	known	27
102	mark					
103		wane	vain	vain	wane	24
104	marry					
105		yend	yand	yend	yand	28
106	loose					
107		night	might	might	night	25
108	third					
109		thimble	symbol	symbol	thimble	26
110	wide					
111		moan	known	known	moan	27
112	throg					
113		yend	yand	yand	yend	28
114	breep					
115		nine	mine	nine	mine	29
116	sound					
117		veck	weck	veck	weck	30
118	white					
119		vest	west	vest	west	31
120	plail					
121		thank	sank	thank	sank	32
122	zoo					
123		thoat	soat	thoat	soat	33
124	flat					
125		mine	nine	nine	mine	29
126	thurks					
127		vendor	fender	vendor	fender	34
128	woone					
129		lop	loop	lop	loop	35
130	copper					
131		selm	thelm	selm	thelm	36
132	wing					
133		weck	veck	veck	weck	30
134	zilked					
135		thank	sank	sank	thank	32
136	shoot					
137		west	vest	vest	west	31
138	krum					

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
139		soat	thoat	thoat	soat	33
140	open					
141		vendor	fender	fender	vendor	34
142	tworsh					
143		yag	yeg	yag	yeg	37
144	glamed					
145		pest	past	pest	past	38
146	native					
147		loop	lop	lop	loop	35
148	krolt					
149		thelm	selm	selm	thelm	36
150	needy					
151		sirp	thirp	sirp	thirp	39
152	krum					
153		vent	went	vent	went	40
154	bruise					
155		pest	past	past	pest	38
156	troot					
157		yag	yeg	yeg	yag	37
158	mask					
159		thwos	swos	thwos	swos	41
160	nurk					
161		sirp	thirp	thirp	sirp	39
162	dirp					
163		leather	lather	leather	lather	42
164	nake					
165		vent	went	went	vent	40
166	thurks					
167		thick	sick	thick	sick	43
168	gossip					
169		prawn	prune	prawn	prune	44
170	nush					
171		wet	vet	wet	vet	45
172	blaned					
173		thwos	swos	swos	thwos	41
174	bottle					
175		lather	leather	leather	lather	42
176	drine					
177		prawn	prune	prune	prawn	44
178	narged					
179		sick	thick	thick	sick	43
180	marching					
181		shap	shep	shap	shep	46
182	marry					
183		vine	wine	vine	wine	47
184	grunkt					
185		wopy	vopy	wopy	vopy	48

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
186	near					
187		mail	nail	mail	nail	49
188	weep					
189		vet	wet	wet	vet	45
190	soup					
191		supy	thupy	supy	thupy	50
192	shone					
193		wine	vine	vine	wine	47
194	moust					
195		shep	shap	shap	shep	46
196	theory					
197		drawl	drool	drool	drawl	51
198	flag					
199		vopy	wopy	wopy	vopy	48
200	sailor					
201		smez	smaz	smez	smaz	52
202	wire					
203		mail	nail	nail	mail	49
204	theory					
205		drawl	drool	drawl	drool	51
206	hot					
207		sum	thumb	sum	thumb	53
208	gicked					
209		thupy	supy	supy	thupy	50
210	loose					
211		smaz	smez	smez	smaz	52
212	bruise					
213		why	vie	why	vie	54
214	fault					
215		sum	thumb	thumb	sum	53
216	shoot					
217		thut	sut	thut	sut	55
218	vouge					
219		thabes	sabes	thabes	sabes	56
220	choaf					
221		fear	veer	fear	veer	57
222	deece					
223		vie	why	why	vie	54
224	dern					
225		wimel	vimel	wimel	vimel	58
226	marine					
227		fault	vault	fault	vault	59
228	foad					
229		nice	mice	nice	mice	60
230	rock					
231		thabes	sabes	sabes	thabes	56
232	flaky					

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
233		fear	veer	veer	fear	57
234	poy					
235		sut	thut	thut	sut	55
236	brulk					
237		fault	vault	vault	fault	59
238	nose					
239		wimel	vimel	vimel	wimel	58
240	flane					
241		spoon	spawn	spoon	spawn	61
242	marble					
243		flew	flaw	flew	flaw	62
244	hades					
245		nice	mice	mice	nice	60
246	nurk					
247		grack	greck	grack	greck	63
248	maybe					
249		veil	wail	veil	wail	64
250	falk					
251		srun	thrun	srun	thrun	65
252	skoal					
253		mat	met	mat	met	66
254	nanched					
255		virp	wirp	virp	wirp	67
256	marble					
257		spoon	spawn	spawn	spoon	61
258	vaps					
259		flew	flaw	flaw	flew	62
260	woose					
261		grack	greck	greck	grack	63
262	vouge					
263		veil	wail	wail	veil	64
264	gnith					
265		tegy	tagy	tegy	tagy	68
266	wums					
267		virp	wirp	wirp	virp	67
268	soup					
269		mat	met	met	mat	66
270	network					
271		srun	thrun	thrun	srun	65
272	mark					
273		knee	me	knee	me	69
274	kiss					
275		neff	naff	neff	naff	70
276	mark					
277		wow	vow	wow	vow	71
278	wums					
279		few	view	few	view	72

	Fillers - all orders	Order 1	Order 2	Order 3	Order 4	Pair
280	woose					
281		tagy	tegy	tegy	tagy	68
282	theory					
283		neff	naff	naff	neff	70
284	treeb					
285		knee	me	me	knee	69
286	thurks					
287		wow	vow	vow	wow	71
288	zicks					
289		view	few	few	view	72
290	third					

Table 3

Shallow Task Stimuli

	English-Specific						Common					
	/æ/ - /ɛ/		/s/ - /θ/		/v/ - /w/		/f/ - /v/		/n/ - /m/		/ɔ/ - /u/	
I	ba_d	ba_d	si_ck	si_ck	ve_st	ve_st	fa_st	fa_st	kno_b	kno_b	lo_p	lo_p
I	ba_t	ba_t	si_n	si_n	ve_t	ve_t	fa_t	fa_t	kno_ck	kno_ck	lo_t	lo_t
I	be_d	be_d	thi_ck	thi_ck	we_st	we_st	va_st	va_st	mo_b	mo_b	loo_p	loo_p
I	be_t	be_t	thi_n	thi_n	we_t	we_t	va_t	va_t	mo_ck	mo_ck	loo_t	loo_t
S	ba_d	ba_t	si_ck	si_n	ve_st	ve_t	fa_st	fa_t	kno_b	kno_ck	lo_t	lo_t
S	ba_t	ba_d	si_n	si_ck	ve_t	ve_st	fa_t	fa_st	kno_ck	kno_b	lo_p	lo_p
S	be_d	be_t	thi_ck	thi_n	we_st	we_t	va_st	va_t	mo_b	mo_ck	loo_t	loo_t
S	be_t	be_d	thi_n	thi_ck	we_t	we_st	va_t	va_st	mo_ck	mo_b	loo_p	loo_p
D	ba_d	be_d	si_ck	thi_ck	ve_st	we_st	fa_st	va_st	kno_b	mo_b	loo_p	loo_p
D	ba_d	be_t	si_ck	thi_n	ve_st	we_t	fa_st	va_t	kno_b	mo_ck	loo_t	loo_t
D	ba_t	be_d	si_n	thi_ck	ve_t	we_st	fa_t	va_st	kno_ck	mo_b	loo_p	loo_p
D	ba_t	be_t	si_n	thi_n	ve_t	we_t	fa_t	va_t	kno_ck	mo_ck	loo_t	loo_t
D	be_d	ba_d	thi_ck	si_ck	we_st	ve_st	va_st	fa_st	mo_b	kno_b	lo_p	lo_p
D	be_t	ba_d	thi_n	si_ck	we_t	ve_st	va_t	fa_st	mo_ck	kno_b	lo_p	lo_p
D	be_d	ba_t	thi_ck	si_n	we_st	ve_t	va_st	fa_t	mo_b	kno_ck	lo_t	lo_t
D	be_t	ba_t	thi_n	si_n	we_t	ve_t	va_t	fa_t	mo_ck	kno_ck	lo_t	lo_t

I = Identical pairs, S = Same category pairs, D = Different category pairs.
ba_d denotes a *ba* token that has been excised from the deep task token *bad.*

Figure Captions

Figure 1. Repetition priming: Difference reaction times in response to repeats and common contrast minimal pairs.

Figure 2. Repetition priming: Difference reaction times in response to repeats and English-specific contrast minimal pairs.

Figure 3. AX (Same-different): Accuracy rates to common contrast and English-specific contrast syllable pairs.

Figure 4. AX (Same-different): Reaction times to common contrast and English-specific contrast syllable pairs.

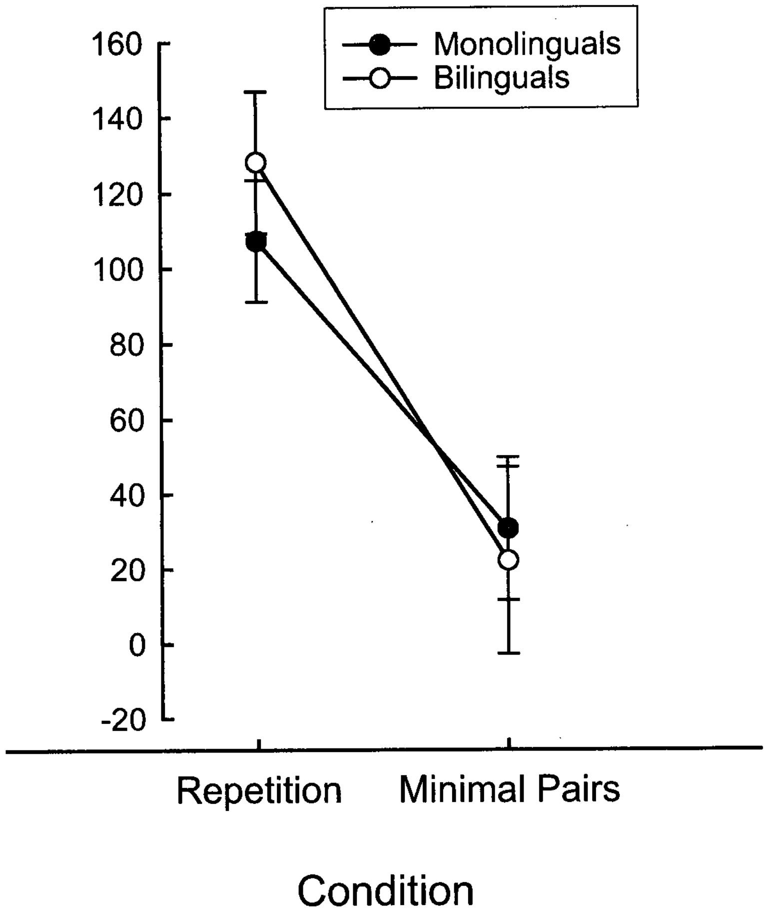

Figure 1. Repetition priming: Difference reaction times in response to repeats and common contrast minimal pairs.

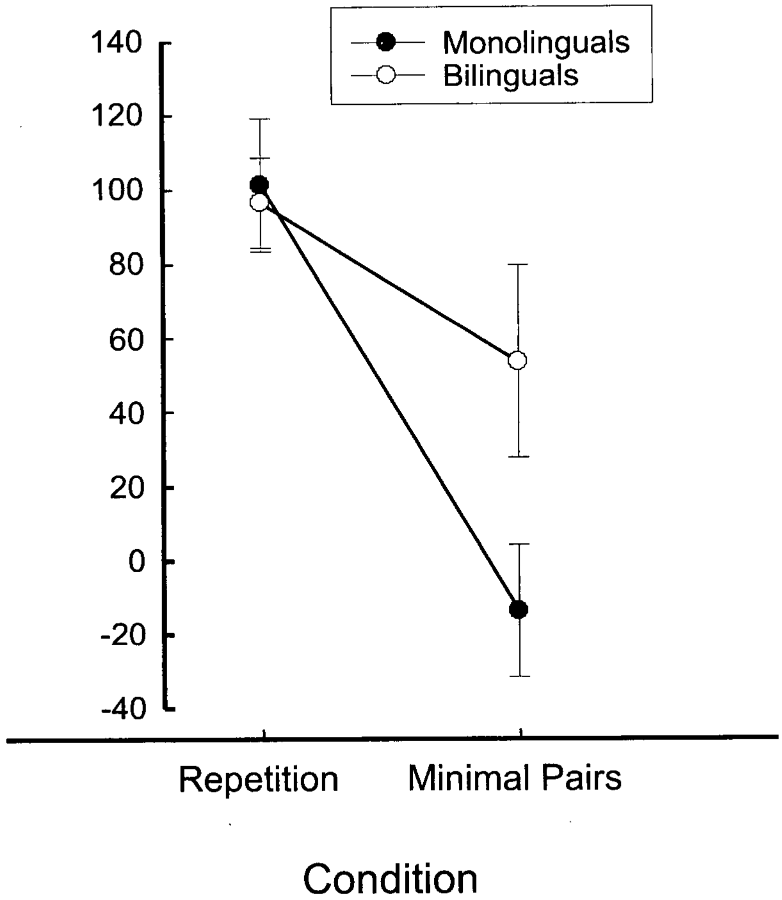

Figure 2. Repetition priming: Difference reaction times in response to repeats and English-specific contrast minimal pairs.

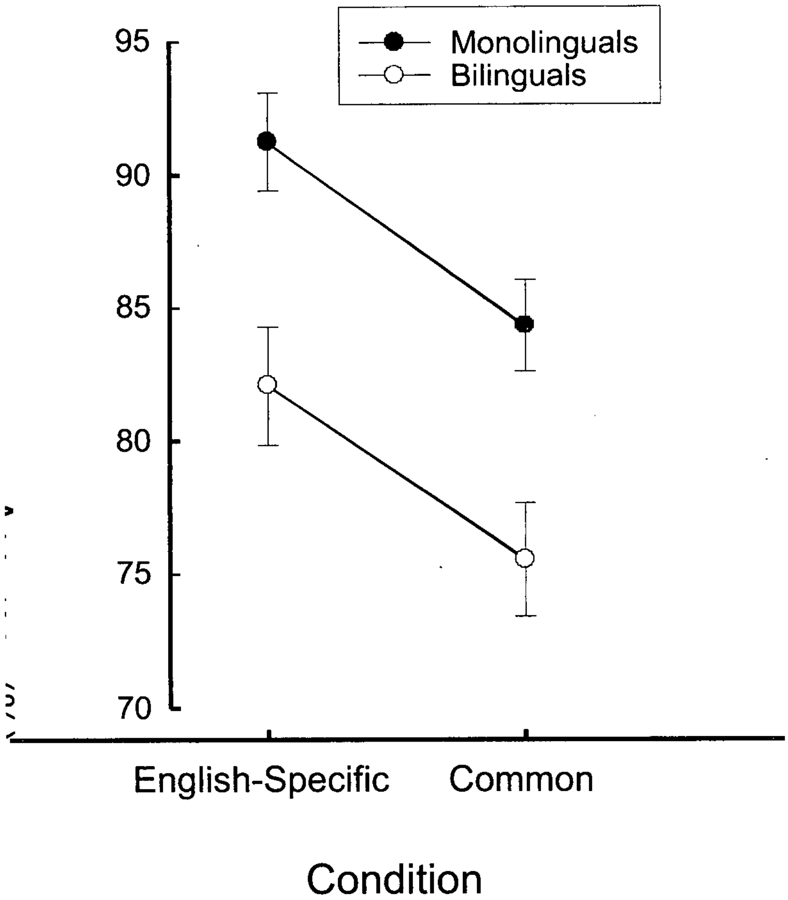

Figure 3. AX (Same-different): Accuracy rates to common contrast and English-specific contrast syllable pairs.

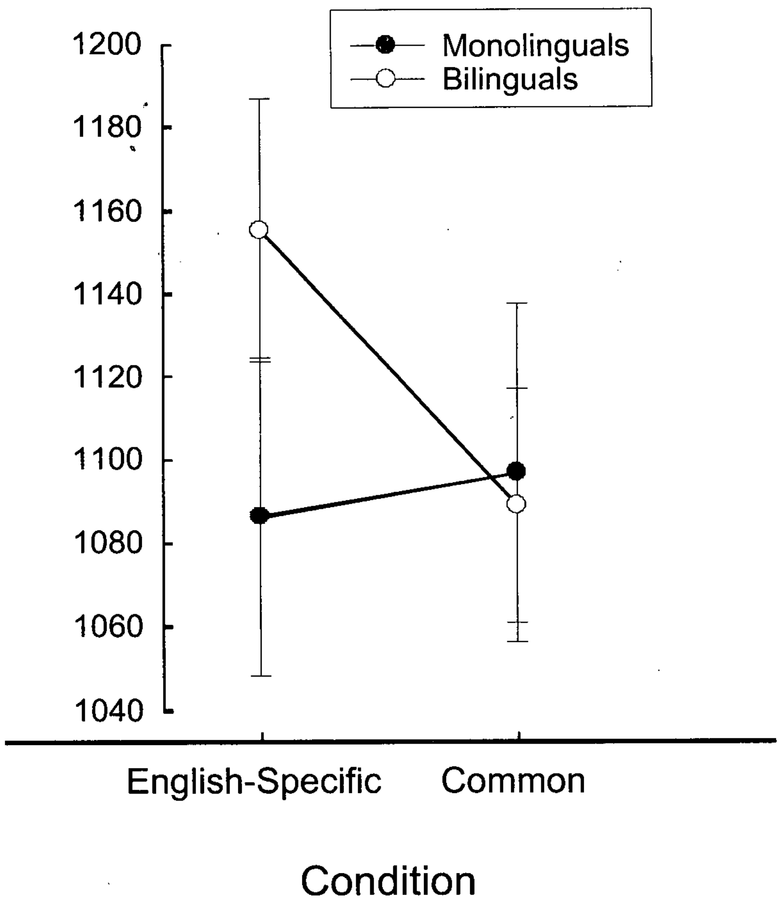

Figure 4. AX (Same-different): Reaction times to common contrast and English-specific contrast syllable pairs.