Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners[a)]

Satomi Imai[b)]
Division of Speech and Hearing Sciences, University of Alabama at Birmingham, 1530 Third Avenue South, Birmingham, Alabama, 35294

Amanda C. Walley
Department of Psychology, University of Alabama at Birmingham, 1530 Third Avenue South, Birmingham, Alabama, 35294

James E. Flege
Division of Speech and Hearing Sciences, University of Alabama at Birmingham, 1530 Third Avenue South, Birmingham, Alabama, 35294

(Received 5 May 2004; revised 28 September 2004; accepted 5 October 2004)

This study examined the effect of presumed mismatches between speech input and the phonological representations of English words by native speakers of English (NE) and Spanish (NS). The English test words, which were produced by a NE speaker and a NS speaker, varied orthogonally in lexical frequency and neighborhood density and were presented to NE listeners and to NS listeners who differed in English pronunciation proficiency. It was hypothesized that mismatches between phonological representations and speech input would impair word recognition, especially for items from dense lexical neighborhoods which are phonologically similar to many other words and require finer sound discrimination. Further, it was assumed that L2 phonological representations would change with L2 proficiency. The results showed the expected mismatch effect only for words from dense neighborhoods. For Spanish-accented stimuli, the NS groups recognized more words from dense neighborhoods than the NE group did. For native-produced stimuli, the low-proficiency NS group recognized fewer words than the other two groups. The-high proficiency NS participants' performance was as good as the NE group's for words from sparse neighborhoods, but not for words from dense neighborhoods. These results are discussed in relation to the development of phonological representations of L2 words. (200 words).
[DOI: 10.1121/1.1823291]

PACS numbers: 43.71.Es, 43.71.Gv [RLD] Pages: 896–907

I. INTRODUCTION

Speech perception becomes attuned to native or first language (L1) sounds during infancy and childhood (e.g., Best, 1995; Jusczyk, 1993; Werker and Polka, 1993). Because of differences in the phonetic inventories between any two languages, individuals who learn a second language (L2) in adulthood may have L2 speech representations that are affected by their L1 phonology, and so differ from those of native speakers (cf. Flege, 1995; Kuhl and Iverson, 1995). A failure to reattune speech representations to specific features of the L2 explains, in part, why L2 learners often display poor speech recognition performance, especially in nonideal listening conditions (e.g., Bradlow and Bent, 2002; Meador, Flege, and MacKay, 2000). However, the word recognition difficulty that L2 listeners sometimes experience may be the result of insufficient higher-level knowledge of the L2, including weaker lexical constraints, in addition to differences in bottom-up processing. The overall purpose of the present study was to provide a better understanding of the interface between the phonological and lexical levels in L2 word recognition.

To date, most L2 speech perception research has focused on demonstrating differences in the processing of phonemic and/or phonetic information by L2 listeners (for review, see Strange, 1995). At the phonemic level, it has been well documented that L2 listeners do not perceive certain sounds in the same way as L1 listeners in categorical identification and discrimination tasks (e.g., Best, McRoberts, and Goodell, 2001; Flege, Munro, and Fox, 1994). According to Best (1995), foreign language (or L2) phonemes that can be assimilated to a single L1 phonemic category present the greatest discrimination difficulty (e.g., native Spanish speakers categorize both English /i/ and /ɪ/ as instances of Spanish /i/), a claim that is generally supported across different languages. More recently, Pallier, Colome, and Sebastian-Galles (2001) showed that in a repetition priming task, some Spanish-dominant bilinguals of Catalan and Spanish processed nonwords containing phonemes that were contrastive

[a)]Portions of this study were presented at the 144th Meeting of the Acoustical Society of America, Nashville, TN, May 2003, and in ''Spoken word recognition of accented and native speech: Lexical factors affecting native and non-native listeners,'' Proceedings from the 15th International Congress of Phonetics Sciences, Barcelona, Spain, August 2003.
[b)]Electronic mail: imais@uab.edu

only in Catalan as if they were Spanish phonemes, without distinguishing the two Catalan phonemes.

At the phonetic level, two languages may exhibit subtle variations, and previous research has demonstrated that adult L2 learners are often not sensitive to the relevant differences. For example, word-initial /p/, /t/, and /k/ are produced with more aspiration in English than Italian, and word-final stops are less likely to be released in English. Thus, native Italian speakers who have learned English as an L2 have difficulty identifying English stop consonants in noise (MacKay, Meador, and Flege, 2001). In a gating experiment using nonwords, Sebastian-Galles and Soto-Faraco (1999) showed that Spanish-dominant bilinguals required more phonetic information to identify Catalan-specific phonemes than Catalan-dominant bilinguals did.

One limitation of previous L2 speech perception research is that we currently know little about how differences in phonemic and phonetic perception between native and non-native listeners impact higher or later levels of speech processing—specifically, word recognition. Several recent studies have examined L2 word recognition in sentence context and found poorer performance by non-native listeners as compared to native listeners (Bradlow and Bent, 2002; Mayo, Florentine, and Buus, 1997; Meador *et al.*, 2000; van Wijngaarden, 2001; van Wijngaarden, Steeneken, and Hougast, 2002). However, in these studies, characteristics of the stimulus words to be identified (e.g., word frequency) and the carrier sentences have not been controlled or systematically manipulated. Thus, the extent to which non-native listeners' poorer performance is due to inaccurate segmental perception or insufficient semantic/syntactic knowledge is unclear.

In addition to the lack of empirical evidence regarding the interface between the phonological and lexical levels in L2 perception, there has been a paucity of theoretical attention directed toward this issue. In the L1 acquisition literature, it has been suggested that young children first establish holistic (e.g., syllable-based) representations of words, and only gradually develop phonological representations that include more detailed, segmental information (e.g., Ferguson and Farwell, 1975; Metsala and Walley, 1998). A major impetus for this developmental shift is vocabulary growth. On this view, an increasing number of items overlap with one another as the lexicon expands, and more fine-grained representations are needed for fast, accurate word recognition (see also Charles-Luce and Luce, 1990). Perhaps with greater L2 exposure and additional word learning, L2 learners' lexical representations may also become more fine-grained or fully specified *vis-à-vis* their new language.

There is some evidence relating L2 learners' segmental perception and word recognition. In Meador *et al.*'s (2000) study, early and late Italian-English bilinguals were asked to repeat semantically unpredictable English sentences that were presented in different levels of noise. The number of correctly repeated words was analyzed and compared to identification scores for vowels and consonants (assessed in other experiments). Regression analyses showed that accuracy scores for segment identification accounted for 17% of the variance in the number of words correctly repeated after demographic variables (e.g., age of arrival, use of the L1, length of residence in Canada) were partialled out. Although these results are correlational, they suggest that accurate perception of L2 vowels and consonants contributes to differences in word recognition among bilinguals.

A study by Bradlow and Pisoni (1999) provided a fuller picture of the lexical factors that influence word recognition. These researchers presented "easy" and "hard" words to native and non-native English speakers of varying L1 backgrounds. The easy words had a high frequency of occurrence and were from sparse neighborhoods (i.e., were phonologically similar to few other words); the hard words had a low frequency of occurrence and were from dense neighborhoods (i.e., were confusable with many other words). In addition, the neighbors of the easy words were, on average, of low frequency, and those of the hard words were of high frequency. The results showed that both listener groups recognized fewer hard than easy words, primarily because words from dense neighborhoods are confusable with one another (e.g., Luce and Pisoni, 1998). However, the difference between the hard and easy words was larger for the non-native group (cf. Takayanagi, Dirks, and Moshfegh, 2002). Bradlow and Pisoni therefore concluded that non-native listeners may have reduced sensitivity to some of the phonetic details that are necessary for lexical discrimination.

The present study extended the Bradlow and Pisoni (1999) study in several ways. First, we compared the performance of native English (NE) speakers and two groups of native Spanish (NS) speakers who differed in pronunciation proficiency (as defined by overall degree of foreign accent). Second, the participants of this study were tested on English words that were spoken by both a NE and a NS speaker. The two NS groups and two types of speech input (native-produced and foreign-accented stimuli) were included to assess how the representation/processing of lexical items might vary at different stages of L2 acquisition.

We assumed that our NS participants' phonological representations would not be optimally matched to English speech input, especially those of the less proficient NS participants ("the phonological mismatch hypothesis"). We expected NS participants to be better able to recognize English words spoken with a Spanish accent than words spoken by a native English speaker. That is, a better match should exist between the NS participants' phonological representations and the acoustic phonetic specification of English words as spoken by other NS speakers, in part because they often hear such renditions and because the NS participants themselves pronounce English words in a similar manner. We assumed that L2 phonological representations are initially affected by the L1 sound system, but that as L2 learning proceeds and L2 pronunciation proficiency increases, these representations become more attuned to the L2 and thus more similar to those of native speakers. Any advantage for Spanish-accented stimuli should therefore be smaller for more proficient NS participants than less proficient ones. If such a phonological mismatch effect were found, L2 learners' phonological representations could be understood as being organized differently than those of native speakers.

There is, in fact, some evidence that L2 listeners benefit from hearing L2 speech produced with L1 accent. In a study by Bent and Bradlow (2003), NE and non-native listeners of different L1 backgrounds listened to English sentences (all embedded in white noise) that were spoken by a NE talker and by proficient and nonproficient Chinese and Korean talkers. The researchers assumed that L2 learners develop a shared knowledge base for the L2 sound system which helps them understand the L2 speech of other talkers of the same language background (the ''matched interlanguage speech intelligibility benefit'') and even of different language backgrounds (the ''mismatched interlanguage speech intelligibility benefit''). In support, it was found that NE listeners recognized more words in the sentences spoken by the NE talker than by the L2 talkers, whereas Chinese and Korean listeners' performance was as good for sentences spoken by proficient talkers of their L1 as it was for sentences spoken by the NE talker. Further, another group of non-native listeners from various language backgrounds (ten different languages) recognized as many words in the sentences spoken by proficient Chinese and Korean talkers as by the NE talker.

Similar results were obtained by van Wijngaarden (2001), who examined the recognition of both native-produced and English-accented Dutch sentences by native Dutch listeners and by native English listeners. The native English listeners had spoken Dutch as an L2 for an average of 20 years and were highly proficient. Using a speech reception threshold method, it was found that the native Dutch listeners generally tolerated a higher level of noise than the English listeners did, and they tolerated more noise for native-produced than English-accented speech. However, the native English listeners tolerated noise for English-accented speech to the same extent as for the native Dutch speech. Thus, non-native speech was less intelligible to native listeners, but it was as intelligible as native speech to non-native listeners.

In a subsequent study (van Wijngaarden *et al.*, 2002), native Dutch participants were asked to identify words in Dutch, English, and German sentences that were produced by native talkers of each language, as well as by native Dutch talkers (thus including Dutch-accented English and German sentences). Again using a speech reception threshold method, the researchers found that the Dutch listeners, who were more proficient in English than German, tolerated more noise for native-produced as compared to Dutch-accented English sentences, whereas more noise was tolerated for Dutch-accented vs native-produced German sentences.

The results from the three studies just cited are, however, only suggestive with respect to our phonological mismatch hypothesis. Lexical characteristics such as word frequency were not controlled or examined in these studies. In addition, the proficiency of non-native listeners in the L2 varied substantially across studies and/or was not systematically evaluated. Further, sentence recognition is affected by various suprasegmental factors including prosody and speaking rate (e.g., Bradlow, Torretta, and Pisoni, 1996; Cohen, Douaire, and Elsabbagh, 2001; Cutler and van Donselaar, 2001), but the influence of these factors was not considered either. As a result, it is not clear how any advantage in perceiving foreign-accented speech might be related specifically to L2 listeners' segmental representations for spoken words.

According to the phonological mismatch hypothesis evaluated in the present study, the accuracy or efficiency of spoken word recognition should be reduced when listeners' lexical representations do not match input speech stimuli. Thus, our hypothesis bears not only on listeners' perception of L2 words, but also on the perception of foreign-accented speech by native listeners. The NE participants in our study were expected to recognize more English words spoken by another NE speaker than English words spoken with a Spanish accent. Apart from the findings already reviewed, the only other relevant research on foreign-accented speech perception by native listeners has used sentence stimuli to show reduced performance as compared to native-produced speech—a difference that has been attributed to normalization processes (Derwing and Munro, 1997; Munro and Derwing, 1995) and processing cost (Derwing and Munro, 2001; Munro, 1998; Munro and Derwing, 2001). If our hypothesis were upheld, it could offer a further explanation at the level of segmental representations for the difficulty that native speakers have in perceiving non-native speech.

Like Bradlow and Pisoni (1999), we examined the effect of word frequency and neighborhood density on word recognition. However, our study differs from that of Bradlow and Pisoni in a third important respect. In Bradlow and Pisoni's study, lexical factors covaried—i.e., easy words were of high frequency and were from sparse neighborhoods, whereas hard words were of low frequency and were from dense neighborhoods. As the researchers themselves acknowledged, their non-native listeners' recognition of fewer hard than easy words might have reflected the influence of word frequency, apart from any effect of neighborhood density. Indeed, in contrast to native listeners, the non-native listeners also rated hard words as less familiar than easy words. Bradlow and Pisoni used a 7-point familiarity scale, but only a rating of ''7'' (''I know this word'') was taken as clear evidence that a word was known (see Balota, Pilotti, and Cortese, 2001, for discussion). A follow-up analysis examining the subset of lexical items given a rating of ''6'' or ''7'' yielded essentially the same results as the analysis of all items. It is nevertheless possible that the non-native participants actually knew fewer hard than easy words, and that this contributed to their difficulty in recognizing hard words.

To address these limitations, the stimulus set examined in the present study consisted entirely of known English words that varied orthogonally in word frequency and neighborhood density.[1] Our results should therefore help pinpoint the causes of any word recognition difficulty by non-native listeners. We hypothesized that the effects of word frequency and neighborhood density would be greater when there is a mismatch between phonological representations and incoming speech stimuli. NE participants should thus display larger effects of these lexical factors for Spanish-accented speech than for native-produced stimuli, whereas NS participants should show larger effects for native-produced speech than Spanish-accented stimuli. Specifically, the impact of frequency (better recognition for high- than low-frequency words) and neighborhood density (better recognition for words from sparse than dense neighborhoods, or a ''competition effect'') should be greater in the mismatch than the match conditions.

Finally, we also collected subjective frequency estimates for the test words from both NE and NS participants to further explore differences in lexical organization. Specifically, we wanted to examine how strongly subjective and objective word frequency measures might be correlated, and whether or not subjective frequency is related to word recognition performance. Experienced word frequency has typically been operationalized by reference to the number of occurrences of lexical items in written corpora such as that of Kučera and Francis (1967), which is based on written materials. However, several previous studies have indicated that such estimates of word frequency may be confounded with age of acquisition and other dimensions of word familiarity that affect various stages of word recognition (e.g., Connine *et al.*, 1990; Garlock, Walley, and Metsala, 2001; Morrison and Ellis, 1995). In addition, objective frequency counts are subject to sampling biases, including the under-representation of low-frequency words and restricted range effects (Carroll, 1971; Lachman, Shaffer, and Hennrikus, 1974). For these reasons, subjective estimates may actually be better predictors of psychological familiarity, and thus word recognition performance (Gernsbacher, 1984; Gordon, 1985). Past research has also revealed modality-specific effects for subjective frequency estimates, suggesting that people can differentiate between auditory and visual experiences with particular words (Amano, Kondo, and Kakehi, 1995; Balota *et al.* 2001; Gaygen and Luce, 1998). In our study then, we asked participants to provide estimates for the relative frequency with which the test words were spoken and/or heard.

In summary, the present study differed from previous studies in several ways: (i) two lexical factors, word frequency and neighborhood density, were orthogonally varied; (ii) native-produced and foreign, Spanish-accented stimuli were presented to listeners; (iii) our non-native listeners had similar L1 backgrounds, but varied in English pronunciation proficiency; (iv) steps were taken to ensure that all the test words were known to the listeners; and (v) subjective frequency ratings were collected.

The specific aims of this study were (a) to determine if there is an advantage for recognizing words when listeners' presumed phonological representations match incoming speech stimuli; (b) to determine how such an advantage might change as L2 pronunciation proficiency increases; (c) to examine the separate effects of word frequency and neighborhood density on L2 word recognition; and (d) to evaluate the relation between subjective and objective word frequency for NE and NS participants, as well as the effect of subjective frequency on recognition performance.

II. EXPERIMENT

A. Method

1. Stimuli and design

a. Stimulus selection. To ensure that all participants, including those in the NS groups, knew all the test words, we began by developing a list of 365 words that were likely to be acquired by adults in the early stages of learning English as an L2. All were one syllable in length, and most had a CVC structure (e.g., *bird*, *cake*, *love*). In a pilot test, six NS-speaking listeners (none of whom subsequently participated in the study) indicated if they knew each word and, if so, rated it for subjective familiarity using a scale that ranged from 1 (''least familiar'') to 7 (''most familiar'').

TABLE I. Mean word frequency and neighborhood density of test words (ranges are in parentheses). Note. WF: word frequency, or occurrence per million words (Kučera and Francis, 1967); ND: neighborhood density, or number of words that differ by a one segment, addition, deletion, or substitution (Luce and Pisoni, 1998).

	High WF		Low WF	
ND	Dense	Sparse	Dense	Sparse
Example	''bed''	''bring''	''bell''	''boss''
WF	169.5 (54-500)	177.8 (58-591)	18.0 (3-37)	22.4 (1-41)
ND	23.5 (17-39)	10.0 (3-15)	23.8 (19-30)	10.4 (4-14)

Four sets of 20 words that differed orthogonally in text word frequency (Kučera and Francis, 1967) and neighborhood density (Luce and Pisoni, 1998) were formed from the 332 items that were known by at least five of the six pilot listeners and that received a familiarity rating greater than 5.0. All 80 test words are listed in the Appendix. The four sets of test words consisted of: (a) words with relatively high text frequency from dense neighborhoods; (b) high-frequency words from sparse neighborhoods; (c) low-frequency words from dense neighborhoods; and (d) low-frequency words from sparse neighborhoods. As shown in Table I, words differing in frequency had comparable neighborhood density values, and words differing in neighborhood density had comparable frequency values.[2]

b. Recording. The 80 test words were recorded by a male NE talker who was born and raised in the Midwest but had lived in Birmingham, Alabama for 21 years. The stimuli produced by this talker are referred to as the ''native-produced'' (or simply ''native'') stimuli. The test words were also recorded by an adult male speaker of Spanish who was born in Chile, had resided in the U.S. for 2 months, and was judged by the authors to speak English with a strong foreign accent. These stimuli are referred to as ''Spanish-accented'' (or ''accented'') stimuli. The words produced by both talkers were digitized at 22.05 kHz and normalized for peak intensity (50% of full scale).

The NE talker read the test words from a list, one at a time. For accented stimuli, in order to generate a range of tokens to select from, the NS talker produced the 80 test words four times each using different elicitation methods (see Flege, Munro, and MacKay, 1995). The four tokens of each test word produced by the NS talker were auditorily evaluated by the third author, a NE speaker with extensive training in phonetics. A few tokens contained a segmental substitution that resulted in a perceived change in lexical identity (e.g., a *hide* token that sounded like *hi*, a *bed* token that sounded like *Beth*). These tokens were eliminated and the token of each word that was judged to have the strongest foreign accent among the remaining tokens was retained for the study.

A single NE pilot listener listened to the 80 native stimuli in the clear and recognized all of them. Three NE pilot listeners evaluated the 80 Spanish-accented stimuli. At least one listener recognized 69 stimuli. Of the 11 stimuli not recognized by any listener, three stimuli were misheard in the same way. Specifically, all three listeners heard the accented *fish* stimulus (pronounced /fitʃ/) as *feet* /fit/; the accented *voice* stimulus (pronounced /bɔɪs/) as *boy*s /bɔɪz/; and the accented *boss* stimulus (pronounced /bɔɪs/) as *bus* /bʌs/. These misperceptions by the native English pilot listeners do not indicate that the aforementioned Spanish-accented stimuli were inappropriately chosen. The *fish*, *voice*, and *boss* stimuli were correctly recognized by 57% of the NS-speaking listeners when presented in noise in the experiment proper. This suggested that while the Spanish-accented stimuli might not have been realized in such a way as to promote correct lexical access by NE-speaking listeners, their phonetic realization was sufficient for lexical access by at least some people who themselves spoke English with a Spanish accent.

The third author carried out a broad phonetic transcription of the 80 Spanish-accented stimuli. A total of 17 segmental substitutions was noted, the most common being /s/ for /z/ (in *cheese, noise, nose*), /tʃ/ for /ʃ/ (in *push, shake, shine*), and /ɔ/ for /ɑ/ (in *boss, lost, sock*). The substitutions were somewhat more common in words drawn from sparse than dense neighborhoods (12 vs 5), but about equally common in high- and low-frequency words (8 vs 9). The imbalance in number of substitutions between high- and low-neighborhood word stimuli is probably due to the fact that segmental substitutions result in a perceived change in lexical identity less often for sparse word stimuli than dense word stimuli. As expected from previous research (e.g., Flege and Munro, 1994), all stimuli contained subsegmental divergences from the phonetic norms of English and therefore conveyed foreign accent. These included the shortening of voice-onset time in word-initial tokens of /p t k/, the partial devoicing of word-final tokens of voiced obstruents, the prolongation of initial /h/ tokens, and the weakening of final /n/ segments.

The stimuli were measured for duration and the duration values were submitted to a 2 (talker) × 2 (neighborhood density) × 2 (word frequency) ANOVA. Words from dense neighborhoods were significantly shorter than words from sparse neighborhoods ($M = 473$ vs 528 ms), $F(1,76) = 8.4$, $p < 0.01$, and the Spanish-accented stimuli were shorter than the native-produced stimuli ($M = 448$ vs 552 ms), $F(1,76) = 80.0$, $p < 0.01$. The effect of frequency was nonsignificant, however; $F(1,76) = 1.1$, $p > 0.10$, and no significant interactions were found.

The 160 stimuli (80 test words × 2 talkers) were mixed with noise to bring word recognition scores off ceiling. Multitalker babble was created by adding the voices of six NE and six NS talkers. The noise segment was set to 10% of full scale of peak intensity, and then added to the 160 stimuli described earlier (which had been normalized at 50% peak intensity of full scale). This yielded stimuli having S/N ratios of approximately 14 dB.

c. Counterbalancing procedure. Two blocks of 40 test words, designated "A" and "B" (with ten items randomly selected from each of the four lexical sets in which word frequency and neighborhood density varied) were formed. The order of words was randomized within each block. Blocks A and B were presented as both native and Spanish-accented stimuli, and the order of presentation for stimulus type (native vs Spanish-accented) was counterbalanced. Thus, the participants were randomly assigned to one of four counterbalanced conditions: A-native/B-accented, A-accented/B-native, B-native/A-accented, or B-accented/A-native. Across participants, each test word was presented in both native and Spanish-accented form. A given listener heard each of the 80 test words (half native, half Spanish-accented) only once.

2. Procedure

Participants were tested individually in a sound booth in a session lasting about 1 h. They completed three primary tasks in the following order: a word recognition task, a written lexical knowledge test, and a sentence production task. These tasks were preceded by a pure-tone hearing screening and followed by a language background questionnaire in which demographic information such as age of arrival and length of residence in the United States, and use of English and Spanish was collected (see Table II).

In the word recognition task, the stimulus words were presented one at a time via loudspeakers. The participants were asked to write down each word on a prepared answer sheet. They were given a short demonstration of the task, and a 16-item practice session preceded each of the two blocks. The participants were encouraged to adjust loudness to a comfortable level during the practice session. The practice words differed from the 80 test words. They were recorded by a female NE and a female NS talker and presented with the same multitalker babble as the test stimuli. Each test block began with five extra items which were not analyzed.

In the lexical knowledge test, participants were asked to indicate whether or not they knew each of the 80 test words, which were presented in written form. Five nonword foils were included to ensure that participants would veridically report any test items they did not know. If they knew an item, participants provided an estimate of frequency of usage on a scale that ranged from 1 ("seldom hear/say this word") to 7 ("often hear/say this word"). To encourage them to use the entire rating scale when rating the 80 items, participants were given examples of seven words that were presumed to span the full 7-point scale (e.g., *dog* is heard/said much more often than *octopus*). Thus, our rating scale emphasized the relative frequency of word usage. Next, two 7-item practice sets were presented. When participants demonstrated an understanding of the instructions, the test items were presented.

The sentence production task involved saying five simple English sentences three times each using a delayed repetition technique (see Flege *et al.*, 1995). The sentences were later used to assign NS participants to subgroups differing in English pronunciation proficiency. Only the third, fluent repetition of each sentence was used for group assignment (see below).

Participants' written responses in the word recognition task were scored "correct" when they exactly matched the

TABLE II. Demographic information for the HP and LP groups. Shown for each characteristic are means (standard deviations and ranges are in parentheses). Note. HP: high pronunciation proficiency group; LP: low pronunciation proficiency group. FA (foreign accent) ratings are based on a 9-point scale from 1 (strongest accent) to 9 (least accent). AOA: age-of-arrival in the U.S. in years; LOR: length of U.S. residence in years. Estimates of hours of English and Spanish use per day are based on a 6-point scale from 1 (none), 2 (0.5 h), 3 (1 h), 4 (2 h), 5 (4 h), and 6 (6 h or more). Estimates of hearing Spanish-accented speech are based on a scale from 1 (very seldom) to 7 (very often). Estimates of ability to speak and understand English and Spanish are based on a 7-point scale (1—poor; 4—OK; 7—good).

	HP	LP
Chronological age	30 (8.3; 19–46)	33 (8.5; 20–47)
FA rating	6.3 (1.8; 3.6–8.8)	2.1 (0.5; 1.4–2.9)
AOA	19.8 (8.8; 1.9–38.5)	29.0 (8.2; 13.4–41.9)
LOR	9.7 (6.7; .9–24.2)	4.1 (3.4; 0.02–1.2)
Age English study began	8.9 (4.3; 0–16)	10.4 (2.6; 6–14)
Years of English study	6.5 (4.5; 0–14)	6.5 (4.1; 0.1–13)
Education in home country	13.5 (5.8; 3–22)	15.8 (4.9; 4–25)
Education in US	5.2 (4.6; 0–14)	1.5 (1.8; 0–5)
Spoken English use	5.9 (0.3; 5–6)	4.6 (1.4; 2–6)
Spoken Spanish use	5.1 (1.2; 2–6)	5.4 (1.1; 2–6)
Hearing Spanish-accented speech	2.4 (1.7; 1–6)	3.3 (2.5; 0–7)
Ability to speak English	5.9 (0.9; 4–7)	4.3 (1.1; 2–6)
Ability to speak Spanish	6.8 (0.5; 5–7)	6.9 (0.3; 6–7)
Ability to understand English	6.0 (0.8; 4–7)	4.8 (1.2; 3–7)
Ability to understand Spanish	6.8 (0.5; 5–7)	7.0 (0; 7)

target words. However, we anticipated that the NS participants might misspell some words. Given that our interest was in spoken word recognition rather than in spelling ability, the experimenter examined each participant's written responses after practice sessions and after each block of 40 test stimuli. She asked for clarification if the handwriting was not legible. For responses that were not exact, she asked what word the participant had intended to write. If the explanation corresponded to the target (e.g., ''being noisy'' for the written response ⟨laud⟩ for the stimulus *loud*), the written response was scored as correct.

3. Participants

Sixty NS speakers living in or near Birmingham, AL and 16 NE speakers, who were born and raised in Alabama, were recruited through advertisements in a university paper and through personal contacts. The NS participants were required to have been born in a predominantly Spanish-speaking country, to have learned Spanish as their L1 from native Spanish-speaking parents, and to be between the ages of 19–50 years. Nine of the 60 NS participants were excluded: two failed the pure-tone hearing screening; one did not complete the experiment; three failed to report three or more nonword foils in the lexical knowledge test as unknown words; and three knew fewer than 96% (77/80) of the test words.

As mentioned earlier, participants recorded sentences after the lexical knowledge test. Three sentences produced by each of the 51 retained NS participants and the 16 NE participants were digitized (22.05 kHz), and then randomly presented to seven NE listeners from Alabama. Sentences produced by the NS and NE speakers who had recorded the stimuli for the word recognition task were also presented. The NE listeners rated the sentences using a scale that ranged from 1 (''strong accent'') to 9 (''no accent'').

An average foreign accent (FA) rating was computed for each NS participant. The 16 NS participants with the mildest foreign accents were assigned to a ''high pronunciation proficiency'' (HP) group, with the restriction that there were four participants from each of the four counterbalancing conditions in this group. The 16 NS participants with the strongest foreign accents were assigned to a ''low pronunciation proficiency'' (LP) group with the same restriction. The average age, in years, of the NE, HP, and LP groups was 34 (range: 22–47), 30 (range: 19–46), and 33 (range: 20–47), respectively, and the average FA rating was 8.67 (range: 7.5–9.0), 6.34 (range: 3.6–8.8), and 2.13 (range: 1.4–2.9). Demographic information for the two NS groups is shown in Table II.

B. Results

1. Word recognition scores

Percent-correct scores for the ten items in each of the eight conditions of the stimulus design were obtained for each participant and submitted to a 3(group: NE, HP, LP)×2 (stimulus type: native, Spanish-accented)×2 (word frequency: high, low)×2 (neighborhood density: dense, sparse) ANOVA. Group served as a between-subjects factor, and stimulus type, word frequency and neighborhood density as within-subjects factors. Only significant main effects or interactions will be reported. Significant interactions were followed up by simple effects tests with Bonferroni's correction ($p<0.05$).

The ANOVA yielded significant main effects of group, $F(2,45)=15.27$, $p<0.001$, stimulus type, $F(1,45)=98.82$, $p<0.001$, and neighborhood density, $F(1,45)=67.06$, $p<0.001$. Both the NE and HP groups performed better than the LP group (66.6% and 65.0% vs 53.0%). Also, more native than accented stimuli were correctly recognized (70.4% vs 52.8%), as were more words from sparse than dense neighborhoods (66.6% vs 56.6%).

The two-way group×stimulus type interaction was significant, $F(2,45)=29.73$, $p<0.001$. The NE and HP groups performed better for native than accented stimuli, whereas

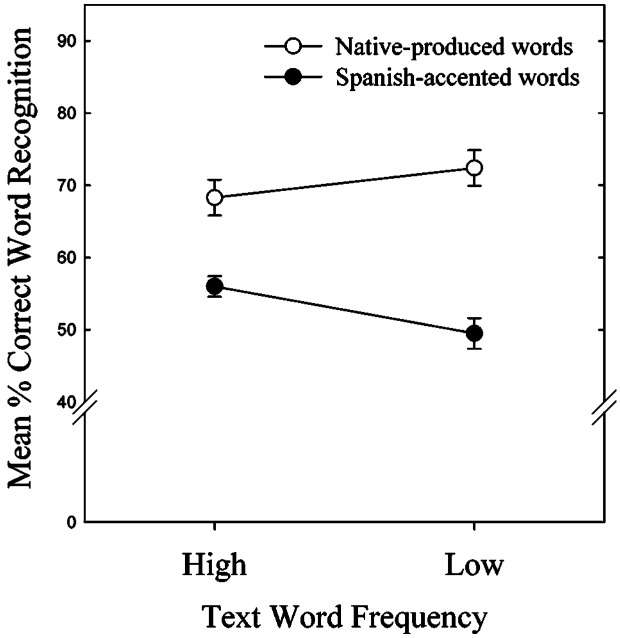

FIG. 1. Mean percent-correct word recognition scores as a function of stimulus type (native, accented speech) and text word frequency. Error bars represent one standard error of the mean.

the LP group's performance was similar for the two types of stimuli. This finding only partially supported the phonological mismatch hypothesis. That is, although the NE group did recognize more native than accented stimuli, the LP group did not recognize more Spanish-accented than native stimuli.

The two-way stimulus type × word frequency interaction was also significant, $F(2,45)=14.99$, $p<0.001$. This was the only significant effect involving word frequency (see Fig. 1). *Post hoc* tests revealed that more high- than low-frequency words were recognized in the accented condition, $F(1,47)=9.03$. The opposite trend for native stimuli did not reach significance when Bonferroni's correction was applied, $F(1,47)=5.11$. This result might reflect subtle differences in the way the NE and the NS talker produced the stimuli. The NE talker might have produced the low-frequency words more clearly than the high-frequency words (e.g., Geffin and Luszcz, 1983; Wright, 1979), whereas the NS talker might have produced the high-frequency words more accurately (with less foreign accent because of the frequent use) than low-frequency words. However, as noted in the Method section, when duration and segmental substitutions were analyzed, no differences were found for high- vs low-frequency words by either the NE or NS talker.

The significant three-way group × stimulus type × neighborhood density interaction, $F(2,45)=5.56$, $p<0.01$, is shown in Fig. 2. For words from sparse neighborhoods (right panel), there was no difference between the NE, HP, and LP groups' performance when the stimuli were accented (59.4%, 59.1%, and 57.8%, respectively), $F(2,45)=0.04$, whereas both the NE and HP groups performed better than the LP group for native stimuli (83.4% and 82.2% vs 57.5%), $F(2,45)=36.54$. For words from dense neighborhoods (left panel), the HP and LP groups performed better than the NE group when the stimuli were accented (51.6% and 48.1% vs 40.6%), $F(2,45)=5.82$. In contrast, for native-produced stimuli, the NE group performed better than the HP group, who performed better than the LP group (83.1% vs 67.5% vs 48.4%), $F(2,45)=43.64$. Thus, for both native and accented words from sparse neighborhoods, the NE and HP groups performed similarly. Interestingly, however, for words from dense neighborhoods, the NE group recognized more native stimuli than the HP and LP groups, whereas the HP and LP groups recognized more accented stimuli than the NE group did.

The significant three-way interaction points to a differential influence of neighborhood density (i.e., better word recognition for words from sparse than dense neighborhoods) for the three groups. The HP and LP groups displayed a competition effect only for the native stimuli, $F(1,15)=56.07$ and $F(1,15)=12.23$ ($p<0.05$, Bonferroni), whereas the NE group displayed this effect only for the accented stimuli, $F(1,15)=22.35$ ($p<0.05$, Bonferroni). The difference between words from sparse and dense neighborhoods for the accented stimuli did not reach significance for the LP

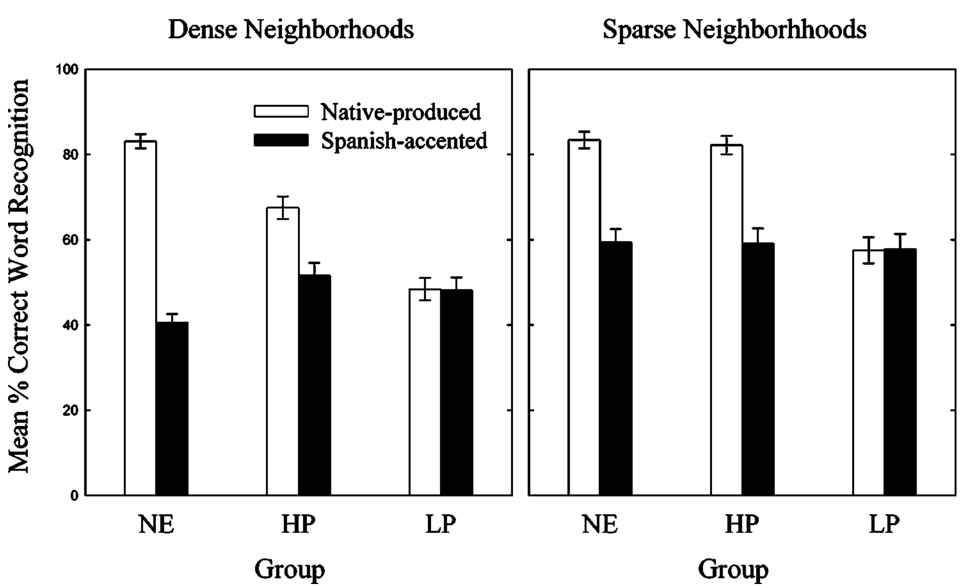

FIG. 2. Mean percent-correct word recognition scores as a function of stimulus type and neighborhood density for the native English (NE), high pronunciation proficiency (HP), and low pronunciation proficiency (LP) groups. Error bars represent one standard error of the mean.

group (57.8% vs 48.1%), $F(1,15) = 5.21$ ($p > 0.05$, Bonferroni), or the HP group (59.1% vs 51.6%), $F(1,15) = 2.21$ ($p > 0.10$, Bonferroni). Thus, the results show interesting interactions by neighborhood density and stimulus type. The competition effect is evident when there were mismatches between speech input and phonological representations. That is, when fine-grained phonological discrimination was required (for words from dense neighborhoods), the mismatch between phonological representations and speech stimuli seems to have impeded performance.

The phonological mismatch hypothesis predicts that fewer words should be recognized when there is a large discrepancy between listeners' phonological representations and the speech input they hear. We further anticipated that this might be especially true for low- vs high-frequency words and/or words from dense vs sparse neighbors. The results show that the NE and HP groups recognized more native than accented stimuli, suggesting that the HP group has developed representations of English words that were close to those of the NE speaker. The LP group, however, did not recognize more Spanish-accented than native stimuli, thus failing to support our hypothesis. Also, word frequency did not interact with group. Possible reasons for this lack of support will be provided in the Discussion section.

Although the effect of word frequency was minimal, the three groups did recognize words differently according to stimulus type (native vs accented) and neighborhood density. For words from dense and sparse neighborhoods, no group recognized more Spanish-accented than native stimuli. However, there were between-group differences, such that the NS groups recognized more Spanish-accented stimuli from dense neighborhoods than the NE group, whereas the NS groups' performance was similar to the NE group's for accented stimuli from sparse neighborhoods. These findings partially support the phonological mismatch hypothesis; i.e., the effect of mismatched conditions was observed when differences between groups were considered.

As noted in the Methods section, analyses of stimulus duration and segmental substitution of Spanish accented stimuli showed differences between words from dense and sparse neighborhoods. The words from dense neighborhoods were shorter in duration and had fewer substitutions than the words from sparse neighborhoods. There were no differences in duration or the number of substitutions between high- and low-frequency words. Our focus in this study was on group differences in recognition performance, and the duration difference would have affected the three groups equally. As there were more substitutions for sparse than dense, if there were an effect of substitution, it would have counteracted the mismatch effect we obtained. Thus, the confounding of the characteristics of Spanish-accented speech and stimulus selection was probably minimal.

2. Correlational data for NS participants

Pearson's r correlations between the word recognition scores for native and accented stimuli, FA ratings, and other demographic information for the 32 NS participants were obtained ($p < 0.001$, unless otherwise noted). As in previous studies (e.g., Flege, 1988; Yeni-Komshian, Flege, and Liu, 2000), we found a negative correlation between mean FA ratings and age of arrival (AOA) in the U.S. ($r = -0.63$), as well as a positive correlation between FA ratings and length of residence, or LOR ($r = 0.72$). AOA was strongly correlated with years of education in the home country and in the U.S. ($r = 0.73$ and 0.72) and also English use ($r = 0.56$). FA ratings was highly correlated with estimates of English use ($r = 0.56$). Estimates of Spanish use were not correlated with any of these variables.

Recognition scores for native stimuli were significantly correlated with FA ratings, AOA, years of education in the U.S. ($r = 0.64, -0.47$ and 0.66), and with estimates of English use ($r = 0.41$, $p < 0.05$), but only marginally with LOR ($r = 0.33$, $p = 0.07$). The moderately high correlation between recognition of the native stimuli and the FA ratings suggests there is indeed a link between good L2 perception and pronunciation.

Notably, word recognition scores for the accented stimuli were not significantly correlated with any of the measures examined here. In particular, having an L2 phonological system that is substantially affected by the L1, as implied by having a strong foreign accent, was not correlated with better recognition of Spanish-accented words. This null finding is consistent with the ANOVA results, which showed little difference in performance for Spanish-accented stimuli between the HP and LP groups.

3. Subjective frequency ratings

We next calculated Pearson's r correlations between the NE, HP, and LP groups' subjective frequency ratings for the 80 test words: $r = 0.82$ for NE and HP; $r = 0.78$ for NE and LP; $r = 0.88$ for HP and LP ($p < 0.001$, in each instance). These strong correlations indicated that for the test words selected for this study, the estimated frequency of word use was similar across listener groups, regardless of proficiency in English. The correlations between the subjective frequency ratings of our three groups and text word frequency were in the moderate range: $r = 0.36$, 0.42, and 0.52, for the NE, HP, and LP groups, respectively ($p < 0.001$, in each instance).

4. The effect of subjective frequency on word recognition

Thus far, text word frequency does not seem to have had a strong effect on spoken word recognition performance. Previous work suggests that subjective measures of experienced word frequency may be more sensitive than objective ones. We therefore reanalyzed participants' word recognition scores using their subjective frequency ratings. The 40 words from dense neighborhoods were rank ordered according to mean subjective frequency ratings from the 54 participants, and then divided into a set of 20 high subjective frequency words (M rating $= 5.0$) and a set of 20 low subjective frequency words ($M = 3.0$). Similarly, sparse words were subdivided into high and low subjective frequency categories ($M = 5.1$ vs 3.6).[3] See the Appendix for average subjective frequency ratings for each word.

Percent-correct scores for these new conditions were submitted to a 3 (group: NE, HP, LP)$\times$2 (stimulus type: na-

TABLE III. Effects obtained for ANOVAs of word recognition scores. Note. In the original analysis, test words were categorized according to text word frequency; in the reanalysis, they were categorized according to subjective frequency ratings.

Effect	Original analysis (text word frequency)	Reanalysis (subjective frequency)
Frequency	$F(1,45)<1$ n.s.	$F(1,45)=36.1^{a}$
Frequency× stimulus type	$F(1,45)=15.0^{b}$	$F(1,45)=2.4$ n.s.
Frequency× neighborhood density	$F(1,45)<1$ n.s.	$F(1,45)=35.0^{a}$
Frequency×stimulus type× neighborhood density	$F(1,45)<1$ n.s.	$F(1,45)=29.8^{a}$

[a] $p<0.01$.
[b] $p<0.05$.

tive, Spanish-accented)×2 (subjective frequency: high, low) ×2 (neighborhood density: dense, sparse) mixed-design ANOVA. Because the reclassification of test words did not affect the results obtained for other variables, we report only those effects related to subjective frequency.

The new analysis yielded four effects that differed from the original ANOVA (see Table III). As in the original analysis, there was no interaction of group with subjective frequency, suggesting subjective frequency affected word recognition by all listeners similarly. In the new analysis, there was a significant main effect of subjective frequency, $F(1,45)=36.1$, $p<0.001$. More high- than low-frequency words were correctly recognized (67% vs 57%). Also, there was a significant two-way interaction between subjective frequency and neighborhood density, $F(1,45)=35.0$, $p<0.001$. More words from sparse than dense neighborhoods were recognized when the words were of high subjective frequency (57.3% vs 75.5%), whereas no neighborhood density effect was found for low subjective frequency words (55.8% vs 57.6%). Last, the three-way subjective frequency ×neighborhood density×stimulus type interaction was significant, $F(1,45)=29.8$ $p<0.001$.

The three-way interaction (see Fig. 3) was explored by tests of simple main effects (Bonferroni, $p<0.05$). For both

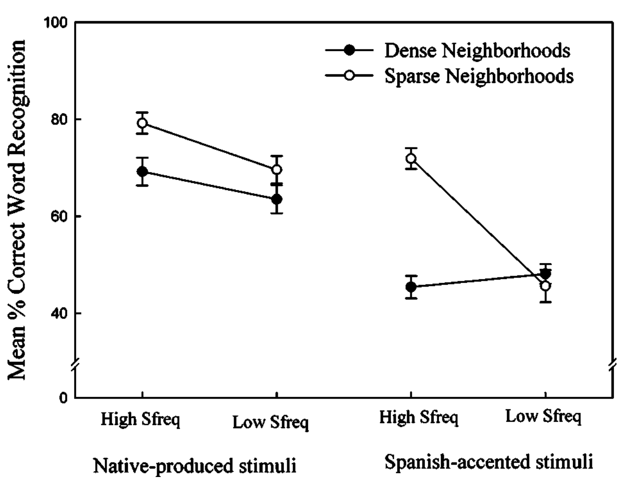

FIG. 3. Mean percent-correct word recognition scores as a function of stimulus type, subjective frequency, and neighborhood density. Error bars represent one standard error of the mean.

native and accented stimuli, the differences between words from sparse and dense neighborhoods were significant only for high, not low subjective frequency items. However, this neighborhood density effect was much larger for accented, as opposed to native stimuli.

In sum, when subjective frequency was used to classify the test words, there was an interactive effect of neighborhood density on word recognition. For native stimuli, performance was in the expected direction (i.e., better for high- than low-frequency items) for words from both dense and sparse neighborhoods. However, for Spanish-accented stimuli, only words from sparse neighborhoods followed this pattern; the recognition of less phonologically confusable items was enhanced when they were of high frequency, whereas the recognition of highly confusable items was not affected by word frequency.

III. DISCUSSION

Previous studies have shown that L2 learners have difficulty recognizing L2 words (e.g., Bradlow and Bent, 2002; Mayo *et al.*, 1997). One likely source of this difficulty is that perceptual representations for vowels and consonants are not optimally attuned to the L2; another is insufficient higher-level, lexical knowledge. To date however, only a handful of studies have considered the relations between L2 learners' phonological representations and their higher-level, lexical knowledge (e.g., Bradlow and Pisoni, 1999; Meador *et al.*, 2000). The present study evaluated the influence of two lexical factors (word frequency and neighborhood density) on word recognition by NE listeners and two groups of NS listeners who differed in overall degree of foreign accent (high pronunciation proficiency vs low pronunciation proficiency). Listeners were asked to identify words that were spoken by a NE speaker (native-produced stimuli) and by a NS speaker (Spanish-accented stimuli).

According to our phonological mismatch hypothesis, differences between the phonetic specification of vowels and consonants in English words and listeners' lexical representations should lead to decreased word recognition performance. We hypothesized that NS adults tend to use "Spanish-like" phonological representations when processing spoken English words. We also hypothesized that, concurrent with L2 learners' improved L2 pronunciation, their L2 lexical representations would more closely resemble those of native speakers. We assumed that NE listeners' lexical representations are optimally attuned to the typical pronunciation of English words. Thus, we expected mismatches to occur when the NE listeners attempted to recognize the Spanish-accented stimuli and when the NS listeners attempted to recognize the native stimuli. The latter mismatch effect was expected to be larger for the LP vs HP group.

This mismatch effect was expected to be influenced by lexical factors such as word frequency and neighborhood density. Specifically, we anticipated that low-frequency words might have higher thresholds for activation than high-frequency words (see the models of Becker, 1980; Forster, 1981), and thus be harder to recognize in the mismatch conditions where the bottom-up input was not optimal. We also expected that fewer words from dense than sparse neighborhoods would be recognized in the mismatch conditions because the former require more fine-grained segmental perception (see also Bradlow and Pisoni, 1999).

Our results provided some support for the phonological mismatch hypothesis. Although the NE and HP groups performed consistently better for native than Spanish-accented stimuli, the LP group's performance for Spanish-accented stimuli was no better than for native stimuli. There are several possible reasons why we did not obtain the expected mismatch effect for the LP group. First, although there are "typical" foreign-accent patterns by talkers of a particular L1, there are likely idiosyncrasies for individuals that are difficult to tease apart from these general patterns. Differential L2 experience in terms of the quality and quantity of input might produce foreign accents that are more varied across L2 learners than individual variations among L1 talkers. Thus, phonological "matches" for native-produced stimuli with native listeners might be greater than matches especially for accented stimuli with the LP group. Second, Spanish has regional accents and the NS participants in our study were from various Central and South American countries. This heterogeneous background might have reduced the effect of exact phonological matches. Third, we added noise to our stimuli, and it has been shown that noise can have more adverse effects on non-native vs native listeners (e.g., van Winjingarden, 2001). Perhaps our LP participants needed clearer segmental information for accurate recognition.

The effect of word frequency on word recognition in our study was similar across groups, which might be due to our having employed only known words. Another possible reason why we did not observe strong text frequency effects is that text word frequency does not constitute a very sensitive measure of the psychological familiarity of words (Balota *et al.*, 2001; Garlock *et al.*, 2001; Gernsbacher, 1984; Gordon, 1985). Indeed, when our stimulus sets were reorganized and analyzed according to subjective frequency estimates, more high- than low-frequency words were recognized overall by all three groups (67% vs 57%). Further, there was a boost in recognition scores for high-frequency words from sparse neighborhoods, and this effect was larger for Spanish-accented stimuli. This may be because recently used words have higher activation potentials and are easier to recognize with limited phonetic information than words of infrequent use.

There were group differences in word recognition according to stimulus type and neighborhood density. For native stimuli, lower word recognition scores were obtained for the LP group, who had strong foreign accents, than for the NE and HP groups. This held true for words drawn from both dense and sparse neighborhoods. Lower scores were also obtained for the HP than NE group, but only when accurate segmental perception was especially important—i.e., for words from dense neighborhoods. In contrast, for Spanish-accented stimuli, both NS groups outperformed the NE group for words from dense neighborhoods that required more fine-grained speech perception.

In the present study, pronunciation proficiency was used to define high- and low-proficiency NS groups. We assumed that good L2 pronunciation corresponds to more native-like L2 lexical representations. In agreement with previous foreign accent research (e.g., Flege *et al.*, 1995), the demographic information indicated the HP participants tended to have arrived in the U.S. at an earlier age than the LP participants ($M=19.8$ vs 29.0 years), to use English more (about 6 or more hours a day vs 3 hours), and to have lived longer in the U.S. ($M=9.7$ vs 4.1 years). These differences might well have contributed to the HP participants' superior English pronunciation, and presumably to the differences observed in word recognition performance.

The present results confirm that during spoken word recognition, higher-level processes interact with more bottom-up, segmental perception. The performance differences between the HP and LP groups for native stimuli that we observed further indicated that the lexical representations of the HP group more closely resembled those of the NE group than the LP group, but that they are not identical.

Most models of spoken word recognition assume a matching process between incoming speech sounds and phonological representations in the mental lexicon (e.g., Gaskell and Marslen-Wilson, 1997; McClelland, 1988; Norris, McQueen, and Cutler, 1995). The nature of these representations has been investigated in many previous studies (e.g., Hintzman, 1986), but few studies have considered such representations for those learning an L2, or how they are related to the L1 phonology (Pallier *et al.*, 2001). Further, to our knowledge, no studies have examined whether L2 lexical representations change as L2 pronunciation improves. The observed differences between the LP and HP groups in word recognition performance provided indirect evidence that improvements in English pronunciation correspond to more native-like representations that more closely match speech of NE speakers.

Similar to our phonological mismatch hypothesis, Bent and Bradlow (2003) have proposed that there is an interlanguage speech intelligibility benefit that is enjoyed by L2 listeners. That is, L2 learners have a shared knowledge base that aids the recognition of speech produced by non-native vs native talkers. This shared knowledge base encompasses numerous features of speech, including stress patterns, intonation, phonotactics, as well as segmental features. In contrast, our mismatch hypothesis focuses on segmental perception and attempts to control these other potentially relevant factors. Additional studies at the sentence level will be needed to elucidate the ways in which such factors affect L2 word recognition in a more precise manner.

Further research using time-sensitive tasks, such as priming, will also be needed to define the time course of the effects of word frequency and neighborhood density, especially regarding L2 speech processing (e.g., Pallier *et al.*, 2001). Perhaps word frequency influences an early stage of recognition by activating candidate words, while neighborhood density affects a later stage, where matching processes between speech input and lexical representation occur (e.g., Dahan, Magnuson, and Tanenhaus, 2001). The fact that we did not find group differences in the effects of either text word frequency or subjective frequency suggests that the initial stage of word activation may be similar for the NS and NE groups. However, as suggested by the differential effects of neighborhood density, the later matching stage may differ between NE and NS speakers.

In conclusion, the results of this study supported the phonological mismatch hypothesis, at least when the neighborhood density effect is considered. There was a greater effect of neighborhood density on spoken word recognition when there were presumed mismatches between lexical representations and incoming speech sounds (i.e., native stimuli for the NS groups and Spanish-accented stimuli for the NE group) than when there were matches. This may be because bottom-up processing of speech segments in words from dense neighborhoods was more affected by differences in phonological representations as compared to words from sparse neighborhoods. The HP group showed better recognition of native stimuli than the LP group, and their performance for words from sparse neighborhoods was as good as that of the NE group, indicating changes in phonological representations as L2 pronunciation improves. However, both NS groups showed reduced recognition of words from dense neighborhoods, suggesting that even when L2 learners' segmental perception improves, their performance under conditions requiring more fine-grained perception may still be compromised.

ACKNOWLEDGMENTS

This study was supported by NIH Grant DC00257. We thank Ana Castaneda for the recruitment of participants and Mitchell Sommers for information about the Washington University in St. Louis Speech and Hearing Lab Neighborhood Database. In addition, we appreciate the comments of two anonymous reviewers and those of Dr. R. Diehl on a previous version of this paper.

APPENDIX

Test words. Note. WF: word frequency; ND: neighborhood density; SF-AVE: average subjective frequency rating across our three listener groups. ND values were obtained from the Washington University in St. Louis Speech and Hearing Lab Neighborhood Database (see the references).

Word	WF	ND	SF-AVE	Word	WF	ND	SF-AVE	Word	WF	ND	SF-AVE	Word	WF	ND	SF-AVE
bed	127	25	5.85	bring	158	8	5.69	bell	19	27	2.17	boss	20	11	4.27
call	188	26	6.25	choice	113	3	4.46	bird	31	22	3.50	cart	5	14	2.44
date	103	24	5.33	cold	171	15	5.38	burn	15	22	2.33	cheese	9	13	5.23
face	371	21	4.79	faith	111	11	4.23	cake	13	26	3.79	coin	10	14	4.10
fall	147	26	3.71	fast	78	15	5.04	corn	34	20	3.17	cute	5	6	4.77
heard	269	20	4.83	five	286	12	4.83	duck	9	25	1.92	fish	35	13	3.83
job	238	19	5.92	foot	70	10	4.27	ham	19	26	3.17	fork	14	13	4.65
lake	54	32	3.19	house	591	7	6.31	hide	22	21	2.31	frog	1	4	1.73
list	133	19	4.67	join	65	8	3.79	hurt	37	23	4.54	jump	24	8	3.29
nose	60	18	3.81	kind	313	7	4.69	nail	6	26	3.02	kiss	17	13	5.10
note	127	26	4.94	lost	173	9	3.85	noon	25	19	5.17	lamp	18	11	3.90
park	94	18	4.77	love	232	11	6.06	peach	3	22	2.67	loud	20	12	4.42
part	500	17	4.78	mouth	103	7	4.52	pen	18	29	5.83	match	41	14	3.42
pass	89	24	3.77	move	171	8	5.35	pet	8	30	3.79	mouse	10	14	2.77
phone	54	27	6.58	post	84	15	2.54	sad	35	25	3.65	noise	37	4	4.75
rate	209	39	3.21	safe	58	11	4.10	shake	17	24	3.40	push	37	5	4.25
rest	164	20	5.15	smile	58	5	5.60	sheep	23	20	1.58	smell	34	7	4.67
save	62	22	5.06	voice	226	7	3.65	shine	5	21	3.15	teach	41	13	4.96
ten	165	27	4.63	white	365	11	4.46	sock	4	26	3.94	van	32	12	3.02
west	235	19	3.25	wrong	129	13	4.65	soup	16	22	3.77	wash	37	7	5.81

[1]Word frequency and neighborhood density information was obtained from the Washington University in St. Louis Speech and Hearing Lab Neighborhood Database (see the references).

[2]A between-items analysis of variance (ANOVA) indicated that the four cells were similar in terms of phonotactic probability, $F(3,76)=2.53$, $p=0.06$. In computing phonotactic probability, we used biphone frequency (segment-to-segment co-occurrence probability) for adult speech, according to Carterette and Hubbard (1974).

[3]A 2 (subjective frequency: high, low)$\times$2 (neighborhood density: dense, sparse) ANOVA was performed on average subjective frequency ratings for the 80 test words to confirm that ratings were balanced across neighborhood conditions. There was a main effect of subjective frequency, $F(1,76)=128.58$, $p<0.0001$. The high subjective frequency condition had higher ratings than the low subjective frequency one (5.08 vs 3.33). There was also a main effect of neighborhood density, $F(1,76)=4.76$, $p<0.05$. The sparse condition had higher frequency ratings than the dense one (4.03 vs 4.37). However, the difference in ratings was small, and the two-way interaction was not significant, $p>0.10$.

Amano, S., Kondo, T., and Kakehi, K. (**1995**). ''Modality dependency of familiarity ratings of Japanese words,'' Percept. Psychophys. **57**, 598–603.

Balota, D. A., Pilotti, M., and Cortese, M. J. (**2001**). ''Subjective frequency estimates for 2,938 monosyllabic words,'' Mem. Cognit. **29**, 639–647.

Becker, C. A. (**1980**). ''Semantic context and word frequency effects in visual word recognition,'' Mem. Cognit. **8**, 493–512.

Bent, T., and Bradlow, A. R. (**2003**). ''The interlanguage speech intelligibility benefit,'' J. Acoust. Soc. Am. **114**, 1600–1610.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–206.

Best, C. T., McRoberts, G., and Goodell, E. (**2001**). ''Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Bradlow, A. R., and Bent, T. (**2002**). ''The clear speech effect for non-native listeners,'' J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). ''Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Carroll, J. B. (**1971**). ''Measurement properties of subjective magnitude estimates of word frequency,'' J. Verbal Learn. Verbal Behav. **10**, 722–729.

Carterette, E. C., and Hubbard Jones, M. (**1974**). *Informal Speech* (University of California Press, Berkeley, CA).

Charles-Luce, J., and Luce, P. A. (**1990**). ‘‘An examination of similarity neighborhoods in young children’s receptive vocabularies,’’ J. Child Lang **22**, 727–735.

Cohen, H., Douaire, J., and Elsabbagh, M. (**2001**). ‘‘The role of prosody in discourse processing,’’ Brain Cogn **46**, 73–82.

Connine, C. M., Mullennix, J., Shernoff, E., and Yelen, J. (**1990**). ‘‘Word familiarity and frequency in visual and auditory word recognition,’’ J. Exp. Psychol. Learn. Mem. Cogn. **16**, 1084–1096.

Cutler, A., and van Donselaar, W. (**2001**). ‘‘Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch,’’ Lang. Speech **44**, 171–195.

Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (**2001**). ‘‘Time course of frequency effects in spoken-word recognition: Evidence from eye movements,’’ Cognit. Psychol. **42**, 317–367.

Derwing, T. M., and Munro, M. J. (**1997**). ‘‘Accent, intelligibility, and comprehensibility: Evidence from four L1s,’’ Stud. Second Lang. Acquis. **19**, 1–16.

Derwing, T. M., and Munro, M. J. (**2001**). ‘‘What speaking rates do non-native listeners prefer?’’ Appl. Linguist. **22**, 324–337.

Ferguson, C. A., and Farwell, C. B. (**1975**). ‘‘Words and sounds in early language acquisition,’’ Language **51**, 419–439.

Flege, J. E. (**1988**). ‘‘Factors affecting degree of perceived foreign accent in English sentences,’’ J. Acoust. Soc. Am. **84**, 70–79.

Flege, J. E. (**1995**). ‘‘Second language speech learning: Theory, findings and problems,’’ in *Speech Perception and Linguistic Experience*, edited byW. Strange (York, Baltimore), pp. 233–272.

Flege, J. E., and Munro, M. J. (**1994**). ‘‘The word unit in second-language speech production and perception,’’ Stud. Second Lang. Acquis. **16**, 381–411.

Flege, J. E., Munro, M. J., and Fox, R. (**1994**). ‘‘Auditory and categorical effects on cross-language vowel perception,’’ J. Acoust. Soc. Am. **95**, 3623–3641.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). ‘‘Factors affecting degree of perceived foreign accent in a second language’’ J. Acoust. Soc. Am. **97**, 3125–3134.

Forster, K. I. (**1981**). ‘‘Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing,’’ J. Exp. Psychol. **33A**, 465–496.

Garlock, V. M., Walley, A. C., and Metsala, J. L. (**2001**). ‘‘Age-of-acquisition, word frequency and neighborhood density effects on spoken word recognition by children and adults,’’ J. Mem. Lang. **45**, 468–492.

Gaygen, D. E., and Luce, P. A. (**1998**). ‘‘Effects of modality on subjective frequency estimates and processing of spoken and printed words,’’ Percept. Psychophys. **60**, 465–483.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1997**). ‘‘Integrating form and meaning: A distributed model of speech perception,’’ Lang. Cognit. Processes **12**, 613–656.

Geffin, G., and Luszcz, M. A. (**1983**). ‘‘Are the spoken durations of rare words longer than those of common words?’’ Mem. Cognit. **11**, 13–15.

Gernsbacher, M. A. (**1984**). ‘‘Resolving 20 years of inconsistent interaction between lexical familiarity and orthography, concreteness, and polysemy,’’ J. Exp. Psychol. Gen. **113**, 256–281.

Gordon, B. (**1985**). ‘‘Subjective frequency and lexical decision latency function: Implications for mechanisms of lexical access,’’ J. Mem. Lang. **24**, 631–645.

Hintzman, D. L. (**1986**). ‘‘ ‘Schema abstraction’ in a multiple-trace memory model,’’ Psychol. Rev. **93**, 411–428.

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

Kučera, F., and Francis, W. (**1967**). *Computational Analysis of Present Day American English* (Brown University Press, Providence).

Kuhl, P. K., and Iverson, P. (**1995**). ‘‘Linguistic experience and the perceptual magnet effect,’’ in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 121–154.

Lachman, R., Shaffer, J. P., and Hennrikus, D. (**1974**). ‘‘Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time,’’ J. Verbal Learn. Verbal Behav. **13**, 613–625.

Luce, P. A., and Pisoni, D. B. (**1998**). ‘‘Recognizing spoken words: The neighborhood activation model,’’ Ear Hear. **19**, 1–36.

MacKay, I. R. A., Meador, D., and Flege, J. E. (**2001**). ‘‘The identification of English consonants by native speakers of Italian,’’ Phonetica **58**, 103–125.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). ‘‘Age of second-language acquisition and perception of speech in noise,’’ J. Speech Lang. Hear. Res. **40**, 686–693.

McClelland, J. L. (**1988**). ‘‘Connectionist models and psychological evidence,’’ J. Mem. Lang. **27**, 107–123.

Meador, D., Flege, J. E., and MacKay, I. R. A. (**2000**). ‘‘Factors affecting the recognition of words in a second language,’’ Bilingualism: Lang. Cogn. **3**, 55–67.

Metsala, J. L., and Walley, A. C. (**1998**). ‘‘Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability,’’ in *Word Recognition in Beginning Literacy*, edited by J. L. Metsala and L. C. Ehri (Erlbaum, Hillsdale, NJ), pp. 89–120.

Morrison, C. M., and Ellis, A. W. (**1995**). ‘‘Roles of word frequency and age of acquisition in word naming and lexical decision,’’ J. Exp. Psychol. Learn. Mem. Cogn. **20**, 116–133.

Munro, M. J. (**1998**). ‘‘The effects of noise on the intelligibility of foreign-accented speech,’’ Stud. Second Lang. Acquis. **20**, 139–154.

Munro, M. J., and Derwing, T. M. (**1995**). ‘‘Foreign accent comprehensibility, and intelligibility in the speech of second language learners,’’ Lang. Learn. **49**, 285–310.

Munro, M. J., and Derwing, T. M. (**2001**). ‘‘Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate,’’ Stud. Second Lang. Acquis. **23**, 451–468.

Norris, D., McQueen, J. M., and Cutler, A. (**1995**). ‘‘Competition and segementation in spoken-word recognition,’’ J. Exp. Psychol. Learn. Mem. Cogn. **21**, 1209–1228.

Pallier, C., Colome, A., and Sebastian-Galles, N. (**2001**). ‘‘The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries,’’ Psychol. Sci. **12**, 445–449.

Sebastian-Galles, N., and Soto-Faraco, S. (**1999**). ‘‘Online processing of native and non-native phonemic contrasts in early bilinguals,’’ Cognition **72**, 111–123.

Strange, W. (**1995**). ‘‘Cross-language studies of speech perception: A historical review,’’ in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 3–48.

Takayanagi, S., Dirks, D. D., and Moshfegh, A. (**2002**). ‘‘Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing,’’ J. Speech Lang. Hear. Res. **45**, 585–597.

van Wijngaarden, S. J. (**2001**). ‘‘Intelligibility of native and non-native Dutch speech,’’ Speech Commun. **35**, 103–113.

van Wijngaarden, S. J., Steeneken, H. J. M., and Hougast, T. (**2002**). ‘‘Quantifying the intelligibility of speech in noise for non-native listeners,’’ J. Acoust. Soc. Am. **111**, 1906–1916.

Washington University in St. Louis Speech and Hearing Lab Neighborhood Database as a web-based implementation of the 20 000-word Hoosier Mental Lexicon (2002, March; accessed). http://128.252.27.56/neighborhood/Home.asp

Werker, J. F., and Polka, L. (**1993**). ‘‘Developmental changes in speech perception: New challenges and new directions,’’ J. Phonetics **21**, 83–101.

Wright, C. E. (**1979**). ‘‘Duration differences between rare and common words and their implications for the interpretation of word frequency effects,’’ Mem. Cognit. **7**, 411–419.

Yeni-Komshian, G. H., Flege, J. E., and Liu, S. (**2000**). ‘‘Pronunciation proficiency in the first and second languages of Korean-English bilinguals,’’ Bilingualism: Lang. Cogn. **3**, 131–149.