Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners

Satomi Imai†, James E. Flege† and Amanda Walley‡
† Div. of Speech and Hearing Sciences, Univ. of Alabama at Birmingham, Birmingham, AL, USA
‡ Dept. of Psychology, Univ. of Alabama at Birmingham, Birmingham, AL, USA
E-mail: imais@uab.edu

ABSTRACT

We examined the effect of a “mismatch” between a listener's phonological representations and speech input on spoken word recognition. Native Spanish (NS) and native English (NE) listeners were asked to write down 80 English words that were presented in noise. The words differed in neighborhood density (ND: dense versus sparse); half were Spanish-accented (produced by a NS talker), the other half were unaccented (i.e., produced by a NE talker). We hypothesized that phonological mismatches would occur when NS listeners responded to unaccented words and when NE listeners responded to Spanish-accented words. Further, the effect of the mismatch was expected to be greater for words from dense versus sparse neighborhoods because these words can be confused with many minimally-paired neighbors. The results supported the mismatch hypothesis. NS listeners showed a larger ND effect for unaccented than Spanish-accented stimuli, whereas NE listeners showed a larger ND effect for Spanish-accented than unaccented stimuli.

1. INTRODUCTION

Speech perception becomes attuned to the vowels and consonants of the native language (L1) during infancy and childhood. Long-term memory representations based on the language-specific properties of L1 vowels and consonants then facilitate the recognition of L1 words. However, the phonetic inventories of any two languages are likely to differ in terms of the number and nature of the contrastive phonetic units used to distinguish word meanings. This may cause difficulty for individuals who learn a second language (L2) in adulthood. L2 learners may have L2 phonological representations that are influenced by their L1 phonology, and so differ from native speakers' representations. A failure by L2 learners to re-attune their perception of speech to the language-specific properties of L2 vowels and consonants explains, in part, why they are less successful in recognizing L2 words than native speakers are, especially in non-ideal listening conditions [1].

In addition to segmental perception differences, lexical factors may affect spoken word recognition. Word frequency and neighborhood density are known to affect the speed and accuracy of spoken word recognition [2]. Word frequency refers to how often a word has been encountered (here defined by text frequency, TF). Neighborhood density (ND) is a measure of the number of phonologically similar, and thus potentially confusable, neighbors for a particular word (i.e., that differ by only one segment). High frequency words are recognized faster and more accurately than low frequency words; words of low ND are recognized faster and more accurately than high ND words.

Bradlow and Pisoni [3] examined the recognition of “easy” English words (i.e., high frequency words of low ND) and “hard” English words (low frequency word of high ND) by NE and non-native listeners. The hard words were expected to be more difficult to recognize than the easy words because they occur less frequently and have more neighboring words than easy words. In contrast, easy words occur more frequently than hard words. As easy words occur in sparse lexical neighborhoods, they are more phonologically distinctive. Hard words were found to be less intelligible than easy words for both listener groups, but the lexical effect (i.e., the word recognition difference between hard and easy words) was stronger for non-native than NE listeners. Bradlow and Pisoni concluded that the lexical effect was stronger for non-native than NE listeners because the nonnative listeners had reduced sensitivity in the fine phonetic details needed to discriminate the hard words from the many other English words differing by a single segment.

The aim of this study was to further investigate interactions between top-down processing (lexical factors) and bottom-up processing (segmental perception) on word recognition. We hypothesized that a mismatch between speech input and listeners' phonological representations would reduce word recognition accuracy. Our stimuli were English words produced by a NE talker and by a NS talker with a Spanish-accent (designated “unaccented” and “accented”). Two groups of listeners participated: NE listeners from Alabama, USA, and NS listeners from Spanish-speaking countries who were living in Alabama. The stimulus set consisted of 80 familiar English words that varied orthogonally in TF and ND. The NE and NS participants were asked to identify 80 stimulus words (half of which were unaccented, half of which were accented).

We started with the assumption that NS listeners would

ISBN 1-876346-48-5 © 2003 UAB

possess phonological representations for English words that were influenced by their L1 (Spanish), and thus differed from NE listeners' representations. We also assumed that there would be a poorer phonological match between NS listeners' phonological representations and segments in the unaccented (native-produced) word stimuli than in the Spanish-accented stimuli, whereas the reverse would hold true for NE listeners. It seemed likely that NS listeners would have more difficulty recognizing unaccented than Spanish-accented words, whereas NE listeners would have more difficulty in recognizing English words spoken with a Spanish-accent than words spoken without an accent.

We hypothesized that a "phonological mismatch" would impair word recognition, and that the effect of phonological mismatches would be greater for hard than easy words. We reasoned that the recognition of a word with many high-frequency neighbors would require an accurate specification of its constituent vowels and consonants because, without accurate phonetic specifications, the word would be confused with one or more of its minimally paired neighbors. This difficulty in spoken word recognition would be exacerbated for low frequency words with many neighbors as compared to high frequency words with few neighbors. If a NS listener lacked phonological representations that were fine-tuned to English (e.g., if the sound representations for /b/ and /æ/ were not English-like) it should be difficult for a listener to recognize a hard word like "bad" as opposed to words like "pad" (/b/ vs. /p/), "dad" (/b/ vs. /d/) or "bed" (/æ/ vs. /ɛ/). Difficulty in fine sound discrimination might not be as important for the recognition of easy words (e.g., "desk" versus "bad"), however.

If our phonological mismatch hypothesis is correct, NS listeners should recognize fewer English words produced without accent than with a Spanish-accent. Conversely, NE listeners should recognize fewer English words spoken with a Spanish-accent than without accent. The phonological mismatch effect should be greater for hard words than easy words for both listener groups.

2. METHOD

2.1. Participants

Twenty NE speakers, all born and raised in Alabama, had an average age of 32 years (range: 22 to 47). The NS participants were born in predominantly Spanish-speaking countries, and had an average age of 32 years (range: 23 to 47). The 20 NS participants arrived in the U.S. at a mean age of 28 years (range: 20 to 41) and had lived in Alabama for an average of 5 years (range: 2 to 8). All participants passed a pure-tone hearing screening and demonstrated knowledge of more than 95 % of the 80 test words (see below).

2.2. Stimuli and Task

The 80 English test words varied orthogonally in TF (high vs. low) [4] and ND (high vs. low) [5], yielding four lexical sets. All 80 items were familiar words [6]. To ensure that the stimulus words were known by all listeners, a lexical knowledge test was administered following the word recognition task.

Table 1 presents summary statistics for the four lexical sets. The 80 words were recorded by a male NE speaker (unaccented speech) and a male NS speaker (Spanish-accented speech). The 160 stimuli (80 x 2 talkers) were digitized, then normalized for peak intensity (50% of full scale). Multi-talker babbling noise (scaled to 10% of the full scale) was added to the stimuli to prevent ceiling effects.

	High ND		Low ND	
	High TF	Low TF	High TF	Low TF
Mean ND	23.5 (17-39)	23.8 (19-30)	10.0 (3-15)	10.4 (4-14)
Mean TF	169.5 (54-500)	18.0 (3-37)	177.8 (58-591)	22.4 (1-41)
Mean familiarity	6.7 (5.5-7.0)	6.5 (5.2-7.0)	6.8 (5.7-7.0)	6.6 (5.0-7.0)

Table 1: Test word characteristics. Ranges are shown in parentheses.

Two blocks (A, B) of 40 test words (10 randomly selected from each of the four lexical sets) were chosen. Within each block, order of stimulus presentation was randomized. Blocks A and B consisted of either all unaccented stimuli (A-unacc, B-unacc) or all accented stimuli (A-acc, B-acc), respectively. Order of presentation of Stimulus Type (unaccented vs. accented) was counterbalanced. Participants were randomly assigned to one of four counterbalancing conditions: A-unacc/B-acc, A-acc/B-unacc, B-unacc/A-acc, or B-acc/A-unacc. Thus, each listener heard all 80 test words, half Spanish-accented and half unaccented, once each.

The listeners were tested individually in a sound booth, where the stimuli were presented one at a time via loudspeakers. The listeners were told to write down each word they heard. There was a short demonstration of the task by the experimenter and then a 16-item practice session before each block. None of the practice words were among the 80 test words. After the word recognition task was completed, the participants were asked if they knew each of the 80 test items (a binary choice) and, if so, to rate it on a scale ranging from 1 ("seldom hear/say this word") to 7 ("often hear/say this word"). Non-word foils were included among the test items to ensure that participants would veridically report any test items they did not know.

3. RESULTS

Figure 1 shows the mean percent correct recognition scores obtained by the NE and NS listeners for the unaccented and Spanish-accented stimuli as a function of ND. The four scores obtained for each participant were submitted to a mixed-design ANOVA. Listener Group (NE, NS) served as a between-subjects factor and Stimulus Type (unaccented, Spanish-accented), TF (high, low) and ND (high, low) were within-subjects factors. The ANOVA yielded significant main effects of Group, F(1, 38)=10.7, p<.005, Stimulus Type, F(1, 38)=86.8, p<.001, and ND, F(1, 38)=37.5, p<.001, but not TF. The NE listeners recognized more words overall than the NS listeners did (66% vs 59%). The recognition scores were higher for unaccented than Spanish-accented stimuli (73% vs 52%), and higher for low ND than for high ND words (67% vs 57%). A two-way interaction, Group x Stimulus Type, was significant, F(1, 38)=37.5, p<.001. Simple effects tests revealed that the NE listeners recognized more unaccented words than the NS listeners did (p<.001), whereas the NE listeners recognized fewer accented words than the NS listeners did (p<.10).

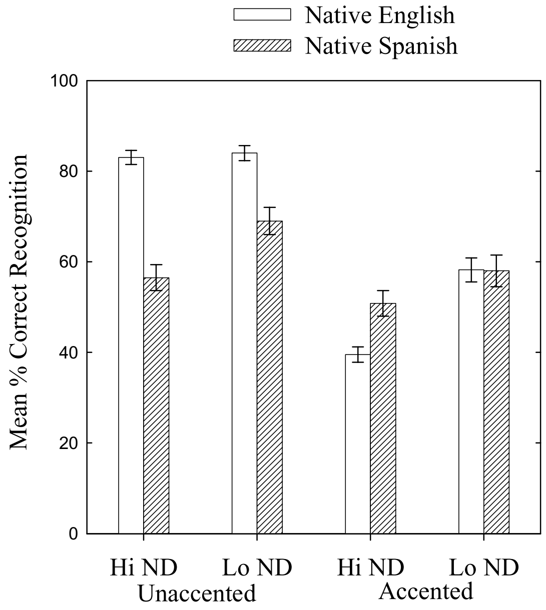

Figure 1: Mean % correct scores as a function of ND and Stimulus Type for the NE and NS listeners.

The two-way Stimulus Type x TF interaction was also significant, F(1, 38)=14.0, p<.001. Figure 2 shows this effect averaged across listener groups (NE, NS). Simple effect tests revealed that the TF effect was significant for the Spanish-accented stimuli. Fewer accented stimuli with low TF were recognized than accented stimuli with high TF (p<.01). For unaccented words, a countervailing TF effect was marginally significant (p<.06).

Finally, the three-way interaction between Group, Stimulus Type and ND was significant, F(1, 38)=10.28, p<.01 (see Figure 1). Simple effect tests indicated that there were significant differences between high versus low ND accented stimuli for the NE listeners (p<.001) and between high versus low ND unaccented stimuli for the NS listeners (p<.001). As predicted by the phonological mismatch hypothesis, the NE listeners showed a larger effect of ND for Spanish-accented than unaccented stimuli; the reverse held true for the NS listeners. The ND effect was present only when there were, presumably, mismatches in phonological representations and the phonetic realizations of stimuli (i.e., the NE listeners identifying accented stimuli, the NS listeners identifying unaccented stimuli). Surprisingly, the difference between high and low ND unaccented stimuli for the NE listeners was non-significant, contrary to the results of Bradlow and Pisoni [3]. This cross-study difference will be discussed below. No other main effects or interactions reached significance.

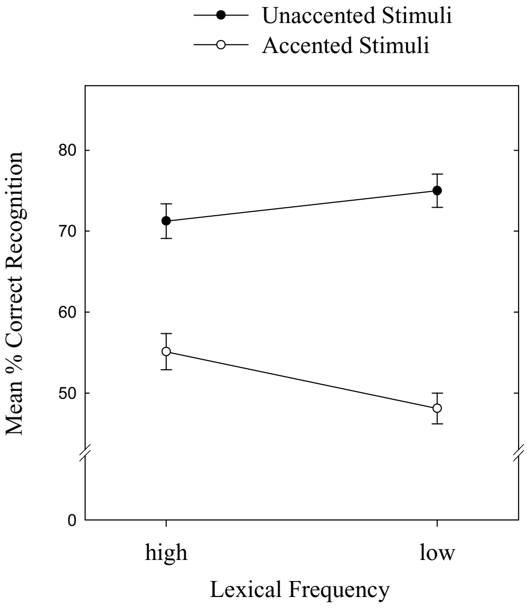

Figure 2: Mean % correct scores in the high and low TF conditions.

4. DISCUSSION

The present study examined the relation between speech input and listeners' phonological representations by evaluating interactions between lexical factors (TF, ND) and the type of speech input (unaccented, Spanish-accented). We tested a "phonological mismatch" hypothesis. According to this hypothesis, word recognition accuracy will be reduced when there is a mismatch between the phonetic specification of incoming words and a listener's phonological representations. We expected the effect of mismatches to be greater for hard than easy words because more fine-grained phonetic representations are needed to accurately recognize hard words owing to their

ISBN 1-876346-48-5 © 2003 UAB

high level of confusability with lexical neighbors.

The NE listeners recognized more words overall than the NS listeners did. There were no significant interactions between TF and ND, suggesting that these lexical factors had independent effects. In general, more low ND words were recognized than high ND words. High TF words were correctly recognized more often than low TF words only when words were Spanish-accented. The lack of a TF effect on unaccented words might be due to the fact that our 80 word stimuli did not include words of very high or low frequency. Also, there is some evidence that the effect of word frequency on spoken word processing may be minimal [7]. Unlike the effect of ND, the TF effect did not interact with listener group. Perhaps an effect of TF was not obtained here for unaccented stimuli because the NE talker who produced them hyperarticulated low frequency words compared to high frequency words, thereby making them easier to recognize.

The results obtained here supported the phonological mismatch hypothesis. The results also showed that the two listener groups exhibited different mismatch patterns. A greater ND effect was found when the NS listeners heard unaccented words than Spanish-accented words. In contrast, the ND effect was greater when the NE listeners heard Spanish-accented words than unaccented words. High ND words have more similar-sounding neighbors, and so may require more accurate segmental representations than less confusable, low ND words. The pattern of results obtained here showed that the effect of ND was greatest when the sound segments of stimuli presumably did not match the listeners' long-term memory representations.

Previous studies have shown that nonnative listeners have difficulty recognizing sentences in an L2 in noise as compared with native listeners [1,8]. Our study demonstrated that this difficulty might originate in the recognition of words at the segmental level due to their L2 phonological representations under influence of their L1 phonology.

The present finding replicated the findings of Bradlow and Pisoni [3], showing that the ND effect was greater for NS than NE listeners for unaccented words. However, in our study, the recognition scores of NE listeners for low ND and high ND were virtually identical, whereas Bradlow and Pisoni found a significant difference between the two conditions for NE listeners. This difference might be due to stimulus word selection. The stimulus words used in this study were pre-screened in pilot studies to ensure the test words would be known by NS listeners who were non-proficient in English. As a result, infrequent words were excluded from the test words. The mean and the median of frequency count of hard words used by Bradlow and Pisoni was 12.1 and 3, respectively, whereas in this study the hard words (in the high ND-low TF list condition) had a mean frequency count of 18.0 and a median of 18. Although the recognition scores for the NE group were below ceiling because noise was added to the stimuli, the ability of NE speakers to discern fine phonological differences among similar-sounding words for familiar items may be robust.

In conclusion, the present study demonstrated that a lexical factor, ND, had an effect on spoken English word recognition in noise for both NS and NE listeners. When recognition of words required finer segmental discrimination (i.e., high ND words), a mismatch between speech input and listeners' phonological representations reduced the accuracy of word recognition in noise. Additional work will be needed to determine if NS listeners' English phonological representations change as their proficiency in L2 increases, and if support for the phonological mismatch hypothesis can be obtained for learners of an L2 other than English.

ACKNOWLEDGEMENTS

This study was supported by NIH grant DC00257. We thank Ana Castaneda for the recruitment of participants.

REFERENCES

[1] D. Meador, J.E. Flege and I.R. MacKay, "Factors affecting the recognition of words in a second language", *Bilingualism: Language and Cognition*, vol. 3, pp. 55-67, 2000.

[2] P.A. Luce and D.B. Pisoni, "Recognizing spoken words: The neighborhood activation model", *Ear and Hearing*, vol. 19, pp. 1-36, 1998.

[3] A.R. Bradlow and D.B. Pisoni, "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors", *Journal of the Acoustical Society of America*, vol. 106(4), pp. 2074-2085, 1999.

[4] F. Kucera and W. Francis*, Computational Analysis of Present Day American English*, Providence, RI: Brown U. P., 1967.

[5] *Neighborhood Database*, Speech and Hearing Lab, Washington University in St. Louis.

[6] H.C. Nusbaum, D.B. Pisoni and C.K. Davis, "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words", *Research in Speech Perception, Progress Report 10,* Bloomington, IN: Indiana University, 1984.

[7] V.M. Garlock, A.C Walley and J.L. Metsala, "Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults", *Journal of Memory and Language*, vol. 45, pp. 468-492, 2001.

[8] A.R. Bradlow and T. Bent, "The clear speech effect for non-native listeners", *Journal of the Acoustical Society of America*, vol. 112, pp. 272-284, 2002.