The effect of word learning on the perception of non-native consonant sequences

Lisa Davidson,[a] Jason Shaw, and Tuuli Adams
New York University, New York, New York 10003

(Received 4 December 2006; revised 28 August 2007; accepted 1 October 2007)

Previous research in cross-language perception has shown that non-native listeners often assimilate both single phonemes and phonotactic sequences to native language categories. This study examined whether associating meaning with words containing non-native phonotactics assists listeners in distinguishing the non-native sequences from native ones. In the first experiment, American English listeners learned word-picture pairings including words that contained a phonological contrast between CC and CVC sequences, but which were not minimal pairs (e.g., [ftake], [fətalu]). In the second experiment, the word-picture pairings specifically consisted of minimal pairs (e.g., [ftake], [fətake]). Results showed that the ability to learn non-native CC was significantly improved when listeners learned minimal pairs as opposed to phonological contrast alone. Subsequent investigation of individual listeners revealed that there are both high and low performing participants, where the high performers were much more capable of learning the contrast between native and non-native words. Implications of these findings for second language lexical representations and loanword adaptation are discussed.
[DOI: 10.1121/1.2801548]

PACS number(s): 43.71.Hw, 43.70.Mn, 43.71.Es [ARB] Pages: 3697–3709

I. INTRODUCTION

Research in cross-language speech perception has shown that listeners faced with perceiving non-native sounds often have trouble distinguishing between a native phoneme and a non-native phoneme that is similar to it, or between two similar unfamiliar sounds that are both close to a native phoneme that is acoustically similar (e.g., Best, 1995; Cebrian, 2000; Flege *et al.*, 2003; Kuhl and Iverson (1995)). For example, a classic finding by Werker and Tees (1984) demonstrated that English-speaking adults have considerable difficulty in distinguishing between a dental /t/ and a retroflex /ʈ/ or the velar ejective /k′/ and the uvular ejective /q′/. Models of cross-language speech perception such as the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995) claim that the phonetic similarities between non-native segments and native ones, defined in terms of acoustic and articulatory characteristics, are predictive of whether or not listeners will be able to develop new perceptual categories for these sounds. Other studies demonstrate that the perception of a non-native segment is affected by prosodic position or surrounding segmental context (e.g., Strange *et al.*, 2001; 2004). For example, it is more difficult for Japanese speakers to distinguish between English /r/ and /l/ in prevocalic than in postvocalic contexts (Lively *et al.*, 1993).

In general, the focus of the cross-language speech perception literature has been on the perception of individual phonemes. A smaller but important set of studies have examined the perception of non-native phonotactics; that is, cases where the phonemes in question may be allowed by the native phonology of the participant, but not in the same sequential combinations allowed in the language being learned (or tested). As pointed out by Peperkamp (2007), there are a number of ways in which listeners may perceive non-native consonant sequences. Some studies have shown that listeners whose language contains at least one type of obstruent +approximant sequence are more likely to perceive an unattested obstruent+approximant sequence as an instance of a legal phonotactic pattern (Hallé *et al.*, 1998; Massaro and Cohen (1983); Moreton, 2002; Pitt, 1998). For example, Hallé *et al.* (1998) found that French listeners perceived natural tokens of /tl/ and /dl/ sequences as /kl/ and gl/, respectively. Massaro and Cohen (1983) demonstrated that English speakers listening to a sound intermediate between /l/ and /r/ more often labeled the sound as /r/ when in the environment of /t_i/ and as /l/ in /s_i/.

Studies that have examined obstruent-obstruent or obstruent-nasal sequences have indicated a different type of perceptual response. A series of studies by Dupoux and colleagues (Dehaene-Lambertz, Dupoux, and Gout, 2000; Dupoux *et al.*, 1999, 2001) and follow-up studies by Kabak and Idsardi (2007) and Berent *et al.* (2007) using both identification and discrimination tasks indicate that another way in which listeners may process non-native CC sequences is to assimilate them to phonotactically permissible CVC sequences. For example, Japanese listeners frequently categorize items like [ebzo] as [ebuzo]. However, the likelihood of occurrence for "perceptual epenthesis" may be conditioned by the native language of the listener and the particular composition of the sequence. For example, whereas perceptual epenthesis was high for Japanese speakers in most obstruent-obstruent contexts, Davidson (2007) showed that English speakers transcribing non-native fricative-initial consonant

[a]Electronic mail: lisa.davidson@nyu.edu

sequences wrote a vowel between the consonants less than 15% of the time, which is much lower than the rate of perceptual epenthesis for Japanese listeners.

The fact that non-native listeners have difficulty distinguishing certain contrasts not found in their own language also has important ramifications for the nature of bilingual or second language lexical entries. Previous research on the phonetic characteristics of lexical representations of experienced second language learners has shown that the phonetic specifications of these representations are not as detailed as native language lexical entries (MacKay *et al.*, 2001; Sebastián-Gallés and Soto-Faraco, 1999). For example, Bradlow and Pisoni (1999) demonstrated that non-native listeners performing word recognition tasks are negatively affected when factors such as phonological neighborhood density are manipulated; non-native listeners have difficulty recognizing even familiar words that have many phonetically similar neighbors (see also Imai *et al.*, 2005).

In the present experiments, English-speaking adults were taught sets of word-meaning pairs that differ only with respect to whether or not they contained a phonotactically possible initial CVC (e.g., [zəgamo]) or an impossible CC sequence (e.g., [zgamo]). Unlike previous studies with adults that provided little incentive to distinguish between these minimally different words (e.g., Davidson, 2007; Dupoux *et al.*, 1999), it was hypothesized that associating pictures to unfamiliar and non-native words would actually assist the adults in attending to the critical acoustic detail differentiating these words. Two previous adult picture-word learning tasks using both native and non-native stimuli demonstrate that it is possible for adults to learn new words and even new non-native contrasts (Curtin *et al.*, 1998; Storkel *et al.*, 2006), but these studies did not explicitly test whether picture-word learning causes discrimination to improve over auditory-only tasks. A more relevant study by Hayes-Harb (2007) indicated that when adult learners are trained on a novel phonemic contrast (e.g., voiced [g] versus unaspirated [k] for English listeners), those participants who were taught minimal pairs in conjunction with pictures were more accurate in a following AX discrimination task than those who were trained using only statistical information.

Following the phonemic category results from Hayes-Harb (2007), it is expected that adults will still show phonological interference from native phonotactics, but that by matching words with meaning, they will have incentive to tease apart the fine phonetic differences between attested CVC and phonotactically illegal CC sequences. The stimuli used in this study and the comparison of word-learning with transcription and discrimination tasks are based on previous studies that also investigated the production and perception of words with fricative-initial CC and CVC sequences (Davidson, 2006, 2007). In the following section, these studies will be briefly reviewed to introduce the stimuli that are employed in the present experiments and to situate them within a larger picture of language acquisition, loanword adaptation, and cross-language speech perception.

Production and perception of fricative-initial sequences. The purpose of Davidson (2006) was to examine the nature of native English speakers' repairs of phonotactically unattested initial CC sequences. Based on previous research demonstrating that speakers often insert vocalic material into non-native consonant clusters (Broselow and Finer, 1991; Davidson *et al.*, 2004), the study was designed to determine whether the vocalic material produced between the two consonants is (a) phonological vowel epenthesis, or (b) a result of insufficient articulatory overlap of the two consonant articulations, which is realized as a transitional schwa due to a period of open vocal tract between the consonants. English speakers were asked to repeat fricative-initial CC and matching CəC sequences produced by a Slovak speaker (e.g., [zgamo] and [zəgamo]). Accurate production of the initial cluster depended on whether the sequence began with [f], [z], or [v], but participants produced 45% of the tokens without vocalic material between the consonants and approximately 30% of the tokens with insertion of vocalic material (these tokens will be referred to in the text as C^{ə}C). Acoustic measurements of the inserted schwas (C^{ə}C) and lexical schwas (CəC) showed that inserted C^{ə}C schwas were shorter and had a lower F1 midpoint, which is consistent with a brief transition between two constrictions but not with movement toward an open vowel target.

Given that English speakers produced C^{ə}C tokens as a considerable proportion of their attempts to produce CC sequences, the output of Davidson (2006) is an interesting test case for the role of perception in loanword adaptation (e.g., Broselow, 2004; Kang, 2003; Peperkamp and Dupoux, 2003). That is, when presented with CC sequences, English speakers produce a mixture of CC and C^{ə}C utterances. If these English speakers are considered as a group of initial borrowers ("disseminators"), then we can ask how a larger group of monolingual speakers ("recipients") learning these loanwords from the disseminators would interpret their productions of CC and C^{ə}C. Davidson (2007) used the output of the speakers of the production study to investigate how recipients perceive phonotactically illegal CC sequences produced by native English speakers, and how they classify C^{ə}C sequences. They may treat any vocalic material as a vowel, or they may recognize that the acoustic properties of [^{ə}] are different from those of [ə], leading them to recover the disseminators' attempt to produce a target CC sequence.

Davidson (2007) used both a transcription task and an AX discrimination task to examine these questions. In the transcription task, participants were presented with CCVCV, CəCVCV, and C^{ə}CVCV utterances produced by one of the English speakers from Davidson (2006). Participants wrote CC tokens as a cluster and CəC with a vowel each about 70% of the time, but C^{ə}C tokens were written with a vowel only 54% of the time. These results suggest that recipients recognize that C^{ə}C tokens tend to be different from tokens with a lexical vowel, but nevertheless may have a slight tendency to categorize the vocalic material as a vowel since the task encourages them to assume only two categories. In the AX discrimination task, listeners had trouble distinguishing between all of the possible pairings (i.e., CC/C^{ə}C, CəC/C^{ə}C, and even CC/CəC), but were significantly worse in the discrimination of CəC/C^{ə}C as opposed to CC/C^{ə}C.

While both results suggest that recipients treat C^{ə}C as a variant of both CəC and CC, the findings of Davidson (2007) indicate that speakers may even extend this confusion to contrasting CəC and CC pairs if they are not presented in a task that forces them to provide the stimuli with a linguistic representation. In the present set of experiments using an experimental paradigm other than a discrimination task, this hypothesis is tested with a picture-word learning task. In the first experiment, participants are taught CəC and CC-initial words that contain the same consonants but that are not minimal pairs (e.g., [zdati] and [zədanu]) in conjunction with pictures. This experiment tests whether the presence of phonological contrast alone, in conjunction with meaning, is enough to improve discrimination between CəC and CC. In the second experiment, the CəC and CC stimuli are taught as minimal pairs. In both experiments, only CəC and CC are explicitly taught in the training phases, but C^{ə}C tokens are added as a condition in the test phase in order to determine how listeners classify these sequences once they have established representations of the CəC and CC words. The results of the current experiments will be discussed with respect to the AX discrimination task from Davidson (2007) to evaluate whether learning sound in conjunction with meaning improves discrimination. Note that with respect to the C^{ə}C items, the purpose of these experiments is not to try to teach a three-way distinction between phonotactic categories, but rather to further study how recipients may treat disseminators' C^{ə}C tokens in loanword adaptation.

II. EXPERIMENT 1

A. Method

1. Participants

The participants were 40 native American English speakers who were recruited from Craigslist and from flyers posted around New York University. No participant had any experience with Slavic, Hebrew, or any other languages that have the initial obstruent clusters used in the experiment. No participants were phonetically trained. None reported any history of speech or hearing impairments. Eight further participants began the experiment but did not complete the training phase (see Sec. II A 3).

2. Materials

Two sets of stimuli were used in Experiment 1; one for the training phase and one for the test phase. Both sets of stimuli were the same subset of the utterances produced by native English speakers who were participants in Davidson (2006), but there was a different speaker for the training and test phases in Experiment 1. In Davidson (2006), participants were asked to produce both words with fricative-initial sequences (i.e., [zgamo], [fnada]) and their counterparts with a schwa (i.e., [zəgamo], [fənada]) that were originally presented to them as spoken by a native Slovak speaker. Although Slovak does not allow word-initial CəC sequences, the speaker was a phonetician and proficient English speaker, and was able to produce CəC with a reduced schwa.

In Davidson (2006), the Slovak pseudowords were presented to the English speakers over computer speakers, accompanied by a written representation of the word in English orthography on a computer screen. The participants then repeated each word aloud once. The participants' responses that were used for both the training and test stimuli were recorded at 22 kHz in a soundproof booth with a Marantz PMD-680 digital solid state recorder and an Audio-Technica AT813 microphone. All of the CC words were bisyllabic, with main stress on the first syllable (e.g., [zgámo]). The CəC words were all trisyllabic, with stress on the second syllable (e.g., [zəgámo]). Participants produced the CəC without any difficulty, but alternately produced the CC sequence accurately (no vocalic material between the consonants), or with a transitional schwa between the consonants (C^{ə}C).

	CC	CəC
Block 1	[vbano]	[vəbaki]
	[vgane]	[vəgapo]
	[fnagu]	[fənapi]
	[zbagi]	[zəbasi]
Block 2	[zmapi]	[zəmagu]
	[zdati]	[zədanu]
	[vzaku]	[vəzaba]
	[ftake]	[fətalu]

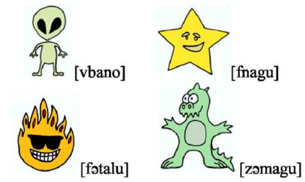

FIG. 1. (Color online) CC and CəC stimuli (left) and example picture-word pairings (right) used in the training and test phase.

a. Training stimuli In the current experiment, the training stimuli consisted of eight CC-initial words and eight CəC-initial words, each of which were paired with a unique cartoon character. For the training stimuli in the current study, if the speaker produced CC with any vocalic material between the consonant cluster, it was excised out in Praat by cutting at zero crossings to prevent acoustic artifacts in the signal. The list of stimuli and some example picture-word pairings are shown in Fig. 1.

In addition to the picture-word pairs, there were also minimal pair distracters in the form of CC-initial words that matched the words learned as CəC, and CəC-initial words that matched the CC words. For example, the word [zəmagu] was learned as the name of the dragon cartoon character, so [zmagu] was presented as a minimal pair distracters. The minimal pair distracters were produced by the same speaker as the target items. The minimal pair distracters were intended to teach the participants that words that differed with respect to the presence or absence of a schwa are not variations in production of the same lexical item. However, no picture was ever associated with the minimal pair distracter item.

b. Test stimuli The test stimuli were a set of words related to the learned items. The stimuli were divided into four conditions: Condition A, where the learned word was paired with a totally unfamiliar CC or CəC word; Condition B, where the C^{ə}C-initial word corresponding to the learned target word was paired with an unfamiliar word; Condition C, where the minimal pair distracter word was paired with an unfamiliar word; and Condition D, where two unfamiliar words were presented. The division of each stimulus item into experimental conditions is given in the Appendix. As explained further in Sec. II A 3, participants were played both words in conjunction with a picture and chose either one of the words or a "neither" option. Each of these test productions, including the C^{ə}C words, were taken from the productions of a different English speaker from Davidson (2006) than the speaker used in the training. This speaker was chosen because all of his repetitions of the CC clusters used in the current experiment were repaired by inserting a transitional schwa. Thus, the matching CC words were obtained by splicing out the schwa as described earlier. An acoustic analysis of the words demonstrated a substantial difference between the duration of the lexical [ə] and inserted [ᵊ] (lexical: $M = 50$ ms, s.d. = 15.5 ms; inserted: $M = 41$ ms, s.d. = 12.7 ms) and the F1 value at the midpoint (lexical: $M = 438$ Hz, s.d. = 67 Hz; inserted: $M = 390$ Hz, s.d. = 68 Hz).

It should be noted that the stimuli shown in Fig. 1 represent a diverse set of consonant sequence types. The particular CC and CəC sequences were chosen for two main reasons. First, since most word-learning studies have small stimulus sets and it takes some time for participants to be adequately trained, we did not want to use extremely similar stimuli (e.g., all /f/-initial sequences) which may have made the task even more difficult. Second, the results of the transcription task in Davidson (2007) showed no effect of the first consonant type on transcribers' accuracy on /f/, /z/, and /v/-initial consonant sequences. While Davidson (2006) did show a significant effect of C1 type on production accuracy, this factor did not seem to carry over to how speakers perceive words with the same non-native sequences. Thus, we chose to use stimuli beginning with a variety of consonant sequence types.

c. Practice stimuli Finally, a set of stimuli were also recorded to use in a practice session intended to familiarize the participants with the task. Four Spanish words were chosen for the training phase: *botas*, *mapa*, *casa*, and *mano*. These words were recorded by a native speaker of English who had studied Spanish in high school and college. The test stimuli, recorded by a different native English speaker who had also studied Spanish in high school and college, consisted of the four training words plus four words that were similar except for a different first phoneme: *notas*, *papa*, *taza*, and *sano*.

3. Procedure

There were three phases in the experiment: A familiarization phase, a training phase, and a test phase. In order to simplify the learning task, participants learned eight word-picture pairings at a time. After the first familiarization-training-test sequence, the participants were given a break and returned to complete the second half. A total of 16 words were learned. The experiment was programmed using E-PRIME.

In order to ensure that participants could complete the task in no more than the hour they were expected to participate for, E-PRIME tracked how long the participants were taking. If they could not complete the first training phase within 45 min, they were thanked for their participation and told they had completed the requirement.

a. Familiarization phase Each of the eight picture-word pairs appeared on the screen twice in random order. While the cartoon character appeared on the screen, the sound file containing the name of the character played once. In the instructions for the familiarization phase, listeners were told that they would hear words spoken by an English speaker, some of which were more plausible English words than others, and that they should try to memorize each character-word pairing. Nevertheless, the familiarization stage was intended to introduce the picture-word pairs, while learning the names actually occurred during the training phase.

b. Training phase Once the familiarization phase was completed, participants automatically moved on to the training phase. The eight cartoon pictures appeared on the screen in two rows, along with a black X on the top right side of the screen. Subjects then heard a word, and had to use the mouse to click on the picture that corresponded to that word from the familiarization phase. When a participant clicked on the correct picture, feedback was given by displaying a green checkmark visually on the screen. If a participant clicked on the incorrect picture, the feedback displayed a red "X," and the cartoon picture was shown again in conjunction with the correct sound file. When the minimal pair distracters were presented, participants were expected to click on the black X. For example, if the participant heard the word [fnagu], the correct response would be to click on the picture of the yellow star. However, if the participant heard [fənagu], then the correct response was to click on the black X. If they did not choose the black X for the minimal pair distracters, they received the red X feedback.

In addition to the feedback, there was also an "escape hatch" provided to allow participants to review the picture-word pairs. By clicking on the word "Practice" at the bottom of the screen, participants were brought to a screen where they could click on a picture and hear the corresponding name. There was no limit to how long they could stay on the review screen.

The training session continued until all eight picture-word pairs were correctly matched in a row. After getting positive feedback, participants were told how many they had gotten right in a row. Choosing the wrong picture for a learned word or using the escape hatch reset the count, and participants had to start over again to get all 8 pairs in a row. To decrease the time spent in the training phase, incorrectly choosing a picture when presented with a minimal pair distracter did not reset the count, but negative feedback was given.

c. Test phase In this phase, a cartoon character appeared on the screen. Participants heard two words and were told to press the first key on a button box if the first word matched the picture, the second key if the second word matched, and the third key if neither word matched. The stimuli were divided up among the conditions, so that there were two CC items and two CəC items in each condition. Thus, after each of the familiarization and training phases, there are 8 test trials, for a total of 16 trials at the end of the experiment. Each condition was counterbalanced for stimulus order. A complete list of the conditions, learned words, stimuli pairs in the test phases, and the correct answers is in the Appendix.

In addition to the instructions regarding the button box, the participants were told that the English speaker for the test trials would be different from the one in the training phase. Though they had learned in training to reject the minimal pair distracters, they were also provided with the following information: "The choices may include exactly the same word that you learned for the character, may be completely different from any of the words that you learned, or may differ slightly from the word that you learned. If one option does not sound exactly like the word that you learned (e.g. a consonant or vowel sound is different), you should choose 'neither.' "

d. Practice The first familiarization/training/test sequence was preceded by a practice session using real Spanish words and the pictures they corresponded to. This was intended to introduce participants to the structure of the task. The familiarization phase consisted of showing each of the four picture-word pairs twice, followed by a training phase in which the participant heard one of the Spanish words they had just been familiarized with (e.g., *casa*), and had to click on the correct picture (e.g., a house) that is within an array of four pictures. Finally, in the test phase, a picture appeared followed immediately by two words (e.g., [kasa] [tasa]) and participants chose the appropriate button on the button box.

After the practice, the first familiarization/training/test sequence was presented. Participants were then given a short break and then continued to finish the second familiarization/training/test sequence.

In the following section, predictions for the experimental conditions are presented.

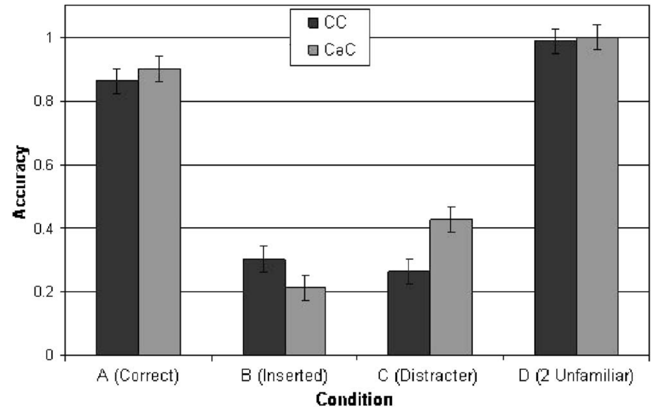

FIG. 2. Accuracy for each of the four conditions of auditory word pairs. The black bars represent the CC words that were learned, and the gray bars are the CəC words. Error bars indicate standard error.

B. Predictions

1. *Condition A: Correct Name*, *Unfamiliar Word*. Participants should choose the correct word on most trials.

2. *Condition B*: *CᵊC*, *Unfamiliar Word*. There are three possibilities:

a. If listeners identify CᵊC with CC, they should choose the CᵊC utterance more often for words learned as CC than for words learned as CəC.
b. If listeners identify CᵊC with CəC, they should choose the CᵊC utterance more often for words learned as CəC than for words learned as CC.
c. If listeners cannot distinguish CᵊC from either CC or CəC, then they should choose CᵊC equally often regardless of the phonotactic type of the learned word.

3. *Condition C: Minimal Pair Distracter*, *Unfamiliar Word*. Since participants should have learned to distinguish between learned words and their minimal pair distracter counterparts in the training phase, they should be reasonably accurate regardless of the phonotactic type of the learned word. However, since minimal pairs are not explicitly presented in the test phase, it is likely that the participants could be led astray in this condition.

4. *Condition D: Two Unfamiliar Words*. Participants should choose "neither" on most trials.

C. Results

Here, we report on results for accuracy in the four conditions presented earlier. Although reaction time data were collected, it is not reported on because the only significant difference was for Condition D (Two Unfamiliar Words), which was significantly faster than all other conditions [$F(3,632)=5.65$, $p<0.001$].

1. Accuracy

The accuracy on each condition was assessed with a univariate analysis of variance (ANOVA). The independent variables were Condition (A, B, C, D) and Sequence (CC words vs. CəC words). The dependent variable was the proportion of each subject's accurate responses for each Condition and Sequence, which was arcsin transformed. Results showed a significant main effect of Condition [$F(3,117)=105.57$, $p<0.001$], no main effect of Sequence [$F<1$], and a significant interaction between Sequence and Condition [$F(3,117)=2.74$, $p=0.047$]. Accuracy results are shown in Fig. 2. An examination of the incorrect responses shows that when participants chose the incorrect response in Conditions B and C (the correct response is "neither" for both), they always chose the inserted form and the minimal pair distracter, respectively.

Student Newman Keuls post-hoc tests indicated that collapsing over sequence type, Conditions A (Correct) and D (2 Unfamiliar Words) are not significantly different from each other, but both are significantly different from Conditions B (Inserted) and C (Minimal Pair Distracter) ($p<0.001$). Conditions B and C are also significantly different from one another ($p<0.05$). The significant interaction of Condition and Sequence is a result of a near-significant difference between CC and CəC words in Condition C (Minimal Pair Distracter) [$F(1,78)=2.98$, $p=0.088$].

These results show that listeners were very consistent in choosing the correct word for Condition A and "neither" for Condition D, as predicted. They were significantly better at rejecting the minimal pair distracters in Condition C than the inserted words in Condition B, but only by a small amount. In Condition C, participants have a tendency to choose the minimal pair distracter more often for words learned as CC than words learned as CəC.

2. High vs. Low performers

After the data were collected, we decided to examine the individual participants according to their performance on Condition C (Minimal Pair Distracter) to investigate whether all participants could learn and generalize the distinction between phonologically permitted (e.g., [zəmagu]) and unattested (e.g. [zmagu]) nonwords to another speaker. There are many indications in the literature suggesting that both first and second language learners have differing phonological abilities that may be attributable to factors such as phonological awareness (deJong *et al.*, 2000; Hu and Schuele, 2005; Speciale *et al.*, 2004) or phonological short term

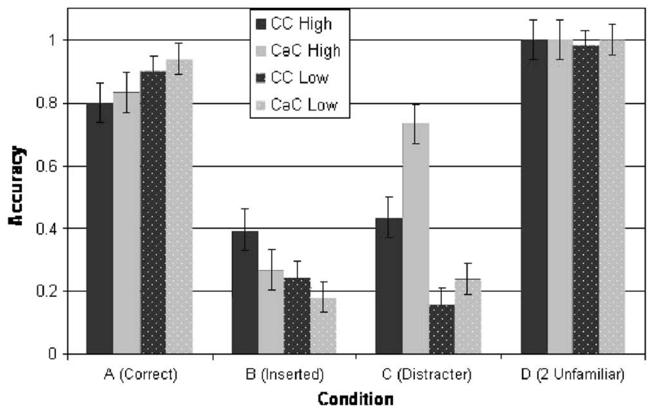

FIG. 3. Accuracy for each of the four conditions of auditory word pairs comparing High and Low performers (solid and dotted bars, respectively). The black bars represent the CC words that were learned, and the gray bars are the CəC words.

memory (Gathercole and Pickering, 1999). Similarly, Wong and Perrachione (2007) showed a distinction between highly successful and less successful English learners of Mandarin tone in a word-learning task. Because of the *a posteriori* nature of this investigation, we have not collected relevant phonological awareness or short term memory data that might allow us to determine why some participants are better at learning the CC/CəC distinction than others. Nevertheless, it is instructive to examine the performance of the participants to determine whether different types of learners are evident. The presence of such a division in these participants would provide motivation for a future study examining what factors might affect second language learners' perceptual discrimination abilities.

All participants who moved on to the test phase passed the training phase, indicating that they could learn the difference between one speaker's productions of target words and minimal pair distracters. However, this does not necessarily mean that mastery of these minimal pairs carried over to the second speaker's productions in the test phase. In dividing the participants into two groups, chance performance in Condition C (Minimal Pair Distracter) was used as the criteria. Since there were three choices for each test trial (word 1, word 2, and "neither"), chance was taken to be 33%. Those participants with average accuracy on Condition C below 33% were considered the Low performers ($N=23$), and those with accuracy above 33% were the High performers ($N=17$).

A univariate ANOVA was performed with the independent variables of Condition, Sequence, and Performance (High performers vs. Low performers). The dependent variable was accuracy. Results show that there is a significant effect of Performance [$F(1,304)=14.29$, $p<0.001$] and Condition [$F(3,304)=108.15$, $p<0.001$], but no effect of Sequence [$F<1$]. The two-way interactions between Performance and Condition [$F(3,304)=10.85$, $p<0.001$] and Sequence and Condition [$F(1,304)=3.62$, $p=0.014$] were significant. The three-way interaction was not significant. The results comparing High and Low performers are shown in Fig. 3.

Separate ANOVAs for each condition were conducted to examine the main effects and interaction of Sequence and Performance. The results for Condition A (Correct) show a significant main effect only for Performance [$F(1,76)=4.84$, $p=0.037$]. The Low performers are significantly more accurate on Condition A than the High performers (the latter respond "neither" more often), but average performance on Condition A for all participants is 80% or greater. For Conditions B (Inserted) and D (2 Unfamiliar), there are no significant effects of Sequence or Performance, nor is there an interaction. For Condition C (Minimal Pair Distracter), there is a significant effect of both Sequence [$F(1,76)=4.95$, $p=0.029$] and Performance [$F(1,76)=24.40$, $p<0.001$], but no significant interaction [$F(1,76)=1.57$, $p=0.214$]. For both High and Low performers, participants were significantly more accurate on words learned as CəC than on words learned as CC. Furthermore, collapsing across sequence, Student Newman Keuls post-hoc tests showed that high performers were significantly more accurate on Condition C than on Condition B ($p=0.002$), whereas Low performers were not ($p=0.99$).

D. Discussion

The results of Experiment 1 conform to the predictions expressed in Sec. II B. In Condition A, in which the correct word was paired with an unfamiliar word, all participants chose the correct word with greater than 80% accuracy. Likewise, for Condition D, which paired two unfamiliar words with the learned cartoon character, participants chose "neither" almost 100% of the time. The unexpected finding that Low performers were significantly better than High performers on Condition A may be explained by their being less capable of distinguishing between the correct target word and the minimal pair distracter word that they had learned in the training. That is, High performers may have missed the correct answer because they were more aware of the possibility that what they were hearing was the minimal pair distracter, whereas Low performers were more often correct because they would have accepted a large range of variants for the target word.

In Condition C, participants were presented with the minimal pair distracter for the word that they had learned. Thus, when the star appeared on the screen, participants heard the word [fənagu] and an unrelated foil, and the correct response was to choose "neither." This distinction had been learned in the training phase, since participants were required to choose a black X for the minimal pair distracter in order to receive positive feedback. When divided into High and Low performers, results for this condition demonstrate that roughly half of the participants could generalize to a new voice with greater than chance accuracy, whereas the other half could not. Still, both High and Low performers were more accurate on choosing "neither" when they were presented with a CC word for a word that they had learned with a schwa as compared to words that had been learned with a consonant cluster.

Finally, in Condition B, in which participants were presented with the C^{ə}C "version" of the word they had learned, both groups were equally likely to choose the C^{ə}C word instead of the "neither" option. High performers were numerically more accurate (more likely to reject C^{ə}C) for words learned as CC. Although it is not significant, this result, which is suggestive of Prediction 2b for the High performers, is consistent with the hypothesis that they associate C^{ə}C words with CəC and are therefore likely to be more accurate on words learned as CC in Condition B. This point will be revisited in Sec. IV in conjunction with results from Experiment 2.

It is also relevant that there was no significant difference in accuracy between Conditions B and C for the Low performers. This indicates that they are unable to generalize the CC/CəC distinction that they ostensibly learned in the training phase to a new speaker. The overall low accuracy in both conditions suggests that these participants are willing to accept acoustic variants of learned words regardless of the presence or absence of a vowel, possibly because they cannot overcome the phonotactic restrictions on the CC sequences, leading them to equate CC and CəC.

In this experiment, participants were explicitly introduced to phonological contrast by having to learn CC and CəC words which had the same initial consonants but which differed in the remainder of the word. They were exposed to minimal pairs in that they were required to reject those acoustic forms in the training session, but they did not ever learn minimal pairs where each word picked out a unique object. In the next experiment, we examine whether learning minimal pairs, where each of the lexical forms is associated with a meaning, enhances the participants' ability to discriminate between CC and CəC forms. We also ask whether learning minimal pairs clarifies how the participants treat the C^{ə}C stimuli.

III. EXPERIMENT 2

A. Methods

1. Participants

The participants were 41 native American English speakers who were recruited from Craigslist and from flyers posted around New York University. None of these participants had taken part in Experiment 1. No participant had any experience with Slavic languages, Hebrew, or any other language that has the initial obstruent clusters used in the experiment. No participants were phonetically trained, and none reported any history of speech or hearing impairments. Eleven more participants were tested but their data were not used; nine failed to complete the training phase, and two failed to respond to more than 15% of the test items.

2. Materials

The materials for Experiment 2 are very similar to those for Experiment 1; however, since the purpose of this experiment is to have the participants learn minimal pairs, the words are CC-CəC pairs. The stimuli are shown in Fig. 4.

The training and test stimuli were taken from the same two speakers as in the training and test conditions of Experiment 1. The stimuli for the training phase contained only the minimal pairs shown in Fig. 4, and stimuli for the test phase also included the corresponding C^{ə}C words. The same practice stimuli used in Experiment 1 were also used in this experiment.

	CC	CəC
Block 1	[vbano]	[vəbano]
	[zbasi]	[zəbasi]
	[zdati]	[zədati]
	[fnapi]	[fənapi]
Block 2	[vgane]	[vəgane]
	[zmagu]	[zəmagu]
	[vzaku]	[vəzaku]
	[ftake]	[fətake]

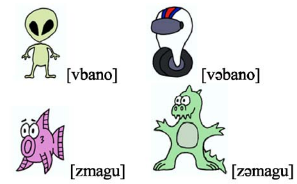

FIG. 4. (Color online) CC and CəC minimal pairs (left) and example picture-word pairings (right) used in the training and test phase.

3. Procedure

As in Experiment 1, there were three phases in the experiment: A familiarization phase, a training phase, and a test phase. Participants again learned 8 word-picture pairings at a time (4 CC/CəC pairs) for a total of 16 words learned. After the first familiarization-training-test sequence, the participants were given a break and returned to complete the second half.

The familiarization phase was the same as for Experiment 1. The training phase differed in that there was only a 4×2 array of cartoon characters but no black X, since there were no minimal pair distracters in this experiment. Participants had to accurately match the CC stimulus (e.g., /zəmagu/) with its picture (the fish) and the CəC stimulus (e.g., /zəmagu/) with its picture (the dragon). Once all 8 pictures were correctly matched in a row, participants moved into the test phase. Again, the count of correctly matched picture-word pairs was reset when a participant made an incorrect choice, and the escape hatch was available for the participant to review the cartoon-name pairings.

In the test phase, all learned pictures were presented in three different conditions. This is different from Experiment 1, where each picture was only presented once with an auditory pair of options. After completing Experiment 1, it was determined that the test conditions should be changed for Experiment 2 to make the task more statistically robust (more observations per participant) and more comparable to previous AX discrimination tasks which did not include a word-learning component (Davidson, 2007). The conditions are shown in Table I. In Condition A trials, participants heard the correct word that they had learned and the distracter word, or the minimal pair which was the name of another cartoon character. In Condition B, participants heard the correct word and the matching word which contained the in-

TABLE I. Test conditions and examples of auditory stimuli Experiment 2. See Fig. 4 for the pictures represented here by text descriptions.

Learned word	Condition	Picture	Auditory stimuli	
C^{ə}C: [zəmagu]	A: Correct vs. Distracter	Dragon	[zəmagu]	[zmagu]
	B: Correct vs. Inserted		[zəmagu]	[z^{ə}magu]
	C: Distracter vs. Inserted		[zmagu]	[z^{ə}magu]
CC: [zmagu]	A: Correct vs. Distracter	Fish	[zmagu]	[zəmagu]
	B: Correct vs. Inserted		[zmagu]	[z^{ə}magu]
	C: Distracter vs. Inserted		[zəmagu]	[z^{ə}magu]

serted schwa. Condition C, consisting of the distracter minimal pair and the inserted word, was a mismatch condition which did not contain the correct name at all.

Participants were instructed to press the first button on the button box if the correct answer was the first response and the second button on the button box if the correct answer was the second response. There was no "neither" option in this experiment. Each subject heard all conditions in the test phase; the stimuli were presented in random order and the experimental halves were counterbalanced. There were a total of 48 test trials, with 24 in each half. The practice phase for Experiment 2 was the same as in Experiment 1.

B. Predictions

In this experiment, speakers must learn that minimal pairs of CC/CəC items cannot be acoustic variants of one another because they denote different pictures. The conditions given in Table I lead to several hypotheses. The participants in this experiment are also divided into High and Low performers on the basis of Condition A (Correct vs. Distracter), which is the counterpart of Condition C (Minimal Pair Distracter) in Experiment 1.

If learning minimal pairs where each word is associated with meaning does not help Low performers discriminate between the lexical items, then it is assumed that they will essentially perform at chance for all conditions. However, if learning minimal pairs improves their performance, then all participants should conform to the following predictions (based on the results of Experiment 1).

1. *Condition A: Correct vs. Distracter*. Accuracy should be greater than chance. It is also possible that accuracy on Condition A for words learned as CəC is significantly greater than for words learned as CC [cf. Condition C (Minimal Pair Distracter) in Fig. 3].

2. *Condition B: Correct vs. Inserted*. It is not expected that participants will be very accurate in this condition. The results of Experiment 1 suggest the following two as the most likely outcomes (i.e., it is not likely that listeners will identify C^{ə}C only with CC):

a. If listeners identify C^{ə}C with CəC, they will choose Correct more often for pictures learned as CC. They will be at chance for pictures learned as CəC because they will be confused by the presentation of C^{ə}C and CC on the same trial.

b. If C^{ə}C is treated as a variant of both types of words, they should be at chance regardless of whether the target word was CC or CəC.

3. *Condition C: Distracter vs. Inserted*. This is a "mismatch" condition that was intended to provide further insight both into how participants categorize C^{ə}C utterances and how robust the representation of the CC and CəC words that they learned is.

a. If listeners identify C^{ə}C with CəC, they will choose Inserted more often than chance for pictures learned as CəC. They will be at chance for pictures learned as CC.

b. If C^{ə}C is treated as a variant of both types of words, and participants recognize that the Distracter word is not the right match for the picture, then the C^{ə}C answer should be chosen more often regardless of whether the target word was CC or CəC.

C. Results

Following Experiment 1, the first step was to divide the participants into High and Low performing groups on the basis of Condition A (Correct vs. Distracter), which is the counterpart of Condition C (Minimal Pair Distracter) in Experiment 1. Again, it was expected that since the participants could not leave the training without accurately matching the CC and CəC minimal pairs with the correct picture, they should do well on Condition A. In this experiment, chance was taken to be 50% since there were only two choices for each test trial. There were 19 participants in the Low performing group and 22 participants in the High performing group.

Because there is no "correct" response in Condition C, it is impossible to directly compare Conditions A and B to Condition C on the basis of accuracy. Thus, an omnibus ANOVA with condition as an independent variable cannot be carried out. Instead, each condition is examined in a separate ANOVA to examine the effects of sequence and performance (High vs. Low). For Conditions A and B, the dependent variable is accuracy. For Condition C, the response type reflects how often the participants chose the "Inserted" (CəC) word instead of the distracter word in the minimal pair. Proportions of responses for Experiment 2 are shown in Table II.

In Condition A (Correct vs. Distracter), there was a significant main effect of Performance [$F(1,78)=81.78$, $p<0.001$] but no significant effect of Sequence [$F<1$]. There was no significant interaction between Sequence and Performance [$F<1$]. Regardless of whether the learned target word was CC or CəC, the High performing participants chose the correct word more often than the minimal pair word.

In Condition B (Correct vs. Inserted), there was no main effect of either Sequence [$F<1$] or Performance [$F(1,78)=2.60$, $p=0.11$]. The interaction between Sequence and Performance is marginally significant [$F(1,78)=3.43$, $p=0.07$]. The interaction results are due to slightly more accurate performance by High performers for words learned as CC as compared to words learned as CəC, but it is not significant ($p=0.27$).

In Condition C (Distracter vs. Inserted), there was a significant main effect of Sequence [$F(1,78)=21.81$, $p<0.001$], but no main effect of Performance [$F<1$]. The interaction between Sequence and Performance was significant [$F(1,78)=4.96$, $p=0.029$]. The interaction is due to High performers choosing the C^{ə}C stimulus significantly more often for words learned as CəC than for words learned as CC, and more often than Low learners ever choose C^{ə}C ($p<0.001$).

D. Discussion

The results for Experiment 2 indicate that there are differences in performance for High performers vs. Low performers. Collapsing over Sequence, the Low performers are

TABLE II. Response types for each of the three conditions of auditory word pairs comparing High and Low performers. Bold numbers represent the accurate response for Conditions A and B. Although there is no "correct" response for Condition C, C^{ə}C is the more expected response. The dash (—) indicates that the response type was not available for that condition.

		Response type for High and Low learners					
		CC		CəC		C^{ə}C	
	Word learned as	High	Low	High	Low	High	Low
Condition A	CC	**0.71**	**0.46**	0.29	0.54	—	—
(Correct vs. Distracter)	CəC	0.36	0.52	**0.74**	**0.48**	—	—
Condition B	CC	**0.64**	**0.52**	—	—	0.36	0.48
(Correct vs. Inserted)	CəC	—	—	**0.59**	**0.59**	0.41	0.41
Condition C	CC	—	—	0.53	0.47	0.47	0.53
(Distracter vs. Inserted)	CəC	0.27	0.37	—	—	0.73	0.63

at chance for all conditions except Condition C (Distracter vs. Inserted) (comparison to chance using the binomial distribution and an alpha of 0.05, the probability for Condition A=0.86, Condition B=0.34, Condition C=0.003). That is, although they were able to learn the distinctions between the minimal pairs and associate them with the correct cartoon characters during the training phase, they could not easily generalize that knowledge to the parallel distinction produced by a different speaker. In Condition C, the mismatch condition, the Low performers pattern like the High performers, preferring to choose the C^{ə}C token for words learned as CəC.

With an accuracy of greater than 70% in Condition A (Correct vs. Distracter), the High performers are fairly successful in generalizing the distinction between CC and CəC that they learned in the training phase. The lack of any difference between words learned as CC and words learned as CəC indicates that High performers' ability to learn and encode both native and minimally different non-native phonotactics is enhanced by learning actual minimal pairs. This finding is in contrast to Experiment 1, where performance on Condition C, in which participants were expected to choose "neither" when presented with a minimal pair distracter (which had not been learned in conjunction with a picture), was better for words learned as CəC than for words learned as CC.

The results for Condition B (Correct vs. Inserted) indicate that the High performing participants had some difficulty distinguishing between the words they had learned and the C^{ə}C variant for both types of sequences. However, there was a small tendency for the High performing participants to choose the correct word more often for words that had been learned as CC. This is consistent with Hypothesis 2a in Sec. III B, which stated that if participants associated C^{ə}C with CəC, then they would be more confused by trials in which these words were paired together.

This trend is better confirmed in Condition C (Distracter vs. Inserted), in which High performers chose the C^{ə}C option much more often for words learned as CəC than for words learned as CC. It was hypothesized that if participants rejected the distracter word equally for both CC and CəC words, then they should choose the C^{ə}C option at equal (and perhaps above chance) rates. However, the results demonstrate that the participants equate C^{ə}C with CəC much more often, and they are at chance for the words learned as CC. In the latter case, it is possible that neither utterance is an acceptable version of CC for the participants, and therefore they randomly choose between the two substandard options.

IV. GENERAL DISCUSSION

These experiments were designed to address three main questions: (1) Does learning sound-meaning pairs assist learners in perceiving and distinguishing the fine-grained acoustic differences between phonotactically native CəC and non-native CC sequences, (2) do minimal pairs further enhance the ability to distinguish these sequences, and (3) how does the picture-matching task affect the interpretation of a speaker's repair of non-native phonotactics?

A. Learning to distinguish CC from CəC

With respect to the first and second questions, the experimental results indicate that it is primarily when listeners are explicitly taught sound-meaning relationships with minimal pairs that they improve their ability to distinguish between CC and CəC tokens across the board. Thus, this study extends Hayes-Harb's (2007) finding regarding the usefulness of minimal pairs in learning to distinguish non-native phonemes to phonotactic sequences. The results from Experiment 1, which required participants to reject the minimal pair distracter words in the training phase, show that Low performers are unable to generalize the distinction between CC and CəC to a new speaker's voice. They only correctly choose "neither' about 20% of the time in Condition C in the test phase, where the minimal pair distracter is paired with a totally unrelated word. High performers are somewhat more accurate, but show a large difference between words learned as CC and those learned as CəC. While they are able to reject the CC distracter for words learned as CəC approximately 73% of the time, they only reject the CəC distracter about 40% of the time.

This result is the first indication that the phonetic encoding or representation formed for phonotactically legal words learned as CəC is more robust and leads to greater accuracy when generalizing to another voice. For words learned as CC, participants seem to accept a greater variety of productions, suggesting that the native language phonological prohibition on the CC sequences used in this study hampers a detailed phonetic encoding of these items, at least at an early stage of exposure. This experiment indicates that simply introducing participants to the phonological contrast between CC and CəC does not sufficiently underscore the lexical relevance of the $ə \sim \emptyset$ alternation. If listeners do not need to prevent lexical confusion because there are no minimal pairs, then perhaps they are willing to treat CC and CəC as variants of the same word.

In Experiment 2, which explicitly relies on minimal pairs, participants show an overall improvement in accuracy, though the results are more evident for the High performers. In Condition A, which directly pits CC against CəC words that were learned in association with a picture, the accuracy of High performers is the same for both CC and CəC words and is similar to their performance for words learned as CəC in Condition C (Minimal Pair Distracter) in Experiment 1. In order to further evaluate whether the minimal pairs contributed to better discrimination than phonological contrast alone, it is appropriate to examine whether an accuracy of 71% in Condition A of Experiment 2 where chance is 50% and 43% when chance performance is 33% (Condition C of Experiment 1) are both above chance. Using the binomial distribution to calculate the cumulative probability of the participants' performance in both studies, the probability of obtaining an average accuracy of 43% or greater given a chance level of 33% in Experiment 1 is 0.16. In Experiment 2, the probability of scoring 71% or greater when chance is 50% is less than 1.17×10^{-8}. Thus, while the results for Experiment 1 were not above chance, the probability of scoring 71% in Experiment 2 is much less.

These findings suggest that when minimal pairs are matched with meaning, participants have greater incentive to attend to the subtle differences between the words, even when spoken by a different talker. These results suggest that effectiveness of minimal pairs in adult word learning tasks is different than for infants, who have more difficulty discriminating unfamiliar minimal pairs when paired with meaning than in purely auditory tasks (Pater *et al.*, 2004; Stager and Werker, 1997). The difference between adult and infant performance is consistent with Werker and colleagues' interpretation that greater task demands prevent infants from integrating the acoustic information with their newly formed lexical representation; adults are not assumed to have infant-like limitations on their processing resources, so the picture matching task helps focus attention on the phonetic differences rather than obscure them. Likewise, minimal pair comparisons in older children with phonological delay have been used successfully to help children learn to correctly produce new sounds, especially when learning two new sounds which were previously both erroneously produced [see Gierut (1998) for an overview of this issue].

Two limitations of the current study should be taken into consideration when interpreting the results. First, there is no baseline condition showing how well the participants would have been able to identify CC or CəC tokens without the benefit of lexical training with pictures. However, the AX discrimination task from Davidson (2007) provides information about English speakers' performance in the absence of training. In the AX experiment, English speaking participants were presented with the same types of CCVCV and CəCVCV pairs produced by the same English speaker as in the test phase of the current experiments (there were also CC/C^{ə}C and CəC/C^{ə}C conditions). To encourage a response based on an acoustic comparison of the contrastive sequences, the discrimination pairs were presented with a small ISI of 250 ms and the participants were asked to determine whether the sound files were "exactly the same" or not. Results for the CCVCV/CəCVCV comparison showed that participants labeled them as "different" only 49% of the time; thus, when asked to determine whether the CəC and CC sound files were exactly the same, participants were at chance (chance level=50%).

The second limitation is that the experimental conditions for Experiment 2 are not exactly the same as those for Experiment 1. This makes it difficult to directly compare the results of the two studies, although the probabilities of the outcomes determined using the binomial distribution assist in interpreting the results. This limitation is particularly evident when comparing Experiment 1 with the AX discrimination task, because there was no condition explicitly pitting CC against CəC. Still, it is useful to evaluate the performance on Condition C (Minimal Pair Distracter) in light of the AX discrimination task. While the Low performers in the current study correctly rejected the minimal pair distracter for both CC and CəC stimuli only approximately 20% of the time, the High performers accurately rejected the minimal pair distracter 43% of the time for CC stimuli and 73% of the time for CəC stimuli. In the absence of direct comparison between the two types of tokens (as in the AX task), the Low performers in this study simply accept the minimal pair distracter as a variant for the word they learned. The High performers, however, already show some benefit from the lexical training for CəC as compared to the AX performance results, but not for the phonotactically unattested CC.

Condition A (Correct vs. Distracter) from Experiment 2 is more directly comparable to the AX discrimination task, since participants were presented with the minimal pair in the trial. High performers were significantly better than chance for both CC and CəC stimuli, and these were not different from one another. Low performers did not improve over the AX discrimination results, despite the addition of the minimal pair training. Although not reported in Davidson (2007), a post-hoc examination of individual subject performance for the AX discrimination task showed that only 8 out of 31 participants performed above chance, so dividing the group into High and Low performers would not have been motivated.

Comparing Experiments 1 and 2 from this study to the AX discrimination task confirms that while forming sound-meaning pairs might improve accuracy in discrimination for the words that already have a more detailed phonetic representation (i.e., CəC words), it is only when lexical contrast is added that participants are really required to attend to the presence versus absence of the vowel. This result is consistent with the general idea outlined by Hawkins (2003) and Coleman (2003) that subtle but systematic phonetic differences play an important role in phonemic (and by extension, phonotactic) differentiation, and that meaning and experience affect the interpretation of fine phonetic detail. While the idea that minimal pairs are crucial for the formation of phonemic and phonotactic inventories is hardly a new idea in phonology, this study does provide new insight into the amount of phonetic detail that speakers might be willing to overlook—such as a reduced vowel—in the absence of a minimal pair.

Though the current study suggests that there are conditions under which participants can acquire the CC/CəC contrast, studies of the lexical representations of highly fluent bilinguals suggest that whether or not contrasts can be learned and robustly encoded in the lexicon may depend on the specific phonemes or phonotactic sequences in question (e.g., Curtin *et al.*, 1998; Weber and Cutler, 2004). For example, Pallier *et al.* (2001) showed that Spanish-dominant Spanish/Catalan bilinguals showed repetition priming for Catalan minimal pairs differing only in [ɛ]/[e], [o]/[ɔ], or [s]/[z], whereas Catalan-dominant bilinguals did not (e.g., [nɛta]/[neta], "clean/granddaughter"). They concluded that a lack of sensitivity to these contrasts for Spanish-dominant bilinguals leads to representation of these minimal pairs as homophones.

The current study, taken together with previous findings, suggests that the phonetic content of some kinds of phonological elements may be either more or less salient to the learner than other types are. Whereas the distinction between vowels may be particularly difficult to represent lexically due to the perceptual organization of the vowel space (Best, 1995; Flege and MacKay, 2004; Kuhl and Iverson, 1995), it may be easier for learners to ultimately learn new phonotactic contrasts, especially where the consonants in question are in the inventory of the listeners but are not allowed in CC sequences in word-initial position. Furthermore, the phonetic detail in the representation of non-native items may not be as robustly encoded as English-possible items (cf. Gathercole *et al.*, 1999 for phonotactic effects in short-term recall by children). As a result, even the High performing listeners accept CəC as a variant of CC more often than they accept CC for CəC.

The division between participants who can distinguish between CC and CəC after training and those who cannot is another issue raised by this study. The Low performers in this study never manage to perform better than chance even in Experiment 2, and it is not clear from this study whether they would continue to have difficulty in distinguishing the CC/CəC contrast, or whether they would be able to carry out the task at the same level as the High performers if they were given more practice. The results of these experiments suggest that not all adult language learners are necessarily equipped with the same ability to perceive and distinguish between non-native sounds, at least at the earliest stages of acquisition (cf. Speciale *et al.*, 2004). Why this is the case, and whether all participants would eventually show equal ability on these tasks over time, is an interesting question for future research.

B. The interpretation of fine phonetic detail

The third question addressed by this study pertained to the effect that associating meaning with the learned words would have on the interpretation of $C^{ə}C$ stimuli. Although results of the AX discrimination task for comparisons containing the $C^{ə}C$ tokens were universally poor in Davidson (2007), there was nevertheless a significant difference between performance on the $C^{ə}C$/CəC discrimination (20% "different" responses) versus the $C^{ə}C$/CC discrimination (34% "different"responses). Similarly, in the transcription task, participants wrote a vowel between the consonants for over half of the $C^{ə}C$ tokens. These findings already suggested a bias for perceiving the inserted schwa as a vowel, despite the fact that several acoustic properties of the inserted schwa, including duration and F1 midpoint, are significantly different from a lexical vowel.

The results of both Experiments 1 and 2 confirm that while confusability with both CC and CəC is generally high, listeners are more likely to interpret $C^{ə}C$ as containing a vowel. In Experiment 1, High performers in Condition B, in which $C^{ə}C$ was presented with an unrelated word, were more likely to choose "neither" for words that were learned as CC than for words learned as CəC. This indicates that they were more likely to equate $C^{ə}C$ with the lexical vowel, and thus were less capable of rejecting the $C^{ə}C$ foil for words learned as CəC. Similarly, in Condition C of Experiment 2, which paired the distracter word with the $C^{ə}C$ word, High performers were more likely to choose $C^{ə}C$ over CC for words that had been learned as CəC, as compared to words that had been learned as CC. Low performers showed a similar tendency.

Although there are inevitable differences between this type of laboratory study and real language contact, the results of these experiments can help inform predictions about loanword adaptation. If we assume that these stimuli represent the actual types of utterances that a recipient might hear from disseminators during the process of loanword adaptation, then the findings of this study, in conjunction with Davidson (2006), suggest that two phonotactic categories distinguishing between CC and CəC might eventually be established over time if English were to borrow large numbers of words with initial obstruent-obstruent sequences. That is, when the conditions are right—when learners must provide some kind of representation to a new word, and if they are fortunate enough to learn minimal pairs—they are better able to perceive and encode CC sequences. As noted in Sec. IA, it was not the purpose of these studies to attempt to teach listeners a three-way contrast between CC, CəC, and $C^{ə}C$. In a borrowing situation, $C^{ə}C$ sequences would likely be divided up among the CC and CəC categories, with a larger proportion of them being assigned a CəC representation. Furthermore, results of Davidson (2007) demonstrate that disseminators sometimes produce fricative-initial CC sequences accurately, so some proportion of these sequences would be both represented and produced accurately by borrowers. It is when these items are passed down to the next generation of children that they will ultimately lead to phonological change such that the relevant CC sequences become permitted by the phonology (Lahiri *et al.*, in press).

V. CONCLUSION

The process of learning new sounds and new contrasts, whether segmental or phonotactic, is affected by factors both internal and external to the phonetic string itself. It has been shown that certain phonemes or phonotactic sequences are more difficult than other to accurately perceive, but regardless of what kind of unit is being acquired, the acquisition of a new sound system or the interpretation of loanword adaptations does not simply require listeners to distinguish between two auditory strings. In language learning and contact situations, lexical items are acquired with their associated meanings. In this sense, a picture matching task may be a better laboratory analogy for loanword adaptation or second language learning than simple discrimination tasks, since learning meaning is an integral part of language contact. The results of these experiments suggest that the more incentive that adult listeners have to distinguish fine phonetic detail, the more they will attend to it and incorporate it into their lexical representations.

ACKNOWLEDGMENTS

The authors thank the three anonymous reviewers, and audiences at Yale University and the New York University Phonetics/Phonology Lab for their comments on this work. We would also like to thank David Goldberg for creating the cartoon characters used in the experiments. This work was supported by the National Science Foundation CAREER Grant No. BCS-0449560.

APPENDIX

Stimuli used in each experimental condition in Experiment 1. Block refers to whether the stimulus was presented in the first or second half of the experiment. See Fig. 1 for examples of the pictures represented here by text descriptions.

Condition	Picture	Learned word	Stimulus pair in test	Correct response
A (Correct)	Alien	vbano	vbano zgade	vbano
	Fly	zdati	zdati fəsaga	zdati
	Octopus	vəbaki	vəbaki fmasa	vəbaki
	Turtle	zədanu	zədanu vəgalu	zədanu
B (Inserted)	Robot	vgane	v^{ə}gane fpami	neither
	Helmet	zmapi	z^{ə}mapi vəmabu	neither
	Fish	vəgapo	v^{ə}gapo fətano	neither
	Dragon	zəmagu	z^{ə}magu vdapi	neither
C (Minimal Pair Distracter)	Star	fnagu	fənagu znaʃo	neither
	Skull	vzaku	vəzaku zədaba	neither
	Flower	fənapi	fnapi zbatu	neither
	Snake	vəzaba	vzaba zəgano	neither
D (2 Unfamiliar)	Ghost	zbagi	fkada vənali	neither
	Bird	ftake	zmafo vəzamo	neither
	Cat	zəbasi	vəmala fkabe	neither
	Flame	fətalu	zəvapa vdagu	neither

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (**2007**). "What we know about what we have never heard: Evidence from perceptual illusions," Cognition **104**, 591–630.
Best, C. (**1995**). "A direct-realist view of cross-language perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* edited by W. Strange (York, Baltimore), pp. 171–204.
Bradlow, A., and Pisoni, D. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**, 2074–2085.
Broselow, E. (**2004**). "Language contact phonology: Richness of the stimulus, poverty of the base," in *Proceedings of NELS 34* edited by K. Moulton and M. Wolf (Booksurge, North Charleston, SC), pp. 1–21.
Broselow, E., and Finer, D. (**1991**). "Parameter setting in second language phonology and syntax," Second Lang. Res. **7**, 35–59.
Cebrian, J. (**2000**). "Transferability and productivity of L1 rules in Catalan-English interlanguage," Stud. Second Lang. Acquis. **22**, 1–26.
Coleman, J. (**2003**). "Discovering the acoustic correlates of phonological contrasts," J. Phonetics **31**, 351–372.
Curtin, S., Goad, H., and Pater, J. (**1998**). "Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers," Second Lang. Res. **14**, 389–405.
Davidson, L. (**2006**). "Phonology, phonetics, or frequency: Influences on the production of non-native sequences," J. Phonetics **34**, 104–137.
Davidson, L. (**2007**). "The relationship between the perception of non-native phonotactics and loanword adaptation," Phonology **24**, 261–286.
Davidson, L., Jusczyk, P., and Smolensky, P. (**2004**). "The initial and final states: Theoretical implications and experimental explorations of richness of the base," *Fixing Priorities: Constraints in Phonological Acquisition*, edited by R. Kager, W. Zonneveld, and J. Pater (Cambridge University Press, Cambridge), pp. 321–368.
Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (**2000**). "Electrophysiological correlates of phonological processing: A cross-linguistic study," J. Cogn Neurosci. **12**, 635–647.
deJong, P. F., Seveke, M.-J., and van Veen, M. (**2000**). "Phonological sensitivity and the acquisition of new words in children," J. Exp. Child Psychol. **76**, 275–301.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (**1999**). "Epenthetic vowels in Japanese: A perceptual illusion?," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1568–1578.
Dupoux, E., Pallier, C., Kakehi, K., and Mehler, J. (**2001**). "New evidence for prelexical phonological processing in word recognition," Lang. Cognit.

Processes **16**, 491–505.
Flege, J. E. (**1995**). "Second-language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD) pp. 229–273.
Flege, J. E., and MacKay, I. (**2004**). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.
Flege, J. E., Schirru, C., and MacKay, I. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.
Gathercole, S., and Pickering, S., J. (**1999**). "Estimating the capacity of phonological short-term memory," Int. J. Psychol. **34**, 378–382.
Gathercole, S., Frankish, C., Pickering, S., and Peaker, S. (**1999**). "Phonotactic influences on short-term memory," J. Exp. Psychol. Learn. Mem. Cogn. **25**, 84–95.
Gierut, J. (**1998**). "Treatment efficacy: Functional phonological disorders in children," J. Speech Lang. Hear. Res. **41**, 85–101.
Hallé, P., Segui, J., Frauenfelder, U., and Meunier, C. (**1998**). "Processing of illegal consonant clusters: A case of perceptual assimilation?," J. Exp. Psychol. Hum. Percept. Perform. **24**, 592–608.
Hawkins, S. (**2003**). "Roles and representations of systematic fine phonetic detail in speech understanding," J. Phonetics **31**, 373–405.
Hayes-Harb, R. (**2007**). "Lexical and statistical evidence in the acquisition of second language phonemes," Second Lang. Res. **23**, 61–90.
Hu, C.-F., and Schuele, M. (**2005**). "Learning nonnative names: The effect of poor native phonological awareness," Appl. Psycholinguist. **26**, 343–362.
Imai, S., Walley, A., and Flege, J. E. (**2005**). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.
Kabak, B., and Idsardi, W. (**2007**). "Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints?," Lang Speech **50**, 23–52.
Kang, Y. (**2003**). "Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean," Phonology **20**, 219–273.
Kuhl, P., and Iverson, P. (**1995**). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, edited by W. Strange (York, Baltimore).
Lahiri, A., Wetterlin, A., and Jönsson-Steiner, E. (in press). "Tones and loans in the history of Scandinavia," in *Phonetic and Behavioural Studies in Word and Sentence Prosody, Tones and Tunes*, Vol. II, edited by C. Gussenhoven and T. Riad (Mouton de Gruyter, Berlin).
Lively, S., Logan, J. S., and Pisoni, D. (**1993**). "Training Japanese listeners to identify English /r/ and /l/ II. The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.
MacKay, I., Meador, D., and Flege, J. E. (**2001**). "The identification of English consonants by native speakers of Italian," Phonetica **58**, 103–125.
Massaro, D., and Cohen, M. (**1983**). "Phonological context in speech perception," Percept. Psychophys. **34**, 338–348.
Moreton, E. (**2002**). "Structural constraints in the perception of English stop-sonorant clusters," Cognition **84**, 55–71.
Pallier, C., Colomé, A., and Sebastián-Gallés, N. (**2001**). "The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries," Psychol. Sci. **12**, 445–449.
Pater, J., Stager, C., and Werker, J. (**2004**). "The perceptual acquisition of phonological contrasts," Language **80**, 384–402.
Peperkamp, S. (**2007**). "Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin," Cognition **104**, 631–637.
Peperkamp, S., and Dupoux, E. (**2003**). "Reinterpreting loanword adaptations: The role of perception," *Proceedings of the 15th International Congress of Phonetic Sciences*, in edited by M. J. Solé, D. Recasens, and J. Romero, Barcelona, Universitat Autónoma de Barcelona, pp. 367–370.
Pitt, M. (**1998**). "Phonological processes and the perception of phonotactically illegal consonant clusters," Percept. Psychophys. **60**, 941–951.
Sebastián-Gallés, N., and Soto-Faraco, S. (**1999**). "Online processing of native and non-native phonemic contrasts in early bilinguals," Cognition **72**, 111–123.
Speciale, G., Ellis, N., and Bywater, T. (**2004**). "Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition," Appl. Psycholinguist. **25**, 293–321.
Stager, C., and Werker, J. (**1997**). "Infants listen for more phonetic detail in speech perception than in word-learning tasks," Nature (London) **388**, 381–382.
Storkel, H., Armbruster, J., and Hogan, T. P. (**2006**). "Differentiating phonotactic probability and neighborhood density in adult word learning," J. Speech Lang. Hear. Res. **49**, 1175–1192.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–1704.
Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–1807.
Weber, A., and Cutler, A. (**2004**). "Lexical competition in non-native spoken-word recognition," J. Mem. Lang. **50**, 1–25.
Werker, J., and Tees, R. (**1984**). "Phonemic and phonetic factors in adult croos-language speech perception," J. Acoust. Soc. Am. **75**, 1866–1878.
Wong, P., and Perrachione, T. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**, 565–585.