SSLA, **27**, 387–414. Printed in the United States of America.
DOI: 10.1017/S0272263105050175

EFFECTS OF ACOUSTIC VARIABILITY ON SECOND LANGUAGE VOCABULARY LEARNING

Joe Barcroft and Mitchell S. Sommers
Washington University in St. Louis

This study examined the effects of acoustic variability on second language vocabulary learning. English native speakers learned new words in Spanish. Exposure frequency to the words was constant. Dependent measures were accuracy and latency of picture-to-Spanish and Spanish-to-English recall. Experiment 1 compared presentation formats of neutral (conversational) voice only, three voice types, and six voice types. No significant differences emerged. Experiment 2 compared presentation formats of one speaker, three speakers, and six speakers. Vocabulary learning was superior in the higher-variability conditions. Experiment 3 partially replicated Experiment 1 while rotating voice types across subjects in moderate and no-variability conditions. Vocabulary learning was superior in the higher variability conditions. These results are consistent with an exemplar-based theory of initial lexical learning and representation.

Successful speech comprehension in any language depends on the ability to perceive individual linguistic forms when presented with acoustically variant examples of those forms in the speech stream. Among the sources of acoustic variability that produce these variant examples are (a) speech produced by different speakers and (b) speech produced by one talker in different voice

Portions of these data were presented at the 143rd meeting of the Acoustical Society of America in Cancun, Mexico and at the Fourth International Conference on the Mental Lexicon in Windsor, Canada. The authors would like to thank Paola Rijos for help in data collection and scoring and the anonymous *SSLA* reviewers.

Address correspondence to: Joe Barcroft, Department of Romance Languages and Literatures, Campus Box 1077, One Brookings Drive, Washington University, St. Louis, MO 63130-4899; e-mail: barcroft@wustl.edu.

types, such as neutral, excited, or whispered voices. To perceive a given linguistic form correctly when presented with acoustic variability from these two sources, a listener must be able to process multiple acoustic manifestations of the form and in each case be able to perceive the same lexical item. For example, to perceive the word "apple" correctly each time when spoken by six different talkers, a listener must be able to process acoustic properties specific to the voice type of each speaker in a manner that allows the listener to perceive six different realizations of "apple" as the same word. The same principle holds true if the listener hears the word "apple" spoken by the same speaker in different voice types (e.g., in neutral voice, excited voice, and whispered voice).

To date, research on acoustic variability of this nature has been limited primarily to investigations of how acoustic-phonetic variability affects perception and memory in a listener's first language (L1) (see Johnson & Mullennix, 1997). Studies in this area have found that L1 speech produced by multiple as opposed to single talkers can impair vowel perception (Assmann, Nearey, & Hogan, 1982), word recognition, word naming (Mullennix, Pisoni, & Martin, 1989), and memory for previously acquired words (Martin, Mullennix, Pisoni, & Summers, 1989). A much more limited amount of research on acoustic variability and second language (L2) learning has been conducted as well. Studies on talker variability during L2 phonemic training suggest that presentation formats involving multiple talkers are more effective for teaching L2 phonemic contrasts than presentation formats with single talkers only (e.g., Lively, Logan, & Pisoni, 1993). In the realm of vocabulary learning, however, Barcroft (2001) found that presentation formats with multiple voice types produced by a single talker had no effect on learning rates for L2 vocabulary as compared with a presentation format using neutral voice only.

The purpose of the present study was to replicate and expand upon previous research on acoustic variability and L2 vocabulary learning in light of theoretical and practical questions that remain in this area. How does presenting new words in an acoustically varied manner affect the representation of those new items in memory? Is the effect of acoustic variability on vocabulary learning modulated by the nature of the variability? If so, what specific types of acoustic variability—if any—can affect vocabulary learning? The three experiments in this study were designed to address specific aspects of these larger questions. The first experiment examined the effect of presenting target words in an acoustically varied format using speech produced by one talker in multiple voice types during L2 vocabulary learning. The second experiment examined the effect of presenting target words in an acoustically varied format using speech produced by multiple talkers during L2 vocabulary learning. The third experiment examined the effect of presenting target words in multiple voice types while rotating voice types across subjects in order to control for the potential effects of differential voice type intelligibility. Before presenting details of these experiments, in the next three sections we review the main findings and theoretical issues in research on acoustic variability and L1 speech processing, acoustic variability and L2 phonemic training, and acoustic variability and L2 vocabulary learning.

ACOUSTIC VARIABILITY AND L1 SPEECH PROCESSING

The findings of many studies on acoustic variability and speech processing point to the cognitive costs of processing acoustically varied stimuli as compared to acoustically consistent stimuli. Performance decrements associated with multiple-talker presentation formats have been observed in studies on perception and memory for individual phonemes and isolated words (Assmann et al., 1982; Martin et al., 1989; Mullennix et al., 1989; Peters, 1955, cited in Pisoni, 1997; Ryalls & Pisoni, 1997; Schacter & Church, 1992; see also Creelman, 1957). Because multiple-talker conditions involve more acoustic variability than single-talker conditions, the prevailing interpretation of these findings has been that processing for acoustic variability during speech perception comes at a cognitive cost. The cost, according to current explanations (see Sommers, Nygaard, & Pisoni, 1994), is incurred because processing for acoustically varied input depletes cognitive resources that could otherwise be used to perform the tasks investigated in these studies.

The finding that some forms of acoustic variability can reduce L1 spoken word recognition is consistent with a theoretical perspective of spoken language processing in which representations of spoken words are abstracted without retaining information about indexical or nonlinguistic properties of the original speech signal, such as talker characteristics or speaking rate. Within this abstractionist framework, acoustic variability is removed through a process of acoustic normalization in which speech signals are converted to canonical forms that match those stored in lexical memory (see Pisoni, 1997, for an explanation and discussion of the need to reexamine the abstractionist perspective). From this perspective, negative effects of acoustic variability result from greater demands on the normalization system for acoustically varied, as compared with acoustically consistent, speech signals.

Goldinger (1998) proposed an alternative to the abstractionist perspective in which indexical properties of speech signals are not normalized or removed prior to identification. Goldinger's exemplar-based model suggests that lexical and indexical properties are components of an integrated representation in which both types of property can contribute to lexical processing. Consistent with the exemplar-based position, Martin et al. (1989) found that listeners encode details about the voices of different talkers into long-term memory and use this information as an aide to retrieval. Mullennix and Pisoni (1990) also found that subjects could not attend selectively either to the domain of voice (indexical) or phoneme (linguistic) without interference from the other domain. From the exemplar-based perspective, performance decrements associated with increased acoustic variability might be due to the cognitive demands of attending to and storing additional indexical information during speech processing as opposed to normalizing it per se. It is also important to note that although studies within the exemplar perspective have been restricted primarily to examining indexical features associated with the acoustic signal (e.g., talker characteristics or speaking rate), this perspective does not preclude the use and representation of articulatory gestures or other features associated with speech production as part of the mechanisms mediating spoken word recognition.

If indexical features in acoustically varied stimuli are encoded and stored during perceptual analysis, the potential exists for access to these features to facilitate some types of learning. The findings of Goldinger (1990), Palmeri, Goldinger, and Pisoni (1993), and Goldinger, Pisoni, and Logan (1991) are consistent with this line of thinking. Goldinger found that when allowed to self-pace word presentation rate, subjects selected a slower pace for multiple-speaker word lists as compared to single-speaker word lists (reflecting increased encoding requirements for the multiple-talker condition) but correctly recalled more of the words on the multiple-speaker lists (reflecting improved access to lexical forms with increased acoustic variability). Similarly, Palmeri et al. found better recognition of same-voice than of different-voice repetitions in a continuous recognition paradigm. Finally, Goldinger et al. found that the effect of speech variability on serial recall was moderated by presentation rate. At faster presentation rates that reduced the opportunity to encode indexical features, subjects recalled more words when spoken by a single speaker than when spoken by 10 different speakers. However, at slower presentation rates that increased the opportunity for encoding item-specific indexical features, subjects recalled more words when spoken by 10 speakers than when spoken by a single speaker. The findings of these studies suggest that when listeners have sufficient time to encode and store indexical properties of speech stimuli (e.g., when using longer presentation rates), this information can facilitate later retrieval of those items.

Further support for the benefits of indexical properties during L1 spoken language processing comes from work investigating the effects of talker familiarity on word recognition. In a series of experiments (Nygaard, Sommers, & Pisoni, 1992; Nygaard & Pisoni, 1998), participants were first trained to identify the voices of 10 talkers. Following training on voice identification, listeners were asked to identify single words and sentences. The critical manipulation was that half of the stimuli in the identification phase were produced by familiar talkers (voices the listeners had learned to identify) and half were produced by novel talkers. The results of these experiments indicated that identification performance was significantly better for familiar than for unfamiliar voices. Within the exemplar framework, words produced by familiar talkers provide a better match with stored representations because they are consistent with both indexical and lexical properties of the stored word forms, whereas words spoken by unfamiliar talkers would match only on the lexical dimension. This work is particularly important because it suggests a direct interaction between lexical and indexical properties of speech during online recognition of spoken words. However, the findings to date do not address how increased acoustic variability might affect other types of memory and learning, such as different aspects of L2 acquisition, which is the focus of the next two sections.

ACOUSTIC VARIABILITY AND L2 PHONEMIC TRAINING

With regard to L2 learning, two areas in which acoustic variability has been investigated are phonemic training and vocabulary learning. Studies in the first of these two areas suggest that acoustically varied presentation formats can be useful tools for teaching phonemic contrasts to L2 learners. For example, a phonemic contrast that is often challenging for native Japanese speakers learning L2 English is that of the liquid consonants /r/ and /l/. In the first of a series of studies on training Japanese listeners to discriminate English /r/ and /l/, Logan, Lively, and Pisoni (1991) examined the effectiveness of a training procedure with a higher degree of acoustic variability than those of previously examined training procedures (see Strange & Dittman, 1984). Specifically, Logan et al. trained Japanese listeners to identify minimal pairs with English /r/-/l/ contrasts in different phonetic environments (e.g., *lock-rock*; *array-allay*). The researchers hypothesized that increasing acoustic variability might help listeners to develop "stable and robust phonetic categories that show perceptual constancy across different environments" (p. 876). Consistent with this prediction, the high-variability training procedure produced significant increases in learning the L2 phonemic contrasts, whereas previous training procedures (e.g., Strange & Dittman) using a single talker were less successful at training this distinction. Of particular importance for the present study, these findings suggest that although acoustic variability might be associated with an initial encoding cost (e.g., poorer L1 word identification), it also has benefits for the long-term retention of L2 phonetic forms. As Logan et al. stated in their discussion, "nonnative listeners encode detailed talker-specific information and apparently store this information in long-term memory" (p. 881).

ACOUSTIC VARIABILITY AND L2 VOCABULARY LEARNING

Although studies of L1 speech processing and L2 phonemic contrasts suggest that indexical information is retained in long-term lexical representations, they do not address whether such information is integrated into lexical representations during the earliest stages of vocabulary acquisition. For example, it might be that in the early stages of learning new word forms, listeners attempt to abstract what is consistent across highly variable inputs and only incorporate indexical features once representations underlying new word forms have stabilized. To address the question of whether acoustic variability can affect the earliest stages of L2 vocabulary acquisition, Barcroft (2001) conducted an initial study in which beginning learners of Spanish attempted to learn 24 new Spanish words via word-picture repetition learning. The target words in the study were learned in three different conditions: (a) no variability—six repetitions of a neutral voice only; (b) moderate variability—two repetitions each of neutral, loud, and whispered voices, and (c) high variability—one repetition each of neutral, loud, whispered, excited, childlike, and nasal voices.[1] Thus, in each condition, listeners heard each phonological form the same number of times (six) but with a different number of voice types (one, three, or six). The dependent variable in the study was recall of the target words based on presentation of pictures only, a productively oriented measure of vocabulary learning.

Barcroft (2001) examined four hypotheses with regard to the potential effect of increased acoustic variability in this learning paradigm: (a) according to the *degraded input* hypothesis, increased acoustic variability should affect vocabulary learning negatively because the target words presented in nonneutral voices in the varied conditions would be less intelligible than the neutral voices in the nonvaried condition; (b) according to the *elaborative processing* hypothesis, increased acoustic variation should invoke a more elaborative type of processing, producing more robust lexical representations and in this way facilitating vocabulary learning; (c) according to the *independent modulation* hypothesis, increased variation should be normalized independently with no processing cost to the lexical encoding process and therefore should have no effect on vocabulary learning; and (d) according to the *robust versus strong connectivity* hypothesis, increased acoustic variation should result in an increased number of connections for the mental representation of each word. However, given that connections of this nature must be weaker in nature than would be the case with multiple repetitions of a word in one voice type, the combined effects of these two processes might cancel each other out and yield no effect on L2 recall as a measure of vocabulary learning. Note that these hypotheses make predictions specific to the potential effects of acoustic variability on vocabulary learning.

Consistent with the predictions of the last two hypotheses, the results of Barcroft (2001) revealed no significant differences between any of the three learning conditions. Although this finding did not exclude the independent modulation hypothesis, the combination of this finding with other findings demonstrating both costs (Mullennix et al., 1989; Nygaard, Sommers, & Pisoni, 1994) and benefits (Goldinger et al., 1991) for acoustic variability can be interpreted as being more consistent with the robust versus strong connectivity hypothesis. According to this hypothesis, acoustic variability produces a more robust representation of the new word forms, but this benefit is offset by weaker connections between individual exemplars and the new word form. Barcroft also offered several ideas for future research in this area. Among these ideas were (a) operationalizing acoustic variability in new ways, (b) using multiple speakers to create an acoustically varied condition, (c) utilizing the potential for digital manipulation of voice types, and (d) including new dependent measures of vocabulary knowledge, such as more receptively oriented measures. These four provisions, among others, were incorporated in the present study.

THE PRESENT STUDY

The purpose of the present study was to expand upon existing research on acoustic variability and the earliest stages of L2 vocabulary learning (Barcroft, 2001) in order to examine whether other types of acoustically varied formats affect measures of L2 vocabulary learning. To this end, the present study partially replicated and expanded upon Barcroft's earlier study while incorporating new dependent measures related to L2 vocabulary learning as well as alternate types of acoustically varied presentation formats. Whereas the dependent measure in Barcroft's study was accuracy of recall of target L2 words using pictures of the target words as stimuli (picture-to-L2 recall), the four dependent variables examined in each of the three experiments in the present study were (a) accuracy of picture-to-L2 recall, (b) latency for picture-to-L2 recall, (c) accuracy of L2-to-L1 recall (recall of L1 translations when presented with L2 target words), and (d) latency for L2-to-L1 recall. Both picture-to-L2 and L2-to-L1 recall were included as dependent measures in the present study in order to examine the effect of acoustic variability on both productively oriented (picture-to-L2) and more receptively oriented (L2-to-L1) measures of L2 word-form learning. Reaction time (RT) was included in an effort to provide a more sensitive overall metric of the effects of acoustic variability on L2 vocabulary learning.

In the present study, the acoustically varied format used in Experiment 1 was based on the same type of within-speaker acoustic variability examined in Barcroft (2001). Target words were produced by a single speaker in different voice types. However, Experiment 1 also incorporated new voice types and digitally manipulated voice types in an effort to increase the range of acoustic variability present in the L2 stimuli. Additionally, by changing the specific voice types from the earlier Barcroft study, we tested the generalizability of his results. In Experiment 2, the acoustically varied format examined was based on a different source of acoustic variability: between-talker acoustic variability. This experiment examined the effects of acoustic variability in speech produced by multiple talkers as opposed to speech produced by a single talker. Experiment 2 also incorporated a rotation procedure in which the voice of each talker was rotated across subjects in the moderate and no-variability conditions to control for the potential effects of talker intelligibility. Finally, Experiment 3 partially replicated Experiment 1 while using the same type of rotation procedure used in Experiment 2 to control for the potential effects of voice-type intelligibility. The results of these three experiments shed new light on the extent to which both within-talker (voice type) and between-talker sources of acoustic variability can affect L2 vocabulary learning.

Experiment 1

Experiment 1 examined whether variations in voice type produced by a single talker affect L2 vocabulary learning. The experiment used a modified version of the design employed by Barcroft (2001) but included four measures of L2 vocabulary learning performance (compared with a single measure in the earlier study). With regard to presentation format, the following modifications were made to produce the voice types and presentation levels used in the present study as compared to the presentation format used in Barcroft's study: (a) Elongated and pitch-shifted voices were created via digital manipulation of neutral voice and replaced the childlike voice and loud voice used in the earlier study; (b) the root-mean-squared (RMS) amplitude levels of all voice types were equalized with the level of the neutral voice type. These modifications were included to test whether the previously observed null effect of acoustic variability on L2 vocabulary learning would maintain with these stimulus modifications. One difficulty with using amplitude variations (loud voice type) as a source of acoustic variability is that there is evidence (Sommers et al., 1994) to suggest that listeners do not attend to variability in overall level because it does not affect phonetically relevant properties of the speech signal (e.g., formant frequency). It was considered that the new voice types included in Experiment 1 would help assure that the voice types varied (relative to the neutral voice type) along phonetically relevant acoustic properties, potentially yielding different results than those observed previously. Equalization of amplitude levels across different voice types helped to avoid confounds between variability and intelligibility (e.g., voice types such as the whispered voice might be less intelligible simply because they are lower in amplitude).

With regard to measures of vocabulary learning performance, in addition to the picture-to-L2 recall measure used in Barcroft (2001), Experiment 1 also included RT measures for picture-to-L2 recall as well as accuracy and latency measures for L2-to-L1 recall as dependent variables. The rationale for including these new measures was that, in light of inherent differences between productively versus receptively oriented measures of vocabulary learning (see Melka, 1997), one might observe that acoustic variability affects L2-to-L1 recall (a task that does not require production of the target L2 word) but not picture-to-L2 recall (a task that requires recall and production of new L2 word forms). By including L2-to-L1 recall as a dependent measure, we were also able to assess the importance of using two types of recall cues. For some words, participants were cued using target words in neutral voice, and in other instances, the target words were cued using a nasal voice. If nasal voice types represent a type of degraded input compared with neutral productions, differential effects of variability might be observed when using these two voice types as cues. Finally, RT measures were included along with the productively and receptively oriented accuracy performance measures in Experiment 1 in an effort to provide a more sensitive measure of L2 vocabulary learning. All of these modifications helped to test the generalizability of the previously observed null effect for within-speaker acoustic variability on L2 vocabulary learning.

Method.

Participants. Participants in Experiment 1 were 60 native speakers (NSs) of English with no previous formal instruction in Spanish. All of the participants were undergraduate psychology students at a private Midwestern university in the United States. Participants received credit related to requirements of their psychology class for their participation.

Experimental words. The experimental words used were 24 Spanish concrete nouns, divided into 3 groups. Each word group had a total of 20 to 21 syllables. Words 1 through 8 were *hongos* "mushrooms," *oso* "bear," *reloj* "clock," *tijeras* "scissors," *sandía* "watermelon," *cuaderno* "notebook," *ardilla* "squirrel," and *cinta* "tape." Words 9 through 16 were *fresas* "strawberries," *tiza* "chalk," *caballo* "horse," *manzana* "apple," *cebolla* "onion," *elote* "(ear of) corn," *lechuga* "lettuce," and *grapadora* "stapler." Words 17 through 24 were *loro* "parrot," *pato* "duck," *lápiz* "pencil," *conejo* "rabbit," *gato* "cat," *naranja* "orange," *basurero* "garbage can," and *pez* "fish." The experimental words were the same as those used by Barcroft (2001). In the present study, however, the words were grouped so that semantic categories (animals, fruits and vegetables, classroom items) were mixed and so that the average number of syllables in each word group was approximately the same. The average number of syllables was 2.5 for words 1 through 8, 2.63 for words 9 through 16, and 2.63 for words 17 through 24.

All of the target words had been previously recorded by a male NS of Mexican Spanish in a soundproof room. Stimuli were recorded using a sampling rate of 44.1 kHz and 16-bit resolution. The NS was instructed to read the list of the 24 target words in 4 different voice types: neutral voice, excited voice, whispered voice, and nasal voice. To obtain 2 additional voice types for each target word, all of the 24 words produced in neutral voice were subsequently digitally altered using commercially available sound-editing software to produce high-pitched and elongated varieties of each word. For the high-pitched modification, the fundamental frequency of each neutral production was increased 53% without altering any other temporal parameters of the speech signal. In the case of elongated words, duration was increased 53% without altering the fundamental frequency. The RMS amplitude of all productions was equated using digital signal processing software. This process yielded six different voice types for all of the words (neutral, nasal, elongated, whispered, excited, and pitch-shifted), in which the RMS amplitude was equated across voice type.

A separate test was conducted to ensure that the productions were highly intelligible to Spanish NSs. Four NSs of Spanish heard all 144 productions (24 items × 6 voice types) presented over headphones at approximately 75 dB sound pressure level and were asked to write down the word that was presented. Presentation order was randomized for all participants. Scoring was based on exact phonetic matches (i.e., adding, deleting, or substituting a single phoneme was counted as an incorrect response). Mean identification accuracy was 99.3% (*SD* = 0.4), indicating that all of the productions were highly intelligible to NSs of Spanish.

We conducted two additional pilot studies to determine whether naïve L2 listeners perceived six distinct voice types and to assess the quality of each production as an example of a given voice type. In the first study, all of the words (24 tokens for each of 6 voice types) were presented in random order to 4 speech-language pathologists. The four participants had a minimum of 11 years experience as licensed speech-language pathologists, and none had previous formal instruction in Spanish. Each token was presented individually. Participants were first asked to assign the production to one of the six voice-type categories. After this decision, they were asked to rate "how good an example" the token was of their selected voice type, using a seven-point scale. Correct categorization (categorizing consistent with the voice type designated at production) exceeded 95% for all four participants. Mean category goodness ratings (7 indicating an ideal example; 1 indicating a poor example) were as follows: nasal = 6.7, whispered = 6.9, excited = 6.5, neutral = 6.7, elongated = 6.9, and pitch-shifted = 6.6. In the second study, we examined whether results with the speech-language pathologists would generalize to a group of participants similar to those who would participate in the proposed experiments. Ten young adults from the student population at a private Midwestern university were asked to perform the voice-type categorization task. Mean correct categorization for the group of students exceeded 96%, with three participants obtaining perfect scores. Taken together, the results of these two pilot studies indicate that the six voice types were easily distinguished and that the productions represented good examples of each voice type.

Procedures. Participants were tested individually in a quiet room according to the following procedures. In the learning phase, each participant sat in front of the computer screen and viewed six repetitions of 24 pictures while hearing the spoken form of the Spanish word for each picture. Each picture appeared on the screen for 5 seconds. Each target Spanish word was spoken 750 milliseconds (ms) after the picture appeared. The picture remained visible for the entire 5 seconds. Each participant learned eight words in each of the three learning conditions: (a) no variability—or six repetitions of neutral voice; (b) moderate variability—or two repetitions each of neutral, nasal, and elongated voices; and (c) high variability—or one repetition each of neutral, nasal, elongated, whispered, excited, and pitch-shifted voices. Participants were informed that all of the words would be produced by a single talker although sometimes in different voice types. Word groups and presentation orders were counterbalanced across learning conditions so that 20 participants heard each of the word groups in 1 of the 3 variability conditions. Within a word group, order of presentation was constant across the six repetitions (e.g., in the no-variability condition, participants heard all six repetitions of each of the eight words repeated in the same order). In the moderate- and high-variability conditions, participants always heard the words in the same order, but the voice type associated with that item was selected randomly without replacement. We presented six repetitions of the eight words within each condition in light of prior reports (Barcroft, 2001) that indicated performance levels that did not approach either floor or ceiling using this method.

The picture-to-L2 recall posttest was administered immediately after the learning phase. All participants performed the cued recall task prior to the translation task in order to avoid additional exposure to the spoken L2 word forms (cues for the L2-to-L1 recall task were the spoken L2 word forms in nasal and neutral voices). Whereas the inclusion of L2-to-L1 recall provided a new measure for assessing the effects of acoustic variability on L2 vocabulary learning, any differences in overall performance between picture-to-L2 and L2-to-L1 recall should be interpreted in light of the order in which these two tasks were administered, given that the picture-to-L2 task made available a form of additional practice that could affect the L2-to-L1 translation task. For the L2 cued recall posttest, the participants attempted to produce the Spanish word for each picture. The picture for each target word appeared on the screen, with order determined by structured randomization. The first 3 pictures were selected at random from each of the 3 word groups in the learning phase; the second 3 words were selected at random from each of the 3 word groups based on the remaining words; the third set of 3 words was selected at random from each of the 3 word groups based on the remaining words, and so forth until all 24 words had been selected. After each set of three words was selected at random from each of the three word groups, the order in which the three words in question would appear on the posttest also was selected at random. A voice key was used to measure response latencies, which extended from the onset of the picture cue to the participant's initial vocalization of the word. The voice key was calibrated to minimize unintentional triggering while maintaining high sensitivity. Each picture appeared on the screen for a maximum of 10 seconds or until the participant provided a spoken response.

The L2-to-L1 recall posttest was administered after the cued L2 recall. For this posttest, participants heard Spanish words and were asked to provide the English translation of each word. The order of the Spanish words was determined using the same type of structured randomization used to determine the order of the picture-to-L2 recall task. Participants were given a maximum of 10 seconds to provide the appropriate English translation for each word they heard. Half the words used as cues in the translation task were presented in neutral voice. The other half of the words used as cues were presented in nasal voice. The use of neutral versus nasal voice cues was determined as follows: Within each word group (1–8, 9–16, 17–24), four words were selected at random for the neutral cue and four words for the nasal cue. Half of the participants received cues presented in this manner. For the other half of the participants, the cue types were reversed.

Scoring. The criterion used to score the picture-to-L2 recall and L2-to-L1 recall awarded 1 point for a completely correct production, 0.5 points for missing one or more of the target phonemes or using one or more incorrect phonemes within a single syllable, and 0 points for all other responses. In the case of L2-to-L1 recall, the 0.5 score was never used because participants did not make any phonetic errors in their L1.

Analyses. All accuracy scores and RTs for correct responses were submitted to repeated-measures analyses of variance (ANOVAs). When analyzing the RT data, latencies for trials receiving a score of 0 on the accuracy measure were excluded from the analyses (i.e., only fully or partially correct trials were included in the RT analyses). We also conducted the analyses using only RTs from trials that were fully correct (i.e., that received a score of 1), and the pattern of results was unchanged from that obtained when both fully and partially correct trials were included. Latencies exceeding 2.5 standard deviations (*SD*s) of the mean of individual conditions were also excluded from the analyses as outliers. Additionally, as part of a questionnaire administered at the end of the experiment, participants were asked if they knew any of the Spanish words prior to the experiment. Responses to items reported as being known previously were excluded from all analyses. Percentages in these cases were based only on the other items. In Experiment 1, two participants indicated that they knew the Spanish word for "cat" was *gato*.

In each ANOVA, condition (no variability, moderate, high) and recall type (picture-to-L2, L2-to-L1) were treated as within-subjects variables. Accuracy and RT were the dependent variables. To examine the effect of voice type used for cueing during L2-to-L1 recall, accuracy and RT scores for L2-to-L1 recall were submitted to additional repeated-measures ANOVAs. For this analysis, condition (no variability, moderate, high) and cue type (neutral, nasal) were treated as within-subjects independent variables, and accuracy and RT were the dependent variables.

Results. Means and *SD*s for accuracy and RT based on condition and recall type appear in Table 1. Results of the ANOVA for accuracy based on condition and recall type revealed that performance was significantly worse for picture-to-L2 recall than for L2-to-L1 recall, $F(1, 59) = 117.63$, $p < .001$, $\eta^2 = .667$. No other significant main effects or interactions were observed for the accuracy measure. The effects of acoustic variability were not significant for either accuracy of picture-to-L2 recall, $F(2, 118) = 1.58$, $p = .21$, $\eta^2 = .026$, or L2-to-L1 recall, $F(2, 118) = 1.7$, $p = .17$, $\eta^2 = .029$. Results of the ANOVA for RT revealed that listeners were significantly faster recalling from L2-to-L1 than from picture-to-L2, $F(1, 59) = 13.11$, $p < .001$, $\eta^2 = .198$. No other significant main effects or interactions were observed. As with the accuracy results, no significant effect of variability was observed for either the picture-to-L2 recall, $F(2, 118) = 1.2$, $p = .28$, $\eta^2 = .021$, or L2-to-L1 recall, $F(2, 118) = 0.73$, $p = .49$, $\eta^2 = .012$.

In examining whether the effects of acoustic variability differed for the two different cue types (nasal vs. neutral) in the L2-to-L1 translation task, several participants were incorrect for all trials with a particular cue type and, therefore, the sample size was reduced for this analysis. Results of the ANOVA for

Table 1. Means for accuracy and reaction time based on condition by recall type (Experiment 1)

Measure	Recall type	Condition	Mean	*SD*
Accuracy	Picture to L2	No variability	0.46	0.22
		Moderate	0.48	0.22
		High	0.52	0.22
		Total	0.49	0.18
	L2 to L1	No variability	0.68	0.22
		Moderate	0.71	0.23
		High	0.70	0.22
		Total	0.70	0.15
RT (ms)	Picture to L2	No variability	2406	1375
		Moderate	2124	956
		High	2468	1138
		Total	2333	708
	L2 to L1	No variability	1986	750
		Moderate	2011	676
		High	2026	525
		Total	2007	460

accuracy revealed a significant main effect for cue type, $F(1, 54) = 37.1$, $p < .001$, $\eta^2 = .407$, reflecting significantly higher performance using neutral voice as compared to nasal voice as a cue for L2-to-L1 recall. Overall means were 0.73, *SD* = 0.20, for neutral voice and 0.61, *SD* = 0.20, for nasal voice. No other significant main effects or interactions were observed. Results of the ANOVA for RT revealed a significant main effect for cue type, $F(1, 51) = 6.8$, $p < .05$, $\eta^2 = .117$, reflecting significantly longer RTs using nasal voice as compared to neutral voice as a cue. Overall RT means were 2243, *SD* = 947, for neutral voice and 2726, *SD* = 899, for nasal voice. No other significant main effects or interactions were observed.

Discussion. The results of Experiment 1 can be summarized as follows. As compared to neutral voice, voice-type variability did not affect L2 vocabulary learning using accuracy and latency of picture-to-L2 recall or L2-to-L1 translation as dependent measures. Although accuracy and latency of L2-to-L1 recall were not affected by acoustic variability using both neutral and nasal voice types as cues, L2-to-L1 translation performance was better when cued by a neutral voice type as compared to a nasal voice type. These results are consistent with those of Barcroft (2001) and with both the independent modulation and strong versus robust connectivity hypotheses. According to independent modulation, acoustic normalization produces L2 lexical representations that are abstract in nature (i.e., they contain lexical but not indexical information). According to the strong versus robust connectivity hypothesis, increased acoustic variation increases the number of connections between forms of a word and its referent, which facilitates L2 vocabulary learning; however, each of these connections is weaker as compared to the connection derived from multiple repetitions of a word in a single voice, which impairs L2 vocabulary learning. The combined effects of these two opposing processes result in no observable effect of acoustic variability on L2 vocabulary learning. Although the null effects for acoustic variability in Experiment 1 and the earlier Barcroft study are consistent with these two hypotheses, other types of acoustic variability warrant investigation as well. Among these other types of acoustic variability is between-talker acoustic variability, which was the focus of Experiment 2.

Experiment 2

Experiment 2 examined whether between-talker acoustic variability—variations produced by multiple talkers—would have an effect on L2 vocabulary learning based on the same four measures of L2 vocabulary learning performance included in Experiment 1. Although within-speaker changes in voice type constitute one type of acoustic variability, it is possible that properties specific to between-talker acoustic variability might yield differences in vocabulary learning performance not observable using the within-speaker variability paradigm. For example, between-talker variability might be perceptually more important than within-talker variability because in normal conversations, talker changes often serve as a signal that a new topic or new information is about to be presented. Additionally, talker variability has been shown, under certain conditions, to produce improved identification and memory of spoken words (Goldinger et al., 1991). Thus, within the framework of the robust versus strong connectivity hypothesis, talker variability might increase the strength of connections between individual word forms and referents to a greater degree than a within-speaker source of variability such as voice type. The result of such an increase in connectivity strength would be that the sum of individual connection strengths might more than offset the dispersal of connections across multiple forms, leading to improved learning for multiple compared with single talkers.

To investigate the effects of between-talker variability on L2 vocabulary learning, the acoustically varied presentation formats used in Experiment 2 were based on the production of target words by multiple talkers as opposed to multiple voice types. As in Experiment 1, three levels of acoustic variability were examined: no variability, moderate variability, and high variability. Instead of using six different voice types to create the acoustically varied conditions, the neutral voices of six different speakers were used—one speaker only in the no-variability condition, three speakers in the moderate-variability condition, and six speakers in the high-variability condition. Experiment 2 also incorporated a procedure for rotating the voices of different talkers within acoustically consistent conditions (Sommers, 1997). Whereas many studies on acoustic variability and L1 speech processing have included the voice of one speaker only in conditions of no talker variability (see Mullennix et al., 1989), the current design rotated the talkers used in the no-variability and moderate-variability conditions. In this way, the effects of single- versus multiple-speaker presentation formats could be compared while counterbalancing for potential differences related to the specific speakers used in the no-variability and moderate-variability conditions.

Method.

Participants. Participants in Experiment 2 were 60 NSs of English with no previous formal instruction in Spanish and were recruited from the same participant pool as Experiment 1. Participants received credit related to requirements of their psychology class for their participation. None of the participants from Experiment 1 took part in Experiment 2.

Experimental words. The experimental words and word groups used in Experiment 2 were the same as those used in Experiment 1. All of the target words had been previously recorded in a soundproof room by three female and three male NSs of Spanish. The female speakers were from central Spain, Chile, and Mexico. The male speakers were from northern Spain, Chile, and Mexico. Each of the NSs was instructed to read the list of the 24 target words in their normal voice. As in Experiment 1, pilot testing was conducted to determine if all of the words would be perceived correctly by NSs of Spanish. Four NSs of Spanish heard all 144 productions (24 words × 6 speakers) and were asked to write down the word that was produced. Three of the pilot participants obtained perfect scores on the test and the other missed a single item (99.3%).

Procedures. Participants were tested individually in a quiet room based on the same procedures as those used in Experiment 1 with the following exceptions. First, each participant learned eight words in each of the following three learning conditions: (a) no variability—six repetitions of each word in the voice of one speaker; (b) moderate variability—two repetitions of each word in the voice of three different speakers; and (c) high variability—one repetition of each word in the voice of six different speakers. Thus, in all three conditions, each participant heard six repetitions of each target word. Second, participants were informed that some of the words would be produced by different talkers. Next, to ensure that equal numbers of participants heard each speaker in the no-variability and moderate-variability conditions, the 60 participants were divided into 6 groups of 10 participants. In the no-variability condition, each group was assigned to the voice of only one of the six different speakers (Speakers 1 through 6). Speakers 1, 3, and 5 were men; Speakers 2, 4, and 6 were women. The assignment of speakers to the moderate-variability conditions across the six groups was as follows: Group 1 = Speakers 1, 2, 3; Group 2 = Speakers 2, 3, 4; Group 3 = Speakers 3, 4, 5; Group 4 = Speakers 4, 5, 6; Group 5 = Speakers 5, 6, 1; and Group 6 = Speakers 6, 1, 2. Note that for each group, the speaker in the no-variability condition was also included in the moderate-variability condition and that a given speaker was included in the moderate-variability condition for three of the groups. Finally, for the L2-to-L1 recall task, half of the words used as cues in the translation task were spoken by a male voice, and the other half were presented in a female voice.

Scoring. The scoring procedure was the same as in Experiment 1. For the words *manzana* and *lápiz*, however, either of two phonemes (θ or s) was accepted as correct in the place of the letter "z" in light of the regional dialects of the NSs to which the participants had been exposed.

Analyses. The analyses conducted for Experiment 2 paralleled those of Experiment 1, including deletion of outliers and inclusion of only correct or partially correct responses in the analysis of RTs. None of the participants in Experiment 2 indicated on the postexperiment questionnaire that they were familiar with any of the Spanish words. To examine the effect of speaker's voice used for cueing during L2-to-L1 recall, an additional ANOVA was conducted with condition (no variability, moderate, high) and cue type (male voice, female voice) as within-subjects variables, and accuracy and RT as the dependent variables.

Results. Means for accuracy and RT based on condition and recall type appear in Table 2 and are depicted graphically in Figures 1 and 2. Results of the ANOVA for accuracy based on condition and recall type revealed that accuracy differed significantly across the three conditions, $F(2, 118) = 108.08$, $p < .001$, $\eta^2 = .647$, and for the two different recall types, $F(1, 59) = 114.21$, $p < .001$, $\eta^2 = .659$. Additionally, a significant Condition × Recall Type interaction was observed, $F(2, 118) = 6.98$, $p < .001$, $\eta^2 = .106$. To examine the source of this interaction, we analyzed the main effects of condition sepa-

Table 2. Means for accuracy and reaction time based on condition by recall type (Experiment 2)

Measure	Recall type	Condition	Mean	*SD*
Accuracy	Picture to L2	No variability	0.38	0.23
		Moderate	0.45	0.23
		High	0.64	0.27
		Total	0.49	0.19
	L2 to L1	No variability	0.56	0.23
		Moderate	0.68	0.24
		High	0.74	0.25
		Total	0.66	0.19
RT (ms)	Picture to L2	No variability	3131	1941
		Moderate	2683	1456
		High	2218	1351
		Total	2667	1582
	L2 to L1	No variability	2466	903
		Moderate	2059	1117
		High	1723	1051
		Total	2082	1023

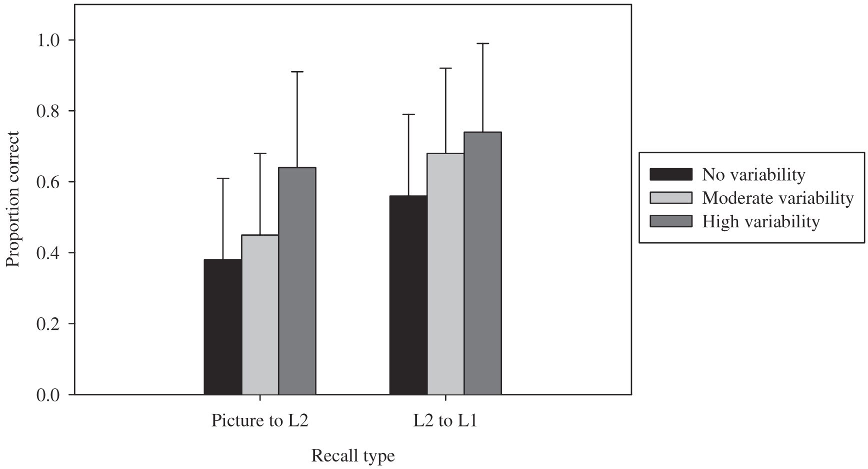

Figure 1. Effects of talker variability on accuracy of picture-to-L2 and L2-to-L1 recall (Experiment 2). Error bars represent standard deviations of the mean.

rately for the two types of recall. The effects of acoustic variability were significant for both picture-to-L2, $F(2, 118) = 62.5$, $p < .001$, $\eta^2 = .51$, and L2-to-L1 translation, $F(2, 118) = 49.1$, $p < .001$, $\eta^2 = .45$. Separate Tukey HSD post hoc comparisons using a Bonferroni correction indicated significant differences

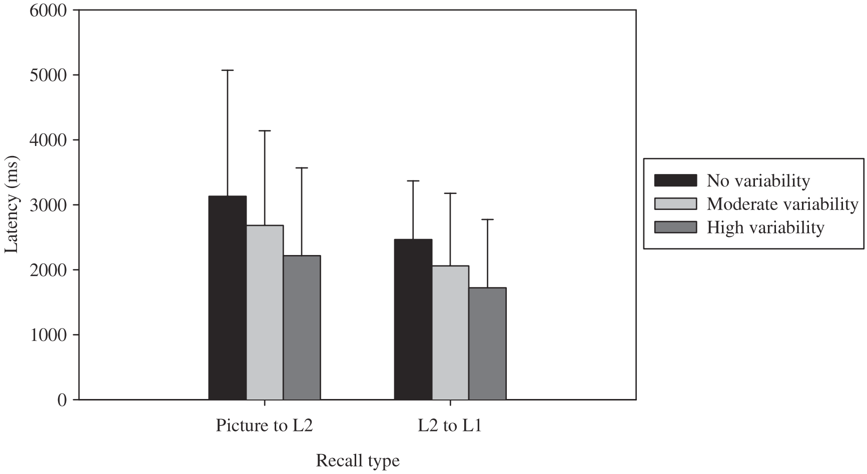

Figure 2. Effects of talker variability on latency of picture-to-L2 and L2-to-L1 recall (Experiment 2). Error bars represent standard deviations of the mean.

among all levels of acoustic variability for both types of recall (picture-to-L2 and L2-to-L1). For the picture-to-L2 recall task, differences between the no-variability and high-variability and between the medium- and high-variability conditions were both significant at $p < .001$. The difference between the no-variability and medium-variability condition was also significant, $p < .01$. For the L2-to-L1 task, differences between the no-variability and medium-variability and between the no-variability and high-variability conditions were significant at $p < .001$. The difference between the moderate- and high-variability conditions was also significant, $p < .01$.

Results of the ANOVA for RT based on condition and recall type revealed significant differences in reaction times across the three conditions, $F(2, 118) = 12.92$, $p < .001$, $\eta^2 = .185$. No other significant main effects or interactions were revealed. Pairwise Tukey HSD analyses with a Bonferroni correction for multiple comparisons indicated the following: (a) lower RTs in the high-variability condition as compared to the no-variability condition, $p < .001$, (b) lower RTs in the high-variability condition as compared to the moderate-variability condition, $p < .05$, and (c) marginally lower RTs in the moderate-variability condition as compared to the no-variability condition, $p = .06$.

Results of the ANOVA for accuracy based on condition and cue type (male or female voice) in the L2-to-L1 task revealed a significant main effect for condition, $F(2, 94) = 56.14$, $p < .001$, $\eta^2 = .548$, but no other significant main effects or interactions. Overall means were 0.67, $SD = 0.22$, for male voice and 0.66, $SD = 0.19$ for female voice. Results of the ANOVA for RT based on condition and cue type revealed a significant main effect for condition, $F(2, 94) = 23.87$, $p < .001$, $\eta^2 = .331$, but no other significant main effects or interactions. Overall RT means were 2054, $SD = 809$, for male voice and 2108, $SD = 1091$, for female voice.

Discussion. The findings from Experiment 2 can be summarized as follows. Talker variability resulted in higher accuracy and lower RTs for both picture-to-L2 and L2-to-L1 recall. The effect of variability was not moderated by the speaker's voice (male voice, female voice) used to cue the translation. To our knowledge, the results of Experiment 2 represent the first demonstration of a positive effect of acoustic variability on L2 vocabulary learning. As such, they have both important theoretical and practical implications. From a theoretical standpoint, the results are consistent with the elaborative processing hypothesis but weigh against predictions of independent modulation. Specifically, the finding that between-talker variability can improve L2 vocabulary learning argues against the independent modulation hypothesis, as this model would predict that sources of variability are normalized prior to encoding and, therefore, have no effect on vocabulary learning. Thus, the independent modulation hypothesis is not consistent with the improved L2 vocabulary learning observed with the high-variability compared with the low-variability condition of Experiment 2. Furthermore, as predicted by the elaborative processing hypothesis, we found systematic improvement in vocabulary learning moving from the no-variability to the moderate-variability to the high-variability conditions.

Thus, the results of Experiment 2 are consistent with the elaborative processing hypothesis, which maintains that talker variability can promote more robust lexical representations of new L2 vocabulary. According to this position, acoustically varied instances of each new lexical item in the input combine to form a representation that is more robust than would have been obtained by an equivalent number of acoustically consistent instances of the same item (i.e., multiple presentations of the same speech signal). The increased robustness might be due to the presence of more indexical information in each representation. Thus, although further research is warranted, the findings of Experiment 2 can be reconciled more easily with a theoretical framework in which—at least for the initial stages of learning new word forms—indexical information is retained as part of lexical representations and can facilitate access to those representations.

The differences between the findings of Experiments 1 and 2 also raise the question of why voice-type variability failed to improve L2 vocabulary acquisition, whereas talker variability produced systematic increases in vocabulary learning. One possible reason for the differences between voice-type and talker variability observed in Experiments 1 and 2 concerns the lack of rotation of voice types across participants in the moderate-variability and no-variability conditions. In Experiment 1 (as in Barcroft, 2001), all target words were presented to all participants in neutral voice in the no-variability condition and in three selected voice types in the moderate-variability condition. A potential consequence of this procedure is that the null effects of voice-type variability might reflect differences related to these specific voice types as opposed to differences between voice-type variability versus voice-type consistency in general. For example, if one or more of the voice types used in the high-variability condition were significantly less intelligible than the neutral voice type and this decrease in intelligibility resulted in worse vocabulary learning, then the potential benefits of voice-type variability might have been offset by this reduced intelligibility. One finding consistent with reduced intelligibility of specific voice types is that performance for the L2-to-L1 translation task in Experiment 1 was significantly reduced when the nasal voice was used as a cue as compared to when the normal voice was used as a cue. If this differential voice-type intelligibility contributed to the null findings in Experiment 1, then positive effects for voice-type variability might emerge using a rotation procedure similar to the one used in Experiment 2. Experiment 3 was conducted to examine this issue directly.

Experiment 3

The purpose of Experiment 3 was to determine whether the different pattern of results in Experiments 1 and 2 are attributable to (a) the use of voice type as opposed to talker as the source of variability or (b) the rotation of talker or voice types across participants in moderate- and no-variability conditions. To address this question, Experiment 3 partially replicated Experiment 1 while rotating voice types across participants in the no-variability and moderate-variability conditions. By rotating voice types in this manner, Experiment 3 counterbalanced potential differences in voice-type intelligibility, thereby permitting assessment of the effects of voice-type variability as compared to voice-type consistency independent of differences related to specific voice types.

Method.

Participants. Participants in the study were 36 NSs of English from the same participant pool as Experiments 1 and 2. They had no previous formal instruction in Spanish. All of the participants were undergraduate psychology students at a private Midwestern university in the United States. Participants received credit related to requirements of their psychology class for their participation. None of the participants from Experiments 1 and 2 took part in Experiment 3.

Experimental words. The experimental words, word recordings, and word groups in Experiment 3 were the same as those in Experiment 1.

Procedures. Procedures for Experiment 3 followed those of Experiment 1 except for the inclusion of the rotation procedure. In Experiment 3, to ensure that equal numbers of participants heard each voice type in the no-variability and moderate-variability conditions, the 36 participants were divided into 6 groups of 6 participants. The assignment of voice types (voice type 1 = normal; 2 = nasal; 3 = elongated; 4 = high pitch; 5 = whispered; 6 = excited) to the moderate-variability conditions across the six groups was as follows: Group 1 = Voice Types 1, 2, 3; Group 2 = Voice Types 2, 3, 4; Group 3 = Voice Types 3, 4, 5; Group 4 = Voice Types 4, 5, 6; Group 5 = Voice Types 5, 6, 1; and Group 6 = Voice Types 6, 1, 2.

Scoring. Scoring was identical to Experiment 1.

Analyses. The analyses conducted for Experiment 3 paralleled those in Experiment 1. On the postexperiment questionnaire, one participant in Experiment 3 indicated having known the Spanish word for “cat,” *gato*, and one indicated having known the Spanish word for “orange,” *naranja*. These two items were excluded from all analyses based on the same procedures as in Experiment 1. All accuracy scores and RTs for correct responses were submitted to two repeated-measures ANOVAs. In both ANOVAs, condition (no variability, moderate, high) and recall type (picture-to-L2, L2-to-L1) were treated as within-subjects independent variables. Accuracy was the dependent variable in the first ANOVA. Reaction time was the dependent variable in the second ANOVA. To examine the effect of the voice type used for cueing during L2-to-L1 recall, accuracy and RT scores for L2-to-L1 recall were submitted to additional repeated-measures ANOVAs. For these analyses, condition and cue type (neutral, nasal) were treated as within-subjects variables, and accuracy and RT were the dependent variables.

Results. Means for accuracy and RT based on condition and recall type appear in Table 3 and are depicted graphically in Figures 3 and 4. Results of the ANOVA for accuracy based on condition and recall type revealed that cued recall scores differed significantly across the three conditions, $F(2, 70) = 32.24$, $p < .001$, $\eta^2 = .481$, and that accuracy was higher for L2-to-L1 recall than for picture-to-L2 recall, $F(1, 35) = 31.04$, $p < .001$, $\eta^2 = .472$. No other significant main effects or interactions were observed. Tukey HSD post hoc pairwise comparisons with a Bonferroni correction for multiple comparisons indicated (a) higher accuracy in the high-variability condition as compared to the no-variability condition, $p < .001$; (b) higher accuracy in the high-variability condition as compared to the moderate-variability condition, $p < .001$; and (c) higher accuracy in the moderate-variability condition as compared to the no-variability condition, $p < .001$. Results for accuracy based on condition and measure appear in Figure 3.

Results of the ANOVA for RT based on condition and recall type revealed that latency differed significantly across the three variability conditions, $F(2, 70) = 57.53$, $p < .001$, $\eta^2 = .622$, and recall type, $F(1, 35) = 53.91$, $p < .001$, $\eta^2 = .603$. No other significant main effects or interactions were observed. Tukey HSD post hoc pairwise analyses with a Bonferroni correction for multiple comparisons indicated that RTs were (a) lower in the high-variability condition as compared to the no-variability condition, $p < .001$; (b) lower in the high-variability condition as compared to the moderate-variability condition, $p < .01$; and (c) lower in the moderate-variability condition as compared to the no-variability condition, $p < .001$. Results for RTs based on condition and recall type appear in Figure 4.

Table 3. Means for accuracy and reaction time based on condition by recall type (Experiment 3)

Measure	Recall type	Condition	Mean	*SD*
Accuracy	Picture to L2	No variability	0.45	0.19
		Moderate	0.51	0.10
		High	0.61	0.14
		Total	0.52	0.08
	L2 to L1	No variability	0.55	0.12
		Moderate	0.64	0.15
		High	0.69	0.15
		Total	0.62	0.10
RT (ms)	Picture to L2	No variability	2637	716
		Moderate	2330	503
		High	2059	647
		Total	2342	451
	L2 to L1	No variability	2316	569
		Moderate	1808	429
		High	1571	425
		Total	1898	377

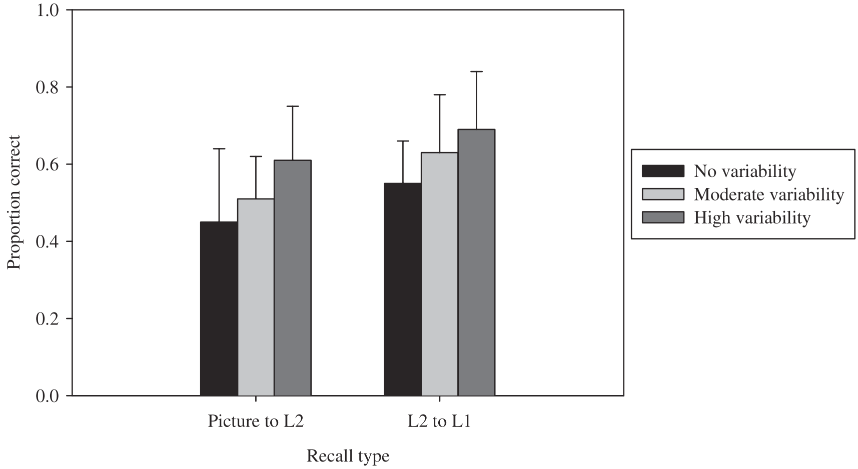

Figure 3. Effects of voice type variability on accuracy of picture-to-L2 and L2-to-L1 recall (Experiment 3). Error bars represent standard deviations of the mean.

Results of the ANOVA for accuracy based on condition and cue type (neutral, nasal) revealed a significant main effect for condition, $F(2, 54) = 5.8$, $p < .01$, $\eta^2 = .179$, and cue type, F(1,27) = 10.2, $p < .01$, $\eta^2 = .27$, but no significant interaction between condition and cue type. Results of the ANOVA for RT also

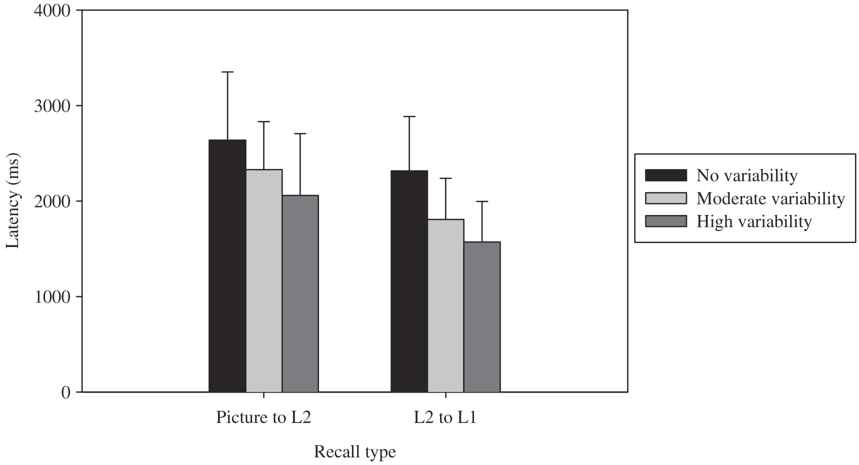

Figure 4. Effects of voice type variability on latency of picture-to-L2 and L2-to-L1 recall (Experiment 3). Error bars represent standard deviations of the mean.

revealed a significant main effect for condition, $F(2, 54) = 52.2$, $p < .001$, $\eta^2 = .658$, and for cue type, $F(1, 27) = 9.4$, $p < .001$, $\eta^2 = .27$, but no significant interaction between condition and cue type.

Discussion. The findings from Experiment 3 can be summarized as follows. Voice-type variability resulted in higher L2 cued recall scores and lower RTs for picture-to-L2 recall as well as for L2-to-L1 translation. On the L2-to-L1 translation task, accuracy was higher and RTs shorter for neutral-voice cues as compared to nasal-voice cues. These results help to reconcile the pattern of results in the study of Barcroft (2001) and Experiment 1 with those of Experiment 2. Specifically, the results of Experiment 3 suggest that the null effects of voice-type variability in Experiment 1 and in Barcroft (2001) are specific to comparisons of voice-type variability with conditions of less variability that maintain (without rotation) specific voice types, such as neutral voice in the no-variability condition and three specific voices in the moderate-variability condition. In contrast to Experiment 1 and Barcroft's study, when voice types were rotated across participants in Experiment 3, a clear positive and additive effect for voice-type variability was obtained. As such, in addition to reconciling the previous mixed findings, the results of Experiment 3 provide a second demonstration of a positive effect of acoustic-phonetic variability on L2 vocabulary learning. Whereas Experiment 2 demonstrated this effect for talker variability, a type of between-talker acoustic variability, Experiment 3 demonstrated this effect using voice type, a type of within-speaker acoustic variability.

GENERAL DISCUSSION

The findings from the present experiments improve our understanding of the effects of acoustic-phonetic variability on L2 vocabulary learning in several ways. First, Experiment 1 partially replicated and extended the findings from an earlier study (Barcroft, 2001) by demonstrating that the null effects of within-talker variability on L2 vocabulary learning in that study were not specific to the particular voice types used nor were they a result of using one specific dependent measure. Second, Experiment 2 demonstrated that at least one type of acoustic variability—talker variability—can improve L2 vocabulary learning. Whereas previous investigations have found improved performance (under some conditions) for identification and memory of L1 words as a result of talker variability (Goldinger et al., 1991), the novel aspect of the findings of Experiment 2 is that similar improvements can be observed for learning new L2 word forms. Finally, the findings of Experiment 3 tied together the previous findings of research in this area by demonstrating positive effects for voice-type variability when specific voice types were rotated across participants in conditions of moderate and no variability. As such, the combined findings of the three experiments indicate that different sources of acoustic variability (talker, voice type) positively affect L2 vocabulary learning, provided that differences in intelligibility resulting from the increased variability are controlled (e.g., by rotating exemplars across participants in the presentation formats with less or no acoustic variability).

From a theoretical standpoint, the combined findings of these experiments are consistent with an exemplar-based view of learning new word forms and the elaborate encoding view of the relationship between acoustic variability and vocabulary learning. According to an exemplar-based framework, indexical features found in acoustically varied stimuli are encoded during perceptual analysis and stored in long-term memory (Goldinger, 1998). The current findings extend these earlier results by demonstrating that retention of indexical features occurs during the earliest stages of vocabulary acquisition and that the representations formed in acoustically varied conditions are a combination of multiple instances that might have helped learners to generate more associative hooks and more robust representations for the target words. Similarly, the elaborative processing hypothesis predicts that the varied mapping between acoustic signals and word forms obligates listeners to engage in more elaborative analysis, resulting in more robust word-form representations.

Note that similar proposals have been made to explain the relationship between acoustic variability and several measures of L1 and L2 learning. For example, Goldinger et al. (1991) reported that when L1 words are presented for serial recall in a self-paced manner, recall performance is better for lists produced by multiple talkers compared with lists produced by single talkers. Within the domain of L2 learning, Logan et al. (1991) found that using multiple talkers improved acquisition of L2 phonemic contrasts compared with training that included only a single talker. All of these studies suggest that the type of processing invoked by acoustic variability can benefit learning, at least under some conditions. Pisoni (1997, p. 21) described this facilitation as follows:

> in some cases, increased stimulus variability in an experiment may actually help listeners to encode items into long-term memory (Goldinger et al., 1991). Listeners encode speech signals in multiple ways along many perceptual dimensions, and the nervous system apparently preserves these perceptual details much more reliably than researchers have believed in the past.

Thus, the picture that is beginning to emerge is that during both initial word-form learning (current study) and online speech perception (e.g., Nygaard & Pisoni, 1998), listeners retain and use indexical features to improve identification and memory for spoken words.

Figure 5 displays a simplified schematic model of the effects of acoustic variability on lexical representations within an exemplar-based framework. The lower layer represents the extent of variability in the input (no, moderate, or high) as operationalized in the current study. The upper layer represents the underlying lexical representation, with the individual ellipses representing variants of the word form. Finally, the darkness of the individual ellipses in the upper layer represents the strength with which each variant form is encoded.

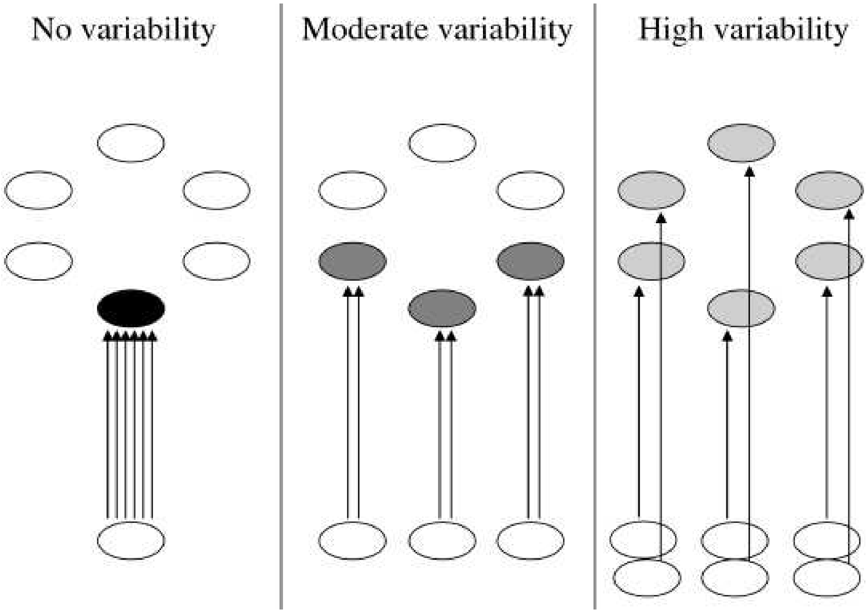

Figure 5. Model of acoustically varied input and lexical representation.

This strength is built upon exposure to multiple repetitions of a word in the same voice type (or by the same talker) over time. In the no-variability condition displayed in the far left panel, six repetitions of the identical voice type (or six repetitions by the same talker) produce a strong but minimally distributed representation of one variant of the word form. In contrast, for the high-variability condition, one repetition of six different voice types (or one repetition each produced by six different talkers) produces six weaker variants of the word form, but these six variants yield a more robust, widely distributed representation in memory. Therefore, within this model, beneficial effects of acoustic variability are obtained because the additional variants of the lexical form in the high-variability condition yield a more distributed representation of the word form with more word-form variants that can be mapped onto the semantic-conceptual representation of the word. Therefore, this model is consistent with demonstration of positive effects for talker variability (Experiment 2) and voice-type variability (Experiment 3) in the present study.

Although we view the current findings as more consistent with an exemplar-based framework for early lexical acquisition, it is important to consider how other theoretical frameworks might accommodate these findings. For example, a modified abstractionist model might propose that, despite the normalization stage, exposure to multiple variants in the high-variability condition allows listeners to extract what is consistent across the various episodes because these variants provide information about what needs to be stored and what does not need to be stored in memory. Within the model schematized in Figure 5, this proposal might be illustrated as a larger ellipse in the upper layer depicting early lexical representations. This broader lexical representation would then confer the same advantage in contacting semantic and conceptual levels as having more distributed but restricted (i.e., smaller ellipses) lexical representations. Although this alternative framework might prove acceptable for long-term lexical representations, it does not take into account the resource-demanding nature of the normalization stage. Specifically, in such a model, the benefits of exposure to multiple variants would be at least partly (and perhaps completely) offset by the additional processing demands imposed by the normalization stage. Therefore, this version of the model would have greater difficulty than one proposing retention of indexical information in accounting for the large effect sizes obtained for acoustic variability in the present study.

Another potential explanation within an expanded abstractionist perspective is that although indexical information is encoded during early word learning, canonical forms develop over time as a result of input frequencies of different word-form variants. These frequencies lead to the development of both canonical forms (e.g., a general tendency toward normal or neutral voice) and episode-specific representations, with the nature of the input determining the relative weights assigned to the different classes of representation. This type of dual-representation model has been proposed for visual (Graf & Ryan, 1990) and auditory (Church & Schacter, 1994) word recognition. Therefore, one challenge for future research is to describe the relationship between the nature of early and long-term lexical representations and—if different—to specify mechanisms by which the former develops from the latter.

In terms of L2 instruction, the present findings provide evidence of the potential utility of using acoustically varied presentation formats as a means of promoting L2 vocabulary acquisition. Holding time of exposure constant, the results of Experiments 2 and 3 revealed substantial positive effects for acoustic variability, and in both cases, the positive effect was additive in nature. In light of these findings, instructors and developers of L2 instructional materials might wish to consider incorporating greater amounts of acoustic variability during the presentation of L2 vocabulary on audiotapes, videotapes, and computer-based presentation programs. Instructors might consider also new techniques for incorporating increased acoustic variability during the presentation of new words as input in the L2 classroom. Future research on the effects of acoustic variability in other contexts of language learning, with L2 learners at different proficiency levels and with both immediate and long-term measures of vocabulary learning, should help to provide further information about the potential benefits of incorporating acoustic variability within L2 instruction.

Although these instructional implications are supported by current findings, one clarification should be made regarding the use of voice-type variability for L2 vocabulary instruction. The clarification concerns real-world implications regarding the rotation of voice types across participants in the moderate and no-variability conditions in the present study. If the study by Barcroft (2001) and Experiment 1 of the present study had revealed positive effects for acoustic variability using normal voice only in the no-variability condition (without rotating voice types), the real-world application of such a finding would be more readily apparent than the findings of Experiment 3. In the real world, vocabulary is more likely to be presented on a regular basis in neutral (conversational) voice as opposed to, for example, in nasal or excited voice. Because the six voice types in Experiment 3 were rotated across subjects in the no-variability condition, the real-world implications of these findings need to be interpreted in this light. The rotation of talkers across subjects in Experiment 2 might be less of a constraint in this regard because each talker in the no-variability condition of that study produced all of the target words in normal voice. Therefore, in terms of real-world practices, the more-readily implemented implication is to incorporate more multiple-talker presentation formats as a technique for promoting L2 vocabulary learning.

(*Received 17 November 2004*)

NOTES

1. In both the present study and the earlier investigation by Barcroft (2001), the term *nasal* is used in its common, less specialized usage to refer to speech produced with the nasal cavity occluded (e.g., as when speaking with a cold). Although based on articulatory considerations, this type of speech is actually denasalized (i.e., without contribution from the nasal cavity); we retain the term *nasal* for two reasons. First, in generating stimuli, Spanish NSs were instructed to "speak in a nasal voice." Second, in evaluating the distinctiveness of different voice types (see pilot studies), participants were asked to rate "how nasal the voice sounded." We elected to use the term *nasal* in both of these instances because this is how the desired voice type is commonly referred to and was easily understood by the participants in both cases.

REFERENCES

Assmann, P., Nearey, T. M., & Hogan, J. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, *71*, 975–989.

Barcroft, J. (2001). Acoustic variation and lexical acquisition. *Language Learning*, *51*, 563–590.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 521–533.

Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, *29*, 655.

Goldinger, S. D. (1990). Effects of talker variability on self-paced serial recall. *Research on Speech Perception* (Progress Report No. 16, pp. 313–326). Bloomington: Indiana University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the locus of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 152–162.

Graf, P., & Ryan, L. (1990). Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 978-992.

Johnson, K., & Mullennix, J. W. (Eds.). (1997). *Talker variability in speech processing.* San Diego, CA: Academic Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.

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

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 676–684.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84–102). New York: Cambridge University Press.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*, 379–390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*, 365–378.

Nygaard, L., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*, 355–376.

Nygaard, L., Sommers, M. S., & Pisoni, D. B. (1992). Effects of speaking rate and talker variability on the representation of spoken words in memory. In J. Ohala (Ed.), *Proceedings of the International Conference on Spoken Language Processing* (pp. 591–594). Edmonton, Canada: University of Alberta Press.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*, 441–452.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 915–930.

Sommers, M. S. (1997). Stimulus variability in spoken word recognition: II. The effects of age and hearing impairment. *Journal of the Acoustical Society of America*, *101*, 2278–2288.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, *96*, 1314–1324.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*, 131–145.