Applied Psycholinguistics **32** (2011), 417–434
doi:10.1017/S0142716410000469

Indexical information, encoding difficulty, and second language vocabulary learning

MITCHELL S. SOMMERS and JOE BARCROFT
Washington University in St. Louis

Received: December 22, 2008 Accepted for publication: August 2, 2009

ADDRESS FOR CORRESPONDENCE
Mitchell S. Sommers, Department of Psychology, Washington University in St. Louis, 417A Psychology Building, St. Louis, MO 63130. E-mail: msommers@wustl.edu

ABSTRACT
Research has demonstrated that second language (L2) vocabulary learning improves when target words are presented in acoustically varied compared with acoustically consistent formats. The present study investigated the extent to which this benefit of acoustic variability is a consequence of difficult encoding demands (cognitive effort hypothesis) versus increased representational quality. Experiment 1 compared L2 vocabulary learning for words produced in normal (easier encoding) or nasal (more difficult encoding) voice. Vocabulary learning was superior in the normal-voice condition, arguing against a simple cognitive effort hypothesis as the basis for improved L2 vocabulary learning with increased acoustic variability. Experiment 2 assessed the resistance of newly acquired L2 word forms to the effects of acoustic degradation. Participants heard six repetitions of each item in either a single-talker or multiple-talker condition. The robustness of the new word-form representations was assessed by measuring the accuracy and latency of L2 to first language (L1) translation as a function of the signal/noise ratio. At all four signal/noise ratios, accuracy and latency of L2 to L1 translation were significantly better for words learned in the multiple-talker as opposed to the single-talker condition. Of particular importance, the difference between single talkers and multiple talkers increased systematically as signal/noise ratio decreased. These findings suggest that the benefits of acoustic variability are a consequence of learners' ability to retain and use indexical information during the earliest stages of word learning and provide support for the representation quality hypothesis.

A main focus of research on memory for verbal materials has been to identify factors that affect retention of spoken material. One factor that has been found to promote memory for individual spoken words is the introduction of certain sources of acoustic variability, including variations in talker characteristics and speaking rate (Goh, 2005; Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Nygaard, Sommers, & Pisoni, 1995; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1993). Nygaard et al. (1995), for example, found that words spoken by multiple talkers or at multiple speaking rates were recalled more accurately than when the identical items were produced by either a single talker or

© Cambridge University Press 2011 0142-7164/11 $15.00

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

at a single speaking rate. One explanation that has been offered to account for the memorial benefits of acoustic variability is that listeners encode and retain both the lexical (word) and indexical (e.g., voice or rate) information of spoken words. In the case of variable input, individuals can use the additional indexical information as a supplemental retrieval cue to increase memory for items, relative to a condition with consistent input (e.g., single talker or single rate). More broadly, compared to conditions with no variability, variable input in the form of talker and rate variability is hypothesized to produce a more distributed representation of the presented items that listeners can use to facilitate retrieval.

In a recent series of studies (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007), the benefits of acoustic variability have been extended to memory for second language (L2) vocabulary learning. In one series of experiments, for example, Barcroft and Sommers (2005) examined the effects of variations in either voice type (e.g., excited or whispered) or talker (single vs. multiple talkers) on the ability to learn new Spanish vocabulary words. The experiments compared three different learning conditions, with exposure to each new L2 word form equivalent across the three conditions: (a) no variability—one speaker or voice type only, with six repetitions of the words presented by the same speaker or in the same voice type; (b) moderate variability—three different speakers or voice types, with two repetitions of each voice type or talker; and (c) high variability—six different speakers or voice types, with one repetition of each. The dependent measures were accuracy and latency (reaction time in milliseconds) of picture-to-L2 recall and L2-to-first language (L1) recall. The results demonstrated clear and additive benefits for variability; participants were faster and more accurate in both the picture-to-L2 and L2-to-L1 recall tasks for moderate compared with no variability and for high variability compared with moderate variability. Moreover, this effect obtained for both the within-speaker source of variability (voice type) and the between-speaker source of variability (talker characteristics).

At least two possibilities exist as to why acoustic variability of this nature has positive effects on L2 vocabulary acquisition. We refer to these possibilities as the representation quality hypothesis and the cognitive effort hypothesis. According to the *representation quality* hypothesis, acoustic variability facilitates L2 vocabulary acquisition because exposure to multiple variants produces a more robust or distributed mental representation of the new vocabulary item and in this way improves the ability of semantic–conceptual representations to contact and activate L1-form links for the word. Thus, within the representation quality framework, acoustic variability improves L2 vocabulary learning because listeners retain indexical (e.g., voice type) information during the early stages of learning lexical forms and can use this information to facilitate later retrieval of those forms.

Figure 1 provides a highly schematized illustration of the representation quality hypothesis. The lower layer represents the extent of variability in the input. The upper layer represents the underlying lexical representation, with the individual ellipses representing variants of the word form. The darkness of the individual ellipses in the upper layer represents the strength with which each variant form is encoded based 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

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

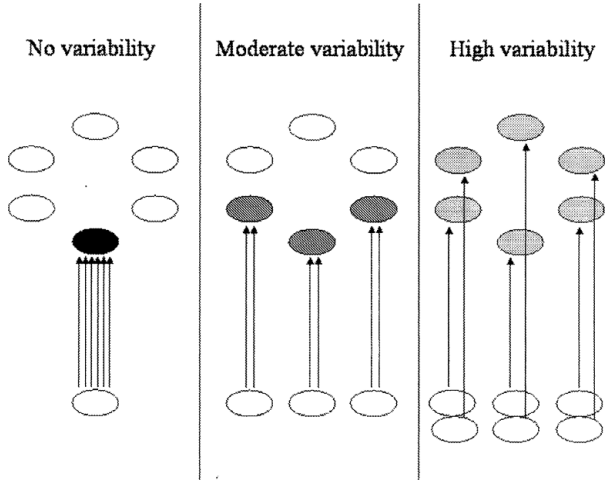


Figure 1. The model of acoustically varied input and lexical representation. The lower layer represents the extent of variability in the input, and the upper layer represents the underlying lexical representation; the individual ellipses represent variants of the word form. The darkness of the individual ellipses in the upper layer represents the strength with which each variant form is encoded, based on 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 or at the same speaking rate) produce a strong but minimally distributed representation of one variant of the word form. Conversely, in the high-variability condition displayed in the far right panel, one repetition of six different voice types (or one repetition each produced by six different talkers or at the same speaking rate) produces six weaker variants of the word form, but these six variants yield a more robust, widely distributed representation in memory.

in the far left panel, six repetitions of the identical voice type (or six repetitions by the same talker or at the same speaking rate) 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 or at the same speaking rate) 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 and L1-form links to the word.

An alternative account of the beneficial effects of acoustic variability, the *cognitive effort* hypothesis, however, maintains that relative to acoustically consistent formats, words presented in acoustically varied formats require additional cognitive processing in order for learners to encode the less intelligible variants, which in turn leads to more extensive encoding of the target words. Although little systematic research has been directed at examining the effects of cognitive effort on L2 vocabulary learning, considerable evidence is available from the L1 memory

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

literature that increased difficulty during initial encoding can produce significant gains in immediate memory tests relative to conditions with less difficult initial encoding (for a review, see Roediger, Gallo, & Gerarci, 2002). Thus, the cognitive effort hypothesis attributes the benefit of acoustic variability to greater encoding effort rather than to more distributed representations.

The results to date therefore suggest that the benefits of acoustic variability on L2 vocabulary acquisition may be a consequence of a more distributed lexical representation, increased encoding difficulty, or a combination of both. The primary goal of the present study was to test the basic tenets of the cognitive effort hypothesis by comparing L2 vocabulary acquisition as a function of voice intelligibility (Experiment 1) and to provide new data pertinent to the representation quality hypothesis by assessing the resistance of newly acquired L2 word forms to the effects of acoustic degradation (Experiment 2). With regard to Experiment 1, evidence for superior word learning when words are spoken with a less intelligible (nasal[1]) voice type compared with a more intelligible (normal) voice would support the cognitive effort hypothesis because the decreased intelligibility of words produced in a nasal voice require greater cognitive effort during acquisition. Conversely, superior word learning for normal compared with nasal voice would be inconsistent with encoding difficulty as the basis for the beneficial effects of acoustic variability because it would demonstrate that greater encoding difficulty did not produce positive effects on L2 vocabulary learning. Sommers and Barcroft (2006) provided evidence that L1 words produced with a nasal voice are less intelligible than words produced with a normal voice among native English speakers. Pilot Study 1 of the present study (see Methods Section) also provided evidence that the target words produced in nasal voice were less intelligible to native Spanish speakers than when produced in normal voice.

With regard to Experiment 2, as noted one way that increased variability in some indexical properties could facilitate L2 vocabulary acquisition is by promoting development of a more robust lexical representation. Experiment 2 tested this proposal by assessing the resistance of newly acquired L2 word forms to the effects of acoustic degradation. If acoustic variability enables listeners to develop more robust lexical representations, then words learned with variable input should be more resistant to distortion than items learned with consistent input. It is of the most importance that we would expect that the benefits of acoustic variability would increase under more adverse listening conditions. That is, if variability serves to promote more robust representations, then the benefits of such variability would be expected to increase as listening conditions become more difficult, such as through decreased signal/noise (S/N) ratios.

To summarize, the purpose of Experiment 1 was to test the basic tenets of the cognitive effort hypothesis by examining whether presenting words in a less intelligible voice (nasal voice) would facilitate L2 vocabulary learning. Because a less intelligible voice is associated with increased encoding difficulty, if learning new words in a nasal (less intelligible) voice produces superior vocabulary learning relative to words spoken in a neutral or conversational voice the cognitive effort hypothesis for the effects of acoustic variability on vocabulary learning would be supported. Evidence to the contrary, however, would weigh strongly against the cognitive effort hypothesis. The purpose of Experiment 2 was to assess

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

whether increases in indexical information associated with L2 vocabulary learning in acoustically varied contexts would help to resist subsequent effects of acoustic degradation. If this is the case, then words learned with variable input should be more resistant to distortion than items learned with consistent input. Moreover, the difference between variable and consistent input should increase as S/N ratios become poorer. Thus, under the representation quality hypothesis we would expect significantly greater L2 vocabulary learning for variable compared with consistent input at all S/N ratios and that the differences between single and variable input would increase with poorer S/N ratios.

EXPERIMENT 1

Participants

Participants were 36 native speakers 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. None of the participants reported a history of speech or hearing difficulties. 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 three groups. Each word group had either 20 or 21 syllables. Words 1–8 were *hongos* (mushrooms), *oso* (bear), *reloj* (clock), *tijeras* (scissors), *sandía* (watermelon), *cuaderno* (notebook), *ardilla* (squirrel), and *cinta* (tape). Words 9–16 were *fresas* (strawberries), *tiza* (chalk), *caballo* (horse), *manzana* (apple), *cebolla* (onion), *elote* ([ear of] corn), *lechuga* (lettuce), and *grapadora* (stapler). Words 17–24 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) would be 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 1–8, 2.63 for Words 9–16, and 2.63 for Words 17–24.

All of the target words had been previously recorded from a male native speaker of Mexican Spanish using a soundproof room. Stimuli were recorded using a 44.1-kHz sampling rate and 16-bit resolution. The native speaker was instructed to read the list of the 24 target words in four different voice types: neutral voice, excited voice, whispered voice, and nasal voice. Only the stimuli produced in neutral and nasal voices were used for this study.

Pilot studies

Pilot Study 1. We conducted Pilot Study 1 to determine whether the experimental words produced in nasal voice were actually more difficult for native speakers of Spanish to encode. Four native speakers of Spanish, who had not recorded any of

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

the experimental stimuli, heard all 48 productions in nasal and normal voices (24 items × 2 voice types) presented in noise over headphones. A small amount of background noise was introduced to avoid ceiling level performance by the native Spanish speakers. The S/N ratio was approximately 0 dB (i.e., the signal and noise were approximately the same amplitude). All 48 word–voice combinations were presented pseudorandomly (without replacement). Thus, voice type was manipulated as a repeated-measures variable. Results of an analysis of variance (ANOVA) revealed a significant main effect for voice type, F (1, 3) = 13.34, $p = .035$, $\eta^2 = 0.816$. Word identification performance was significantly lower for words presented in nasal voice ($M = 27.5$, $SD = 10.91$) compared to words presented in normal voice ($M = 36.00$, $SD = 7.75$), demonstrating that the target words were significantly more difficult to encode when produced in nasal voice than when produced in normal voice.

Pilot Study 2. We conducted Pilot Study 2 to determine whether L2 naive listeners perceived the nasal and normal stimuli as distinct voice types and to assess the quality of each production as an example of one of the two voice types. All 48 tokens (24 words × 2 voice types) were presented (along with tokens in other voice types) in random order to four 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 six voice types.[2] After this decision, they were asked to rate on a 7-point scale "how good an example" the token was of their selected voice type. Correct categorization (categorizing consistent with the voice type designated at production) exceeded 95% for all four participants. Mean category goodness ratings (7 = *an ideal example*, 1 = *a poor example*) were 6.7 for both normal and nasal productions.

Taken together, the pilot studies indicated that the stimuli were more difficult to encode when presented in nasal voice than when presented in normal voice (Pilot Study 1), and that there were clear perceptual distinctions between the nasal and normal voice types used in the study (Pilot Study 2).

Procedures

Participants were tested individually in a quiet room according to three procedures

Procedure 1. 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. The same presentation order was used for all six repetitions. Each picture appeared on the screen for 5 s. In each case, the target Spanish word was spoken 750 ms after the picture appeared. Half of the participants heard the words produced in a normal voice and the other half heard them spoken in a nasal voice.

Procedure 2. The picture-to-L2 recall posttest was administered immediately after the learning phase. All participants performed the picture-to-L2 recall task

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

prior to the L2-to-L1 recall 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 normal voices). For the picture-to-L2 recall task, the participants attempted to produce the Spanish word for each picture. The picture for each target word appeared on the screen in a random order different from the order used during the learning phase. 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. Each picture appeared on the screen for a maximum of 10 s or until the participant provided a spoken response.

Procedure 3. 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. Half of the participants from each encoding group (nasal voice vs. normal voice) heard the translation cue in the same voice type used for the learning phase (normal [or nasal] voice at both study and test) and half heard the translation cue in a different voice type than at study (normal voice at study and nasal voice at test or vice versa). The order of the Spanish words was a random order different from the orders used for the learning phase and for the picture-to-L2 recall task. Participants were given a maximum of 10 s to provide the appropriate English translation for each word they heard. Response latencies were recorded with the same equipment used in the picture-to-L2 recall task.

Scoring

The criteria used to score the picture-to-L2 recall and L2-to-L1 recall was 1 point for a completely correct production, 0.5 points for errors in a single syllable, and 0 for all other responses. Scoring was done by a native speaker of English who was fluent in Spanish. A subset of the data (eight participants) was also scored by a second fluent Spanish speaker and interrater reliability was approximately 0.91. Production of a word or part of a word in nasal voice or any other voice type was never counted as an error as scoring was based only on the presence or absence of target phonemes within each target word. 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

Latencies for trials receiving a score of 0 on the accuracy measure were excluded from the analyses. Latencies exceeding 2.5 *SD* of the mean of individual conditions were also excluded from the analyses as outliers. Approximately 2.3% of the responses in the normal voice at study and 2.9% of the responses in the nasal voice at study conditions were excluded as outliers.

Scores were submitted to two repeated-measures ANOVAs with accuracy as the dependent variable in the first and latency as the dependent variable in the second. For both analyses, voice at study (normal, nasal) was a between-subjects independent variable, and recall type (picture-to-L2, L2-to-L1) was a within-subjects independent variable.

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

Table 1. *Means for accuracy and latency based on voice at study by recall type for Experiment 1*

Measure	Recall Type	Voice	*M*	*SD*
Accuracy	Picture-to-L2	Normal	0.46	0.12
		Nasal	0.27	0.08
		Total	0.36	0.14
	L2-to-L1	Normal	0.62	0.14
		Nasal	0.46	0.14
		Total	0.53	0.16
	Total	Normal	0.54	0.14
		Nasal	0.36	0.14
		Total	0.45	0.10
RT (ms)	Picture-to-L2	Normal	1900	222
		Nasal	2279	225
		Total	2090	292
	L2-to-L1	Normal	1560	320
		Nasal	1848	283
		Total	1704	331
	Total	Normal	1730	314
		Nasal	2063	314
		Total	1897	222

Note: L2, second language; L1, first language; RT, reaction time.

Results

Means and standard deviations for accuracy and latency based on voice at study and recall type appear in Table 1. These results are depicted graphically in Figure 2. Results of the ANOVA on accuracy revealed that performance was significantly better when study items were spoken with a normal, compared with a nasal voice, $F\,(1, 34) = 29.31$, $p < .001$, $\eta^2 = 0.463$. Accuracy was also higher for L2-to-L1 recall than for picture-to-L2 recall, $F\,(1, 34) = 46.86$, $p < .001$, $\eta^2 = 0.580$. The interaction was not significant. Overall means for accuracy were approximately 1.5 times higher for normal voice at study than for nasal voice at study. Results of the ANOVA on latency paralleled those for accuracy; participants responded significantly faster when study words were studied in a normal voice type than when they were studied in a nasal voice type, $F\,(1, 34) = 20.29$, $p < .001$, $\eta^2 = 0.374$, and they were significantly faster for the L2-to-L1 recall than for picture-to-L2 recall, $F\,(1, 34) = 62.46$, $p < .001$, $\eta^2 = 0.648$. Again, the interaction was not significant. The mean reaction times were approximately 19% faster for normal voice at study than for nasal voice at study.

Discussion

The results of Experiment 1 indicate that presenting L2 words in a less intelligible nasal voice during study reduced both the accuracy and latency of later picture-to-L2 recall and L2-to-L1 recall compared with presenting the same words in a

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

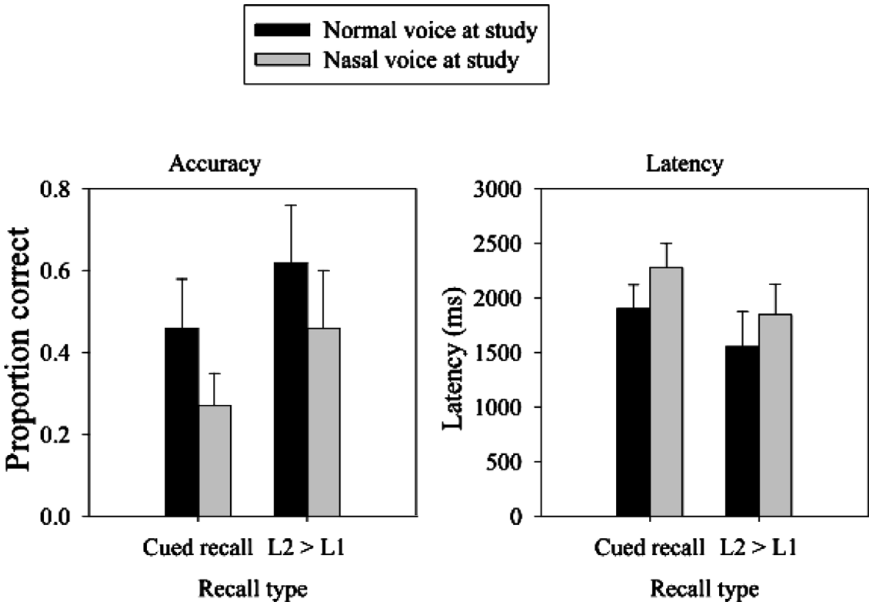


Figure 2. The effects of voice under study (nasal vs. normal) on the accuracy and latency of picture-to-second language (L2) and L2-to-first language (L1) recall in Experiment 1.

normal voice. The theoretical implications of these results provide strong evidence against the proposal that differences in cognitive effort underlie the positive effects of acoustic variability on L2 vocabulary learning. Specifically, the cognitive effort hypothesis predicts positive effects for learning new words in the less intelligible nasal voice because of increased encoding difficulty and consequent greater processing demands for this condition. In contrast to this hypothesis, the present experiment demonstrated that presenting L2 words in a less intelligible nasal voice negatively affected L2 vocabulary learning, based on lower accuracy and longer latencies in the nasal-at-study condition. Although this finding does not provide direct evidence favoring the representation quality hypothesis as the mechanism mediating the benefits of acoustic variability on L2 vocabulary learning, it does provide strong evidence against at least one alternative explanation.

One potential concern with interpreting the findings from Experiment 1 as evidence against the cognitive effort hypothesis is that the poorer accuracy for words spoken in the nasal voice may reflect impaired encoding rather than any effect of cognitive effort. That is, the lower intelligibility of words spoken in a nasal voice may have simply made it more difficult for listeners to encode all of the phonemes in the novel words. Under this scenario, the reduced accuracy for the nasal productions would not index increased cognitive effort, but instead would serve as a metric of intelligibility. A strong argument against this position, however, is the latency data from Experiment 1 (cf. Macht & Buschke, 1983, for evidence of reaction time as a measure of cognitive effort). Recall that the latency data

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

were calculated based only on correct trials and that for both dependent measures latency was significantly slower for the nasal than for the normal productions. Thus, even when encoding was completely accurate, both picture-to-L2 and L2-to-L1 recall were significantly slower for words spoken in a nasal compared with a normal voice type. These findings are therefore consistent with the proposal that recall of the nasal productions required greater cognitive effort than recall of the normal productions and that these additional demands reduced individuals' ability to remember the L2 words.

EXPERIMENT 2

Experiment 2 was designed to assess whether the benefits of multiple- over single-talker input for L2 vocabulary acquisition results from differences in the quality of the underlying representation. As in previous studies, participants were asked to learn novel Spanish words in either a low-variability (one talker) or high-variability (six talkers) condition. Accuracy and latency for L2-to-L1 translation was then measured at four different S/N ratios (+2, 0, −2, −4). Consistent with previous results (Barcroft & Sommers, 2005), we expected increased accuracy and shorter latencies for the multiple compared to the single talker condition at each S/N ratio. Most important, however, is that if the benefits of multiple talker input derive from producing a more robust or higher quality representation, we would expect an interaction between talker condition (single vs. multiple) and S/N ratio, such that differences between single and multiple talker performance would be greater at lower S/N ratios. That is, as the perceptual conditions become more difficult with decreasing S/N ratio, the benefits of a higher quality representation resulting from more variable input should be magnified.

Method

Participants. A total of 72 participants took part in Experiment 2, with half in the single-talker condition and half in the multiple-talker condition. None of the participants from Experiment 1 took part in Experiment 2. All participants were between the ages of 18 and 25 and reported normal hearing with no history of speech or hearing difficulties. Participants had no previous formal training in Spanish and received either course credit or $10/h for participating.

Stimuli. The same 24 words used in Experiment 1 were used in the current experiment. Stimuli were recorded by a total of eight native Spanish speakers (four males, four females). Productions for six of the talkers (three males, three females) were used for training. These are the same six talkers used by Barcroft and Sommers (2005) to demonstrate the benefits of variable input. Additional demographic information about these six talkers is provided in the earlier study. The remaining two talkers (one male, one female) were used for testing so that both the single- and multiple-talker training groups were tested with a novel voice. Both of the talkers who produced the stimuli (cues) for testing were between the ages of 25 and 30. The male talker was from southern Mexico; the female talker was from South America.

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

To generate test stimuli with different S/N ratios, the original versions of all stimuli were equated for overall level at approximately 72 dB SPL. Four different versions of each stimulus were then created by adding white noise at the level needed to achieve each of the four S/N ratios (e.g., white noise at 70 dB SPL was added to achieve the +2 S/N ratio). The noise was generated with Adobe Audition and started 50 ms prior to signal onset and extended 50 ms after signal offset.

Design and procedures. The experiment followed a 2 (Condition: single talker vs. multiple talker) × 4 (S/N ratio = +2, 0, −2, −4) mixed design with condition as a between-participants variable and S/N ratio as a within-participants variable. Participants first completed a training phase followed by a test phase.

Training. All participants heard six repetitions of each of the 24 vocabulary words. In the single-talker condition, participants heard all 24 words presented by the same talker. These words were repeated six times in the same order. To minimize differences in talker intelligibility across single- and multiple-talker conditions, each of the six talkers from the multiple talker condition served as a talker for six participants in the single-talker condition (i.e., talker A served as the single talker for six participants, talker B served as the single talker for a new set of six single-talker participants, and so forth). Training for the multiple-talker condition was exactly the same as for the single-talker condition except that each presentation of the 24 vocabulary words was from a new talker. Participants heard all 24 words by one of the talkers in the same order used for the single-talker condition, and these items were then repeated (in the same order) by the remaining five talkers (i.e., all 24 words were produced by a given talker before a new talker was presented).

For both the single- and multiple-talker conditions, participants 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 s. In each case, the target Spanish word was spoken 750 ms after the picture appeared.

Testing

In the testing phase, participants were presented with one of the 24 Spanish words learned in the training phase and were asked to provide the English equivalent as quickly and accurately as possible. The Spanish words were embedded in white noise that began 50 ms prior to the word and ended 50 ms after word offset. The S/N ratio associated with a given word was counterbalanced across participants in each group such that each word was presented at a given S/N ratio to approximately the same number of participants in each group. A total of six words were presented at each S/N ratio, and participants received all six items at a given S/N ratio consecutively. The order of S/N ratios was determined randomly for each participant.

Results

Accuracy. Figure 3 displays accuracy and latency for the L2-to-L1 translation task as a function of S/N ratio. Means and standard deviations for each condition

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

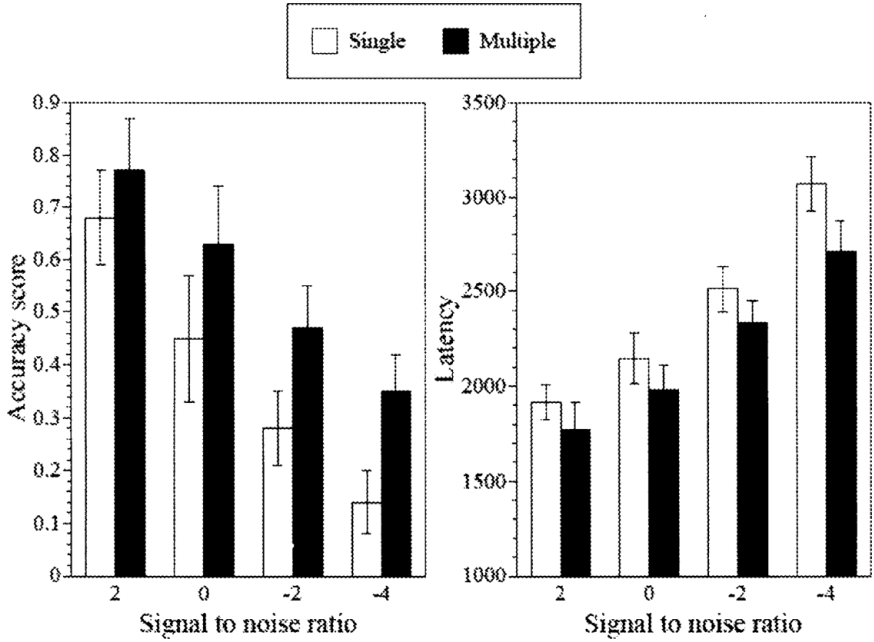


Figure 3. The accuracy and latency of the second language to first language translation as a function of the signal/noise ratio for Experiment 2.

are provided in Table 2. The accuracy data were analyzed using a 2 (Condition: single vs. multiple talker) × 4 (S/N Ratio = +2, 0, −2, −4) mixed ANOVA with condition as a between-participants variable and S/N ratio as a within-participants variable. As expected, accuracy differed significantly across S/N ratio, $F(3, 210) = 319.2$, $p < .001$, $\eta^2 = 0.82$. Consistent with previous results for the effects of talker variability on L2 vocabulary learning, accuracy was significantly higher for multiple- than for single-talker training, $F(1, 70) = 245.6$, $p < .001$, $\eta^2 = 0.78$. Post hoc comparisons of performance at each S/N ratio with a Bonferroni correction for multiple comparisons indicated that accuracy was significantly higher for the multiple-talker condition at each S/N ratio ($p < .001$ for all comparisons). Of particular importance to the current study, the difference between single and multiple talker conditions increased significantly as a function of S/N ratio, as indicated by a Significant Condition × S/N Ratio interaction, $F(3, 210) = 8.4$, $p < .001$, $\eta^2 = 0.11$.

Latency. The right side of Figure 3 displays mean L2-to-L1 translation latencies as a function of S/N ratio for correct trials. Reaction times below 250 ms (less than 1% of all trials) and exceeding 3000 ms (approximately 2.4% of single-talker trials and 3.1% of multiple-talker trials) were removed as outliers. The remaining data were analyzed using a mixed design ANOVA that paralleled the one used for analyzing accuracy. It was not surprising that the latencies were significantly slower

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

Table 2. *Means for accuracy and latency of second language to first language recall as a function of signal/noise ratio (S/N) in Experiment 2*

Measure	Condition	S/N	*M*	*SD*
Accuracy	Single talker	+2	0.66	0.09
	Multiple talker	+2	0.73	0.10
	Total	+2	0.69	0.10
	Single talker	0	0.45	0.12
	Multiple talker	0	0.64	0.11
	Total	0	0.54	0.15
	Single talker	−2	0.28	0.07
	Multiple talker	−2	0.47	0.09
	Total	−2	0.38	0.12
	Single talker	−4	0.14	0.06
	Multiple talker	−4	0.36	0.07
	Total	−4	0.25	0.12
Latency (ms)				
	Single talker	+2	1915	192
	Multiple talker	+2	1770	247
	Total	+2	1843	242
	Single talker	0	2147	233
	Multiple talker	0	1979	234
	Total	0	2063	257
	Single talker	−2	2514	220
	Multiple talker	−2	2336	217
	Total	−2	2425	248
	Single talker	−4	3073	243
	Multiple talker	−4	2712	260
	Total	−4	2893	336

for poorer S/N ratios, $F\,(3, 210) = 890.4$, $p < .001$, $\eta^2 = 0.93$. The analysis also indicated that translation latencies were shorter for multiple- than for single-talker training $F\,(1, 70) = 170.6$, $p < .001$, $\eta^2 = 0.71$. Post hoc comparisons for each S/N ratio after a Bonferroni correction for multiple contrasts indicated significantly shorter latencies for the multiple-talker condition at each S/N ratio, $p < .001$ for all four S/N ratios. Again, of particular importance to the current investigation, the difference between single- and multiple-talker conditions varied as a function of S/N ratio, as indicated by a Significant Condition × S/N Ratio interaction, $F\,(3, 210) = 20.4$, $p < .001$, $\eta^2 = 0.23$.

Discussion

The results of Experiment 2 provide important new evidence with regard to the mechanism underlying the positive effects of acoustic variability on L2 vocabulary learning. Specifically, the significant interaction between learning condition and S/N ratio provides, to our knowledge, the first evidence that the type of lexical representation produced by vocabulary learning with multiple talkers is

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

qualitatively different than the type of lexical representation produced by vocabulary learning with single talkers. If lexical representations stemming from single and multiple talkers were similar in nature from a qualitative standpoint, one would not expect to see the positive effects of multiple-talker vocabulary learning increase as a function of decreasing S/N ratios, but that is precisely what was observed in Experiment 2. This finding provides strong evidence that lexical representations emerging from multiple-talker learning are more robust and provide learners with increasingly better opportunities to access and retrieve recently learned word forms as the quality of the cue provided to retrieve the word form decreases. As such, the results of Experiment 2 provide strong evidence in support of the representation quality hypothesis because it is only the representation quality hypothesis that specifically predicts the type of qualitative differences needed to explain how increasingly degraded stimuli moderated the increasingly strong positive effects of acoustic variability in this experiment.

GENERAL DISCUSSION

The combined findings of Experiments 1 and 2 provide evidence in favor of the representation quality hypothesis over the cognitive effort hypothesis as an explanation for the benefits of acoustic variability on L2 vocabulary learning. First, in direct contradiction to predictions of the cognitive effort hypothesis, in Experiment 1 we observed that a condition of vocabulary learning involving increased cognitive effort (nasal vs. normal voice) did not result in improved vocabulary learning. Instead, the learning condition associated with *more* cognitive effort resulted in substantially *less* L2 vocabulary learning. This finding might be viewed as sufficient, or at least largely sufficient, to rule out the cognitive effort hypothesis and to point us more in the direction of the representation quality hypothesis, which focuses on differences in the quality of lexical representations produced by acoustically consistent versus acoustically varied input.

The results of Experiment 2 provide more direct evidence in support of the representation quality hypothesis, by documenting that words learned with multiple talkers as opposed to a single talker were increasingly more resistant to degradations in stimulus quality (in this case, decreases in S/N ratios). If the representations of words learned via single versus multiple talkers were not qualitatively different, no interaction between learning condition and degree of stimulus degradation should have emerged in Experiment 2. Thus, the combined findings from the two experiments favor the representation quality explanation for the benefits of acoustic variability on L2 vocabulary learning.

Indexical information is encoded in developing lexical representations

If the positive effects of acoustic variability are because of the development of more robust lexical representations, precisely how do the more robust lexical representations emerge from the acoustically varied input? To begin, we propose that the positive effects of talker, voice type, and speaking-rate variability on L2 vocabulary learning are because of indexical information related to these sources of variability being encoded along with developing L2 lexical representations and

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

in this way creating a more distributed, qualitatively different type of representation than would be the case in acoustically consistent learning conditions. In the case of talker, for example, if a leaner is presented with six repetitions of the Spanish word *ardilla* (squirrel) by six different talkers, the learner may encode indexical properties such as [female], [soft voice], and [slight lisp] based on the first presentation of the word; indexical properties such as [male] and [raspy voice] based on the second presentation of the word; and indexical properties such as [female] and [childlike voice] based on the third presentation of the word. As another example, consider the case of speaking-rate variability: if a learner is presented with six repetitions of the word ardilla in six different speaking rates, the learner may encode indexical properties such as [somewhat slow rate] based on the first presentation of the word, [very fast rate] based on the second presentation of the word, [somewhat slow rate] based on the third presentation of the word, and so forth. The encoding of indexical features such as these allows learners to expand the representational space that is used to map semantic–conceptual information onto corresponding lexical forms. Thus, in much the same way that exposure to multiple variants of category exemplars (e.g., exposure to a wide variety of chairs) will aid in classifying novel instances of a category (a bean-bag chair), increased acoustic variability can function to expand the representational space of lexical categories, thereby easing the mapping of semantic-conceptual information onto that category.

The theoretical account that we are describing here is consistent not only with the findings of the present study and previous demonstrations of the benefits of talker, voice type, and speaking-rate variability on L2 vocabulary learning (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007), but also with the findings of other studies on the effects of talker variability on certain types of memory for L1 words. For example, Martin et al. (1989) found that listeners encode information about the voices of different talkers and can use this information to aide in subsequent retrieval. Goldinger (1990) also found that when participants were allowed to self-pace word presentations, they selected a slower pace for word lists presented by multiple talkers compared to word lists presented by a single talker and later correctly recalled more of the words on the multiple-speaker lists (we note the inherent time confound here as an important caveat). These findings suggest that indexical properties of speech can facilitate later retrieval of those items when participants are allowed enough time to encode these properties. The findings are also consistent with Goldinger's (1998) exemplar-based model, which posits that lexical and indexical information form two components of an integrated representation and that both types of information contribute to lexical processing.

In considering the relationship between the L1 memory studies and the L2 vocabulary learning results, it is important to note that the locus of the improvement for variable, compared with consistent input is proposed to be within memory representations. An alternative possibility, however, is that the effects represent an online perceptual advantage for acoustically variable conditions. Previous results from studies of acoustic variability and perception of spoken words (Mullennix, Pisoni, & Martin, 1989; Nygaard et al., 1995; Sommers, Nygaard, & Pisoni, 1994), however, argue strongly against a perceptual basis for the beneficial effects

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

of variability. Specifically, studies of acoustic variability and perception of spoken words consistently report a negative effect of input variability; perceptual identification in noise, word naming, and lexical decisions, all of which assess online perception, are also all negatively affected by increased acoustic variability. Thus, the representation quality hypothesis ascribes the benefits of acoustic variability to changes within the representations stored in long-term memory, rather than online perceptual facilitation.

Variability and the development of new lexical and phonemic categories

The results of the Experiment 2 along with previous studies on acoustic variability and L2 vocabulary learning (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007) suggest that variability can facilitate learning new word form–semantic concept mappings (e.g., that the word-form "ardilla" maps onto semantic concepts associated with squirrels). The findings from recent developmental studies (Rost & McMurray, 2009; Singh, 2008) suggest that variability can also serve to promote learning lexical categories. For example, Rost and McMurray (2009) reported that infants trained to discriminate pairs of lexical neighbors (words differing by a single phoneme such as /bih/ and /dih/) based on multiple-talker input were able to learn the discrimination, but those trained using single-talker input were not. Similarly, Singh (2008) found that affect variability during a familiarization phase promoted infants' ability to recognize novel word forms presented in sentences to a greater extent than when only a single affect was used for familiarization. Finally, Lively, Logan, and Pisoni (1993) investigated the effects of acoustic variability on adult listeners' ability to learn nonnative phonetic categories. In their study, native Japanese listeners trained to discriminate the English /r/–/l/ distinction with multiple talkers generalized this acquired ability to novel voices, whereas those trained with a single talker failed to demonstrate such generalization.

Taken together, the picture that is beginning to emerge from studies of acoustic variability and word-form learning is that sources of variability, including talker characteristics, speaking rate, affective content, and voice type promote the development of a more robust and distributed representation that can have beneficial effects both for the acquisition of new word forms and for generalizing to novel instances of the word forms (i.e., generalizing to novel talkers). Singh (2008) has suggested that one mechanism by which increased variability may lead to more distributed representations is that high-variability input may focus attention away from surface form similarities (e.g., talker characteristics, affective quality) and direct attention to subtle phonological variations. With time and experience, the range of phonological variations that are accepted as instances of a given lexical category broadens, leading to more distributed representations. Thus, input variability may represent a basic mechanism of category formation, whereby variable input results in an expanded set of representations that facilitate mapping new instances onto that category. Future work in our laboratory will further investigate this idea by investigating whether previous evidence for the beneficial effects of variability extend to other aspects of language processing (e.g., learning morphosynactic abilities) and to explore further the mechanisms underlying these beneficial effects.

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

NOTES

1. As in previous studies (Barcroft, 2001; Barcroft & Sommers, 2005), we use the term nasal here in its common, less specialized, usage to refer to speech produced with the nasal cavity occluded (e.g., as when speaking with a cold). This type of speech, however, is actually denasalized based on articulatory considerations (i.e., without contribution from the nasal cavity). We retain the term nasal for two reasons. First, in generating stimuli, the native Spanish speakers were instructed to "speak in a nasal voice." Second, in evaluating the distinctiveness of the two voice types in this study (see Methods Section), 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 referred to commonly and was easily understood by the participants in both cases.
2. We obtained data for six different voice types, including the nasal and normal voices used in the present study. We report only the data for the nasal and normal voice types but similar results were also obtained for the remaining four voice types.

REFERENCES

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

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27, 3*, 387–414.

Goh, W. D. (2005). Talker variability and recognition memory: Instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 40–53.

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, IN: 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 nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152–162.

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*, 5521–2421.

Macht, M. M., & Buschke, H. (1983). Age differences in cognitive effort in recall. *Journal of Gerontology, 38*, 695–700.

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, 4*, 676–684.

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. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989–1001.

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. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication, 13*, 109–125.

Roediger, H. L., Gallo, D. A., and Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory*, *10*, 319–322.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*, 339–349.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*, 833–870.

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press

Sommers, M. S., & Barcroft, J. (2006). Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *Journal of the American Acoustical Society 119*, 2406–2416.

Sommers, M. S., & Barcroft, J. (2007). An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics, 28*, 231–249.

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.

https://doi.org/10.1017/S0142716410000469 Published online by Cambridge University Press