RESEARCH ARTICLE

Effects of Sociophonetic Variability on L2 Vocabulary Learning

Friederike Fichtner[1] (ORCID), Joe Barcroft[2], Mitchell Sommers[3] and Paul Olejarczuk[4]

[1]Languages and Cultures, California State University, Chico, Chico, CA, USA; [2]Romance Languages and Literatures, Washington University in St. Louis, St. Louis, MO, USA; [3]Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA and [4]English, California State University, Chico, Chico, CA, USA

Corresponding author: Friederike Fichtner; Email: ffichtner@csuchico.edu

Abstract

Acoustic variability refers to variations in speech that do not alter linguistic content. Previous studies have demonstrated that acoustic variability improves second language (L2) word learning when varying talker, speaking style, or speaking rate but not amplitude or fundamental frequency (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007). The current study examined the effects of region-based sociophonetic variability. In Experiment 1, English speakers attempted to learn German nouns while viewing pictures and listening to the words with low sociophonetic variability (six speakers of one regional variety, one repetition per speaker) and high sociophonetic variability (six speakers of each of six different regional varieties, one repetition per speaker). Participants completed picture-to-L2 and L2-to-first language (L1) posttests. Experiment 2 replicated Experiment 1 while counterbalancing word groups and learning conditions. Results of both experiments revealed increased accuracy for high over low variability, suggesting that regionally varied exemplars of words lead to more robust developing lexical representations.

Introduction

The speech to which listeners are exposed contains two key types of information. One is linguistic, which includes phonemic contrasts that distinguish between words, such as /bɛr/ 'bear' vs. /pɛr/ 'pear.' The other is indexical, which includes different sources of acoustic variability that do not alter linguistic content, for instance, allophonic differences between different regional varieties, such as how the word *bear* is pronounced as [bɛɹ] in American English and as [beə] in British English. Other examples of indexical information would include when the same word is spoken by different talkers, in different speaking styles (excited voice, whispered voice, and child-like voice, among many others), at different speaking rates, at different amplitudes, at different fundamental frequency (F_0) levels, and so forth. *Acoustic variability* refers to variations of

indexical properties of this kind in the speech signal that provide listeners with information about the nature and context of the speech (e.g., *Who is the speaker? How fast are they speaking? What is the mood of the speaker?*) without altering linguistic content.

One intriguing finding that has emerged over the past two decades in research on second language (L2) acquisition (SLA) and lexical input processing is that increasing acoustic variability in spoken input positively affects L2 vocabulary learning, such as when using talker, speaking style and speaking rate as sources of variability when presenting novel L2 words as input (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007). For example, accuracy scores for learning L2 Spanish vocabulary increased by 68% (relatively, from a mean of .38 to a mean of .64) when changing from no talker variability, or six repetitions of each target word by one talker, to high variability, or one repetition each by six talkers.

The present study explored the extent to which the benefits of acoustic variability on L2 vocabulary learning might extend to *sociophonetic variability*, a source of variability based on variations in speech that emerge from sociolinguistic variables such as region, socio-economic status, gender, age, or language ideology. More specifically, the study investigated the effects of phonetic variation tied to different regional varieties of German, using input samples from first language (L1) speakers of German from different regions of Germany and Austria. Would acoustic variability of this nature increase, detract from, or have no effect on L2 word learning?

Review of previous research

Before providing details about the methods and results of this study, we first consider several areas of previous research related to acoustic variability. The organization of this research review is roughly historical, beginning with studies on acoustic variability and L1 speech processing from the 1980s and 1990s. We then turn to studies on the benefits of talker variability on memory for L1 words presented in lists and the benefits of talker variability as part of instructional programs that target challenging L2 phonemic contrasts, such as the distinction between word-initial /r/ and /l/ in English for L1 speakers of Japanese. Next, we arrive at our primary focus, which is research since the 2000s on the effects of different sources of acoustic variability on L2 vocabulary learning. Finally, we conclude our review by highlighting an interesting pattern of findings regarding the effects of different sources of acoustic variability on L1 speech processing and L2 vocabulary learning, one that may be accounted for by the *extended phonetic relevance hypothesis* (ePRH) (Sommers & Barcroft, 2007). The ePRH proposes that only those sources of acoustic variability that affect properties pertinent to phonological contrasts in a language spoken by the listeners/learners in question will affect word learning. This hypothesis is pertinent to the potential effects of sociophonetic variability, the previously uninvestigated source of acoustic variability upon which this study is focused because properties of sociophonetically varied speech do indeed meet this criterion of the ePRH.

Acoustic variability and speech processing

Much of the early research on acoustic variability focused on the cognitive costs of talker variability during L1 speech processing. Studies indicated less accurate and slower speech processing when participants were presented with lists of words spoken

by multiple talkers as compared to single talkers, as evidenced by decreased performance in vowel perception (Assmann et al., 1982), word recognition (Mullennix et al., 1989; Ryalls & Pisoni, 1997), and word naming (Mullennix et al., 1989). This pattern of negative effects for talker variability was extended to speaking-rate and speaking-style variability, respectively, in word identification studies by Sommers et al. (1994) and Sommers and Barcroft (2006, Experiment 3). Speaking-style variability was based on one talker who produced six speaking styles or voice types: normal, excited, denasalized, elongated (computer-assisted), child-like, and whispered. From a theoretical standpoint, these findings suggest that lower performance in speech processing tasks can be due to the additional cognitive resources needed either (a) to normalize acoustically varied input (to strip away indexical features from the linguistic content), which is an abstractionist perspective, or (b) to encode and retain multiple exemplars of acoustically varied input in memory, which is an exemplar-based perspective.

Sommers et al. (1994), in their exploration of sources of variability other than talker, discovered that overall amplitude had little effect on L1 word identification. Based on this finding, they proposed the *phonetic relevance hypothesis* (PRH), which maintains that only sources of variability pertinent to phonetic discrimination will affect L1 word identification performance. Sommers and Barcroft (2006, Experiment 2) provided additional evidence to support this hypothesis by confirming that F_0 variability also had no significant effect on word identification performance for native speakers of English. In sum, this line of research on L1 speech processing revealed negative effects for talker, speaking-style, and speaking-rate variability but not for amplitude and F_0 variability, at least for the English speakers who participated in these studies.

Acoustic variability and memory for L1 words in lists

In addition to numerous demonstrations of the cognitive costs of acoustic variability during L1 speech processing, one also finds some early evidence of benefits of talker variability when participants were allowed sufficient time to process L1 words in lists. Mullennix et al. (1989), for example, confirmed improved memory for L1 words in lists spoken by multiple talkers as compared to words spoken by a single talker. In another study, Goldinger et al. (1991) compared the effects of multi-talker vs. single-talker presentation of words while using three inter-stimulus intervals (ISIs). At an ISI of .5 seconds, memory for L1 words in lists was greater for the single-talker condition; at an ISI of one second, there was no difference between single-talker and multi-talker conditions; and at an ISI of four seconds, memory was greater for the multi-talker condition than for the single-talker condition (see also Nygaard et al., 1995, for more evidence of the impact of ISI length, as related to other sources of variability). These findings clearly point to the need for sufficient time to encode exemplar-specific information in acoustically varied input if acoustic variability is to serve as a means to improving memory.

Acoustic variability and learning phonemic contrasts

Another body of research, now more within the area of SLA, explored the potential benefits of using acoustically varied input as part of instructional programs designed to teach challenging phonemic contrasts in L2, especially the English contrast between the liquid consonants /r/ and /l/ as in the cases of *read* vs. *lead* and *river* vs. *liver*. Findings of these studies pointed toward the potential benefits of acoustic variability to help

learners improve their perception and production of such contrasts (Logan et al., 1991; Lively et al., 1993; Lively et al., 1994; Bradlow et al., 1997; see also Hardison, 2003). Theoretical accounts of these benefits focused on how varied indexical features may help to establish clearer phonemic categories. As Logan et al. (1991) put it, acoustically varied input during training may help listeners develop "stable and robust phonetic categories that show perceptual constancy across different environments" (p. 876). Note, however, that a recent study by Wiener et al. (2020) indicated no effect of talker variability on training for English speakers' ability to produce L2 Mandarin tone contours except when talker variability interacted with explicit instruction.

Acoustic variability and vocabulary learning

All five sources of variability that have been tested with regard to their effects on different measures of L1 speech processing have also been tested with regard to their effects on L2 vocabulary learning and, in one case, with regard to L1 vocabulary learning. The five sources that have been examined with regard to L2 vocabulary learning are talker, speaking style, speaking rate, amplitude, and F_0 (Barcroft & Sommers, 2005; Sommers & Barcroft, 2007; Barcroft & Sommers, 2014), and the source that has been examined for L1 vocabulary learning was talker (Sommers et al., 2008). In what follows, we review effects observed for each of these sources of variability as they relate to vocabulary learning, noting that the design and methods used in all of these studies were similar to those used in the present study.

To begin, Barcroft and Sommers (2005) explored the potential effects of speaking-style (voice type) variability and talker variability on L2 Spanish word learning in participants with no prior study of Spanish. Effects of speaking-style variability were assessed in two experiments, one in which individual speaking styles (neutral, excited, whispered, denasalized, pitch-shifted, elongated) were not rotated across participants (i.e., the same speaking style was used in the no variability condition for all participants in Experiment 1), partially replicating an earlier study by Barcroft (2001) that found null effects for speaking style (voice type) variability when such a rotation was not used, and one in which individual speaking styles were rotated across an equal number of participants (Experiment 3). Experiment 2, on the other hand, assessed the potential effects of talker variability while rotating individual talkers across an equal number of participants. The purpose of the rotation procedure was to counterbalance any potential effects of any given single talker in Experiment 2 or any given single speaking style in Experiment 3, to focus on the variability-vs.-no variability question only. In all three experiments, the researchers compared vocabulary learning in conditions with no variability (one speaking style or one talker), moderate variability (three speaking styles or three talkers), and high variability (six speaking styles or six talkers). Posttest measures included both accuracy and speed (reaction time [RT] in ms) of picture-to-L2 recall and L2-to-L1 translation. Results of the study indicated positive effects of a graded nature for both speaking-style and talker variability when the rotation procedure was used (Experiments 2 and 3) but null effects for speaking-style variability when the rotation procedure was not used (Experiment 1), in the latter case corroborating the null effects observed by Barcroft (2001). Moreover, the benefits observed for both speaking-style and talker variability were reflected in higher accuracy scores as well as faster RTs.

From a theoretical standpoint, Barcroft and Sommers (2005) offered an account of these findings that focused on the extent to which the formal aspects of developing lexical
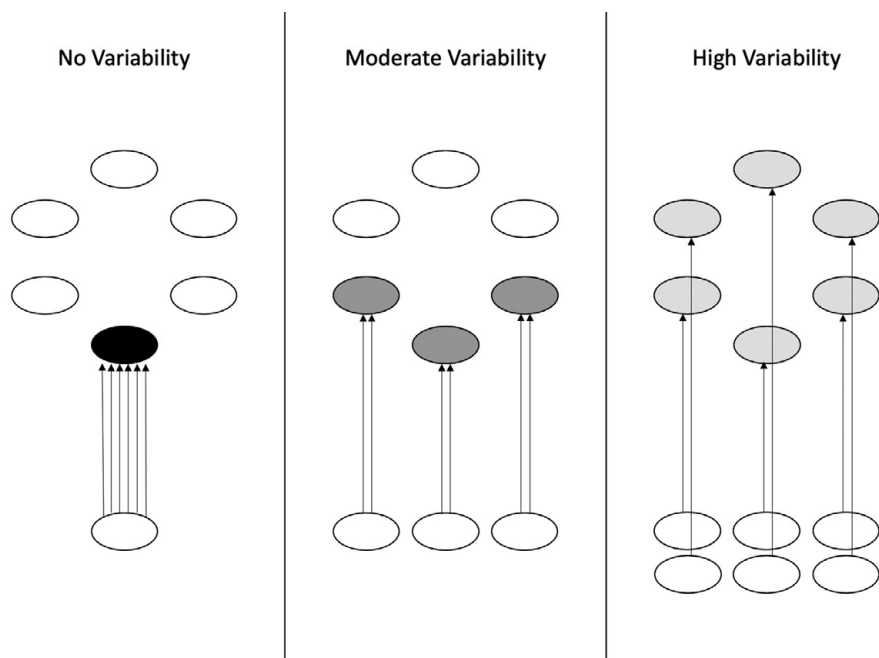
Figure 1. Degree-of-distribution model of the effects of acoustic variability on developing lexical representations.
Note: Adapted from Barcroft & Sommers, 2005.

representations are distributed. As depicted in Figure 1, this degree of (exemplar) distribution (DD) model of the effects of acoustically varied input (referred to as "Model of acoustically varied input and lexical representation" in the 2005 article) posits that acoustically consistent exemplars (without variability) in the input lead to stronger representations (darker ovals in the figure) that are less distributed in nature whereas acoustically varied exemplars (with variability) lead to weaker (lighter ovals in the figure) but more distributed (robust) representations of developing lexical forms. The ultimate degree of distribution obtained is graded in nature, such that the larger the number of acoustically varied exemplars in the input, the greater the distribution (dispersion in lexicosemantic space). With reference to the results of Experiments 2 and 3, the graded increases in speaking-style variability and talker variability increased, in a graded manner, the extent to which the new L2 Spanish word forms were distributed. For this reason, improvements in vocabulary learning (both higher accuracy and faster retrieval speeds in ms) in the moderate variability conditions fell between (and were significantly different from) those in the no variability and high variability conditions.

If acoustic variability leads to more robust developing lexical representations, how is it that the greater distribution supports increased accuracy when cued by a picture of a target word? The answer concerns the likelihood that the activated semantic space for a word might encounter and connect with exemplars of developing word forms. The presentation of a picture of the referent of a word will activate the (L1-based) semantic space for the word, and if the word form in question is more distributed, there is a greater likelihood of connecting to ('hooking onto') one or more of the exemplars in question.

Sommers and Barcroft ([2007](#)) extended research on acoustic variability and L2 vocabulary learning by assessing three new sources of acoustic variability: overall amplitude, F_0, and speaking rate. The study was designed to test the ePRH, which predicts that only phonetically relevant sources of variability would lead to improved L2 vocabulary learning. Considering that amplitude and F_0 do not fall within that category (amplitude affecting the perception of loudness and F_0 affecting perception of pitch without altering other phonetically relevant parameters), at least for the L1 English-speaking participants attempting to learn novel L2 Spanish vocabulary in the study, it was hypothesized that amplitude and F_0 variability would not have significant effects on learning L2 vocabulary. Speaking-rate variability, on the other hand, which included phonetically relevant properties of speech for the participants, was expected to produce positive effects. The results bore out these hypotheses, revealing graded benefits (higher accuracy, faster RTs) for speaking-rate variability when going from no variability to moderate variability to high variability conditions but null effects for both amplitude and F_0.

Are the benefits of acoustic variability we see emerging from this line of research limited to L2 Spanish or do they also emerge in other L2s and extend to L1 vocabulary learning as well? A study by Sommers et al. ([2008](#)) suggests that the benefits are not limited on either front. Their study examined whether the benefits of talker variability would extend to learning novel words in L2 Russian (Experiment 1), novel words (pseudowords) with known vs. novel objects (nonobjects) as referents (Experiment 2), and low-frequency concrete nouns in L1 English. In all three experiments, L1 English speakers attempted to learn eight novel words in each of three conditions: no variability (one talker), moderate variability (three talkers), and high variability (six talkers). The results of all experiments revealed positive and additive effects of talker variability on accuracy and speed of vocabulary recall, suggesting that the positive effects of acoustically varied input emerge in a range of different types of lexical learning.

An alternate explanation of the benefits of acoustic variability concerns the amount of cognitive effort required to encode acoustically varied input. From this perspective, cognitive effort of this nature constitutes a type of desirable difficulty that is beneficial, which at least arguably could be due to the need for additional abstracting of linguistic content from varied exemplars, applying an abstractionist view to vocabulary learning from acoustically varied input. Sommers and Barcroft ([2011](#)) sought to compare this *cognitive effort hypothesis* with the *representation quality hypothesis* based on their previous accounts of the benefits of variability being attributable to more distributed (robust) representations of lexical exemplars encouraged by acoustically varied input. The first experiment in the study compared the effects of learning novel words using input in normal voice (easier encoding) vs. input in denasalized voice (more effortful encoding) and demonstrated—in contrast to the cognitive effort hypothesis and in favor of the representation quality hypothesis—more vocabulary learning in the normal voice condition. The second experiment then compared accuracy and latency of L2-to-L1 translation at four different signal-to-noise ratios (SNRs) as a means of assessing the robustness of developing lexical representations of words learned with and without acoustically varied input. At all four SNRs, words learned with acoustic variability were produced more accurately and faster. Moreover, the degree to which performance was better for words learned with acoustic variability increased as a function of the extent to which SNR decreased. These results provide additional evidence in favor of the representation quality hypothesis while disfavoring the cognitive effort hypothesis.

Barcroft and Sommers (2014) tested predictions of the ePRH by assessing the effects of F_0 variability with L1 speakers of a tone language (Zapotec) for whom F_0 was contrastive as compared to participants who did not speak a tone language. According to the ePRH, only the tone language speakers should have an opportunity to benefit from the F_0 variability. Bilinguals in Zapotec (a tone language in which F_0 is lexically contrastive) and Spanish were compared to bilinguals in Spanish and English, creating a situation in which F_0 should only be phonetically relevant to the first of these two groups. Both groups attempted to learn 24 Russian concrete nouns in three conditions: one level of F_0 for six repetitions (no variability); three levels of F_0 for two repetitions (moderate variability); and six levels of F_0 for one repetition (high variability). Specific levels of F_0 were counterbalanced across participants. The findings of the study demonstrated (based on proportion scores) significantly improved L2 vocabulary learning for speakers of the tone language (Zapotec) but null effects for the non-tone language speakers. These findings are wholly consistent with the predictions of the ePRH.

How might the Fuzzy Lexical Representation (FLR) Hypothesis (Gor et al., 2021) and the Ontogenesis Model (OM) of the L2 Lexical Representation (Bordag et al., 2022) relate theoretically to mechanisms underlying the positive and null effects of acoustic variability on L2 vocabulary learning? While these proposals about evolving L2 lexical representations share a usage-based perspective with the DD model, the latter is different in that it proposes specific mechanisms that account for the benefits of acoustically varied input during lexical input processing (see Barcroft, 2015) and the earliest stages of word form learning. The FLR hypothesis focuses on "fuzzy" (meaning imprecise) encoding of lexical representations and its consequences (such as lexical confusions and slow lexical access). The DD model, in contrast, addresses online, real-time, millisecond-by-millisecond processing of novel word forms and how learners derive *lexical intake* (initial data that becomes available to the developing lexicose-mantic system). The OM, which assumes the same type of imprecision as the FLR hypothesis does, centers on how L2 learners develop incrementally over time with regard to different aspects of word knowledge, something that was never disputed among L2 vocabulary researchers both prior to and after the OM was proposed. The OM does not explain the benefits of acoustically varied input either. One might argue from an OM perspective that acoustically varied input leads to increasingly more precise word form encoding, but the general argument that increased precision happens over time does not explain the mechanisms underlying this particular benefit. From the DD perspective, in contrast, it is the robustness and greater degree of distribution of formal lexical features that account for it.

Do the benefits of acoustically varied input, which is a form-oriented manipulation, imply that variability in the use of referent tokens when presenting target L2 words should also have a positive effect? While it may seem intuitive to answer this question positively, from the perspective of the *type of processing – resource allocation* (TOPRA) model (Barcroft, 2002), the answer is negative because of the critical distinction that must be made between form-oriented vs. semantically oriented processing, the trade-offs that can take place between these two types of processing, and the increases and decreases in different aspects of word learning that occur as a result of such trade-offs. Specifically, the TOPRA model predicts that *referent token variability*, variations in aspects related to the referents of target words (e.g., six different pictures of the referent vs. one picture of the referent), can decrease L2 word form learning by exhausting limited processing resources in the direction of semantic analysis at the expense of encoding novel word forms. Sommers and Barcroft (2013) tested this prediction

directly by examining the effects of three levels of referent token variability: one picture of each referent × six repetitions for no variability; three pictures of each referent × two repetitions each for moderate variability, and six pictures of each referent × six repetitions each for high variability). In contrast to the positive effects of different types of acoustic variability, which are form-oriented, the results of the study indicated negative effects for referent token variability, which is semantically oriented. The negative effects of referent token variability were also graded in nature, producing a mirror image in the opposite direction when compared to the graded positive effects of form-oriented sources of variability such as talker, speaking style, and speaking rate. As such, these findings, in combination with those of studies on form-oriented sources of variability, are consistent with the TOPRA model.

Three potential limits on the benefits of acoustic variability that warrant consideration are (a) how it affects the performance of children learning novel L2 words; (b) how its effects may change when substantially different study phase and testing procedures are used; and (c) whether its positive effects extend to L2 grammar learning. Regarding the first of these, Sinkeviciute et al. (2019) found that whereas adults benefited greatly from talker variability when learning novel L2 words, children did not, suggesting that age may be an important factor to consider. Regarding the use of different study-phase and testing procedures, note that in Uchihara et al.'s (2022) study, Japanese speakers attempted to learn 40 words in L2 English by studying them in blocks of five words each and were tested immediately after each block. This procedure differed greatly from Barcroft and Sommers' (2005) study phase, in which English speakers attempted to learn 24 words in L2 Spanish in blocks of eight words each without being tested until all blocks were completed. The results of Uchihara et al. indicated no significant main effect for talker variability on word learning, which was likely due to the nature of the study-test procedure utilized. This procedure also led to performance levels that approached the ceiling and, as such, restricted the extent to which the positive effects of talker variability could be observed. Additionally, the very low delayed posttest scores in the study can also be attributed to the insufficiently challenging five-item vocabulary learning task, not requiring the type of encoding needed for long-term retention of vocabulary items. Finally, Bulgarelli and Weiss (2021) indicated that a high level of talker variability (eight talkers) neither hindered nor facilitated learning target features of an artificial grammar, whereas a limited level (two talkers) was found to interfere with grammar learning under more difficult conditions of grammar learning.

Acoustic variability, speech processing, and word learning: Is there a pattern?

Research to date on the effects of acoustic variability on L1 speech processing and L2 vocabulary learning has unveiled an intriguing pattern of effects as one considers different sources of acoustic variability. Sources that pose cognitive costs to speech processing are precisely the same as those that improve word learning. As predicted by the ePRH, phonetically relevant sources of variability investigated to date, which include talker, speaking-style, and speaking-rate variability, all decrease performance on tasks related to L1 speech processing (e.g., decreased performance in L1 word identification) and improve L2 (and L1) vocabulary learning. This pattern becomes clear in Table 1 when one considers the first three rows marked as "Yes" in the second column for sources of variability that are phonetically relevant. Also predicted by the ePRH, sources of variability that are not phonetically relevant, which include

Table 1. Effects of different variability sources on L1 word identification and L2 word learning

Source of variability (L=language)	Phonetically relevant?	L1 word identification	L2 word learning
Talker	Yes	Negative[1]	Positive[4]
Speaking style	Yes	Negative[2]	Positive[4]
Speaking rate	Yes	Negative[2,3]	Positive[5]
F_0 for tone-L speakers	Yes	?	Positive[6]
F_0 for non-tone-L speakers	No	Null[2]	Null[5,6]
Amplitude	No	Null[2,3]	Null[5]

Note: [1]Mullennix et al. (1989); [2]Sommers & Barcroft (2006); [3]Sommers et al. (1994); [4]Barcroft & Sommers (2005); also Sommers et al., 2008 for talker and L1 vocabulary learning; [5]Sommers & Barcroft (2007); [6]Barcroft & Sommers (2014).

amplitude, have no effects on tasks related to L1 speech processing (e.g., no difference in performance in L1 word identification) and no effects on L2 vocabulary learning. Finally, most uniquely consistent with the ePRH is the finding that F_0 variability improves L2 word learning for speakers of a tone language (for whom F_0 is phonetically relevant) but not for speakers of non-tone languages (for whom it is generally not). Rows 4 and 5 in Table 1 depict this pattern in the findings, while the effects of F_0 variability on L1 word identification among tone-language speakers remain to be confirmed, the clear prediction of the ePRH being a negative effect on L1 speech processing (e.g., L1 word identification) for this source of variability.

What other sources of acoustic variability might be added to Table 1, and what are the predictions of the ePRH for them? The present study sought to take a new step toward answering this question by considering the effects of one type of sociophonetic variability on L2 vocabulary learning. Several reasons underlie the immediate rationale for testing this type of variability. First, we note the predominance of regional sociophonetic variability in language in general. Second, there is always the potential that the positive effects of phonetically relevant sources of acoustic variability may asymptote, and the presence of sociophonetic variability beyond talker variability, as tested in the present study, constituted a unique opportunity to yield a potential asymptote of this nature. Third, investigating sociophonetic variability provides a research-based foundation for making decisions about the extent to which it should be included in L2 instruction, which is often an unresolved issue. Does exposure to an increased range of regional varieties improve L2 word learning? Should instructors encourage students to work with different L2 regional varieties? Questions such as these are addressed by this study.

How might sociophonetic variability affect L2 vocabulary learning?

As mentioned previously, sociophonetic variability is a type of acoustic variability that is based on sociolinguistic variables such as region, socio-economic status, gender, age, or language ideology. Because phonetic variations resulting from these variables tend to be phonetically relevant to listeners across languages, according to the ePRH, increases in different types of sociophonetic variability should lead to increases in L2 (and L1) vocabulary learning. The type of sociophonetic variability assessed in this study was based on region, which falls within what the ePRH predicts for any type of sociophonetic variability.

Sociophonetic variability includes *phonetic variation*, that is, variations in phones (sounds that need not be linguistically contrastive) without changes in meaning, but not

phonemic variation. *Phonemic variation*, on the other hand, refers to variations in phonemes, which are linguistically contrastive. Different variants of target words included as stimuli in this study represented cases of phonetic variation and not phonemic variation (dialect-specific phone changes in variants of the target words in the study did not lead to different word meanings). Exemplars of target words varied only phonetically, even in cases with large phonological distance between regional variants, such as through increased number of deletions and substitutions between different regional variants of a target word, as with *Birne* 'pear' being pronounced as [biːɐn] in Upper Austrian and [brɐnə] in Hamburg dialect. This study was designed to explore how variability of this nature might affect L2 vocabulary learning.

In addition to potentially large amounts of phonological distance, another potential issue that can arise when working with sociophonetic variability concerns the degree to which variation at the phonetic level may co-occur with variation at the phonemic level. For example, vowel changes tied to different regional varieties can lead to (at least perceived) changes in linguistic content at the lexical level, as in the case of the Canadian English word *about* being interpreted by speakers of other varieties of English as *a boat*. Such cases were avoided in the study. Moreover, this study did not include region-based *lexical variation*, which would involve truly different words for the same referent, such as *soda* vs. *pop* in American English or *Semmel* vs. *Brötchen* for 'bun' in German. In sum, the present study focused on region-based sociophonetic variability, for which the target spoken forms varied phonetically but not phonemically.

Research question

In light of existing findings about the effects of different sources of acoustic variability on vocabulary learning, the purpose of this study was to assess how region-based sociophonetic variability, which has yet to be investigated, affects L2 word learning. Specifically, the study was guided by the following research question:

> What are the effects (if any) of increased region-based sociophonetic variability in the input on L2 vocabulary learning?

An answer to this question will help to determine whether region-based socio-phonetic variability, which is a phonetically relevant source of acoustic variability, positively affects L2 word learning, as predicted by the ePRH. Additionally, the answer will have important implications for language instruction with regard to the extent to which the presence of sociophonetic variability should be encouraged in the L2 classroom and instructional materials. The high ecological validity of the study, which included a range of real-world regional varieties of a language, is a feature that dovetails nicely with the goal of increasing sociolinguistic awareness and linguistic diversity in L2 classrooms (see e.g., Modern Language Association Ad Hoc Committee on Foreign Languages, 2007), including with regard to different varieties of English and other pluricentric languages around the world.

Experiment 1

Participants

Participants in Experiment 1 were 36 English-speaking students with no prior study of German who were undergraduates at a private university in the midwestern USA.

Table 2. Target German words by word group with translations

Word group 1		Word group 2	
Target word	Translation	Target word	Translation
Rauch	smoke	*Ente*	duck
Tisch	table	*Stiefel*	boot
Kirsche	cherry	*Fenster*	window
Löffel	spoon	*Gabel*	fork
Vogel	bird	*Birne*	pear
Nilpferd	hippo	*Pfirsich*	peach
Pilze	mushrooms	*Käfer*	bug
Schlüssel	key	*Staubsauger*	vacuum cleaner
Zahnbürste	toothbrush	*Kaninchen*	bunny
Briefkasten	mailbox	*Mülleimer*	trash can
Schildkröte	turtle	*Regenschirm*	umbrella
Seifenblasen	bubbles	*Kugelschreiber*	pen

Target words

The target words used in the study were 24 concrete nouns in German that were not cognates with English and that would likely exhibit regional variation when spoken. Two word groups of 12 words each were created. The mean number of phonemes was equal (m=6 in word group 1 and word group 2), and the mean number of syllables was nearly equal (m=2.25 in word group 1 and m=2.42 in word group 2). The effort to equalize word length in this manner was taken as one step to minimize differences in learning difficulty in the two learning conditions, in addition to counterbalancing (as discussed below). All target words appear in Table 2. Samples of pictures used for target word referents also appear in Appendix A.

Moreover, the target words were selected to contain segments that would elicit a range of distinctive variety-specific phonetic features so that a sufficient amount of regional phonetic variation would be present in the stimuli.

Recording of target words in different regional varieties

Eighteen female native speakers of German were recorded in Germany and the United States. The recordings of 12 of them were selected as stimuli for two experimental conditions. (1) Six speakers of the same Ore Mountain dialect were used to represent a single regional variety for a no-dialect-variability condition. (2) Six speakers of different regional varieties of German were used to represent a high-dialect-variability condition. These regional varieties were: *Hohenlohisch* (an East Frankonian variety), *Allgäuisch* (an East Swabian variety), Palatine German, Hamburg German, Viennese, and Upper Austrian, representing a range of dialects from Upper, Central, and Low German (see, e.g., Barbour & Stevenson, 1995, and König et al., 2019, for an overview of phonetic/phonological characteristics of these dialect groups and subgroups). Speakers were selected based on the degree to which their exemplars of target words conformed to a single Ore Mountain variety (for the low variability condition) and the degree of difference across other varieties of German (for the high variability condition). Although indexical differences could have been generated digitally in a manner that attempted to simulate regional differences, we opted to allow the variability to come from speakers of real-world varieties of German. In doing so, we maintained a higher level of ecological validity in this study.

Recordings were completed in quiet spaces that avoided ambient noise. Speakers were instructed to read the target words and say them out loud in their normal voice and their regional variety. After the individual words were spliced out of the recordings, the files were normalized to a root mean square (RMS) intensity of 70 dB sound pressure level (SPL).

Confirming the amount of variability in the two conditions

As a means of confirming greater regional sociophonetic variability of the exemplars obtained in the six regional varieties of German (to be used for the high variability condition) as compared to the single Ore Mountain variety (to be used for the no variability condition), *phonological edit distance* was calculated based on phonetic transcriptions of the stimuli. This measure is based on *Levenshtein distance* (the number of one-symbol deletions, additions, and substitutions necessary to turn one string into another), but it is weighted by featural similarity so that substitutions between close sounds (such as [s] and [z]) are treated as more similar than substitutions between very different sounds (such as [s] and [m]). For each exemplar of a given word, the phonological edit distance to each of the remaining five exemplars within the relevant condition was calculated with the Phonological Corpus Tools software suite (Hall et al., 2019) using the phonetic features from Hayes (2009) (additional details available at https://corpustools.readthedocs.io/en/latest/string_similarity.html). These values were then averaged within each condition. Figure 2 shows that words within the Multiple (high variability) condition were indeed more phonetically variable ($M = 5.2$, $SD = 3.3$) than were words in the Single (low variability) condition ($M = .5$, $SD = 1.2$).
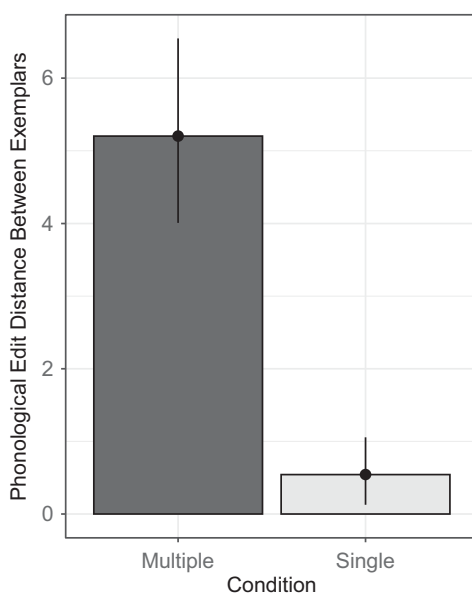

Figure 2. Phonological edit distance by condition shows greater distance in the multiple variety condition than in the single variety condition.
Note: Error bars indicate bootstrapped confidence intervals.

Procedures

All participants were instructed to do their best to learn the 24 target words while viewing pictures of them on a computer screen and listening to them spoken. Participants were tested individually in a noise-attenuating sound booth, and all stimuli were presented over circumaural headphones (Sennheiser 4515). No participant was instructed to repeat words out loud or not to do so. A researcher monitored each participant from a control room equipped with a two-way intercom so that they could communicate with the participant. Participants were told that they would hear 6 repetitions of each picture-word pair and that, in some cases, words would be repeated by a speaker they had heard previously, and in others, six different speakers would produce the words. Word groups and learning conditions were not counterbalanced in Experiment 1 but were in Experiment 2, as reported below. On each trial, participants first viewed the picture of the to-be-learned item for 750 ms, and then the spoken form of the word was presented. After 5 s, the picture disappeared, and a message "please press the spacebar for the next trial" appeared. Each participant attempted to learn 12 words in the low sociophonetic variability condition (six speakers of the same regional variety, one repetition per speaker) and the other 12 words in the high sociophonetic variability condition (six speakers of each of six different regional varieties, one repetition per speaker). The words in word group 1 were always presented in the low variability condition, and the words in word group 2 were always presented in the high variability condition. Additionally, the specific words produced by a given speaker of the Ore Mountain dialect in the no variability condition were rotated across participants (that is, if participant 1 heard the Ore Mountain dialect for a word by speaker 5, the next participant would hear that word by a different speaker of this dialect). By including six speakers in each condition, the study decoupled effects of sociophonetic variability from those of other sources of acoustic variability, in particular the potential effects of indexical features tied to the talker.

After the learning phase (after completion of all learning trials, i.e., all 24 words had been presented six times each), all participants immediately completed two posttests, which were recorded for subsequent scoring. The first posttest (the picture-to-L2 posttest), which was productive, required each participant to produce as much of each target German word as they could when presented with a picture of the word referent. The second posttest (the L2-to-L1 posttest), which was more receptive, required each participant to produce the L1 version of the word when presented with the target German word in spoken form. In the low variability condition, participants were presented with each target word spoken by an Ore Mountain speaker not included in the learning phase, whereas participants in the high variability condition were cued by a novel speaker producing Standard German. We elected to use a novel speaker and a novel dialect in the high variability condition so that there would be no differences in cue familiarity for the L2-to-L1 task across participants.

Scoring

All responses were scored by two native speakers of German as 1, .5, or 0 based on degree of accuracy. Completely correct responses were scored as 1, whereas responses with errors in only one syllable for multisyllabic words were scored as .5. Partially correct responses for monosyllabic words were also scored as .5 when the error affected one phoneme only. All other responses were scored as 0. In the high variability condition, any word form variant to which the participants were exposed was

considered to be a viable correct production. No partial scores were needed for L1 (English) responses for the L2-to-L1 posttest given that the participants already spoke this language. For scoring of picture-to-L2 recall, interrater reliability was assessed by calculating the *Intraclass Correlation Coefficient* (*ICC*, Shrout & Fleiss, 1979) and found to be excellent (*ICC*(A, 1) = .94, *F*(863,666) = 35.7, *p* < .001), exceeding a value of .9 (Koo & Li, 2016).

Data analysis

For the picture-to-L2 recall, learning scores were averaged across the two raters and modeled with a set of linear, multi-level Bayesian models. Each model predicted *learning score* as a function of the fixed effect of *condition*. The predictor was treatment-coded with Multiple (high variability) set as the baseline level. To determine the appropriate random effects structure, four models were fit to the data and compared using the leave-one-out cross-validation procedure. The base model (M1) featured random intercepts for participants and words, M2 added by-participant slopes for condition, M3 added by-word slopes for condition, and M4 added both by-participant and by-word random slopes for condition. All models were fit with the brms package version 2.17.0 (Bürkner, 2017) in R version 4.2.1 (R Core Team, 2022). The priors on all parameters were left at their default, uninformative settings for two main reasons. First, some types of acoustic variability (e.g., talker) have been found to be beneficial to vocabulary learning, and others (e.g., amplitude) have not. Second, assessment of the effects of sociophonetic variability in this study would necessarily have to go beyond those of talker variability, which was a high bar for the type of variability of interest.

Posteriors were sampled with the Hamiltonian Markov Chain Monte Carlo (MCMC) algorithm featuring four chains of 10,000 iterations each (5,000 of which were warm-up, so the total post-warm-up draws consisted of 20,000 samples). All chains converged efficiently with good mixing (all *Rhat* = 1.00, all *ESS* > 1,000).

For the L2-to-L1 task, the outcome variable was binary (correct/incorrect L1 translation of the German prompt). Accordingly, the data were analyzed with a Bayesian, mixed-effects logistic regression. *Accuracy* was modeled as a function of *condition* (single variety vs. multiple varieties), with a random effects structure assessed using the same comparison procedure as that employed in the picture-to-L2 task. As before, all models featured default, uninformative priors on all parameters.

Results (Experiment 1)

Results of Experiment 1 indicated that region-based sociophonetic variability increased early lexical learning and did so beyond increases previously observed for talker variability alone. Results for posttest performance based on picture-to-L2 and L2-to-L1 recall are presented below.

Picture-to-L2 recall

Figure 3 depicts the distribution of picture-to-L2 recall scores by condition, illustrating that the high variability condition (*M* = .42, *SD* = .42) outperformed the no variability condition (*M* = .29, *SD* = .39). (Note that high *SD*s are not uncommon in intentional vocabulary learning studies given the wide range of performance levels of participants/ learners when it comes to this task.) The winning model was M2, which featured random
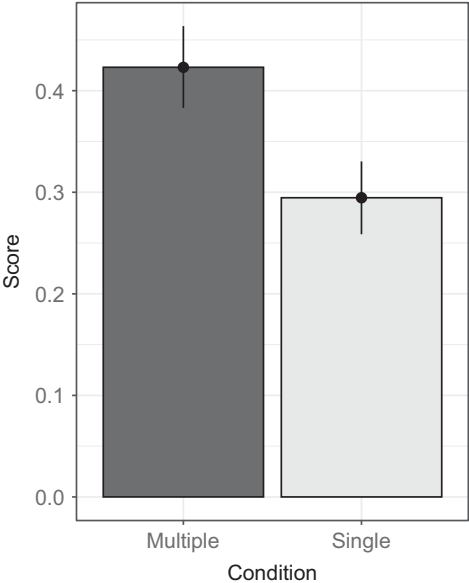
Figure 3. Mean score by condition for picture-to-L2 recall (Experiment 1).
Note: Error bars indicate bootstrapped confidence intervals.

Table 3. Winning model output for picture-to-L2 recall (Experiment 1)

Parameter	Estimate	Est. Error	95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	.423	.031	[.362, .486]	1	10903.79	13571.07
ConditionSingle	−.128	.051	[−.229, -.029]	1	10215.35	12593.37

Note: CI = credible interval; ESS = effective sample size.

intercepts for participants and words, as well as by-participant random slopes for *condition* (*ELDP* difference = .2, *S.E.* = .6, Bayesian R^2 = .12). This model also out-performed a null, intercept-only model (*ELDP* difference = .5, *S.E.* = 1.3). The model output is shown in Table 3. As evidenced by the *Rhat* and *Effective Sample Size (ESS)* values, model convergence was good, and posterior sampling was efficient. The *Intercept* parameter represents the model's estimate for the baseline level of *condition* (recall that this was set to Multiple, the high variability condition). The *conditionSingle* parameter represents the difference in mean scores between Multiple and Single. The model estimated that words presented in a single variety yielded picture-to-L2 recall scores that were .128 lower on average than words presented in multiple varieties. The credible interval around this estimate was [-.229, -.029], indicating a 95% probability that the true difference was within this range. In other words, the analysis revealed a credible effect of condition on score.

L2-to-L1 recall

Figure 4 depicts the mean proportion of correct responses for L2-to-L1 recall by condition, indicating a higher mean score in the high variability condition (*M* = .53,
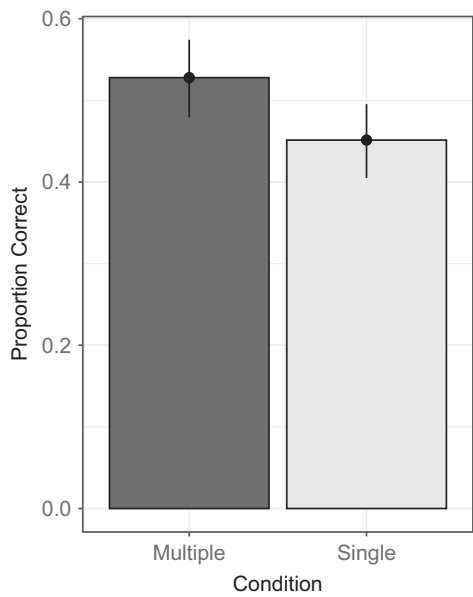
Figure 4. Mean proportion of correct responses by condition for L2-to-L1 recall (Experiment 1).
Note: Error bars indicate bootstrapped confidence intervals.

Table 4. Winning model output for L2-to-L1 recall (Experiment 1)

Parameter	Estimate	Est. Error	95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	.076	.260	[−.461, .597]	1	5713.617	8813.362
ConditionSingle	−.276	.293	[−.840, .323]	1	6865.658	10229.046

Note: CI = credible interval; ESS = effective sample size.

SD = .50) than in the no variability condition (*M* = .45, *SD* = .50). The winning model (*ELDP* difference = .8, *S.E.* = .5) featured random intercepts for both participants and words as well as by-word slopes for *condition*. However, this model was outperformed by an intercept-only model (*ELDP* difference = .4, *S.E.* = .5, Bayesian R^2 = .06), indicating that the *condition* did not reliably improve fit. The model output is shown in Table 4. The *Rhat* and *ESS* values indicated good model convergence and efficient posterior sampling. When converted from log-odds to probabilities, the *Intercept* parameter estimates that words produced in the Multiple condition yielded mean scores of .52. The *conditionSingle* parameter indicates that the scores in the Single condition were lower, at .45 on the probability scale. However, the Bayesian credibility interval around this parameter [-.840, .323] included zero, suggesting that the difference between the two conditions based on L2-to-L1 recall was not credible.

Discussion (Experiment 1)

The findings of Experiment 1 demonstrate that region-based sociophonetic variability can be added to the growing list of sources of acoustic variability that positively affect L2 vocabulary learning, at least based on picture-to-L2 recall, the posttest measure most

sensitive to word form learning in the experiment (because it requires learners to produce the novel word forms themselves). The current list now includes talker, speaking-style, speaking-rate, and sociophonetic variability but not amplitude or F_0 for speakers of non-tone languages. For speakers of tone languages, it also includes F_0 variability. Theoretical and pedagogical implications of these findings are discussed, along with those of Experiment 2, in the General Discussion section.

Experiment 2

Experiment 2 was conducted to confirm whether or not the results of Experiment 1 were tied to word group by counterbalancing word group and learning condition. Recall that in Experiment 1, words in word group 1 were always presented in the single-dialect condition, and words in word group 2 were always presented in the multiple-dialect condition. Individual words might have differed with respect to the effects of input variability (even after efforts to systematically equalize word groups; see Experiment 1, Target Words section) and that not counterbalancing word group (specific lexical items) for the single- and multiple-dialect conditions might have posed a potential confound between word group and input variability. Experiment 2 was designed to address this possibility. All methods for Experiment 2 were the same as those for Experiment 1, with the following exceptions. (1) The total number of participants was 72. (2) We created six versions of the training materials. Within each version, a given item was presented in the low and high variability conditions equally often. In addition, half the participants in each version got the single-dialect condition first and half got the multiple-dialect condition first. Across versions, each speaker served as the talker for the single-dialect condition an equal number of times. Finally, in the multiple-dialect condition, each combination of talker and dialect was heard by an equal number of participants. (3) Interrater reliability for the scoring of picture-to-L2 recall was excellent (ICC(A, 1) = .97, F(1727,521) = 66.6, p < .001) and also slightly higher than in Experiment 1.

Results (Experiment 2)

Results of Experiment 2 again revealed benefits of sociophonetic variability on L2 word learning and confirmed that the benefits observed in Experiment 1 were independent of the potential effects of word group. Results for posttest performance based on picture-to-L2 and L2-to-L1 recall are presented below.

Picture-to-L2 recall

Figure 5 displays the distribution of picture-to-L2 recall scores by condition, indicating that the high variability condition (M = .42, SD = .40) outperformed the no variability condition (M = .34, SD = .42). The best model featured random intercepts for participants and words as well as by-participant random slopes for *condition* (*ELDP* difference = .8, *S.E.* = .4). This model also outperformed a null, intercept-only model (*ELDP* difference = .1, *S.E.* = 1.1, Bayesian R^2 = .12). The model output appears in Table 5. Model convergence and sampling efficiency were good (see *Rhat* and *ESS* values). The model estimated the mean score for the Multiple condition (*Intercept*) to be .417 and that of the difference between conditions to be -.073. The Bayesian Credibility Interval around the difference estimate was [-.144, -.002], which excludes
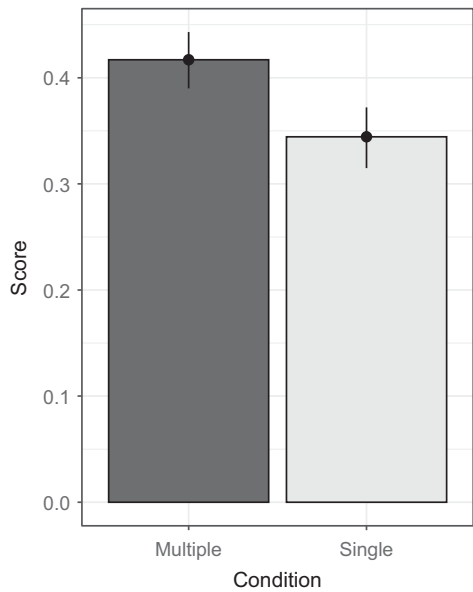
Figure 5. Mean score by condition for picture-to-L2 recall (Experiment 2).
Note: Error bars indicate bootstrapped confidence intervals.

Table 5. Winning model output for picture-to-L2 recall (Experiment 2)

Parameter	Estimate	Est. Error	95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	.417	.022	[.374, .461]	1	11523.12	14186.79
ConditionSingle	−.073	.036	[−.144, −.002]	1	10372.63	13534.09

Note: CI = credible interval; ESS = effective sample size.

zero, suggesting that the difference between the high variability condition and the no variability condition was credible.

L2-to-L1 recall

Figure 6 shows the mean proportion of correct responses for L2-to-L1 recall by condition. The mean was .57 (*SD* = .5) for the high variability (Multiple) condition and .50 (*SD* = .5) for the no variability (Single) condition. The best model (*ELDP* difference = .9, *S.E.* = .7) featured random intercepts for both participants and words. This model also outperformed an intercept-only model (*ELDP* difference = 3.4, *S.E.* = 2.9, Bayesian R^2 = .03), indicating that adding *condition* improved model fit. The model output is shown in Table 6. The *Rhat* and *ESS* values indicate that the model converged on the estimates and that posterior sampling was efficient. When converted from log-odds to probabilities, the *Intercept* parameter estimates that words produced in the Multiple condition yielded mean scores of .57. The *conditionSingle* parameter indicates that the scores in the Single condition were lower (.50 on the probability scale). The Bayesian Credibility Interval around this parameter [-.477, -.098] excluded zero, suggesting that the difference between the conditions in this task was credible.
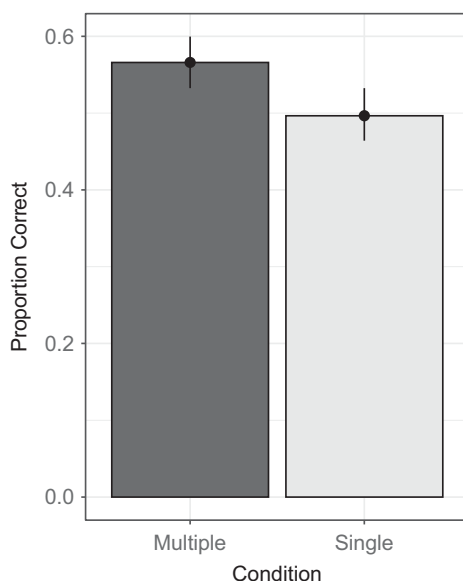
Figure 6. Mean proportion of correct responses by condition for L2-to-L1 recall (Experiment 2).
Note: Error bars indicate bootstrapped confidence intervals.

Table 6. Winning model output for L2-to-L1 recall (Experiment 2)

Parameter	Estimate	Est. Error	95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	.270	.094	[.089, .450]	1	2657.942	2631.12
ConditionSingle	−.285	.097	[−.477, -.098]	1	6725.682	2493.36

Note: CI = credible interval; ESS = effective sample size.

Discussion (Experiment 2)

The results of Experiment 2 confirmed that the benefits of regional sociophonetic variability are independent of word group. Additionally, the positive effects of this type of variability emerged for both picture-to-L2 and L2-to-L1 recall, going beyond the evidence provided by Experiment 1. As such, the combined findings of Experiments 1 and 2 suggest that regional sociophonetic variability should be added to the list of sources of acoustic variability that positively affect vocabulary learning. The current list now includes talker, speaking-style, speaking-rate, and (now based on more evidence) sociophonetic variability, but not amplitude or F_0 for speakers of non-tone languages, and for speakers of tone languages, it also includes F_0 variability. Theoretical and instructional implications of these findings are discussed in the next section.

General discussion

Both experiments in this study demonstrated that regional sociophonetic variability improved L2 word form learning. Experiment 1 indicated mean scores that were 44.8% higher for sociophonetic variability over the no variability condition based on picture-to-L2 recall. Although the mean in L2-to-L1 recall for sociophonetic variability was not

credibly higher than in the no variability condition in Experiment 1, means for both picture-to-L2 and L2-to-L1 recall were credibly higher for sociophonetic variability in Experiment 2. Specifically, mean scores for sociophonetic variability were 23.5% higher based on picture-to-L2 recall and 14% higher based on L2-to-L1 recall. These higher scores obtained for sociophonetic variability should be of particular interest to language instructors as they are statistically credible and not trivial. Notably, the positive effect of sociophonetic variability emerged in both experiments even when (a) talker variability was present in both high and no sociophonetic variability conditions and when (b) there was a high degree of phonological distance between exemplars of the target words in the high variability condition, such as when *Kirsche* (cherry) was [kɪɐʃə] in Viennese and [krɐʃɛ] in Swabian, and *Rauch* (smoke) was [ʁɛːç] in Palatine German and [ʁaʊx] in *Hohenlohisch*. Also, the superiority of the high variability condition emerged even though the low variability condition was given an opportunity to shine in the L2-to-L1 recall by using an Ore Mountain speaker to produce the cues. The theoretical and pedagogical implications of these findings are discussed in the next two sections.

Theoretical implications

Table 7 summarizes research findings, including those of this study, on the effects of different sources of acoustic variability on both speech processing (as reflected by L1 word identification performance) and L2 word learning (as reflected by performance on both picture-to-L2 and L2-to-L1 posttest measures). The unique pattern that continues to emerge is consistent with Sommers and Barcroft's (2007) ePRH in that only phonetically relevant sources of variability positively affect vocabulary learning due to more distributed (robust) developing lexical representations (Barcroft & Sommers, 2005) and pose costs to speech processing. For example, as phonetically relevant sources of variability, talker, speaking-style, and speaking-rate variability positively affect L2 word learning and pose costs to L1 speech processing, whereas the two sources of variability that are not phonetically relevant (amplitude and F_0 for non-tone language speakers) do not affect L2 word learning nor pose costs to L1 speech processing. Importantly, regional sociophonetic variability has now been added to this table. As supported by the findings of this study, this phonetically relevant source of variability has been found to increase L2 word learning. As such, all of the findings summarized in Table 7 are consistent with the ePRH, and future studies may assess whether regional sociophonetic variability and F_0 variability for tone language speakers

Table 7. Effects of different variability sources on L1 word identification and L2 word learning

Source of variability (L=language)	Phonetically relevant?	L1 word identification	L2 word learning
Talker	Yes	Negative[1]	Positive[4]
Speaking style	Yes	Negative[2]	Positive[4]
Speaking rate	Yes	Negative[2,3]	Positive[5]
Regional sociophonetic	Yes	?	Positive[7]
F_0 for tone-L speakers	Yes	?	Positive[6]
F_0 for non-tone L speakers	No	Null[2]	Null[5,6]
Amplitude	No	Null[2,3]	Null[5]

Note: [1]Mullennix et al. (1989); [2]Sommers & Barcroft (2006); [3]Sommers, Nygaard, & Pisoni (1994); [4]Barcroft & Sommers (2005); also Sommers, Barcroft, & Mulqueeny, 2008 for talker and L1 vocabulary learning; [5]Sommers & Barcroft (2007); [6]Barcroft & Sommers (2014); [7]Present study.

negatively affect L1 word identification (as a measure of speech processing), as would be predicted by the ePRH.

The demonstration that the benefits of sociophonetic variability go beyond those of talker variability alone in this study presents another important issue to be considered on the theoretical front. While one need not assume a priori that regional sociophonetic variability involves more acoustic variability than talker variability does, the findings of this study suggest that this likely tends to be the case and that the increased amount of acoustic variability tied to sociophonetic variability is also the explanation why it leads to more word learning than talker variability alone. In fact, the greater phonological distance measured for the high variability condition in the study further corroborates this position. These points suggest that future studies should continue to measure and test the effects of different amounts of acoustic variability in the input, such as when measured based on phonological distance. Will additional increases in the amount of acoustic variability (such as by adding speaking-rate variability to regional sociophonetic variability) continue to increase word learning to a greater degree? If so, considering limits on processing capacity and time at study, at what point will these positive effects asymptote, and at what point will they become negative effects because the input is too acoustically varied? For example, Goldinger et al. (1991) found benefits for talker variability on memory for L1 words when participants were allowed 4 s per word at study, no effect when they were allowed 1 s per word at study, and negative effects when they were allowed only 550 ms per word at study. Sommers and Barcroft (2019) found parallel positive, null, and negative effects for L2 word learning.

Additionally, what role might lexical variation based on truly different word forms, such as *Brötchen* vs. *Semmel* for 'roll,' play in combination with acoustic/phonetic variation of target word forms, such as [ʃʏdkʁøːtn] in Upper Austrian and [ʃildkʁøːde] in Hamburg dialect for the German target *Schildkröte* 'turtle' as acoustically varied exemplars used as stimuli in the present study? Exposing language learners to multiple regional varieties in a naturalistic (uncontrolled) manner is going to involve both types of variation; therefore, understanding the relative impact of each type and the combination of both becomes a topic of interest. Take, for example, the case of L2 Spanish. Exposing learners to six acoustically sociophonetically varied exemplars of *ardilla* ('squirrel') should increase the likelihood that the word form *ardilla* is learned. However, exposing learners to six lexical variants for the Spanish word for 'popcorn' may correspond to *palomitas*, *canguil*, *pochoclo*, *cancha*, *roseta*, and *ñaco*, changing the learning task completely. When considering this issue and potential naturalistic exposure to multiple regional varieties, one needs to weigh the benefits of consistency of target forms in the input with the benefits of acoustic variability and consider the manner in which different languages function with regard to these different types of variation.

Pedagogical implications

The findings of the study also suggest that input from a range of regional varieties can improve early lexical development in adult L2 learners by yielding more distributed (robust) formal lexical representations than is the case for input with less socio-phonetic variability. Sociophonetic variability of this nature can be incorporated into language instruction in a variety of ways. To begin, developers of instructional materials can work to ensure that more sociophonetic variability is included,

particularly when it comes to spoken input presented to L2 learners online or offline in digital formats, such as by using speakers of multiple regional varieties to produce audio-recordings of target words to which L2 learners will listen. Second, contrary to what may be a common misconception, even beginning learners can benefit from exposure to input with increased amounts of sociophonetic variability, including during the early stages of learning new words. While instructors may already seek to expose learners to the types of linguistic diversity that exist among speakers of the target language to increase their sociolinguistic awareness, the present findings suggest that the benefits of doing so extend to the psycholinguistics of language learning as well. Specifically, exposure to sociophonetic variability increases vocabulary learning and does so on top of the benefits of talker variability. In instructional settings, these two types of variability happen at the same time when input from multiple speakers of different regional varieties is used, and hence, one can expect the positive effects to be combined. Because the present study only tested participants immediately after learning, it would be useful for future studies to assess the extent to which the observed benefits remain over longer periods, keeping in mind that Uchihara et al.'s (2022) finding of null effects for both talker variability and exposure frequency on delayed word form recall may be tied to the particular 5-item study phase utilized in that study.

These points being made, a few other clarifications are in order. First, although sociophonetic variability is helpful when L2 vocabulary learning is the goal, it should not be assumed to be helpful if improved L2 speech comprehension is the goal. While (to our knowledge) no studies to date have been conducted on the effects of different sources of acoustic variability on L2 speech processing, findings for acoustic variability and L1 speech processing (see Table 7, column 3) suggest that advanced-level speech processing in L2 is likely to be negatively impacted as it is also negatively impacted in L1.

Second, the manner in which target words are presented with increased sociophonetic variability needs to be sufficiently parallel to the context of learning assessed in the study reported here. The particulars of this context involved that the participants were exposed to sociophonetically varied exemplars of the target words at the word level rather than at the sentence or discourse level while viewing a visual image of the target word referent during the learning phase (*word-picture learning*). In other words, the participants always had access to the word's meaning and ample opportunity to make correct form-meaning connections. While there is no reason to assume that the benefits observed in a laboratory setting would not generalize to classroom settings, one should not overextend the implications of the present findings by generalizing their applicability to discourse-level spoken input given that only word-level spoken input was assessed in this study. Future research is needed to confirm whether vocabulary learning based on discourse-level spoken input also benefits from increases in sociophonetic variability.

Third and lastly, it is important to remember that increased sociophonetic variability is a form-oriented manipulation of how target words are presented in the input. In no way should this be conflated with the potential effects of semantically oriented (and visually oriented) manipulations, such as increased referent token variability. Recall that, as is consistent with the TOPRA model (Barcroft, 2002) and the findings of Sommers and Barcroft (2013), increasing referent token variability in the input can, in fact, decrease L2 word learning.

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.

Barbour, S., & Stevenson, P. (1995). *Variation in German. A critical approach to German sociolinguistics*. Cambridge University Press.

Barcroft, J. (2001). Acoustic variation and lexical acquisition. *Language Learning*, 51(4), 563–590.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323–363.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins.

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

Barcroft, J., & Sommers, M. (2014). Effects of variability in fundamental frequency on L2 vocabulary learning: A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition*, 36(3), 423–449. doi:10.1017/S0272263113000582

Bordag, D., Gor, K., & Opitz, A. (2022). Ontogenesis Model of the L2 Lexical Representation. *Bilingualism: Language and Cognition*, 25(2), 185–201.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310. https://doi.org/10.1121/1.418276

Bulgarelli, F., & Weiss, D. J. (2021). Desirable difficulties in language learning? How talker variability impacts artificial grammar learning. *Language Learning*, 71(4), 1085–1121.

Bürkner, P-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. doi:10.18637/jss.v080.i01

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, & Cognition*, 17(1), 152–162.

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy Lexical Representations in Adult Second Language Speakers. *Frontiers in Psychology*, 12, 732030. doi: 10.3389/fpsyg.2021.732030

Hall, K. C., Mackie, J. S., & Yu-Hsiang Lo, R. (2019). Phonological Corpus Tools: Software for doing phonological analysis on transcribed corpora. *International Journal of Corpus Linguistics*, 24(4), 522–535.

Hardison, D. A. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hayes, B. (2009). *Introductory Phonology*. Blackwell-Wiley.

König, W., Elspaß, S., & Möller, R. (2019). *Dtv-Atlas Deutsche Sprache* (19th ed.). Deutscher Taschenbuch Verlag.

Koo, T., & Li, M. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(March). doi:10.1016/j.jcm.2016.02.012

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(2), 874–886.

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(3), 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96(4), 2076–2087. https://doi.org/10.1121/1.410149

Modern Language Association Ad Hoc Committee on Foreign Languages. (2007). Foreign languages and higher education: New structures for a changed world. *Profession*, 234–245.

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(1), 365–378.

Nygaard, L. C., Sommers, M., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, 57(7), 989–1001.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, 33(3), 441–452.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–28.

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*(4), 795–820.

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*(4), 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*(2), 231–249.

Sommers, M., Barcroft, J., & Mulqueeny, K. (2008, November 13–15). *Further Studies of Acoustic Variability and Vocabulary* [Conference paper]. 49th Annual Meeting of the Psychonomic Society, Chicago, IL, United States.

Sommers, M., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, *32*(2), 417–434.

Sommers, M., & Barcroft, J. (2013). Effects of referent token variability on L2 vocabulary learning. *Language Learning*, *63*(2), 186–210.

Sommer, M., & Barcroft, J. (2019, June 26–28). *The effects of talker variability on L2 vocabulary learning depend on time allowed for encoding* [Conference paper]. Experimental Psycholinguistics Conference, Palma de Mallorca, Spain.

Sommers, M., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, *96*(3), 1314–1324.

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, *44*(2), 357–380.

Wiener, S., Chan, M. K., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *The Modern Language Journal*, *104*(1), 152–168.

Appendix A: Examples of Pictures for Referents of Target German Words Used in the Experiments

Figure A1. Kirsche

Figure A2. Vogel

Figure A3. Kaninchen

Figure A4. Regenschirm