Spelling provides a precise (but sometimes misplaced) phonological target. Orthography and acoustic variability in second language word learning

Pauline Welby, Elsa Spinelli, Audrey Bürki

Journal of Phonetics 94 (2022) 101172

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Spelling provides a precise (but sometimes misplaced) phonological target. Orthography and acoustic variability in second language word learning

Pauline Welby [a,b,*], Elsa Spinelli [c], Audrey Bürki [d]

[a] Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[b] University of New Caledonia, Nouméa, New Caledonia
[c] Univ. Grenoble Alpes, CNRS, LPNC, 38000 Grenoble, France
[d] Department of Linguistics, University of Potsdam, Potsdam, Germany

ARTICLE INFO

Article history:
Received 18 August 2021
Received in revised form 23 June 2022
Accepted 23 June 2022

Keywords:
Orthography
Talker variability
Second language learning
Word learning
Phonological representations
Speech production
Speech perception

ABSTRACT

L1 French participants learned novel L2 English words over two days of learning sessions, with half of the words presented with their orthographic forms (Audio-Ortho) and half without (Audio only). One group heard the words pronounced by a single talker, while another group heard them pronounced by multiple talkers. On the third day, they completed a variety of tasks to evaluate their learning. Our results show a robust influence of orthography, with faster response times in both production (Picture naming) and recognition (Picture mapping) tasks for words learned in the Audio-Ortho condition. Moreover, formant analyses of the Picture naming responses show that orthographic input pulls pronunciations of English novel words towards a non-native (French) phonological target. Words learned with their orthographic forms were pronounced more precisely (with smaller Dispersion Scores), but were misplaced in the vowel space (as reflected by smaller Euclidian distances with respect to French vowels). Finally, we found only limited evidence of an effect of talker-based acoustic variability: novel words learned with multiple talkers showed faster responses times in the Picture naming task, but only in the Audio-only condition, which suggests that orthographic information may have overwhelmed any advantage of talker-based acoustic variability.

1. Introduction

Many people have the intuition that seeing the written form of a new word helps them to learn it, particularly in a second language (L2). Some even say that they *need* the spelling, that hearing the word is not enough. However, the pronunciation of an L2 word can be influenced by its spelling, for example, *talk* or *salmon* said with an 'l' (which is silent for native speakers of English). Another common experience is that different kinds of acoustic variability may help or hinder our comprehension of an L2 – for example, it often seems easier to understand a friend than someone we do not know.

Where these second language experiences may come together is that when faced with speech, which is inherently both highly variable and fleeting, the orthographic form offers L2 speaker-listeners something stable to "grab on to". In spoken language, no utterance or word is ever produced in exactly the same way, even by the same person (e.g. Harrington, 2010). Its production is influenced by a multitude of factors including dialect, speaker, relationship between interlocuters, speaking style and rate, prosodic context, physiological factors, and emotional state. The orthographic form of a word, on the other hand, is immutable and stable – it lasts in time and does not change.[1] Previous studies have shown, however, that relying on orthography may not always be beneficial (see §1.1) and that acoustic variability does not necessarily hinder learning (see §1.2).

The current study builds on an earlier one (Bürki, Welby, Clément, & Spinelli, 2019), in which L1 French participants learned novel English (pseudo)words and their pictured meanings. The words were learned either with acoustic-auditory forms only or with both acoustic-auditory and orthographic

* Corresponding author at: Laboratoire Parole et Langage (LPL), CNRS - Aix Marseille Université, 5 avenue Pasteur, 13100 Aix-en-Provence, France.
E-mail addresses: pauline.welby@univ-amu.fr (P. Welby), elsa.spinelli@univ-grenoble-alpes.fr (E. Spinelli), buerki@uni-potsdam.de (A. Bürki).

[1] This is true synchronically for the many languages with standardized orthographies.

https://doi.org/10.1016/j.wocn.2022.101172
0095-4470/

forms. Note that French and English share a common alphabetic writing system, and that the orthographic forms of the novel words contained only familiar graphemes. Results of a picture naming task showed that for words learned with both the audio and the orthographic form, while participants had better spoken recall and faster response times, they also had more non-native-like (French-like) pronunciations (e.g. for words like *mib* /mɪb/, a vowel more like [i] than [ɪ]) than for words learned with only audio input. In addition, we observed that the vowels of words learned with both audio and orthographic forms were more consistent, more tightly positioned in the formant space (Welby, Bürki, Clément, & Spinelli, 2018).

The current study examines the interplay among orthography, phonological targets, and acoustic variability, and has three main goals. First, we examine the hypothesis that the orthographic form of a word provides learners with phonological targets for subsequent pronunciations. Second, we examine the hypothesis that the influence of orthographic input on pronunciation is greater when the spoken input is more variable. Finally, we examine whether the learning of receptive vocabulary and that of productive vocabulary are both enhanced by orthographic input.

1.1. Influence of orthography on L2 word learning and pronunciation

For literate speakers of an L2, orthography may exert an influence on word learning and production that is mediated by several factors, including the perceived similarity between the L2 sound and sounds in the L1 phoneme inventory (see Best & Tyler, 2007; Flege, 1995; van Leussen & Escudero, 2015), knowledge of the L1 and L2 grapheme-to-phoneme correspondences and phoneme-to-grapheme correspondences (i.e. the mapping between letters or letter combinations in the written form of a word and sounds in the spoken form), and proficiency in the L2. The weighting of these factors may differ across speakers.

Several studies have examined the influence of orthography on receptive vocabulary learning in an L2, with most using tasks involving non-native contrasts, reporting "mixed" results (Bassetti, Escudero, & Hayes-Harb, 2015). Most relevant for the current study are studies using L1-L2 pairs with a shared alphabetic writing system (for a discussion of unfamiliar writing systems, see Mathieu, 2016). In a study of the learning of novel words in L2 Dutch by L1 Spanish speakers, Escudero, Simon, and Mulak (2014) found that learning was facilitated by orthographic forms with grapheme-to-phoneme correspondences (GPCs) that were similar in L1 and L2, while it was hindered by those that were not. Simon, Chambless, and Alves (2010) examined the influence of orthographic input on novel word learning and the acquisition of an unfamiliar non-native phonological contrast (French /u/ vs. /y/) by L1 English speakers, reporting null results. Pattamadilok, Welby, and Tyler (2021) examined the influence on L2 word learning of three learning modalities: audio only, audio accompanied by visible speech articulation gestures, or audio and orthographic form. L1 French participants learned minimal pairs of novel L2 English words beginning with /f/ or /θ/ (e.g. *fedge, thedge*). Immediately after learning, for all three modalities, response accuracy in a picture naming task was well above chance, but no additional benefit was found for any of the modalities. The day after learning, however, performance significantly improved for the Audio-Ortho condition, a pattern of results compatible with the consolidation of lexical knowledge after a night's sleep (Davis & Gaskell, 2009; Earle & Myers, 2014; Gaskell & Dumay, 2003). Using the visual world eye-tracking paradigm to measure word recognition, Escudero, Hayes-Harb, and Mitterer (2008) found that orthographic input helped highly proficient L1 Dutch speakers of L2 English to learn new words, in particular to form phonological representations for a difficult non-native vowel contrast that were then used to build lexical entries (/æ/ vs. /ɛ/, e.g. *tandek*, *tenzer*). Participants displayed an asymmetric pattern of responses for words learned with both audio and orthographic forms, which was explained with respect to L1 and L2 GPCs (<e> ∼ /ɛ/ in both languages, but Dutch <a> ∼ /ɑ/, English <a> ∼ /æ/ (a mid, front vowel, close in the vowel space to Dutch /ɛ/). To our knowledge, only one study has shown that orthographic information can facilitate the learning and on-line retrieval of new productive vocabulary in an L2 (Bürki et al. 2019; see Ehri & Rosenthal, 2007 on L1; also Sadoski, 2005). Moreover, few studies have examined whether orthography contributes to novel word learning in recognition when the material does not involve a non-native contrast (Showalter, 2018 is an exception). Additional evidence regarding the role of orthography on productive and receptive L2 vocabulary learning is needed.

Orthographic information has also been shown to influence pronunciation in an L2 but here again the evidence does not offer a homogenous picture. As speaker-listeners, we regularly encounter examples of "spelling pronunciations" that depart from the pronunciation of native speakers. A number of studies have found that orthographic information can lead to productions that diverge from the audio input and the target L2 phonemes. Bürki et al. (2019) showed an influence of L1 French orthography on pronunciation in L2 English. For words learned with their orthographic form, French participants produced vowels that were more in line with their L1 GPCs (e.g. for words like *mib* /mɪb/, a vowel more like [i]: French <i> ∼ /i/, than [ɪ], English: <i> ∼ /ɪ/). L2 speakers may produce length or other distinctions that are not present in the audio input for vowels represented by digraphs versus singleton letters or for homophonous words with different spellings (Bassetti & Atkinson, 2015; Bassetti, 2017 on L1 Italian speakers of L2 English; see also Rafat, 2016). Several studies have shown that English-speaking learners of German are influenced by their L1 GPCs in producing final obstruents that are voiced rather than voiceless (devoiced), as in the audio input (e.g. *Rad* 'wheel', *Rat* 'advice', both [ʁat] in L1 German; Young-Scholten & Langer, 2015; see also Hayes-Harb, Brown, & Smith, 2018), although a recent study showed that orthographic input can lead to either less or more native-like production, even for the different forms of the same word (e.g. *Rad* [ʁat] 'wheel' and *Raden* [ʁadən] 'wheels', Barrios & Hayes-Harb, 2020). Nimz and Khattab (2019) demonstrated the interplay between knowledge of the L1 and L2 orthographic and phonological systems. Using a picture naming task to examine the influence of orthography on L2 vowel production in known words among L1 Polish L2 German speakers, they found that their L2 speakers more accurately produced the phonological length distinction between long and short German vowels when it was marked in the orthography by a "lengthening h"

than when it was unmarked (e.g. for /aː/, *fahren* 'to drive' vs. *Tafel* 'blackboard'). With respect to vowel quality, they found non-uniform effects of mismatches in the L2 grapheme-to-phoneme correspondences, depending on whether or not there was perceptual overlap between the sounds in L1 and L2 (see also Rafat & Stevenson, 2019). On orthographic input leading to more native-like productions, see also Rafat, 2015, 2016 and Zampini, 1994; as well as Hayes-Harb & Barrios, 2021.

How can these discrepant results be reconciled? One possibility is that experience with orthography, including exposure to the orthographic form of a word and knowledge of GPCs, contributes to the formation of phonological targets that speakers use to generate subsequent pronunciations. In some cases, these phonological targets correspond to sounds in the speaker's native language, in others, to those in the non-native language, in still others to sounds that are neither L1-like nor L2-like (Rafat 2011, 2016; Nimz & Khattab, 2019). There is evidence that some inconsistencies in GPCs between L1 and L2 lead to more phonological transfer than others (Rafat 2011, 2016; Hayes-Harb & Barrios, 2021), but the factors influencing the phonological target formed are not yet fully understood. Alternatively, the difference in findings across studies is due not to an influence of orthography. Rather it arises from differences in how L2 speakers represent sounds learned without orthographic information. In the absence of orthographic input, L2 speakers may form phonological targets based primarily on acoustic-auditory input. Since acoustic speech input is by nature highly variable, transitory, and vulnerable to being obscured, the phonological targets formed may be less focused and more disperse.

1.2. Influence of talker variability on L2 speech sound and word learning

A word can be learned by listening to and interacting with one or more talkers. In classroom settings, new L2 words may often be learned spoken by only one person, the language instructor. L2 speakers may also encounter new words pronounced by several people, in their interactions with others or in video clips and television series. A number of studies have examined the impact of this variability on novel word learning and pronunciation. Barcroft and Sommers (2005) found evidence that talker-based acoustic variability facilitated novel word learning in a non-native language. L1 English speakers learned auditorily presented words in an unfamiliar language, Spanish, and their pictured meanings. The acoustic tokens of the learning input varied, either in number of talkers or in "voice type". Participants showed more accurate responses and faster response times in a picture naming task and a Spanish-to-English oral translation task. Barcroft and Sommers interpret their findings with respect to an exemplar-based model: "beneficial effects of acoustic variability are obtained because the additional variants of the lexical form in the high-variability condition yield a more distributed representation of the word form with more word-form variants that can be mapped onto the semantic-conceptual representation of the word" (p. 411). In a study of L1 English speakers learning novel Lithuanian words, Sinkeviciute, Brown, Brekelmans, and Wonnacott (2019) also found a talker-based variability benefit for adults in production, but not comprehension. No benefits were found for children, either in production or comprehension.

Studies in L1 report costs associated with the processing of information presented by multiple talkers rather than a single talker, at least for particularly demanding tasks, but benefits for variability in the long term. Martin, Mullennix, Pisoni, and Summers (1987), for example, found adverse talker variability effects in recall of (ordered) lists, but not in a free recall task (see also Craik & Kirsner, 1974; Mullennix, Pisoni, & Martin, 1988; Martin, Mullennix, Pisoni, & Summers, 1989; Goldinger, Pisoni, & Logan, 1991). Other studies have examined the effect of presentation with multiple talkers on different aspects of speech processing, in particular the ability to establish new phonological categories or strengthen existing ones, with results varying depending on the population studied and the attentional demands and processing resources required by the task. For example, there is evidence that talker variability in the input helps in the formation of abstract categories for L2 sounds (e.g. Lively, Logan, & Pisoni, 1993, on the identification of the English /r/ and /l/ contrast by Japanese listeners; Sadakata & McQueen, 2013 on the identification by Dutch listeners of a Japanese geminate-singleton fricative contrast). Other studies have shown that the potential benefit of the acoustic variability of multiple talker input is modulated by the difficulty or confusability of the L2 target categories or contrasts with respect to the L1 phonology. Still other studies have found an interaction with L2 proficiency, with talker variability hindering learning for lower proficiency learners, but facilitating it for higher proficiency learners (e.g. Antoniou & Wong, 2015, Wong & Perrachione, 2007; but see Brosseau-Lapré, Rvachew, Clayards, & Dickson, 2013). Zhang, Cheng, Qin, and Zhang (2021) examined the influence of talker-based acoustic variability in the perception and production of the English /i/-/ɪ/ vowel contrast by native speakers of Mandarin Chinese. While they found a benefit in perception for learning with multiple talkers in a canonical high-variability phonetic training (HVPT) paradigm, no benefit was found when the learning input included "adaptive temporal acoustic exaggeration" or visible articulatory gestures.

A few studies have examined the role of talker-based variability in the production of L2 sounds. In Kartushina and Martin (2019), L1 Spanish speakers learned a vowel contrast (/e/ vs. /ɛ/) of an unfamiliar language, French, with audio input based on either multiple talkers or a single talker. Training consisted of a vowel repetition task, with articulatory feedback based on real-time formant analysis of each production. Across the two training conditions, the acoustic input was matched in context and acoustic dispersion in the F1/F2 vowel space but differed in fundamental frequency. After training, in both conditions, productions in a vowel repetition task were more native-like (French-like), with slightly better performance for the single talker condition. Training with multiple talker variability led to less disperse vowel categories and generalization to sounds produced by an unfamiliar talker. The authors conclude that "talker variability supports the establishment of abstract phonemic categories in production". Brosseau-Lapré et al. (2013) examined the performance of native speakers of English who learned minimal pairs of words in French, a language with which they were "minimally familiar", containing a difficult non-native vowel contrast (/ə/ vs. /ø/, e.g. *je* 'I'/ *jeu*

'game'). Participants who had learned words with voices resynthesized from multiple talkers showed better identification performance on a novel minimal word pair, although this advantage did not extend to their production of these vowels. Examining phonological representations beyond segments, in a study of L1 Japanese learners of English, Uchihara, Webb, Saito, & Trofimovich, 2022 found that learning new words pronounced by multiple talkers was associated with greater accuracy in word stress placement but not in spoken word recall. In contrast to their results for perception, Zhang et al. (2021) found no difference in production measures for words learned with multiple talkers and those with a single talker.

To summarize, evidence regarding the role of talker variability on L2 speech sound and word learning is still limited. With respect to novel word learning, there is some evidence that talker-based acoustic variability can be beneficial. By contrast, one study found that talker variability leads to less native-like pronunciations. In the current study, we re-examine the role of talker variability on these two aspects of vocabulary learning. In addition, we investigate whether and how talker-based variability and orthographic information interact to modulate the building of phonological representations.

1.3. The current study

The current study examines together the influence of orthography and of talker variability on L2 learning. Our first aim was to test the reliability of the effects of these two factors on productive and receptive vocabulary[2] learning and to assess their relative contributions. Our second aim was to determine how orthography and variability interact to modulate the phonological representations of sounds whose grapheme-phoneme correspondences differ across languages.

French participants learned novel English (pseudo)words associated with pictures representing their meanings. They later performed a variety of tasks (four in total) to allow us to assess their ability to map between the newly learned words and their pictured meanings, to name the pictures using the new words, and to write the words. Phonetic analyses of the recorded responses were performed, focusing on the vowel.

Our first set of hypotheses concerns the encoding of novel words in memory (the preregistered hypotheses are available at https://osf.io/cdh7n). We hypothesized that memorizing and accessing the association between novel words and their meanings is facilitated by learning with 1) the orthographic forms in addition to the acoustic-auditory forms and/or 2) talker-based acoustic-auditory variability in the input. We reasoned that both would contribute to the formation of more robust phonological representations, albeit by different mechanisms for orthographic input (as described below) and for talker-based variability (following Barcroft & Sommers, 2005; see §1.2). Having more robust or clearer phonological representations should facilitate the mapping of phonological form to meaning, either because it frees up cognitive resources for this aspect of word learning or because there are more "associative hooks" (Barcroft & Sommers, 2005: 410). We therefore predicted higher accuracy and shorter response times in both a recognition task where participants mapped pictures to words and a production task where they named the pictures.

Our second set of hypotheses concerns the phonological representations built and stored for the newly acquired words in the Picture naming task. We hypothesized that seeing the orthographic form of a word during learning contributes to the formation of phonological targets to which the participants converge when they later produce the novel words. This can be reflected in less native-like pronunciations, if the L1 and L2 grapheme-to-phoneme correspondences (GPCs) conflict, as in our materials. We hypothesized that the influence of orthography is greater when the acoustic-auditory input is more variable, since the vowel GPC provides a clearer phonological target than does the audio input. We reasoned that L2 speaker-listeners may use orthographic information to circumvent the processing difficulty associated with talker-based acoustic variability (see §1.2) and to home in on a clear – albeit potentially spurious – phonological target. This predicts that L2 English speakers' productions of new words learned with their orthographic forms will be both acoustically closer to the L1 French targets and less disperse.[3]

2. Methods

2.1. Participants

Forty native speakers of French (28 women, 12 men; age: 18–26 years, mean: 21.2) participated in the experiment. All were from France and were students at AixMarseille University or Sciences Po Aix. Each participant received a €30 gift card as remuneration for their participation, which totaled approximately three hours over three days. All participants reported normal or corrected-to-normal vision and no hearing impairment. All had English as an L2, with varying degrees of proficiency, but reported spending most of their time speaking and listening to French (84%, SD = 14%, vs. English, 14%, SD = 13%).

2.2. Materials

Stimuli consisted of 20 monosyllabic English pseudowords with the structure C(C)VC(C), each of which was paired with a color picture of a rare animal, plant or object. The pseudoword-picture pairs were the same as those developed for Bürki et al. (2019), although for the current study, they were recorded by different talkers. The pseudowords contained only consonant phonemes present in both English and French, and there were no minimal pairs. Half were spelled with <i> (e.g. *lisk*) and half with <o> (e.g. *mog*), which crucially have different GPCs in English and in French: North American English: <i> ∼ /ɪ/ (e.g. *disk* [dɪsk]), <o> ∼ /ɑ/ in general in monosyllabic words (e.g. log [lɑg]) (Carney, 1994); French: <i> ∼ /i/ (e.g. *disque* [disk] 'disk'), <o> ∼ /ɔ/ in closed syllables (e.g. *bogue* [bɔg] 'husk'). In French <o> never corresponds to the low vowel /a/.

[2] The issue of whether word forms are shared between production and recognition or not is debated (e.g. Roelofs, 2003).

[3] In the pre-registration, we further hypothesized that L2 speakers generate their own spellings for novel words learned in the Audio only condition (i.e. without explicit orthographic input) and that these spellings reflect their phonological representation of the words, via the grapheme-phoneme correspondences. To test this hypothesis, we had our participants perform a dictation task (see §2.3), the results of which are presented in Welby, Spinelli, and Bürki (2022) and will serve as pilot data for a follow-up study.

The two English vowels are likely to differ in their perceptual assimilation by French listeners. The English vowel /ɑ/ has no direct counterpart in the French of our participants, who do not maintain the historical distinction between low back /ɑ/ (e.g. *pâte*) and low front /a/ (e.g. *patte*) (see §2.3), but have a single low vowel /a/. English /ɑ/ is likely to be assimilated to French /a/. The vowel /ɪ/ has no counterpart in the French inventory and is known to be difficult for French speakers (e.g. Iverson, Pinet, & Evans, 2012); many participants likely assimilated it to their native /i/ or possibly /ɛ/. While our hypotheses do not differ according to vowel, we will take this into account in our interpretation of certain patterns in the results.

The pseudowords were: <i> *biv, blit, disp, flid, glizz, lisk, mib, nif, vig, zick*; <o> *blop, flob, gosk, losp, mog, skock, sloz, stot, vod, zox*. The word list was read several times by six female native speakers of American English from different regions of the United States. The materials were recorded at the Centre d'Expe´rimentation sur la Parole, Laboratoire Parole et Langage, Aix Marseille University, CNRS, in Aix-en-Provence, France, in a sound-attenuated chamber, using an AKG C520 L head-worn microphone and a Zoom H4nSP Handy recorder at a sampling rate of 44.1 kHz. For each pseudoword and each talker, a total of eight tokens were selected for use in the two learning sessions (four tokens for each session). An additional two tokens were selected for one of the talkers (the talker of the Single talker condition, see §2.3) for use in the test session. The normalized F1 and F2 values at the vowel midpoint from the learning tokens are plotted in Fig. 1. The two vowels are produced as expected (/ɪ/ mid-high front, /ɑ/ low back).

2.3. Procedures

The experiment consisted of three sessions conducted over three days: two learning sessions conducted in a quiet classroom with small groups of participants and one test session conducted individually in a sound-attenuated chamber. For each participant, at least one day and at most three days intervened between any two sessions. We manipulated two factors, learning Modality (a between-items, within-participant factor): Audio only (i.e. the novel words were presented in their audio form only) vs. Audio-Ortho (the novel words were presented in both their audio and orthographic forms) and talker-based acoustic Variability (a within-item, between-participants factor): a Single talker (ST) vs. Multiple talkers (MT) in the learning input.

2.3.1. Learning session: Word learning (Day 1)

Participants were told that they were going to learn new English words to be used in an American mobile phone app under development and that they would be tested on how well they had learned the words. Each of the 20 pseudowords was presented a total of 24 times, with one presentation of each pseudoword in each of 24 blocks. Half of the participants were pseudo-randomly assigned to the Single talker (ST) group (n = 20), in which only the voice of a single talker was heard during the word learning sessions. The other half were assigned to the Multiple talker (MT) group (n = 20), in which the voices of six talkers (including the talker of the ST group) were heard. For both groups, within each of the 24 blocks, the 20 pseudowords were presented in a randomized order.

For all pseudowords, a sound file was played binaurally over headphones (Sennheiser HD 212Pro) and its associated picture was simultaneously displayed at the center of a computer screen for four seconds. Immediately after the offset of the image, the next sound file/picture pair was presented. For each participant, half (10) of the novel words (five <i> and five <o>) were presented with the orthographic form displayed under the picture (Audio-Ortho condition) and half (10: five <i> and five <o>) were presented without the orthographic form (Audio only condition). The two conditions were counterbalanced across two experimental lists.

2.3.2. Learning session: Word learning (Day 2)

The learning session paralleled that of Day 1 except that for each of the two groups, another set of four tokens of each pseudoword was heard. For the Single talker (ST) group, the tokens were produced by the same talker as in Day 1. For the Multiple talker (MT) group, the tokens were produced by

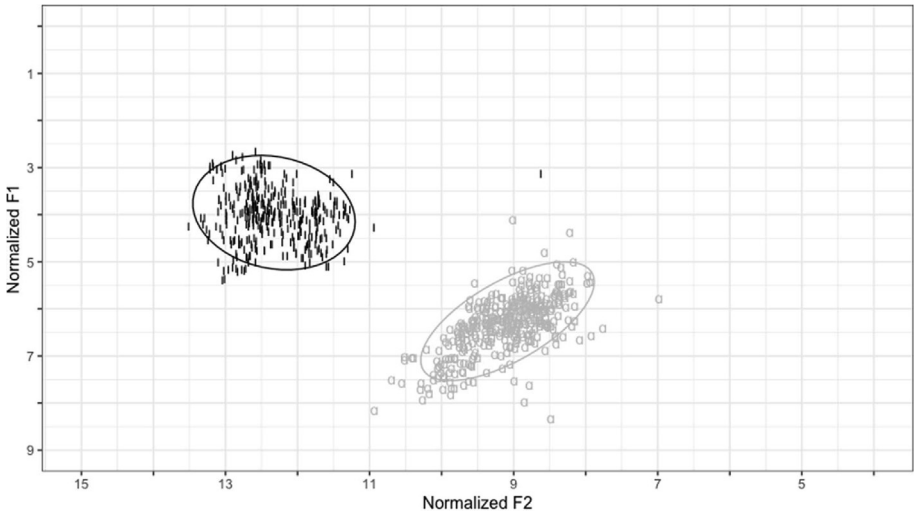


Fig. 1. Normalized F1 and F2 values for the 20 novel words produced by the six L1 English talkers heard in the learning sessions. Ellipses cover a 95% confidence interval around the means.

the same six talkers as in Day 1. Each of the two learning sessions lasted approximately 40 minutes.

2.3.3. Test session: (Day 3)

To assess learning of the new words, each participant was tested individually on four tasks, in the order presented below.

Task 1: Picture naming. Participants were first asked to name each picture as quickly and as accurately as possible, and their responses were recorded. Each picture was named four times (in separate blocks) by each participant. Within each block, the presentation order was random. In this task, we assess lexical access in production to the representations built in the learning phase by measuring spoken recall accuracy and response times to the picture targets. We also use formant-based measures to examine pronunciation of the vowels of the novel L2 words, which provides information about their phonological representations.

Task 2: Picture mapping. Participants were then asked to map the name of each pseudoword to its corresponding picture in a two-alternative forced choice task. Each pseudoword was presented two times in separate blocks, using two previously unheard tokens of each pseudoword produced by the talker from the Single talker condition. In this task, we measure response accuracy and response times to assess lexical access during spoken word recognition to the representations built in the learning phase.

Task 3: Dictation. Each of the 20 pseudowords was presented once over headphones, using the ST condition voice, and participants were asked to type the word they heard. As noted earlier (footnote 3), the results of this task will be discussed in another article.

Task 4: Reading aloud of French word list. Finally, participants were asked to read aloud a list of 24 monosyllabic French words, each presented a single time on the computer screen and containing the vowels /i/, /ɛ/, /a/,[4] or /ɔ/ (e.g. *fiche, bec, fasse, pote*). This task provides a baseline for each participant's pronunciation of the L1 French vowels most closely corresponding to the two English vowels (/ɪ/ and /ɑ/) heard in the pseudowords. It also provides a baseline for each participant's pronunciation of the vowels represented by the French grapheme-to-phoneme correspondences (<i> ∼ /i/, <o> ∼ /ɔ/); recall that we hypothesize that our French participants use L1 GPCs to form phonological targets.

All four tasks were controlled by E-Prime 2.0 software (Psychology Software Tools, 2016), using an AKG C520 head-worn microphone and Roland Rubix22 audio interface for the audio recordings (Tasks 1 and 4). At the end of the test session, each participant completed a language background questionnaire.

2.4. Statistical analyses

Data analyses were conducted in R (R Core Team, 2021). We conducted (generalized) mixed-effects regression models using the "lmerTest" (Kuznetsova, Brockhoff, & Christensen, 2017) and "lme4" packages (Bates, Mächler, Bolker, & Walker, 2015). We used effect coding for all predictors. For all statistical models we started with by Item and by Participant random intercepts and slopes for all contrasts (when supported by the design), but no correlation between random intercepts and slopes. Deviations from this random effect structure were sometimes necessary to avoid singularity warnings and are reported in the results section. When a singularity warning occurred, we removed the random slope or slopes with a value of zero. Note that for all the analyses where the model was simplified, the results are the same with and without the full random effect structure.

Following Baayen, Davidson, and Bates (2008), each model was run twice, with a first model on all data points and a second model without the residuals of the previous statistical model that were larger than 2.5 standard deviations. The output of the second model is reported.

Extreme values, defined based on density plots, were disregarded prior to running the statistical models. The Box-Cox test (Box & Cox, 1964) was run on each dependent variable to decide on the most relevant transformation (choosing one of the following options: natural logarithm, inverse (−1/response times) or no transformation). Unless otherwise specified, the analyses were pre-registered (https://osf.io/cdh7n).

The data and scripts used in these analyses are publicly available (https://osf.io/krsqt/).

3. Results

3.1. Picture mapping

Participants were at ceiling in this task with only 45 (of 3200) incorrect responses. Therefore, no further analyses were conducted on accuracy. Following the visualization of the response time distribution, we disregarded the 43 data points above 2500 ms (1% of correct responses), leaving 3112 data points for the analysis of response times.

The goal of the statistical analyses was to test the hypotheses that (1) novel words learned with both the audio form and the orthographic form are recognized more accurately and more quickly than those learned with audio input only and that (2) novel words learned with multiple talker (MT) input are recognized more accurately and more quickly than those learned with single talker (ST) input. That is, both orthographic information and multiple talkers in the input will facilitate the addition of a lexical entry and its subsequent retrieval. Mean response times (RT) as a function of learning Modality and acoustic Variability are plotted in Fig. 2.

Pooling over the ST and MT conditions, mean RTs were significantly faster in the condition with both the audio and orthographic input (1017 ms, SD = 397) than in the condition with the audio input alone (1044 ms, SD = 396; β = 3.1 * 10^{-5}, SE = 1.1 * 10^{-5}, p = 0.0104), providing support for the hypothesis that the novel words learned with both the audio and orthographic input were recognized more quickly than the novel words learned with the audio input alone. Pooling over the two Modality conditions, mean RTs were numerically shorter for the Single talker condition (1017 ms, SD = 302) than for the Multiple talker condition (1044 ms, SD = 314), but this difference was not significant (β = 3.3 * 10^{-5}, SE = 4.1 * 10^{-5},

[4] On variation in the maintenance of the contrast between the French low vowels /ɑ/ (*pâte*) and /a/ (*patte*), see Berns (2015, 2019), and references therein. In the reading list, we included three words for each vowel (reflected in the orthography by the presence or absence of a circumflex). An examination of the formant values showed overlapping distributions, with a low, front articulation.

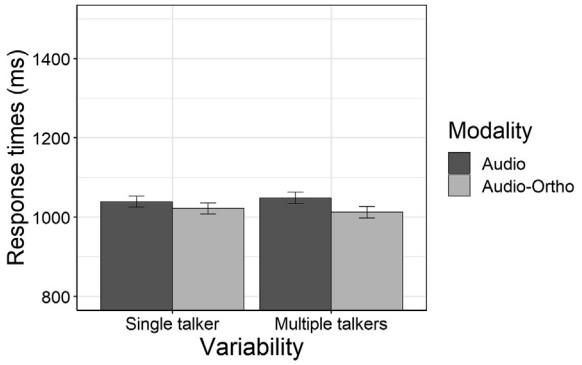


Fig. 2. Observed mean response times and standard errors (values are adjusted for within-Participant designs following Morey, 2008) in the Picture mapping task for each condition.

p = 0.43). The random effect structure of this model does not include a by item random slope for the variable acoustic Variability. This analysis does not provide evidence for or against the hypothesis that talker-based variability facilitates lexical recognition in a perception task. (The details of an exploratory analysis with the interaction between learning Modality and acoustic Variability are given in Appendix A. The estimate for the interaction is not significant.).

3.2. Picture naming

Each of the 3200 response productions was coded as correct or incorrect. A response was coded as correct if and only if all phones of the target pseudoword were produced in the correct order with no additional phones, and the vowel produced was in the same region of the vowel space as the target (e.g. for *mog*: [mɑg], [mag], [mæg] or [mɔg]). Other responses were coded as incorrect (e.g. for *mog*: [mig]). Coding was performed independently by two coders. For cases in which the decisions of the two coders differed (for 166 of 3200 items or 5.2% of the data), the coders performed a second coding, blind to the initial coding. If three of the four codings agreed (136 items), we used this majority coding. The remaining items (30 items) were coded as incorrect. Both initial and final inter-coder agreement were high (κ = 0.984 and κ = 0.991, respectively). The 40 participants gave a total of 1981 correct responses (62%, n = 3200).

3.2.1. Spoken recall accuracy

The goal of the statistical analyses was to test the hypotheses that (1) presentation of the orthographic form along with the audio form facilitates the learning of novel word forms and their retrieval from the production lexicon, and that (2) multiple talker (MT) input facilitates the learning of novel word forms and their retrieval from the production lexicon. Note that these hypotheses for production parallel those for recognition (§3.1).

Accuracy was higher in the Audio-Ortho modality (67.9%) than in the Audio only modality (55.9%), and similar for the Multiple talker and the Single talker conditions (62.2% and 61.6%, respectively). Descriptive statistics are provided in Table 1. The statistical model shows that novel words learned with spoken and written input were named with higher accuracy than those learned with spoken input only (β = 0.87, SE = 0.27, z = 3.25, p = 0.0011, odds ratio = 2.39, 95% CI = 1.41–4.04). There is no statistical difference between the Single and Multiple talker conditions (β = 0.019, SE = 0.50, z = 0.04, p = 0.97, odds ratio = 1.02, 95% CI = 0.38–2.73).

Table 1
Number of correct responses per condition in the Picture naming task.

	Audio-Ortho	Audio	Total
Multiple talker	540	455	995
Single talker	547	439	986
Total	1087	894	1981

This analysis provides support for the hypothesis that the presence of orthographic input facilitates the learning of novel words. It does not provide evidence in favor of (or against) the role of talker-based variability.

3.2.2. Response latencies

The data set was restricted to correct responses to the first presentation of the picture, excluding responses produced with dysfluencies or errors (440 responses) and responses over 2500 ms (15 responses). A total of 425 responses were available for this analysis. Following the Box-Cox, the naming latencies were log transformed.

Response latencies were significantly shorter in the Audio-Ortho (1241 ms, SD = 506) than in the Audio condition (1359 ms, SD = 514); (β = 0.1, SE = 0.03, t = 3.65, p < 0.001), providing support for the hypothesis that the presence of orthographic input in learning facilitates the retrieval of words from the production lexicon. Response latencies were numerically shorter in the Multiple talker condition (1251 ms, SD = 400) than in the Single talker condition (1337 ms, SD = 408) but this difference was not significant (β = 0.06, SE = 0.05, t = 1.36, p = 0.18). These results are illustrated in Fig. 3. The random effect structure of this model does not include a by item random slope for variability or orthography.

We performed an exploratory analysis to see whether there was an interaction between learning Modality and acoustic Variability. The interaction is significant (β = −0.11, SE = 0.06, t = 2.0, p = 0.046, 95% bootstrap CI: −0.21, −0.001; see Appendix B for the full model). Further analyses showed that the interaction was driven by a difference in the Audio learning modality: for words learned with audio input only, RTs were significantly shorter in the Multiple talker condition than in the Single talker condition (β = 0.13, SE = 0.055, p = 0.0297). For words learned with both audio and orthographic input, there was no difference in RT between the Multiple and Single talker conditions (β = 0.011, SE = 0.055, p = 0.842). This suggests that the learning with multiple talkers may benefit learning in the condition without orthographic input.

3.2.3. Acoustic analyses

We performed acoustic analyses to assess the pronunciation of the vowels in the newly learned words, using four formant-based measures. In these analyses, the data from all four repetitions of Task 1 (Picture naming) were considered, including only items that had been coded as correct responses in the accuracy coding (1981 items). For each item, the beginning and end of the word and of the vowel were labelled, using

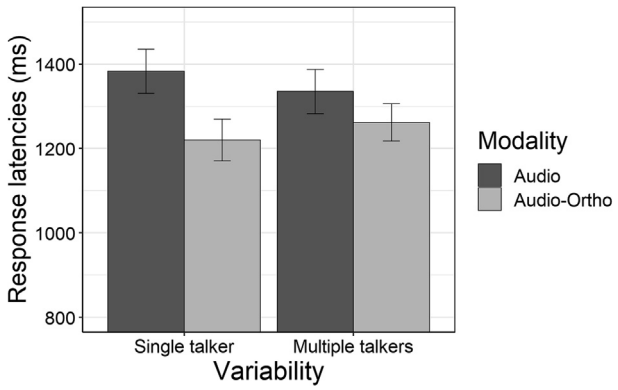


Fig. 3. Observed mean response times and standard errors (values are adjusted for within-participants designs following Morey, 2008) in the Picture naming task for each condition.

forced alignment (Kisler, Reichel, & Schiel, 2017) and hand correction. The analyses were performed in Praat (Boersma & Weenink, 2020), using scripts to semi-automate the process. The first two formants (F1, F2) were extracted from the vowel mid-point and hand corrected, using a Praat script adapted from Styler (2015). Items with unclear formants were excluded (33 items), leaving 1948 data points for these analyses. The formant values were then normalized to allow comparisons across speakers using the Bladon et al. procedure, which is based on the Bark scale and appropriate for data sets with very few vowel categories represented (Bladon, Henton, & Pickering, 1984, evaluated in Flynn & Foulkes, 2011). The same procedures were also used to extract and normalize formant values for the vowels of the French word list (see §2.3). The four measures and their formulas are detailed in Table 2.

The goal of these analyses was to test the hypotheses that (1) presenting the orthographic form along with the audio form during learning leads to less native-like pronunciation, if the critical grapheme-to-phoneme correspondences (GPCs) differ between L1 and L2, as in our materials; and (2) the influence of orthography is greater when the spoken input is more variable. Formant analyses provided information about the articulation of the critical vowels. The *first formant (F1)* is inversely correlated with tongue height. The *second formant (F2)* is inversely correlated with the length of the vocal tract forward of the oral constriction; both a more posterior articulation and lip rounding lengthen this front cavity. In line with the French (L1) GPCs, for <i> words, we expected vowels to be more /i/-like in the Audio-Ortho condition than in the Audio condition, that is, higher and fronter, thus with lower F1 and higher F2. For <o> words, we expected vowels to be more /ɔ/-like in the Audio-Ortho condition, that is, higher and backer and possibly rounded, thus with both lower F1 and lower F2. We further predicted an interaction between learning Modality and acoustic Variability, with a greater influence of orthography in the Multiple talker condition.

Table 2
Formant-based measures used to evaluate production in the Picture naming task.

Measure	Formula
Normalized F1 at vowel midpoint	$F_i^N = 26.81\left(\frac{F_i}{1960+F_i}\right) - 1.53$ (for women)
Normalized F2 at vowel midpoint (Bladon et al., 1984)	$F_i^N = 26.81\left(\frac{F_i}{1960+F_i}\right) - 0.53$ (for men)
Dispersion score (DS) (Kartushina & Frauenfelder, 2014)	$DS = sd_{F1}sd_{F2}\pi$
Euclidean distance (ED)	$ED_{ij} = \sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2}$

The distribution of normalized formant values was visualized, and two extreme values were disregarded. Separate statistical models were run for F1 and F2, using the normalized formant values as the dependent variable. Mean normalized F1 and F2, as a function of learning Modality and acoustic Variability, are plotted in Fig. 4, for each of the two vowels. The formant values for individual data points are displayed in Fig. 5.

Normalized F1. We fitted a linear mixed-effects model to the normalized F1 values (untransformed, following the Box-Cox test), with learning Modality, acoustic Variability, and the interaction between the two as fixed effects. We added Vowel as covariate in the model. The results are displayed in Table 3.

The model for F1 provides support for the hypothesis that presentation of the orthographic form of a word during learning leads to less native-like pronunciations. F1 is lower in the Audio-Ortho than in the Audio only condition. The model does not provide support for the hypothesis that talker-based acoustic Variability interacts with learning Modality. The plot suggests, however, that Variability and Modality interact for <o> but not for <i>. To explore this issue further, we fitted the same model with the three-way interaction (Vowel, acoustic Variability, learning Modality). The detailed results of this exploratory analysis confirm the three-way interaction (see Appendix C). For <o>, learning with the orthography pulls down F1, reflect-

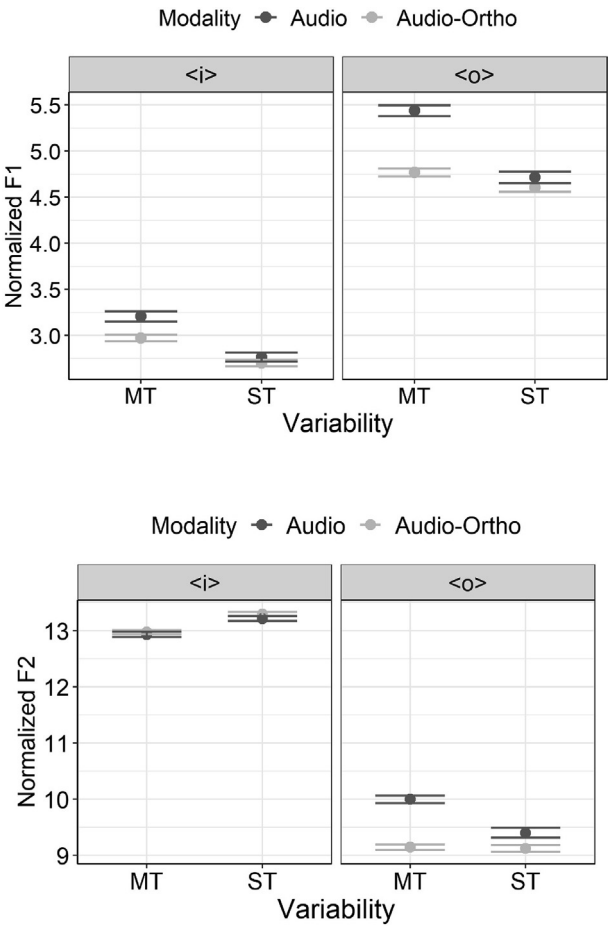


Fig. 4. Mean normalized F1 (upper panel) and F2 (lower panel) values as a function of learning Modality and talker-based acoustic Variability, for each orthographic vowel. MT = Multiple talkers, ST = Single talker.

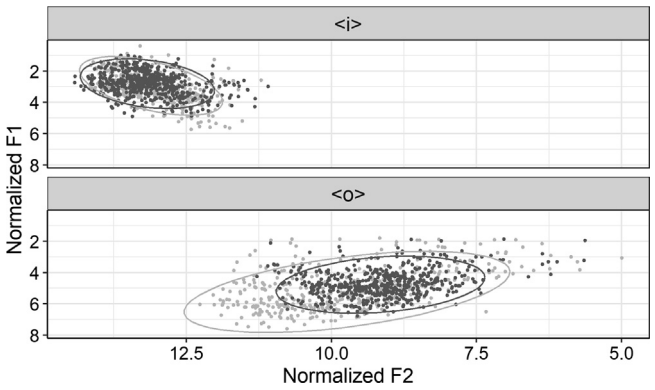


Fig. 5. Normalized F1 and F2 values for L1 French participants' correct responses in the Picture naming task (Task 1), by orthographic vowel (top panel <i>, bottom panel <o>) and learning Modality (Audio: light gray, Audio-Ortho: dark gray). Ellipses cover a 95% confidence interval around the means.

Table 3
Results of statistical model predicting the first formant (normalized value).

	Normalized F1			
Predictors	Estimates	SE	*t*	*P*
(Intercept)	3.88	0.11	36.28	<0.001
Modality	0.26	0.11	2.32	0.027
Vowel	−2.00	0.09	−22.89	<0.001
Variability	0.27	0.20	1.36	0.18
Modality * Variability	0.20	0.16	1.30	0.21

ing a higher vowel closer to a French /ɔ/ target, to a greater extent for words learned with multiple talkers than those learned with a single talker, as predicted by our hypothesis.

Normalized F2. We fitted a linear mixed-effects model to the normalized F2 values (untransformed, following the Box-Cox test), with learning Modality, acoustic Variability, Vowel, and two- and three-way interactions between the three variables (recall that for F2 our predictions go in opposite directions for the two vowels). The results are displayed in Table 4.

To better understand the three-way interaction, separate models were conducted for each vowel. Statistical details for these models are presented in Appendix D. The model for <i> shows no effect of learning Modality, and no interaction between Modality and acoustic Variability. There is a main effect of acoustic Variability, with a higher F2 for participants in the Single talker condition. The model for <o> shows the expected effect of learning Modality, with a lower F2 (more French-like, less native-like) in the Audio-Ortho than in the Audio only condition but no interaction with acoustic Variability.

To summarize, our analyses partly confirm the hypothesis that presentation of the orthographic form of a word during learning leads to less native-like production. They show that this is the case when we consider F1 for <i> and <o> and when we consider F2 for the vowel <o> but not <i>. Our analyses provide some support for the hypothesis that learning Modality and acoustic Variability interact. This is only true when we consider F1 for the vowel <o>. We had hypothesized that the effect of orthography would be greater for tokens learned in the Multiple talker condition. The data suggest however that the interaction arises because of a difference between the Multiple and Single talker conditions in the Audio only modality.

Table 4
Results of statistical model predicting the second formant (normalized value).

	Normalized F2			
Predictors	Estimates	SE	*t*	*p*
(Intercept)	11.30	0.09	121.47	<0.001
Modality	0.29	0.10	3.03	0.002
Vowel	3.60	0.15	23.30	<0.001
Variability	0.07	0.17	0.41	0.681
Modality * Vowel	−0.73	0.16	−4.53	<0.001
Modality * Variability	0.19	0.12	1.62	0.105
Vowel * Variability	−0.74	0.27	−2.77	0.006
Vowel * Modality * Variability	−0.42	0.09	−4.53	<0.001

3.2.4. Dispersion score and Euclidean distance

The normalized formant values for vowels from correct responses in the Picture naming task are plotted in Fig. 5, by orthographic vowel (<i>, <o>) and learning Modality, with ellipses covering a 95% confidence interval around the means. The Audio-Ortho ellipses are largely contained in the Audio ellipsis, indicating a more disperse distribution in the latter condition. To quantify this difference, for each combination of Participant, Vowel, and learning Modality a *dispersion score (DS)* was calculated (Kartushina & Frauenfelder, 2014).[5] The DS is a measure of the relative dispersion in the vowel space, adapted from the formula for the area of an ellipse; a smaller DS indicates less dispersion.

Recall that participants performed a task in which they read aloud a list of French monosyllabic words containing one of four vowels: /i/, /ɛ/, /a/ or /ɔ/ (see §2.3). The normalized formant values for the French words are plotted in Fig. 6 (see also Appendix F). This allowed us to calculate, for each participant, the Euclidean distance (ED) between the normalized F1 and F2 of each correct response in the Picture naming task and the median normalized F1 and F2 of the corresponding French (L1) vowel. The French vowels were /i/ and /ɔ/, following the French GPCs (<i> ∼ /i/, <o> ∼ /ɔ/ in closed syllables), and there were six words for each vowel: /i/: *dites, fiche, pique, piste, quitte, tique*; /ɔ/: *bosse, code, phoque, poche, pote, poste*. Here ED, an application of the Pythagorean theorem, allows a direct comparison of participants' L2 vowels and the corresponding L1 vowels. A small ED indicates that an L2 English vowel produced in the Picture naming task is acoustically close to the corresponding L1 French vowel (e.g. the vowel in a participant's production of the novel English word *mib* is close to that participant's vowel in French words like *piste* 'track, trail').

The goal of the dispersion score (DS) and Euclidean distance (ED) analyses was to test the hypothesis that seeing the orthographic form of a word during learning provides phonological targets to which the participants converge when they later produce the novel words. This hypothesis makes two predictions. The first is that the formant values of the vowels of the novel words learned in the Audio-Ortho condition will be less disperse (will have a smaller DS) than those of the words learned in the Audio condition. The second prediction is that the formant values in the Audio-Ortho condition will be closer to the formant values of vowels corresponding to the

[5] Kartushina and Frauenfelder (2014) refer to this as "compactness score": the less compact the distribution, the higher the compactness score. We have renamed the measure "dispersion score" (DS), leading to a more intuitive description: the more disperse the distribution, the higher the dispersion score.

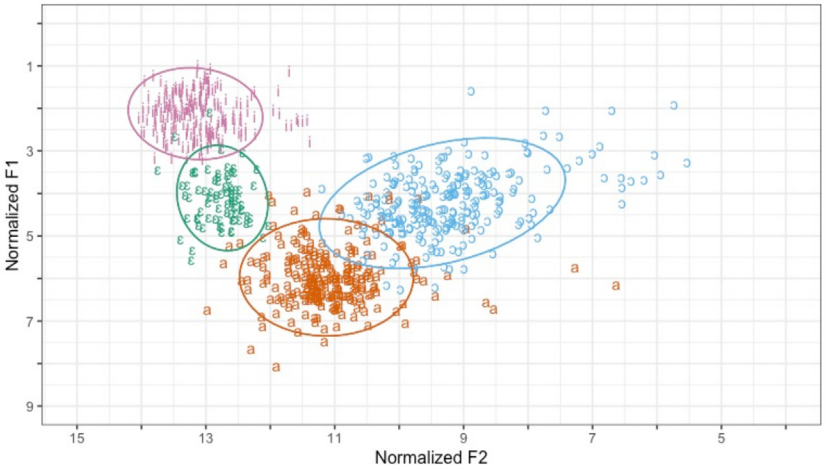


Fig. 6. Normalized F1 and F2 values for the vowels of the words read by the participants in their L1 French (Task 4). Ellipses cover a 95% confidence interval around the means.

L1 GPCs produced by the same participants in their L1 (will have a smaller ED).

Dispersion score. The data set was restricted to the data points for which we had at least five values for the calculation of a Dispersion score (14 data points were removed, leaving 139 data points). Following screening for extreme values, four values above 3.50 for <o> and five data points above 2.00 for <i> were disregarded. Following the Box-Cox test, the DS values were log transformed. The variable Vowel was included in the model as covariate. The final model did not include a by participant random slope for the effect of orthography. The model provides support for the hypothesis that vowels in the Audio-Ortho condition have a smaller DS (0.63, SD = 0.70) than vowels in the Audio condition (1.06, SD = 1.06; $\beta = -0.40$, SE = 0.11, $t = -3.5$, $p < 0.001$). As can be seen in Fig. 5, the formant values are more dispersed in the Audio only than in the Audio-Ortho condition.

Euclidean distance. For the analysis of Euclidean Distance (ED), one outlier (a value above 4) was removed, leaving 1947 data points for the analysis. We fitted a linear mixed-effects model with the logarithm of the ED values as dependent variable and learning Modality as fixed effect. We also included Vowel as a covariate. As predicted, ED in the Audio-Ortho condition is smaller (1.02, SD = 0.76) than in the Audio condition (1.34, SD = 1.1; $\beta = -0.24$, SE = 0.051, $t = 4.90$, $p < 0.0001$). Moreover, ED is also smaller for <i> than for <o> ($\beta = 0.28$, SE = 0.067, $t = 4.16$, $p < 0.00065$). Analyses of Dispersion score and Euclidean distance provide support for the hypothesis that orthographic information provides phonological targets to which the productions converge.

We performed an exploratory analysis to examine whether input from multiple talkers led to smaller Euclidean distance for words learned with both audio and orthographic input. The model with acoustic Variability and the interaction between acoustic Variability and learning Modality (see Appendix E) does not provide support for an influence of acoustic Variability on Euclidean Distance.

4. Discussion

Input in second language learning varies along many dimensions. In this study we examined the influence of two types of variability in the input on word learning and on the building of phonological targets: learning modality (audio only or both audio and orthographic form) and talker-based acoustic variability (one or several voices). To our knowledge, these two types of variability have not been examined together. Because they are inherent to second language learning and are often explicitly manipulated in language teaching and learning, it is important to understand their respective contributions as well as how they might interact.

In this study, L1 French participants learned to associate 20 novel English pseudowords with images representing their meanings. Half of the words were presented with the acoustic-auditory form only, and the other half with both audio and orthographic input. One group of participants heard the words pronounced by a single talker, while another group heard the words pronounced by multiple talkers. Participants completed two learning sessions over two days, then on a third day, they performed a variety of tasks to evaluate their learning.

We consistently find evidence for an influence of orthography, both on the learning of the novel L2 words and their meanings and on their phonological representation. The evidence for an influence of talker-based variability, however, is less clear. We first discuss the findings of the current study, with respect to the addition of new lexical entries, including their phonological representations. We then address several remaining questions and avenues for future research.

In the Picture naming task, participants responded both more quickly and more accurately for words learned with both their audio and orthographic forms, providing evidence for the hypothesis that the presence of orthographic input supports the addition of novel words to the production lexicon and the retrieval of those words, replicating the results of Bürki et al. (2019). In the Picture mapping task, we also found faster response times for words learned with both their audio and orthographic forms, providing evidence that orthography in the learning input can also facilitate recognition. Our finding of a beneficial effect of orthographic input where others have reported no effect (e.g. Simon et al., 2010) may be explained in part by our choice of dependent variable in the forced choice word recognition task. In particular, online measures such as response times are more sensitive than response accuracy, which has been relied on in many studies.

An important question for further studies will be to establish whether the benefit found in the current study for learning with the orthographic form is specific to orthography or whether it would also be found if the audio form were accompanied by another type of information about phonological form. Other studies have compared the impact of different types in information in the learning input (audio form only versus audio form and either orthographic form or articulatory gestures, Pattamadilok et al., 2021; audio form only versus audio form and either articulatory gestures or participants' own productions, Llompart & Reinisch, 2017; audio form only versus audio and orthographic form, Escudero et al., 2008). There are, however, crucial differences between those studies and the current one. For example, in the current study, we did not focus on a contrast between pairs of phonemes. In addition, we assessed learning only after sleep, unlike Llompart and Reinisch (2017) and Escudero et al., 2008, who tested immediately after learning, and Pattamadilok et al. (2021) who tested both immediately after learning and after sleep. Follow-up studies are needed to investigate the nature of the contributions of orthographic information and of articulatory information to L2 word and phonological learning, including whether they enable the consolidation of lexical knowledge associated with sleep.

We find limited evidence that the acoustic variability associated with learning with multiple talkers rather than a single talker supports the addition of new lexical entries. In the Picture mapping task, there was no difference in RTs between the group that learned the words pronounced by multiple talkers (MT group) and the group that learned the words pronounced by only one talker (ST group). In this task, however, the words were presented using the talker of the ST condition, creating a confound between talker-based acoustic Variability and familiarity with the talker's voice. While in the learning sessions, both groups heard tokens produced by the ST talker, the ST group heard each novel word pronounced by this talker many more times than did the MT group (48 vs. eight times, respectively). Despite this imbalance, in the Picture naming task, we observed an interesting pattern. Learning with multiple talkers facilitated the retrieval of new words from the production lexicon, as shown by faster RTs, but only for words that were learned in the Audio only condition, without orthographic forms. The result replicates that of Barcroft and Sommers (2005), who found that multiple talker input in learning led to more accurate spoken recall of novel words in a non-native language. (Barcoft and Sommers did not examine the influence of orthographic information in the learning input; we can therefore compare only the results from our Audio learning condition with those of their study.)

Acoustic variability in the spoken input may help to build more robust representations, but its influence is, at best, fragile. Over 30 years of research has shown that talker-based acoustic variability leads to an initial processing cost and that any longer-term benefit is sensitive to task factors such as demands on cognitive resources associated, for example, with short interstimulus intervals (Goldinger et al., 1991) or the processing of visual articulatory gestures from multiple talkers (Zhang et al., 2021). In addition, a recent large-scale study failed to replicate the results of two classic studies (Logan, Lively, & Pisoni, 1991; Lively et al. 1993) on the role of talker-based acoustic variability in L2 speech sound learning (Brekelmans, Lavan, Saito, Clayards, & Wonnacott, 2022).

If there is indeed an influence of talker-based acoustic variability in the spoken input on the building and reinforcing of phonological representations, the influence of orthographic input may be so strongly weighted that it overwhelms any potential influence of talker-based acoustic variability. Via the GPCs, the orthographic form offers a categorical resolution to variable or "vague" acoustic input. Reliance on information other than highly variable acoustic input was also reported by Wiener, Ito, and Speer (2018). These authors tested spoken word recognition in L1 English learners of L2 Mandarin who were taught words in an artificial tonal language spoken either by a single talker or by multiple talkers. They found that participants who learned with multiple talker input relied more on their knowledge of syllable + tone co-occurrence frequencies as a way of overcoming the greater perceptual uncertainty associated with multiple talkers. In our study, the interaction between acoustic Variability and learning Modality was not pre-registered, and our result must therefore be considered exploratory.

In summary, having the spelling helped L1 French speakers learn new L2 English words, both in production (more accurate and faster responses in the Picture naming task) and in recognition (faster responses in the Picture mapping task). Learning with multiple talkers helped retrieval of new words, but only in production (faster responses in the Picture naming task), and only for words that were not learned with their spellings. Orthographic information may be so heavily weighted that it overwhelms any advantage of talker-based acoustic variability.

We hypothesized that both orthographic information and talker-based acoustic-auditory variability in learning contribute to formation of phonological targets, and that these targets will be less native-like if the L1 and L2 grapheme-to-phoneme correspondences (GPCs) differ, as was the case for the vowels in our materials. Our results confirm the influence of orthography and provide some limited evidence of an influence of talker-based variability. Formant analyses of the responses in the Picture naming task indicate that having the orthographic form in learning pulls L2 speakers' pronunciation to a non-native phonological target, replicating the results of Bürki et al. (2019). As predicted, when the orthographic form was present during learning, F1 is lower for the vowels of both <i> and <o> words, and F2 is lower for the vowels of <o> words. The expected difference was not, however, found for the F2 of <i> words. In Bürki et al. (2019), we found significant differences in the expected directions for both F1 and F2 for both <i> and <o> words. That the results are less robust for <i> than for <o> is not surprising. As noted in §2.2, many participants likely assimilated English /ɪ/ it to their native /i/ vowel category. If our participants' perception and production were already very /i/-like, this may leave less room for an influence of orthography. With respect to the influence of the talker-based variability on the formation of phonological targets, we did not observe the predicted two-way interaction between learning Modality and talker-based acoustic Variability. In an exploratory analysis, however, we found a three-way interaction between acoustic Variability, learning Modality, and Vowel. The effect of orthography on the F1 of the vowel of <o> words, pulling F1

down to a more French-like target, is greater when the learning input includes multiple talkers than when it includes just a single talker, in line with our hypothesis that the influence of orthography will be greater when the acoustic input is more variable. Once again, that the effect is not found for the vowels of <i> words is not surprising. First, a comparison of the formant values of French participants' vowels in their productions of the novel L2 English <i> words (Fig. 5) and their productions of L1 /i/ words (Fig. 6) shows an almost completely overlapping distribution in F2 (the front-back dimension) (see Appendix F), in both learning modalities. In addition, French participants' vowels in English <i> words are higher than those of the L1 English talkers' vowels (/ɪ/) in the learning input (as seen in comparing the formant values in Fig. 5 to those in Fig. 1). If participants assimilate this vowel to French /i/, two learning sessions with acoustically variable input may not be sufficiently strong to influence the phonological target. Second, cross-linguistically, non-low front vowels have been reported be less acoustically and/or articulatorily variable (e.g. Whalen, Chen, Tiede, & Nam, 2018), so speakers may have less flexibility in their articulations.

In summary, learning L2 English words with their spellings pulled the vowels of these words to more French-like pronunciations, replicating the results of Bürki et al. (2019). In addition, we found some evidence, albeit limited, that spelling played a stronger role when words were learned pronounced by multiple talkers.

Additional evidence that orthography provides phonological targets to which L2 speakers' productions converge comes from several patterns in our data. For <o> words in the Audio-Ortho condition, a comparison of the formant values from the Picture naming task shows an almost completely overlapping distribution between our French participants' L2 English vowels (Fig. 5) and their L1 French vowel /ɔ/ (Fig. 6) (see also Appendix F), compatible with the French GPC <o> ∼ /ɔ/ in closed syllables. Further evidence comes from the analyses of the relative dispersion of vowel productions in the F1/F2 vowel space (dispersion score, DS) and of the Euclidean distance (ED) between participants' L2 English vowels and their comparable L1 French vowels. Vowels in words learned with both audio and orthographic forms have smaller DSs than those learned with audio forms only. This reflects the formation of more precise phonological targets when the orthographic form is present in learning. EDs were also smaller, indicating that although these phonological targets were more precise, they were also more French-like (less like the native English acoustic input). Note that while Kartushina and colleagues consider less disperse (more compact) productions to reflect more stable, more native-like production (and perception) (see, for example, Kartushina & Martin, 2019), important differences between their studies and ours prevent such an interpretation. Crucially, in our study we manipulate the presence or absence of orthographic input, and since for our critical vowels the GPCs always differ between L1 and L2, the orthographic input amounts to non-native input. For novel words learned with both audio and orthographic forms, we find less disperse, more stable productions, but whose locus in the vowel space is non-native, e.g. misplaced with respect to the native input. To summarize, the dispersion score (DS) and Euclidean distance (ED) measures allow us to quantify the precision of the vowel productions and their positions in the vowel space. As predicted, for L2 English words learned with both audio and orthographic input, our L1 French participants produced vowels that were both more precise, more tightly clustered around a target (less disperse, smaller DS), and misplaced (with respect to the English learning input) in the vowel space, with loci corresponding to L1 French vowels (smaller ED).

A new word added to the L2 lexicon may have a phonolexical representation that is less precise, "fuzzy", vague or "malleable" (Hayes-Harb & Masuda, 2008; Darcy, Daidone, & Kojima, 2013; Cook, Pandža, Lancaster, & Gor, 2016; Llompart, 2021; Llompart & Reinisch, 2021). In addition, there is evidence that certain types of lexical information, such as phonological information, may be more heavily weighted in the L2 lexicon than other types of information, such as semantic information (for a review, see Cook et al., 2016). Recent findings suggest that the role of orthographic information in the L2 lexicon is also heavily weighted. Mairano, Bassetti, Sokolović-Perović, and Cerni (2018) report a particularly strong influence of orthography, with spurious consonant duration differences induced by L1 Italian orthographic conventions (e.g. a longer [t] in *pretty* than in *city*) being "more resistant to naturalistic L2 [English] exposure" than L1 phonological patterns of VOT not encoded in the orthography. In a study of the influence of different types of training tasks on the production of L2 French vowels by L1 Arabic speakers, Solier, Perret, Baqué, and Soum-Favaro (2019) found that training modalities with a written component led to better pronunciation performance than oral-based ones, with better results for a copy task in which there was explicit orthographic input than in a dictation task in which orthographic information was available only indirectly. (See also, Escudero, Smit, & Angwin, 2021, on the strength of orthographic vs. audio input in cross-situational word learning for L1.) The results of the current study also provide support for the strength of orthographic input. Learning with multiple talkers facilitated the retrieval of new words from the production lexicon, as in Barcroft and Sommers (2005), but only for words that were learned without their orthographic forms. Any benefit of talker-based acoustic variability in the spoken input in building robust phonological representations was overwhelmed by the influence of orthographic input. In summary, there is converging evidence that orthographic representations are heavily weighted in the L2 lexicon.

A further question is the extent to which our results can be generalized. Characteristics of the novel words may modulate the influence of orthography on learning. For example, increased similarity among words in the lexicon or in the learning set may encourage L2 speakers to give more weight to information in addition to the acoustic signal. Considering similarity as a continuum with minimal pairs of words on one extreme and words containing all different phonemes and numbers of syllables on the other, the monosyllabic words in our experiment, with recurring consonant phonemes common to both English and French (e.g. *mog, mib, disp, losp, vig, vod*), fall somewhere between these two extremes. Using minimal pairs of words might also focus participants' attention on one segment (for example, the vowel). As an anonymous reviewer suggests, focusing L2 speakers' attention on phonological form through the use of minimal pairs may lead to less

disperse (more compact) vowels, even in an Audio only learning modality. This hypothesis could be tested using a paradigm similar to that of Llompart and Reinisch (2021), who examined L1 German speakers' learning of L2 English words containing a difficult non-native vowel contrast (/æ/ vs. /ɛ/), auditorily presented through either "phonologically specific training" (in minimal pairs of novel words, e.g. *tandek/tendek*) or "phonologically vague training" (without minimal pairs). While Llompart and Reinisch assessed learning through categorization (picture mapping), the influence of phonological form-focused learning on the nature of the phonological representations built could be tested using a production task such as picture naming.

A number of other characteristics of the novel words may be relevant. We might expect the benefit of orthography to be greater for longer words, at least in production, where there is more phonological detail to retain. The vowels of the novel words of the current study had grapheme-to-phoneme correspondences (GPCs) that differed between L1 and L2 **(**French: <o> ∼ /ɔ/, English: <o> ∼ /ɑ/; French: <i> ∼ /i/, English: <i> ∼ /ɪ/), which led the L2 participants to form precise phonological targets close to those of their native language via the L1 GPCs. With respect to L2 word form learning, which includes both building phonological representations and memorizing the associations between form and meaning, any benefit of orthographic input may be modulated by the degree of consistency, the relationship between the L1 and L2 grapheme-phoneme correspondences at the lexical level (see Ziegler, Jacobs, & Stone, 1996). Despite GPC mismatches for the vowels, having orthographic as well as audio input helped our participants to learn novel L2 words – to form a phonological representation (albeit one that differed from that of native speakers), and to memorize the association between the novel label and its meaning. In our study, the degree of grapheme-phoneme consistency was high: while the L1 and L2 GPCs differed for the vowels of the novel words, the consonants were all present in both English and French and had the same grapheme-phoneme correspondences in the two languages. In other cases, there will be little or no overlap. Consider, for example, a native speaker of French learning an L2 Drehu[6] word like *jidr* /äidʒ/ 'night' (both French and Drehu: <i> ∼ /i/; but French <j> ∼ /j/, <d> ∼ /d/, <r> ∼ /ʁ/; Drehu: <j> ∼ /ä/, <dr> ∼ /dʒ/). In these cases, having the orthographic form as well as the audio form in learning may be of no benefit. It may even interfere in building phonological representations and memorizing the associations between form and meaning. In addition, it is clear that not all differences in GPCs between L1 and L2 have the same influence on pronunciation and phonological representations, that not all L1-L2 GPC mismatches interfere to the same extent or are treated as conflicts (Rafat, 2011, 2011; Nimz & Khattab, 2019). The factors contributing to differences in effects likely include the nature of the L1 and L2 grapheme-phoneme correspondences, the perceptual salience of the L2 sound categories or contrasts, the relationship between the L1 and L2 phoneme inventories, and that between the L1 and L2 writing systems (Hayes-Harb & Barrios, 2021, and references therein; for a review of the many factors affecting L2 word learning, see Peters, 2020).

[6] Drehu (/dʒehu/) is a Southern Oceanic language of the indigenous Kanak people of New Caledonia.

Further research is needed to establish whether certain classes of phonemes or graphemes are more sensitive to orthographic effects. In the current study, we focused on vowels, and mismatches in vowel GPCs may interfere to a lesser extent than those in consonant GPCs (at least for certain types of consonants). There is evidence from L2 word and speech sound learning that vowels may be encoded with less phonological detail than consonants (Escudero, Mulak, & Vlach, 2016; Mulak, Vlach, & Escudero, 2019). In the case of a L1-L2 GPC mismatch, it may then be easier for orthographic input to overwhelm the less precise or weaker phonological representation of a vowel than the more detailed, stronger phonological representation of a consonant. This is in line with observations in the literature on processing differences between consonants and vowels. As Bundgaard-Nielsen, Best, and Tyler (2011) note about L2: "Perception of vowels is more continuous rather than strictly categorical. Within-category discrimination remains possible to some extent, and vowels afford a more rapid readjustment than consonants. . ." (p. 459) (on the classic findings of categorical perception in consonants vs. vowels, see Fry et al., 1962; Stevens et al., 1969; Pisoni, 1973; see also Tyler, Best, Faber, & Levitt, 2014 on perception of non-native consonant versus vowel contrasts). Second, consonants may contribute more to lexical processing than vowels (see Escudero et al., 2016 and references therein), particularly when words are encountered in citation style (see Mulak et al., 2019 and references therein). Finally, across varieties of the same language, there is arguably more variability in the vowel systems than in the consonants: a phonetic realization that "counts" as a given vowel in one variety may count as a different one in another variety. Consider, for example, the word *bat*, pronounced in Australian English with the vowel /æ/ (the TRAP vowel, according to the Wells, 1982 classification system) and in New Zealand English with a more raised vowel, close to the /ɛ/ vowel of the Australian DRESS set (Harrington, Cox, & Evans, 1997; Hay, Maclagan, & Gordon, 2008). In their study of vowel perception across regional varieties of the same language, Shaw, Best, Docherty, Evans, Foulkes, Hay, and Mulak (2018), note "In spoken word recognition, English listeners tolerate more variation in vowels than in consonants as demonstrated, for example, by the word reconstruction paradigm, in which English listeners presented with a non-word such as *eltimate* are more likely to make a word by changing the vowel *eltimate* → *ultimate* than by changing a consonant *eltimate* → *estimate* (Van Ooijen, 1996)" (p. 5). It may therefore be an advantage for speakers to remain flexible when faced with an L1-L2 GPC mismatch for vowels in neighboring regions of the vowel space. Another possibility is that the relevant dimension is not the consonant vs. vowel distinction *per se*, but degree of contrastiveness. As Scobbie and Stuart-Smith (2008) write, "From the point of view of phonology, are all phonemes equal? We think the answer is that some contrasts are more contrastive than others. . ." (p. 107).

5. Conclusion

Our results contribute to our understanding of the influence of two types of variability in the input on word learning and on the building of phonological targets in a second language. Orthographic input has a consistent and strong influence on a number of levels. We find, however, only limited evidence for a benefit of

talker-based acoustic variability. Words learned with multiple talkers are retrieved faster from the production lexicon, but only when they are learned without their orthographic form. This exploratory analysis suggests that orthographic input may overwhelm any benefit of multi-talker acoustic variability. Further studies are needed to explore this possibility.

With respect to the learning of phonological forms, word meanings, and their associations, we find that orthographic input facilitates the addition of new words to both the input and the output lexicons. The finding of an effect in production replicates the results of Bürki et al. (2019), while the finding of an effect in recognition extends the results of this earlier study. With respect the nature of the phonological representations built, formant analyses reveal more French-like (L1) pronunciations when orthographic forms are present in learning. A new finding is that although these pronunciations reflect more precise phonological targets (smaller dispersion scores), these targets are misplaced in the vowel space. That is, their locus is more French-like than English-like (smaller Euclidean distances between the L1 and L2 vowels). In other words, spelling provides a target that is very focused, but that may be focused on the wrong target.

Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project number 317633480 – SFB 1287, Project B05, PI. Audrey Bürki) as well as by the University of Potsdam International Cooperation Program (KoUP) and the Kommission für Forschung und wissenschaftlichen Nachwuchs at the University of Potsdam. This research was supported by the Institute for Language, Communication, and the Brain (https://www.ilcb.fr). We thank Jasmin Sadat for her assistance with data collection and coding, and Nadéra Bureau, James Sneed German, Alain Ghio, and Catherine Perrot for finding solutions at key points in the project. We thank two anonymous reviewers for their insightful comments.

Data statement

The study was preregistered. The preregistered hypotheses and analyses are publicly available (https://osf.io/cdh7n), as are the data and scripts used to perform the analyses and generate the graphs (https://osf.io/krsqt/).

Declarations of interest

None.

Appendix

Appendxi A
Output of mixed-effects model for response times in the Picture mapping task, with the interaction between learning Modality and talker-based acoustic Variability (exploratory analysis).

	−1/RT		
Predictors	Estimates	*t*	*p*
(Intercept)	−0.0010	−40.1	<0.001
Modality	−0.000031	−2.74	0.006
Variability	−0.000032	−0.77	0.44
Modality * Variability	0.000019	0.96	0.34

Appendix B
Output of mixed-effects model for response times in the Picture naming task, with the interaction between learning Modality and talker-based acoustic Variability (exploratory analysis).

	−1/RT			
Predictors	Estimates	SE	*t*	*p*
(Intercept)	7.14	0.02	300.09	<0.001
Modality	−0.10	0.03	−3.81	<0.001
Variability	0.07	0.04	1.62	0.106
Interaction	−0.11	0.06	−2.01	0.045
$N_{Participant}$	40			
N_{Item}	20			

Appendix C
Output of statistical models for F1 (exploratory analyses). Model with interaction between learning Modality, talker-based acoustic Variability and Vowel.

	Normalised F1			
Predictors	Estimates	SE	*t*	*p*
(Intercept)	3.86	0.11	35.51	<0.001
Modality	0.29	0.09	3.22	0.001
Vowel	−1.92	0.14	−13.94	<0.001
Variability	0.35	0.21	1.72	0.086
Modality * Vowel	−0.49	0.15	−3.30	0.001
Modality * Variability	0.15	0.12	1.25	0.212
Vowel * Variability	−0.04	0.24	−0.15	0.878
Vowel * Modality * Variability	−0.28	0.09	−3.11	0.002

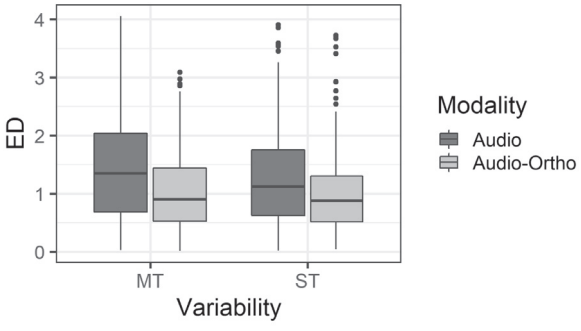
Appendix D
Output of statistical models for normalized F2 (exploratory analyses) for each orthographic vowel.

Vowel <i>

	Normalized F2			
Predictors	Estimates	SE	*t*	*p*
(Intercept)	13.12	0.08	169.16	<0.001
Modality	−0.04	0.06	−0.69	0.490
Variability	−0.30	0.14	−2.14	0.033
Modality * Variability	−0.02	0.08	−0.22	0.825

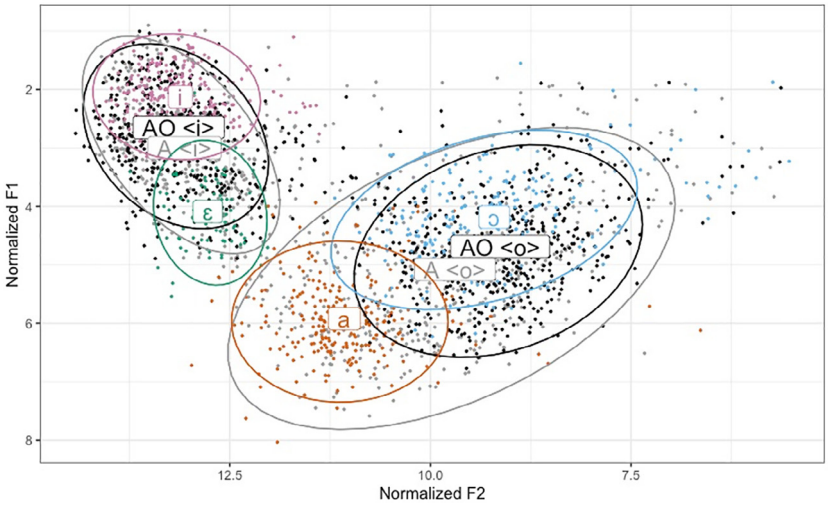
Vowel <o>

	Normalized F2			
Predictors	Estimates	SE	*t*	*p*
(Intercept)	9.47	0.15	61.34	<0.001
Modality	0.65	0.19	3.40	0.001
Variability	0.40	0.26	1.52	0.128
Modality * Variability	0.49	0.23	2.12	0.034



	ED			
Predictors	*Estimates*	*SE*	*t*	*p*
(Intercept)	-0.04	0.08	-0.54	0.590
Modality	-0.25	0.05	-4.95	<0.001
Variability	-0.10	0.12	-0.77	0.443
Modality * Variability	0.09	0.10	0.89	0.373

Appendix E. Exploratory analysis of the interaction between talker-based acoustic Variability and learning Modality for the Euclidean Distance. Graphical representation and output of statistical model.



Appendix F. Normalized F1 and F2 values for the vowels in participants' correct responses in the Picture naming task (Task 1), by orthographic vowel (<i>, <o>) and learning Modality (Audio (A), Audio-Orthography (AO)), together with formant values for the vowels of the words in the French word list (Task 4) (/i/, /ɛ/, /a/, /ɔ/). Ellipses cover a 95% confidence interval around the means.

References

Antoniou, M., & Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America, 138*, 571–574.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

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

Barrios, S., & Hayes-Harb, R. (2020). L2 learning of phonological alternations with and without orthographic input: Evidence from the acquisition of a German-like voicing alternation. *Applied Psycholinguistics, 41*, 517–545.

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology Learning, Memory, and Cognition, 43*, 1835–1842.

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics, 36*, 67–91.

Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics, 36*, 1–6.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Berns, J. (2015). Merging low vowels in metropolitan French. *Journal of French Language Studies, 25*, 317–338.

Berns, J. (2019). Low vowel variation in three French-speaking countries. *The Canadian Journal of Linguistics/La Revue Canadienne deLinguistique, 64*, 1–31.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.

Bladon, R. A. W., Henton, C. G., & Pickering, J. B. (1984). Towards an auditory theory of speaker normalization. *Language and Communication, 4*, 59–69.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.0, retrieved from http://www.praat.org/.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, 26*, 211–252.

Brosseau-Lapré, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and perceptual learning of non-native vowel categories. *Applied Psycholinguistics, 34*, 419–441.

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*, 433–461.

Brekelmans, G., Lavan, N., Saito, H., Clayards, M., & Wonnacott, E. (2022). Is there a multi-talker advantage for learning non-native phoneme contrasts? A large scale representation. Poster presented at LabPhon 18 [online conference].

Bürki, A., Welby, P., Clément, M., & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond "friend or foe?". *Journal of the Acoustical Society of America,* 145, EL265–271.

Carney, E. (1994). *A survey of English spelling*. Psychology Press.

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology, 7*, 1345.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26*, 274–284.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8*, 372–420.

Davis, M., & Gaskell, M. (2009). A complementary systems account of word learning: Neural and behavioural Evidence. *Philosophical Transactions: Biological Sciences, 364*, 3773–3800.

Earle, F. S., & Myers, E. B. (2014). Building phonetic categories: An argument for the role of sleep. *Frontiers in Psychology, 5*, 1192.

Ehri, L. C., & Rosenthal, J. (2007). Spellings of words: A neglected facilitator of vocabulary learning. *Journal of Literacy Research, 39*, 389–409.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*, 345–360.

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016). Cross-situational learning of minimal word pairs. *Cognitive Science, 40*, 455–465.

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition, 17*, 384–395.

Escudero, P., Smit, E., & Angwin, A. (2021). Investigating orthographic versus auditory cross-situational word learning with online and lab-based research. Manuscript. Retrieved from https://psyarxiv.com/tpn5e/.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flynn, N., & Foulkes, P. (2011). *Comparing vowel formant normalization methods*. Hong Kong: International Congress of Phonetic Sciences.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171–189.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132.

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.

Harrington, J. (2010). *The phonetic analysis of speech corpora*. Wiley-Blackwell.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics, 17*, 155–184.

Hay, J., Maclagan, M., & Gordon, E. (2008). New Zealand English. In Dialects of English. Edinburgh University Press.

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching, 54*, 297–326.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research, 24*, 5–33.

Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech, 61*, 547–564.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*, 145–160.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers inPsychology, 5*, 1246.

Kartushina, N., & Martin, C. (2019). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning, 69*, 71–105.

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech and Language, 45*, 326–347.
Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.
Lively, S. E., Logan, J. D., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.
Llompart, M. (2021). Lexical and phonetic influences on the phonolexical encoding of difficult second-language contrasts: Insights from nonword rejection. *Frontiers in Psychology, 12*, 659852.
Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1040–1056.
Llompart, M., & Reinisch, E. (2021). Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics, 88*, 101080.
Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–886.
Mairano, P., Bassetti, B., Sokolović-Perović, M., & Cerni, T. (2018) Effects of L1 orthography and L1 phonology on L2 English pronunciation. *Revue Française de Linguistique Appliquée, XXIII,* 45–57.
Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1987). Effects of talker variability on recall of spoken word lists. *Research on Speech Perception Progress Report, 13.* Bloomington, IN: Indiana University.
Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 676–684.
Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research, 32*, 145–170.
Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Tutorials in Quantitative Methods for Psychology, 4*, 61–64.
Mulak, K. E., Vlach, H. A., & Escudero, P. (2019). Cross-situational learning of phonologically overlapping words across degrees of ambiguity. *Cognitive Science, 43*, e12731.
Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1988). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365–378.
Nimz, K., & Khattab, G. (2019). On the role of orthography in L2 vowel production: The case of Polish learners of German. *Second Language Research, 36*, 623–652.
Pattamadilok, C., Welby, P., & Tyler, M. D. (2021). The contribution of visual articulatory gestures and orthography to speech processing: Evidence from novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication*. https://doi.org/10.1037/xlm0001036.
Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 125–142). Routledge.
Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253–260.
Psychology Software Tools, Inc. (2016). E-Prime 2.0. Retrieved from http://www.pstnet.com.
R Core Team. (2021). R: A language and environment for statistical computing. V. 4.0.5. [computer software manual]. Vienna. Retrieved from http://www.r-project.org/.
Rafat, Y. (2011). *Orthography-induced transfer in the production of novice adult English-speaking learners of Spanish*. University of Toronto. Doctoral dissertation.
Rafat, Y. (2015). The interaction of acoustic and orthographic input in the L2 production of assibilated/fricative rhotics. *Applied Psycholinguistics, 36*, 43–64.
Rafat, Y. (2016). Orthography-induced transfer in the production of English-speaking learners of Spanish. *The Language Learning Journal, 44*, 197–213.
Rafat, Y., & Stevenson, R. A. (2019). Auditory-orthographic integration at the onset of L2 speech acquisition. *Language and Speech, 62*, 427–451.
Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production. Differences and similarities* (pp. 115–158). Mouton de Gruyter.
Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustical Society of America, 134*, 1324–1335.
Sadoski, M. (2005). A dual-coding view of vocabulary learning. *Reading and Writing Quarterly, 21*, 221–238.
Scobbie, J. M., & Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 87–113). Mouton de Gruyter.
Shaw, J. A., Best, C. T., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., & Mulak, K. E. (2018). Resilience of English vowel perception across regional accent variation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*, 11.
Showalter, C. E. (2018). Impact of Cyrillic on native English speakers' phono-lexical acquisition of Russian. *Language and Speech, 61*, 565–576.
Simon, E., Chambless, D., & Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32*, 380–394.
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*, 795–820.
Solier, C., Perret, C., Baqué, L., & Soum-Favaro, C. (2019). Written training tasks are better than oral training tasks at improving L2 learners' speech production. *Applied Psycholinguistics, 40*, 1455–1480.
Stevens, K. N., Liberman, A. M., Öhman, S. E. G., & Studdert-Kennedy, M. (1969). Crosslanguage study of vowel perception. *Language and Speech, 12*, 1–23.
Styler, W. (2015). FormantMeasureVerifyv3.praat. v. 3.1. https://github.com/stylerw/styler_praat_scripts.
Tyler, M., Best, C., Faber, A., & Levitt, A. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*, 4–21.
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*, 357–380.
van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology, 6*, 1000.
Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*, 573–583.
Welby, P., Bürki, A., Clément, M., & Spinelli, E. (2018). In *Positive and negative influences of orthography on L2 word learning and phonological encoding*, Poster presented at the Institute of Language, Communication, and the Brain conference, Marseille, France. .
Welby, P., Spinelli, E., & Bürki, A. (2022). Does spelling provide a phonological target even when spelling is not provided? Poster presented at LabPhon18 [online conference].
Wells, J. C. (1982). *Accents of English, 3.* Cambridge University Press.
Whalen, D. H., Chen, W.-R., Tiede, M. K., & Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics, 68*, 1–14.
Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech, 61*, 632–656.
Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*, 565–585.
Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in L2 German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics, 36*, 93–114.
Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania, 77*, 470–481.
Zhang, X., Cheng, B., Qin, D., & Zhang, Y. (2021). Is talker variability a critical component of effective phonetic training for nonnative speech? *Journal of Phonetics, 87*, 101071.
Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers, 28*, 504–515.