Original Article

The role of L2 input in developing a novel L2 contrast phonetically and phonologically: Production evidence from a residence abroad context

James Turner [ID]
University of Southampton, UK

Abstract
This study analyses the production of French /y/ and /u/ by 42 native English learners of French (ELoF) at the start and end of a Residence Abroad (RA) in a French-speaking country. As an approximation of both phonological and phonetic development, categorical change is teased apart from gradient change using k-medoid clustering of acoustic data and different input measures are tested as predictors of both types of development. Results of the phonological analysis reveal that while no change occurs for /y/, the proportion that learners correctly use their back vowel in /u/ contexts increases over the RA. The quantity of French input declared over the RA is a significant predictor of this categorical change, especially the amount of auditory and visual engagement (e.g. listening and reading in French). Results of the phonetic analysis indicate, instead, that /y/ becomes more target-like over the RA, while /u/ makes less progress, suggesting a partial mismatch between the phonological and phonetic levels. Nevertheless, the phonetic development of /u/ is more substantial for individuals who: (1) have been learning French longer, (2) have yet to experience naturalistic exposure by the start of the RA, and (3) are English language teachers over the RA rather than students at a foreign institution. Taken together, these results have implications for the link between second language (L2) input and L2 production, the assumptions of L2 speech models, and the relationship between different levels of linguistic representation in L2 speech learning.

Corresponding author:
James Turner, University of Southampton, B65 Avenue Campus, Highfield Road, Southampton, SO17 1BF, UK
Email: James.Turner@soton.ac.uk

I Introduction

It is well known that adult language learners struggle to produce contrasts in a foreign (L2) language that do not exist in their native (L1) inventory. For example, numerous studies have highlighted that L1 Japanese learners of L2 English experience difficulty in differentiating the phonemes /r/ and /l/ in English, a distinction that is not present in Japanese (Aoyama et al., 2004; Sheldon and Strange, 1982; Shinohara and Iverson, 2021). The reasons for non target-like productions can be wide-ranging, and include – but are not limited to – the nature of input that the learner experiences, performance factors, perceptual encoding errors, motivational factors, and age-related constraints, among others (see, for example, Archibald, 2021a). However, relatively little attention has been given to the fact that there are also different criteria by which productions of L2 sounds can be classed as 'non target-like'.

For example, when it comes to acquiring a new contrast, language learners are presented with two learning tasks. The first is more phonetic in nature: adapting the acoustic quality of each sound towards the target present in the input, while the second is more phonological: learning the distribution of each phoneme so that the two representations are rarely, if ever, confused. Both challenges can yield productions that might be considered non target-like and the two processes are not necessarily interdependent. For example, L1 Japanese speakers may demonstrate high phonological accuracy by consistently retrieving – what is for them – the phoneme /r/, and producing it in contexts that require this sound, e.g. 'rat', and /l/ in contexts that require /l/ such as 'left', even if the phonetic quality of both sounds bears little similarity to the acoustic detail found in the input. In contrast, they may produce /r/ and /l/ with phonetic detail easily within the native English range and yet often retrieve their /l/ in contexts which require /r/, e.g. '[l]at' instead of '[ɹ]at', or produce their /r/ in contexts which require the lateral, e.g. '[ɹ]eft' instead of '[l]eft'.

Presumably, both of these phonological and phonetic learning tasks are facilitated by extensive amounts of high-quality L2 input, as is generally assumed by studies of L2 speech learning (Flege, 2018; Flege and Bohn, 2021; Moyer, 2008; Muñoz and Llanes, 2014; Trofimovich, 2011) and other areas of second language acquisition (SLA) research (Carroll, 1999; Rankin and Unsworth, 2016; Slabakova, 2013; Yang, 2018). However, what precisely constitutes high-quality input merits further investigation, particularly when reliable measurements of L2 input have yet to be established (Flege and Wayland, 2019: 36).

The present research addresses this gap with a novel statistical method of analysing input using the Language Engagement Questionnaire (LEQ) (Mitchell et al., 2017), a tool which provides information about how often 26 different activities in the foreign language are completed. These data are tested as predictors of the real time development of French /y/–/u/ vowel productions by 42 adult L1 English learners of French (34 British, 8 North American) over a residence abroad (RA) in a French-speaking country. Given that participants were predominantly university students undertaking a residence abroad as part of their course, destinations were not limited to a single location: 26 applied to reside in South East France (predominantly Lyon), 14 went elsewhere in France, one stayed in Geneva and one resided in Dakar.[1] To ensure that the development

of the contrast was analysed comprehensively, phonological development over the RA was teased apart from phonetic change using k-medoid clustering analyses of the acoustic data, and the input measures were tested as predictors of both types of change. Overall, the objectives of this study are threefold:

- to determine whether participants confuse French /y/ and /u/ less after a 6-month RA;
- to analyse whether the quality of either vowel changes over the RA; and
- to investigate which measures of input most closely correlate with both types of development.

II Background

1 L2 input and the effect of Age of Arrival

One conceptualization of L2 input is the raw acoustic and visual information belonging to the language of the ambient environment (Carroll, 1999). This includes linguistic forms that have not necessarily been encountered (Carroll, 2017). However, because not all L2 speech studies make a clear distinction between input, exposure, language use and language experience, the present research will use the term 'input' to describe 'all situations in which learners engage with a language in a meaningful way' (Trofimovich, 2011; see also Mitchell et al., 2017). While this may not align with all previous definitions, Carroll (2017: 4) highlights that 'there is no consensus on the right word to describe the learner's interactions with the environment'.

Despite L2 input being considered an important factor in studies of L2 speech learning and SLA research more broadly, difficulties have arisen in its examination because of a well-known confound with Age of Arrival (AoA) (DeKeyser and Larson-Hall, 2005; Flege and Wayland, 2019; Mayberry and Kluender, 2018; Moyer, 2008; Muñoz, 2006; Slabakova, 2013; Trofimovich, 2011). Indeed, these authors highlight that learners who move to an L2 environment and learn the second language at a younger age are also likely to have learnt the language for a longer period of time, and potentially, therefore, also accumulated more – and possibly even better – L2 input than individuals who move to the L2 environment in adulthood. Research comparing the role of AoA and input in L2 speech learning has found mixed results, however: while in some circumstances, input appears to have a stronger effect than age (e.g. Flege and Mackay, 2004; Flege and Wayland, 2019), AoA has also been shown to account for more variability than measures of input (e.g. Flege et al., 1995). One of the main reasons hypothesized for the inconsistency behind these L2 input findings is the validity of the input measure itself.

2 Measuring L2 input

Length of residence (LoR) has often been used as a proxy for the quantity of L2 input and does in certain circumstances reveal a moderate effect. For example, Saito (2015) analyses the effect of LoR on the speech of late L1 Japanese learners of English in Canada and

finds that it correlates with comprehensibility in English pronunciation. However, several studies have found little effect of LoR (e.g. Flege and Liu, 2001; Flege et al., 2006) and others yet provide evidence that the influence of LoR is only apparent within the first year of immersion (Winitz et al., 1995). Flege and Wayland (2019) and Flege and Bohn (2021) are particularly critical of LoR as an index of L2 input because, as they point out, the amount of time in a new linguistic environment does not necessarily indicate how often the L2 is used or heard, nor does it reveal how/where/with whom.

Another common instrument for indexing input is language background questionnaires. Several studies, in particular those analysing temporary residences abroad, have demonstrated an effect of the amount of declared input on L2 speech learning. For example, Stevens (2011), Díaz-Campos (2004) and Muñoz and Llanes (2014) all examine language learners' phonetic and phonological gains in brief sojourns abroad and, across these studies, significant predictors of development include: the amount of exposure to L2 television, use of L1 spoken English, years of formal L2 instruction, reported L2 use, hours spent in class learning the L2, and hours spent speaking with native speakers. Furthermore, these studies demonstrate the benefit of residences abroad on L2 speech learning given that development observed from naturalistic learning often surpasses the gains occurring in classroom settings at home (e.g. Muñoz and Llanes, 2014; Stevens, 2011), presumably due to a higher quality and quantity of input.

Language background questionnaires can also be useful in residence abroad studies as they provide more information about the RA context. Indeed, university language students in the UK who undertake a year abroad are likely to become either English language assistants in French schools or university exchange students, although some do secure different work placements. Nevertheless, it is unclear whether the RA type itself determines the extent of L2 gains. For example, Mitchell et al. (2017: 228) analyse the linguistic gains of 52 L1 Anglophone students on a residence abroad in a French-speaking country, and find that among the top 10 individuals for linguistic development, an equal number of participants were English language teachers in French schools and exchange students at a French university. Furthermore, both of these contexts were found to provide adequate environments for the development of new L2 social networks (Mitchell et al., 2017: 163).

From this brief overview, it appears that the specific measurement of input that is used may well affect the extent to which it predicts L2 speech learning. However, the type of input measure, i.e. the independent variable, is not the only important consideration in such studies. Indeed, the way in which L2 speech learning is understood and measured, i.e. the dependent variable, must also be problematized. For example, if L2 input affects foreign accentedness (for an overview, see, for example, Muñoz and Llanes, 2014), it is unclear whether this is at a phonetic level, i.e. the quality of different phones; at a phonological level, i.e. the phonemic distribution; or both. Analysing both types of development in production is rarely undertaken, especially given that the majority of L2 speech models are couched in either phonetic or phonological theory. Nevertheless, it has been argued that 'neither an abstract, phonemic analysis nor an acoustic description of L2 segments will adequately predict the way in which an L2 segment is learned' (Levy, 2009: 1150). As such, aspects of different L2 speech models are important to consider if L2 development is to be treated as a multi-level process.

3 Phonetic and phonological approaches to L2 speech learning

The revised Speech Learning Model, the *SLM-r* (Flege and Bohn, 2021) – based on its influential predecessor the *SLM* (Flege, 1995, 2007) – is predominantly considered a phonetic model because it is based on the assumption that L1 and L2 segments are stored within a 'common phonetic space' (Flege and Bohn, 2021: 21, 42). The model suggests that L2 segments which are perceptually similar to those in the L1 are most difficult to learn, while for dissimilar L2 sounds, language learners have a better chance at hearing the cross-linguistic phonetic differences, thus facilitating the formation of new phonetic categories (Flege and Bohn, 2021: 33).

On the other hand, certain phonetic studies analysing perception (Elvin et al., 2021; Escudero, 2005, 2009) and even a number of production studies (e.g. Lang and Davidson, 2019; Levy and Law, 2010; Turner, 2022) have found that substantial cross-linguistic acoustic overlap appears facilitative in L2 segmental learning, provided both individual variation and the effects of surrounding phonetic environment are accounted for. While one interpretation is that these findings are in opposition to the fundamentals of the *SLM-r*, it remains possible that the acoustically similar L2 sounds examined in these studies are perceived to be similar enough to the L1 as to be processed as instances of an already-existing L1 category. Therefore, no learning may in fact be necessary.

More abstract accounts of L2 speech learning suggest that the phonetic detail relevant for maintaining phonological distinctions is prioritized during processing. Indeed, phonetic detail may be filtered out in favour of the phonologically relevant information (e.g. Trubetzkoy, 1969: 36–37). Brown's (1998, 2000) Phonological Interference Model, *PIM*, can be classed as an abstract model; it assumes the existence of a phonological 'filter which funnels acoustically distinct stimuli into a single phonemic category' (Brown, 1998: 147). *PIM* asserts that L2 speech learning is predominantly a feature-based task with matching distinctive features in the L1 and the L2 facilitating the acquisition of new contrasts. Phonological learning is therefore achieved through assembling new features but, as Archibald (2005) points out, existing L1 features can also be reassembled to satisfy the demands of the L2. It must be noted, however, that there are different ways to conceptualize distinctive features in phonology. For example, the term 'feature' can refer to the presence or absence of an articulatory property or it can be used to refer to a category or 'class' of sound that behaves in a certain way. Indeed, Uffmann (2021: 182–183) notes: 'a feature like [+voice] denotes sounds that are produced with vocal fold vibration (voicing), but also makes a class of segments . . . available for phonological processes like voicing assimilation or final devoicing.' As will be discussed in the following section, the importance of this distinction between articulatory-based and abstract accounts of features is particularly relevant for the acquisition of French high rounded vowels by L1 English speakers.

4 French /y/–/u/ for L1 English speakers

The predictions of the models and theories discussed above do not always complement each other and, as noted by Colantoni et al. (2015: 37), one of the variables that poses challenges for certain theories is the acquisition of the French /y/–/u/ contrast by L1 English speakers.

The French front vowel, /y/, is traditionally treated as a phonetically 'new' vowel for L1 American English speakers, while /u/ is more similar and therefore more difficult to acquire (Flege, 1987; Flege and Hillenbrand, 1984; Strange et al., 2007). However, because the GOOSE (Wells, 1982) vowel in English has become progressively fronted due to a sound change (see, amongst others, Harrington et al., 2008; Holmes-Elliott, 2021; Jansen, 2019; Labov et al. 2006; Strycharczuk and Scobbie, 2017), it has also been suggested that /y/ may actually now be a closer phonetic fit to the English high back rounded vowel for both Standard Southern British English Speakers (Turner, 2022: 5–6) and certain varieties of American English (see, for example, Lang and Davidson, 2019: 18). Indeed, the prolific nature of this sound change across many English varieties has been well documented and includes not only the varieties spoken by the present study's participants such as Englishes in the UK and the USA, but also those of Australia, New Zealand and South Africa, among others (for an overview, see, for example, Jansen, 2019: 1–2). Given that not all varieties of English show the same degree of GOOSE-fronting, however, the effect of L1 English variety on French /y/ and /u/ development is an important consideration, and one that must be controlled for (see Section IV.4).

Previous research has also suggested that because GOOSE-fronting is more common in certain contexts, such as after alveolar consonants, variation in pronunciation of French /y/ and /u/ is also likely determined by preceding environment. For example, Levy and Law (2010) find that American English learners of French produce /y/ more accurately in alveolar contexts than bilabial contexts and /u/ more accurately in bilabial contexts than alveolar contexts. Preceding environment is also, therefore, a variable that must be taken into account (again, see Section IV.4).

Different phonological accounts of GOOSE-fronting are also important to consider. As mentioned earlier, features are sometimes thought to reflect articulatory gestures, but according to 'substance-free phonology' (e.g. Iosad, 2012; Uffmann, 2021), the role of a feature is, instead, to group together sounds that behave similarly in many respects, but that are not necessarily characterized by any particular articulatory configuration or acoustic realization. For example, Uffmann (2010) highlights that after a high front English vowel, a palatal glide [j] is inserted before another vowel, e.g. 'see [j] it', while after a high back English vowel, [w] insertion occurs, e.g. 'do [w] it'. The author finds that regardless of the degree of phonetic GOOSE frontedness, speakers continue to use [w] insertion. Evidence such as this has been used in support of the claim that GOOSE-fronting is a predominantly phonetic process (Hamann, 2015: 5), with the phonological status of English /u/ as a [+back] vowel remaining unaffected. In contrast, more phonetically-grounded accounts of phonology (e.g. Harrington et al., 2008) have suggested that due to GOOSE-fronting, phonological backness is lost: [+back] becomes [−back].[2] If phonological overlap is facilitative (Archibald, 2005; Brown, 2000), the two theoretical accounts above have different predictions for English speakers with a fronted GOOSE vowel learning French /y/ and /u/. The first approach, which treats fronted English GOOSE as belonging to a [+back] phoneme, arguably predicts an advantage for French /u/ because of the cross-linguistic phonological similarity. The second approach is likely to predict an advantage for French /y/, instead, because both /y/ and the English GOOSE phoneme are characterized by [−back].

Overall, production research has generally found /y/ to be more target-like and acquired more quickly for L1 Anglophones. Flege and Hillenbrand (1984), for example, analyse the French /y/ and /u/ productions of two groups of L1 American English participants who vary in terms of French experience. Among the less experienced group, productions of the front vowel, /y/, are more accurate than the back vowel according to native French speakers' identifications. However, only the back vowel /u/ appears to develop in line with French experience. Acoustic analyses also support these results and lead the authors to conclude that /y/ is acquired more quickly due to it being a more phonetically dissimilar vowel. In a similar study by Flege (1987), experienced French speakers produce /y/ no different than monolingual French speakers according to the second formant, while none of the L1 English groups' /u/ reach such a nativelike level. These results underline the advantage for the front vowel once more, a conclusion further enhanced by the findings of Lang and Davidson (2019). In the latter study, the authors analyse the speech of native American English speakers living in Paris for 5 years and find substantial overlap between the participants' L2 French /y/ productions and the French /y/ productions of native speakers, while the learners' L2 French /u/ productions are significantly different from those of native speakers. The authors' results also suggest that the American English GOOSE vowel overlaps with French /y/ more than French /u/. Similarly, Turner (2022) finds that not only is /y/ production more nativelike for Standard Southern British English (SSBE) speakers compared to their /u/ production, but also that this French front vowel is most acoustically similar to these speakers' native SSBE /u/ vowel, suggesting that phonetic overlap may be facilitative. However, all of these studies analyse the phonetic quality of /y/ tokens and /u/ tokens without accounting for phonemic distributional errors (i.e. that learners might sometimes use their front vowel in /u/ contexts and their back vowel in /y/ contexts). The current study attempts to tease apart phonetic and phonological gains in order to provide a more nuanced analysis of L2 pronunciation development.

III Overview of the study

This research analyses the French vowel productions of /y/ and /u/ by L1 English learners of French (ELoF) both at the start and end of a 6-month residence abroad in a French-speaking country. Phonological development is assumed to occur if ELoF increase the proportion of times they correctly use, what is for them, a back vowel for /u/ items, and increase the proportion of times they correctly supply their front vowel in /y/ items. Phonetic development is said to occur if ELoF's phonologically accurate /y/ and /u/ vowels become more acoustically target-like over the RA.

Based on the above criteria, our first research question analyses whether categorical and gradient change occur between the pre-RA and post-RA sessions.

- Research question 1: Do ELoF's /y/ and /u/ vowels develop both phonologically and phonetically over the RA?

In light of French /y/–/u/ production research for L1 English speakers (see Section II.4), the learners' front vowel is predicted to show more evident signs of acquisition than the

learners' back vowel (Flege, 1987; Flege and Hillenbrand, 1984; Lang and Davidson, 2019; Turner, 2022). Given that participants will undoubtedly experience very different RAs, it is also expected that there will be a substantial amount of individual variation in both phonetic and phonological development. Our second research question focuses on the extent to which this variation can be explained by different measures of L2 input.

- Research question 2: Are both phonological and phonetic development determined by the quantity and quality of L2 input?

In line with the theoretical considerations discussed above, we expect little to no effect for LoR, but a significant effect for the quantity of L2 input over the RA. When quantity of L2 use is broken down into modality, conversational and aural input are hypothesized to be more influential on L2 speech learning than engagement with visual activities (e.g. reading) (see, for example, Flege, 2008: 175). AoA is not predicted to be significant because the range in AoA is small, which ensures that there is little confound with the effect of input. Nevertheless, years of L2 instruction prior to the RA is recorded and is also predicted to play a role (see, for example, Díaz-Campos, 2004; Saito, 2019; Stevens, 2011). Whether participants have had previous naturalistic L2 exposure is also recorded in case participants with previous extended stays abroad already have more target-like productions (Muñoz and Llanes, 2014; Stevens, 2011). Finally, the type of RA scheme (English teachers/assistants vs. university students) is investigated, but this is not predicted to have an effect on L2 gains (e.g. Mitchell et al., 2017).

IV Methodology

1 Participants

Forty-three adult L1 English learners of French took part in this study. Data from one participant were excluded because the learner did not return for the post-test, leaving a sizeable sample of 42 speakers (28 female, 13 male, 1 non-binary; mean age: 20.5 years, SD: 0.88).[3] Overall, these were intermediate-advanced foreign language learners with the vast majority (*n*=38) studying French at university. However, before their RA, relatively few had received any naturalistic French exposure (*n*=8). Given that language learners have substantially different schooling experiences and learning backgrounds, a LEXTALE vocabulary test (Brysbaert, 2013) was also included. The test consists of both real and non-words and participants are instructed to respond 'yes' or 'no' as to whether they know each item. The original English LEXTALE has been shown to significantly correlate with standardized proficiency tests, especially among advanced learners (Lemhöfer and Broersma, 2012; Puig-Mayenco et al., 2023), and the French LEXTALE also correlates with various indicators of language proficiency (Brysbaert, 2013; Cromheecke and Brysbaert, 2022). The test is used in the present study to offer an indication of French proficiency among the current learners at the start of their RA. Finally, participants self-reported their L1 English variety in case this determined either phonological or phonetic variation. Responses such as 'American'/'USA' or 'Southern'/'South England' were grouped into broad categories as 'General American' and 'Southern

British English'. This yielded 26 × Southern British English, 8 × General American, 4 × Midlands (England), 2 × Scottish and 2 × Northern (England).[4] Although there is likely to be more nuanced variation within these groups, native English data obtained from a wider research project (see Turner, 2023) suggest that all participants' GOOSE vowels are fronted to the extent that every participant's mean second formant of GOOSE is closer to the mean F2 of the native French controls' /y/ than the native French controls' mean F2 of /u/. This is consistent with the well-documented prevalence of GOOSE-fronting across different varieties of English (see Jansen, 2019).

The residence abroad was, for the most part, organized under the ERASMUS scheme, with participants either exchange students in a French-speaking university (*n* = 25), or English teachers/assistants in French schools (*n* = 16). One participant also undertook a work placement, but this individual was excluded when RA Type was analysed as an input measure. On average, the learners' RAs lasted around 6 months (mean: 6.27, SD: 0.65).

Finally, 10 native French speakers (NFS) (6 female, 4 male; mean age: 22.4 years, SD: 2.62) were also included as a control group. These participants were mostly from Paris, although two of the females were also from Belgium.[5]

2 Procedure and items

Twenty-two participants were recorded for the first data collection session in the late summer of 2019 in the three months leading up to the RA (mean: 127 days before the start of the RA; SD: 84). The other 20 were recorded at the start of October 2019, within the first three months of arriving for their RA in Lyon, France (mean: 36 days after RA start; SD: 17). At first, these two groups were treated individually by including Time of First Recording (Pre-RA vs. Start of RA) as a fixed effect in both the phonological statistical analyses and the phonetic analyses.[6] However, this variable never improved model fit as a main effect, nor did it interact with Time Point, suggesting neither phonological nor phonetic change over the RA was affected by whether the individual was recorded just before or at the start of their RA. Given that both groups patterned similarly, they were collapsed into one experimental group. Nevertheless, because of the variability in each participant's length of RA, this was also tested in models (alongside other individual differences, see Section IV.5).

Participants undertook a number of tasks including an English production task, a French perception task, a French production task and the LEXTALE vocabulary test (Brysbaert, 2013). At the end of the RA, participants also completed a Language Engagement Questionnaire (LEQ) (Mitchell et al., 2017), from which the input measurements were calculated. For the purposes of space in the present study, only the data from the French production task, the LEXTALE test, and the LEQ are reported.

The production task involved reading aloud short sentences that were presented in isolation and randomized using Experiment Builder (SR Research, 2011). French minimal pairs (*n* = 8; see Appendix A) contrasting the /y/–/u/ vowels were embedded in the carrier phrase: *C'est qui est le mot que tu cherches* ('It's which is the word you're looking for'). These lexical items were monosyllabic with a simple consonant + vowel

composition and were mostly high frequency items, which increased the likelihood that participants would be familiar with their pronunciation.[7] The consonant preceding both vowels within lexical items was coded post-hoc as Coronal or Non-Coronal in light of previous research that reports English speakers' productions of French /y/ and /u/ are modulated by preceding context (e.g. Levy and Law, 2010).

Several other French phrases were included in the design (for a comprehensive list, see Turner, 2023). These were adapted from those used in Lang and Davidson (2019), and featured lexical items containing one of 10 oral French vowels. The production of these items yielded six tokens for each of the 10 French vowels uttered by each participant. Eight of these vowels (French /y/ and /u/ were excluded) were then used to normalize the high rounded vowels taken from the minimal pairs, as described in the following section. This process also helped to reduce any slight differences in recording quality given that the first data collection session took place in person while the second occurred remotely in March–April 2020 due to Covid-19. At the pre-RA test, a Sennheiser e845S dynamic microphone was used while at the post-test participants used the best microphone available to them.[8] Participants were compensated for their time with a small monetary token.

3 Data analysis

All phrases were automatically transcribed using a script designed by the author in *Praat* (Boersma and Weenink, 2020), then force-aligned phonemically in *SPPAS* (Bigi and Meunier, 2018), with the vowel boundaries checked manually using a second *Praat* script. The onset of the vowel was said to be the zero crossing preceding the first reliable glottal pulse in which formant structure was also visible (Chang, 2019). For preceding laterals, vowel onset was labelled as soon as vertical striations became suddenly darker and/or regular patterning in the waveform changed. The end of the vowel was labelled at the point where one of the first three formants was no longer visible or reliable.

Acoustic extraction consisted of three parts. First, the duration of the vowel was measured in milliseconds. Second, the first (F1), second (F2) and third (F3) formants were extracted from the vowels at 10%, 30%, 50%, 70% and 90% of their duration using the standard Burg LPC settings (Childers, 1979). The maximum number of formants was set at 5; the window length at 25 ms; the pre-emphasis at 50 Hz; and the maximum formant frequency at 5,000 Hz for lower voices and 5,500 Hz for higher voices. Finally, vowel normalization was conducted using the Lobanov (1971) method to reduce the influence of interspeaker anatomical differences such as vocal tract size on the formant measures (Adank et al., 2004; Flynn, 2011). This particular normalization method was also chosen based on Rathcke et al. (2017), in which the Lobanov method was shown to be one of the most effective normalization techniques for minimizing the effect of different recording qualities on formant measurements.

4 Statistical modelling

This study employed a series of binary logistic mixed effects regression models (for phonological accuracy) and linear mixed effects regression models (for phonetic

quality). These were fitted using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2020) in *R* (R Core Team, 2018). All models included random intercepts of Speaker and Word (Baayen et al., 2008; Linck and Cunnings, 2015), and preliminary models featured maximal random structures with random slopes adjustments for all appropriate fixed effects (Barr et al., 2013). Time Point (effects coded: Pre-RA vs. Post-RA) was always included as a fixed effect, and Segment (dummy coded: /u/, /y/) for models that analysed data from both vowels. Alongside the fixed effects fitted in order to answer the present study's research questions, four extraneous variables were tested in separate models in case they too determined the phonological accuracy and phonetic variation observed for both segments over the RA. These include Preceding Phonological Environment (effects coded); French Proficiency; L1 Variety (deviation coded) and Time of First Recording (effects coded). These models were backwards stepwise reduced using the drop1 and step functions from the *lmerTest* package (Kuznetsova et al., 2020). Any of these variables that were found to be significant were included together in one final model (again with a maximal random structure) and backwards stepwise reduced. Final model selection was determined by likelihood ratio tests for logistic regression models and *F* tests for linear models (Kuznetsova et al., 2020). Where post-hoc pairwise comparisons are also reported in the text, these are Tukey-corrected estimated marginal means (Lenth et al., 2020) and computed using the Kenward–Roger degrees of freedom method (Kenward and Roger, 1997). The data, final model outputs and respective *R* formulae are publicly available (https://osf.io/rz7wp).

5 Calculating L2 input measurements

The present study uses the LEQ to measure L2 French input over the RA. This questionnaire asked participants to declare how often they performed 26 different activities (for details, see Appendix B) in French over the RA and to select a response on an ordinal scale of: never, rarely, a couple of times a month, a few times a week, several times a week, or every day. First, the Cronbach's Alpha was calculated for the LEQ ($\alpha = .84$), suggesting good internal consistency. Next, the percentage that each response was selected across all activities was calculated and is illustrated in Figure 1.

To calculate an overall input score from these data for each participant, the responses from the LEQ were modelled using an ordinal mixed effects logistic regression analysis with a flexible threshold that was computed with the Ordinal package (Christensen, 2019). Unlike the traditional linear coding scheme, the present approach accounts for the fact that distances between responses categories are unlikely to be equal. Random intercepts of Participant and Activity were also included, but no fixed effects were specified to yield a 'null' model. The by-participant Beta intercepts were then extracted using the ranef(MODEL) function in *R* as a score for overall quantity of input and transformed from the logit scale to the probability scale using the formula: $\mathrm{Exp}(x)/(1 + \mathrm{Exp}(x))$.

Next, each French activity from the LEQ was coded in accordance with the modality of the input (see Appendix B). This factor was coded three ways: Aural, Conversational or Visual. To acquire values for the different input modalities, a second ordinal mixed effects logistic regression model was fitted which, once more, included random factors of Speaker and Activity. However, on this occasion, Input Modality was fitted as a fixed
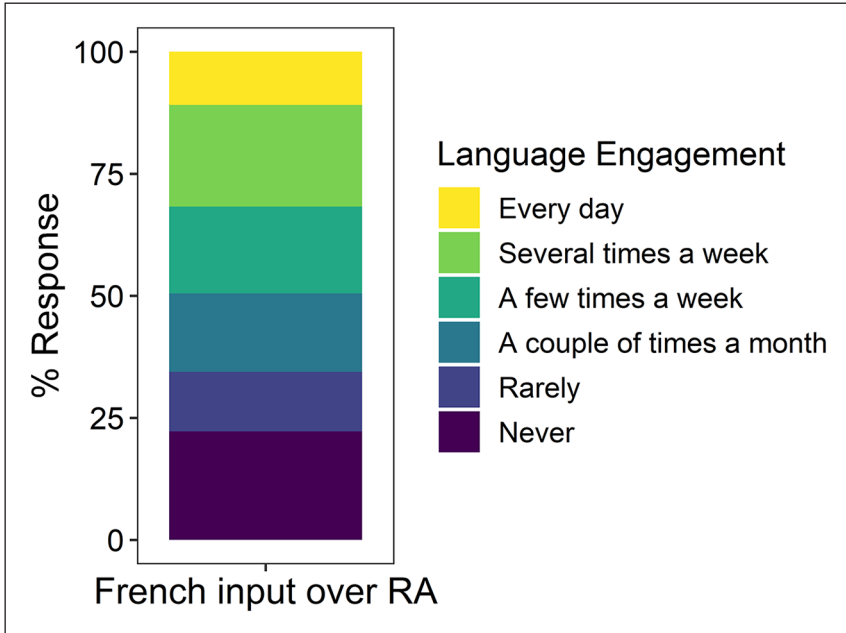
**Figure 1.** Break down of responses for all 26 French activities over the residence abroad.

effect as well and a by-Speaker random slope. Each speaker's coefficients for the three input modalities were then calculated. Because these values were once again on the logit scale, they were transformed to the probability scale using the same formula as above.

The LEQ was also adapted in order to record responses for Length of Residence (LoR); Age of Arrival (AoA); type of residence abroad, i.e. university student, English language assistant, or work placement (RA Type); Length of Learning French previously in instructed settings (LoLF); and whether a participant had already spent time in a French-speaking country before starting their RA (Naturalistic Exposure).[9]

6 K-medoid clustering method

The first stage of analyses determined which of the acoustic dimensions (of those mentioned in Section IV.3) are cues used by ELoF to differentiate /u/ and /y/ in production. Each acoustic measurement was fitted as the dependent variable in separate linear mixed effects regression models, with Segment (/y/ vs. /u/) included as a fixed effect. Speaker and Word were included as random intercepts and a by-Speaker random slope was included for Segment where model convergence would allow. Segment was only found to be significant in models of F2 at 10%, 30%, 50%, 70% and 90% of the vowel duration. Descriptive density plotting for the F2 of French /y/ and /u/ suggested a bimodal distribution for both vowels, which justified the use of k-medoid clustering.

The unsupervised clustering was performed using these five measurements of F2 with Speaker included as a sixth variable. To carry out this process, the phonemes were first
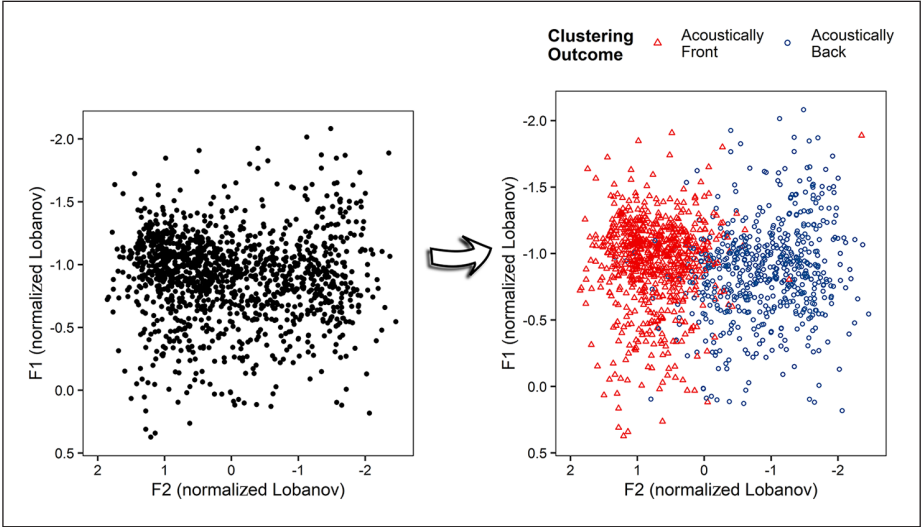
Figure 2. Left-hand side: Productions of both /y/ and /u/ of English learners of French (ELoF) from both pre-RA and post-RA (all pooled together). Right-hand side: Data clustered into an acoustically front vowel (red, triangle) and an acoustically back vowel (blue, circle), irrespective of which vowel each lexical item typically contains.

stripped of their Segment and Time Point labels (see Figure 2, left-hand panel). Subsequently, using functions from both the *Rtsne* (Krijthe, 2015) and *Cluster* (Maechler et al., 2021) packages in *R*, the Gower (1971) distance between data points was calculated and the 'Partition Around Medoid' algorithm was set to work on establishing two clusters. K-medoid clustering is similar to k-means clustering apart from being somewhat less sensitive to outliers (Madhulatha, 2011), and can be performed on mixed data sets, i.e. both categorical variables and continuous variables. The resultant clusters are visualized in Figure 2 (right-hand panel) as an acoustically front vowel, and an acoustically back vowel.

Next, a new Phonological Accuracy variable was coded by comparing these clusters to their original vowel labels. If a token was attributed to the acoustically front cluster and was originally from an item containing /y/, this was coded as 1 (accurate), but 0 (inaccurate) if the token was from a /u/ word. Similarly, if a token was attributed to the back cluster and was originally from an item containing /u/ this was coded as 1 (accurate), but 0 (inaccurate) if the token was from a /y/ word.[10]

V Results

1 L2 phonological development over the RA

From the clustering analyses in the previous section, it was possible to calculate the percentage of times that ELoF correctly used an acoustically front vowel for a /y/ item and
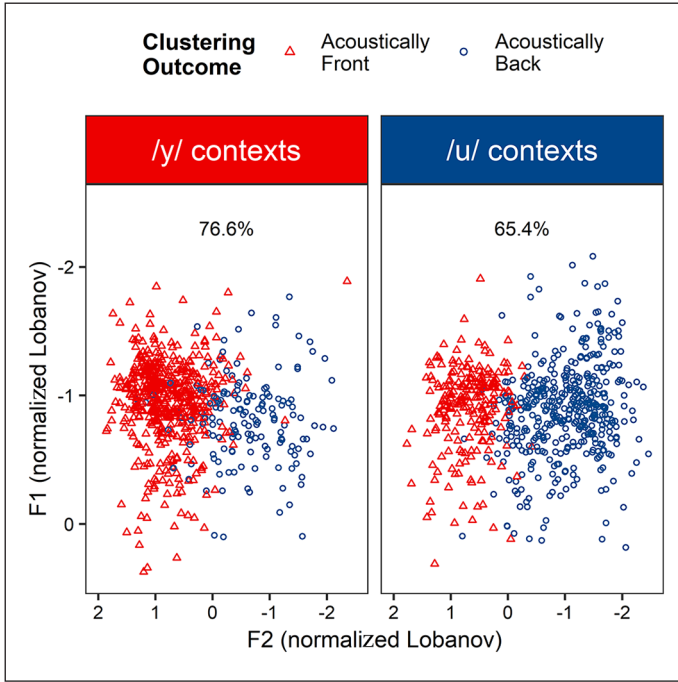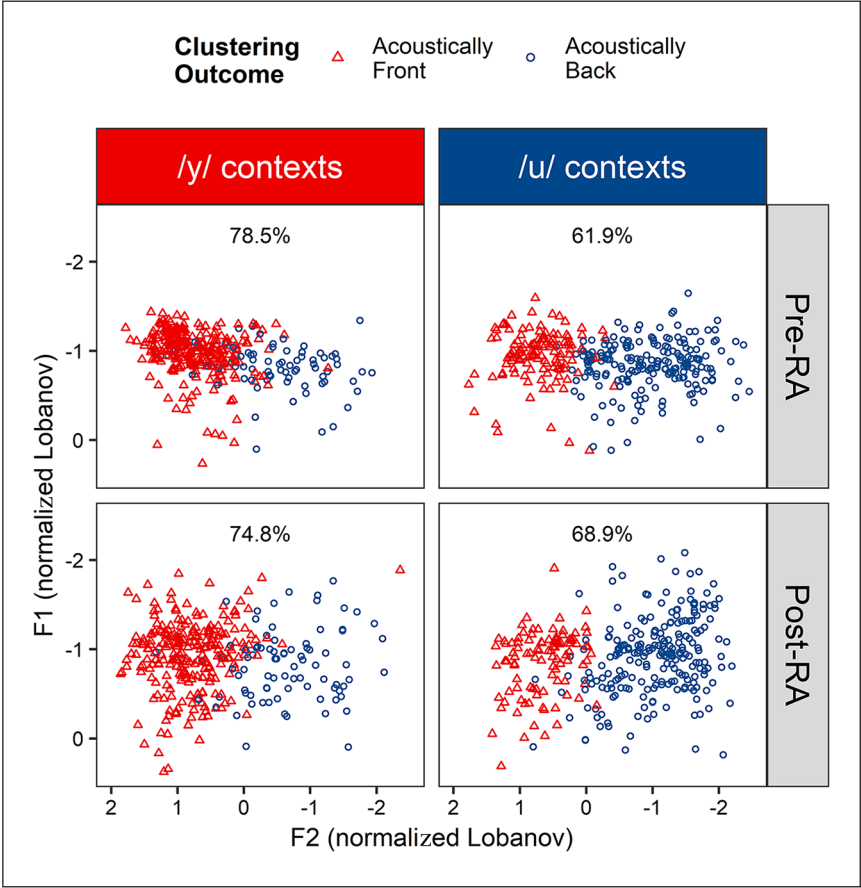
Figure 3. Left-hand panel: Percentage that English learners of French (ELoF) correctly produce their front vowel (red, triangle) for /y/ items. Right-hand panel: Percentage ELoF correctly produce their back vowel (blue, circle) for /u/ items.
Notes. RA = Residence Abroad.

the percentage of instances that an acoustically back vowel was uttered for a /u/ item (Figure 3). This was then broken down by Time Point (Pre-RA vs. Post-RA) to analyse how phonological accuracy changed for both /y/ and /u/ words over the residence abroad (Figure 4).

While there appears to be little phonological progress for /y/ items over the RA (a 3.7% reduction in fact), /u/ accuracy increases by 7%. A binary mixed effects logistic regression model was run to analyse these findings statistically with Phonological Accuracy as the binary dependent variable and an interaction fitted between Time Point and Segment as a fixed effect. The model also controlled for a number of variables. First, Proficiency was retained ($\chi^2(1)=11.68$, $p<.001$) after it was found that higher proficiency levels equate to better phonological accuracy ($\beta=8.46$, $SE=2.56$, $z=3.30$, $p<.001$). Second, an interaction between Preceding Environment (effects coded −0.5, +0.5) and Segment ($\chi^2(1)=24.29$, $p<.001$) was included because phonological accuracy of /u/ items was lower for preceding coronal consonant contexts compared to non-coronal contexts ($\beta=-2.00$, $SE=0.42$, $z=-4.73$, $p<.001$), but phonological accuracy of /y/ items was higher for preceding coronal consonants ($\beta=4.34$, $SE=0.60$, $z=7.25$, $p<.001$). Finally, the model controlled for participants' variety of L1 English

Figure 4. Top-left panel: Percentage that English learners of French (ELoF) correctly produce their front vowel (red, triangle) for /y/ items at the pre-RA time point. Bottom-left panel: Percentage that ELoF correctly produce their front vowel for /y/ items at the post-RA time point. Top-right panel: Percentage that ELoF correctly produce their back vowel (blue, circle) for /u/ items at the pre-RA time point. Bottom-right panel: Percentage that ELoF correctly produce their back vowel for /u/ items at the post-RA time point.
Notes. RA = Residence Abroad.

$(\chi^2(3)=9.82, p<.05)$, because Southern British English speakers were less accurate than the average accuracy of all the English varieties ($\beta=-0.77$, $SE=0.26$, $z=-2.96$, $p<.01$). Random intercepts were included for Speaker and Word and by-Speaker random slopes were also included for the Time Point*Segment interaction.

This model revealed that while /u/ items became more phonologically accurate from pre-RA to post-RA ($\beta=1.13$, $SE=0.48$, $z=2.33$, $p<.05$), the phonological accuracy of /y/ items did not change ($\beta=0.01$, $SE=0.40$, $z=0.04$, $p=.97$).[11] Nevertheless, the difference between the amount of change observed for each segment only approaches

significance ($\beta=-1.12$, $SE=0.67$, $z=-1.68$, $p=.09$). Finally, the phonological accuracy of /y/ was not found to be significantly higher than /u/ across time points ($\beta=0.07$, $SE=0.66$, $z=0.11$, $p=.91$).

In summary, these results suggest that the most substantial phonological improvement was for /u/ items. Nevertheless, phonological accuracy of both vowels increased with proficiency suggesting both are acquirable in the long run. This is modulated by preceding environment, however, with phonological accuracy of /y/ items higher for preceding coronal consonant contexts, and /u/ accuracy higher for preceding non-coronal consonants.

2 The effect of L2 input on L2 phonological development

Next, the French input measures (LEQ input measures, LoR, LoLF, RA type and previous naturalistic French exposure) as well as AoA were tested as predictors of this phonological variation. Separate logistic mixed effects regression models were run and, in each one, a different input variable was fitted as a fixed effect and its interaction with Time Point was included. Once more, Preceding Environment, Proficiency, L1 variety and Time of First recording were also controlled for where significant effects were observed. The first set of models analysed /u/ items and controlled for Preceding Environment ($\chi^2(1)=12.77$, $p<.001$) and Proficiency ($\chi^2(1)=5.99$, $p<.05$), while the second set analysed phonological accuracy in /y/ items and controlled for Preceding Environment ($\chi^2(1)=11.65$, $p<.01$) and L1 English variety ($\chi^2(3)=8.26$, $p<.05$).[12]

Results for the /u/ models revealed a significant interaction between Time Point and French Input Quantity ($\chi^2(1)=4.86$, $p<.05$): participants who declared higher quantities of French input were more likely to increase their phonological accuracy for /u/ items over the RA ($\beta=5.61$, $SE=2.69$, $z=2.09$, $p<.05$), whereas for those experiencing less French input, phonological accuracy for /u/ items increased less substantially or actually decreased. This fitted interaction is plotted in Figure 5.

Subsequently, the effect of input quality was analysed by breaking the effect of L2 input down by modality. A significant interaction with Time Point was observed only for Aural Input ($\chi^2(1)=4.15$, $p<.05$) and Visual Input ($\chi^2(1)=5.70$, $p<.05$), while Conversational Input did not appear to significantly determine phonological development ($\chi^2(1)=2.73$, $p=0.10$). None of the remaining input variables tested revealed an effect: Time Point did not interact with LoR ($\chi^2(1)=1.96$, $p=0.16$), neither did LoLF ($\chi^2(1)=1.73$, $p=0.19$), nor RA type ($\chi^2(1)=0.78$, $p=0.38$), nor previous naturalistic French exposure ($\chi^2(1)=1.17$, $p=.28$), nor AoA ($\chi^2(1)=1.43$, $p=0.23$), and none of these variables significantly improved model fit as simple main effects either.

Finally, all input variables and AoA were tested in /y/ models while controlling for Preceding Environment and L1 English variety. Only RA Type was found to be a significant predictor of the change in phonological accuracy ($\chi^2(1)=3.96$, $p<.05$). The beta coefficient indicated that the proportion of times that the target front vowel quality was used in /y/ items increased more substantially for individuals at French-speaking universities compared to those working as language assistants ($\beta=1.40$, $SE=0.70$, $z=1.99$, $p<.05$).
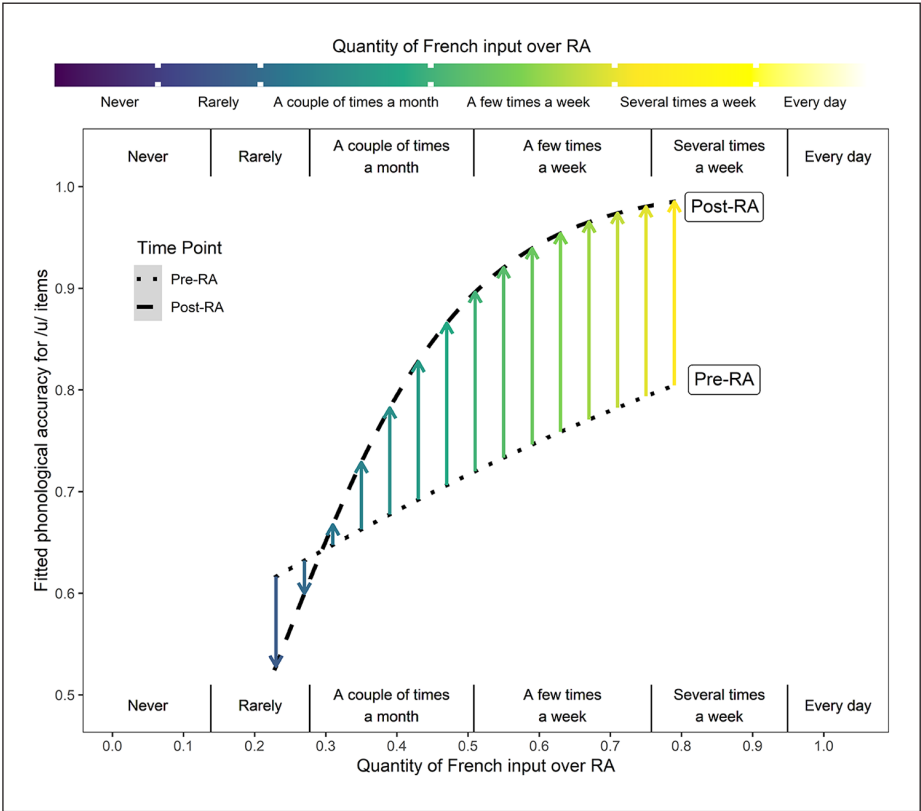
Figure 5. The effect of French input quantity over the Residence Abroad (RA) on the change in phonological accuracy for /u/ items from pre-RA (dotted) to post-RA (dashed).
Note. Category thresholds from ordinal model of French input are marked for guidance (transformed to probabilities).

In summary, the quantity of French input declared over the RA was a predictor of the phonological development observed for /u/, and the input type mattered. Specifically, Aural Input and Visual Input were most determinant, while Conversational Input appeared to play less of a role. Moreover, the input values obtained from the LEQ had a stronger effect than LoR. RA Type also appeared to play a role but primarily for /y/ phonological development. Indeed, more progress was observed for university students compared to individuals working as English assistants or teachers.

3 L2 phonetic development over the RA

Subsequent analyses determined whether the phonetic quality of either vowel became closer to native French norms from pre-RA to post-RA, with the second formant at 50% of the vowel duration employed as the dependent variable. This acoustic dimension was

found to distinguish ELoF's /y/ and /u/ more strongly than F1 and F3 (see Section IV.6) and was therefore the measure used for this stage of modelling. To ensure only phonetic development was being analysed, phonologically 'inaccurate' tokens were excluded from both time points. That is, only productions of a front vowel in /y/ contexts and productions of a back vowel in /u/ contexts were retained. These tokens were then compared to the French /y/ and /u/ of native French group. The factor Group (effects coded 3-ways: ELoF pre-RA, ELoF post-RA and Native French Speakers, NFS) was fitted as an interaction with Segment (/u/ vs. /y/, dummy coded). Random intercepts were included for Speaker and Word, as well as a by-Speaker random slope for Segment.

The model revealed that the interaction was significant ($F_{(2, 99)}$=13.53, $p<.001$). Indeed, the F2 of ELoF's back vowel /u/ was higher than that of the NFS group at both the pre-RA session (β=0.77, SE=0.13, t=5.89, $p<.001$) and the post-RA session (β=0.68, SE=0.13, t=5.25, $p<.001$), while the F2 of ELoF's front vowel /y/ was significantly more nativelike than ELoF's /u/ at both the pre-RA session (β=−0.91, SE=0.19, t=−4.79, $p<.001$) and the post-RA session (β=−0.74, SE=0.19, t=−3.92, $p<.001$). Furthermore, the F2 of ELoF's /y/ was not significantly different from NFS' at both the pre-RA session (β=−0.14, SE=0.10, t=−1.50, p=0.14) and the post-RA session (β=−0.06, SE=0.10, t=−0.65, p=0.52).[13] These results are visualized in Figure 6.

The learner data were then analysed in isolation to test for significant phonetic change over the RA while testing for a proficiency effect. The second formant was employed once more as the dependent variable. Two separate models were run: one for phonologically accurate tokens of /y/ and one for phonologically accurate tokens of /u/. The final /u/ model controlled for Proficiency ($F_{(1,34)}$=24.47, $p<.001$), because F2 decreased as participants' French LEXTALE scores increased (β=−3.31, SE=0.66, t=−5.00, $p<.001$). The same model also controlled for Preceding Environment ($F_{(1,406)}$=47.68, $p<.001$), given that F2 was higher for coronal contexts than other environments (β=0.35, SE=0.06, t=5.71, $p<.01$). No effect of Proficiency or Preceding Environment was observed in the /y/ model and L1 English variety was not found to be significant in either model.

Results reveal that for phonologically accurate /y/ tokens, ELoF's category becomes more fronted over the RA (β=0.08, SE=0.04, t=2.10, $p<.05$), approximating the native French mean, while the backing observed in ELoF's /u/ vowel does not quite reach significance (β=−0.09, SE=0.05, t=−1.88, p=.06).

In summary, these results demonstrate that the phonetic quality of /y/ was already within the native speaker range at the first time point, yet still developed to approximate the NFS mean F2 values more closely by the end of the RA. The phonetic quality of ELoF's phonologically accurate /u/ tokens, however, was distant from the native French values at both time points and did not exhibit a significant phonetic change over the RA. The effect of proficiency for the back vowel suggests that the phonetic detail of this vowel is, nonetheless, acquirable in the long run. Finally, the phonetic quality of /u/ is modulated by the preceding environment, with coronal contexts resulting in fronter productions.

4 The effect of L2 input on L2 phonetic development

All input measures (LEQ input measures, LoR, LoLF, RA type and previous naturalistic French exposure) and AoA were tested as main effects and as interactions with Time
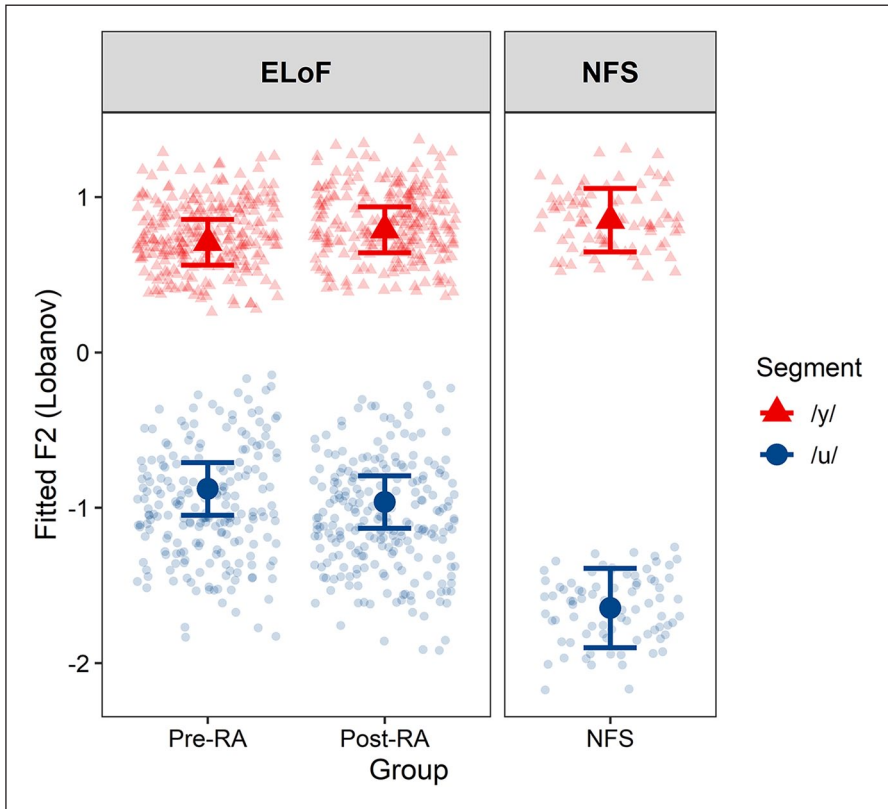
Figure 6. Fitted model output for F2 demonstrating greater phonetic overlap between /y/ productions of English learners of French (ELoF) and native French speakers (NFS) compared to their /u/ productions.

Point in both the phonologically accurate /y/ and /u/ models. The significant effects of Preceding Environment and Proficiency highlighted in the previous section for French /u/ were also controlled for. For the front vowel, /y/, no input factors were found to be significant, nor was AoA. However, in the /u/ model, three notable effects were observed. First, LoLF was found to interact with Time Point ($F_{(1,413)}=4.02, p<.05$), indicating that the most substantial /u/-backing occurred for participants who had learned French the longest ($\beta=-0.03, SE=0.02, t=-2.01, p<.05$). Second, the RA Type appeared to play a role in ELoF's phonetic development of /u/ ($F_{(1,398)}=5.38, p<.05$), with language teachers/assistants more likely to make phonetic gains than the students at a foreign institution ($\beta=-0.24, SE=0.10, t=-2.33, p<.05$). Finally, the effect of previous naturalistic exposure ($F_{(1,403)}=6.53, p<.05$) revealed that participants who had not previously resided in a French-speaking country also experienced most /u/-backing over the RA ($\beta=-0.29, SE=0.11, t=-2.55, p<.05$). Indeed, pairwise comparisons (Lenth et al., 2020) demonstrate that by the post-RA session, this group start to draw level ($\beta=-0.04, SE=0.14,$

$t=-0.26$, $p=.99$) with the group of learners who had already lived in a French-speaking country previously and who had a slightly more retracted /u/ vowel initially.[14]

In summary, none of the declared input variables from the Language Engagement Questionnaire were found to determine the extent of phonetic change over the RA, but the number of years already learning French, the type of RA that learners undertook, and the amount of previous naturalistic exposure did all determine the amount of /u/-backing.

VI Discussion

This study analysed the effect of a number of L2 input measures on both the phonetic and phonological development of French /y/ and /u/ for L1 English learners of French (ELoF) undertaking a residence abroad (RA). Here, we return to the study's research questions.

- Research question 1: Do ELoF's /y/ and /u/ vowels develop both phonologically and phonetically over the RA?

First, results reveal clear evidence of phonological development for /u/ contexts over the RA: ELoF significantly increase the proportion that they correctly utter, what is for them, a back vowel in /u/ items. However, the same cannot be said for /y/ contexts: ELoF do not increase the proportion that they correctly use their front vowel in /y/ words. After controlling for phonological accuracy, however, phonetic analyses indicate an advantage for the front vowel instead. That is, phonologically accurate /y/ tokens become more fronted over the RA and are consistently closer to native French norms than accurate /u/ tokens, while the phonetic change observed for accurate /u/ tokens over the RA does not quite reach significance.

The results for phonological development run counter to the majority of L2 production studies analysing L1 Anglophones' acquisition of the French /y/–/u/ contrast, in which the front vowel /y/ is generally found to be more acquirable (e.g. Flege, 1987; Flege and Hillenbrand, 1984; Lang and Davidson, 2019; Levy and Law, 2010; Turner, 2022). On the other hand, findings at the phonetic level are more consistent with such studies, given that /y/ appears more easily learned than /u/. Taken in combination, these results indicate a discrepancy between phonological development and acquisition of phonetic detail. Nevertheless, the phonological findings appear consistent with more abstract models and the phonetic results align with certain phonetic-oriented accounts.

For example, if cross-linguistic overlap in phonological features is facilitative (e.g. Archibald, 2005; Brown, 1998, 2000) the observation that /u/ items are more easily acquirable than /y/ items can be explained by the fact that the [+back] feature required for French /u/ (Dell, 1980) also characterizes L1 English /u/ (e.g. Chomsky and Halle, 1968). In contrast, French /y/ bears [−back] which is not directly available from the phonological specification of English /u/ and therefore poses a substantial learning challenge. These results also imply that GOOSE-fronting is a phonetic process that does not affect the vowel's phonological status as [+back] (see, for example, Hamann, 2015). Indeed, if the phonological specification of GOOSE had changed from [+back] to [−back] in line with the phonetic change, it would be difficult to explain the advantage for the French back vowel observed here.

However, after removing phonological errors (i.e. front tokens used in /u/ contexts and back tokens in /y/ items), the present research demonstrates a phonetic advantage for accurate /y/ tokens in terms of F2. Given that /y/ is also presumed to be the most acoustically similar sound to ELoF's L1 English GOOSE counterpart (e.g. Lang and Davidson, 2019; Turner, 2022), these results support the phonetic-based accounts of L2 speech learning that suggest acoustically similar sounds may, in some circumstances, be more easily acquired. As discussed earlier, multiple perception studies (Elvin et al., 2021; Escudero, 2005, 2009) and certain production studies (Lang and Davidson, 2019; Levy and Law, 2010; Turner, 2022) find that substantial cross-linguistic phonetic overlap can facilitate L2 speech learning. For French /u/, however, new acoustic detail needs to be internalized and reproduced through unfamiliar articulatory configurations (see, for example, Oakley, 2019). The findings for proficiency in the present study also support the contention that /y/ requires little phonetic development: learners of all French proficiencies appear to produce French /y/ with relatively target-like phonetic detail, while the phonetic quality of /u/ is only backed among the most advanced French speakers.

Although the present results at the phonetic level are consistent with previous research, the findings appear somewhat in opposition to the *SLM-r* (Flege and Bohn, 2021). This is because the *SLM-r* suggests that perceptually similar sounds are likely to pose the greatest learning challenge at the phonetic level, while the present research suggests the phonetic detail of the similar vowel, /y/, is more readily acquirable. However, the *SLM-r* also highlights that cross-linguistic phonetic similarity cannot be determined by acoustic comparisons alone (Flege and Bohn, 2021: 33). Indeed, although perceptual assimilation tests were outside the scope of the present study, if French /u/ is perceptually similar to English /u/ – despite the presumed substantial cross-linguistic acoustic differences – these findings could yet offer support for the *SLM-r*. Indeed, Chang (2015) points out that it is relatively common for L1 Anglophones to perceive /u/ in languages such as French, Mandarin and German as similar to English /u/, despite little acoustic overlap. Given the difficulties noted in the acquisition of /u/ by L1 Anglophones, he implies that phonological similarity between L1 and L2 sounds may inhibit the perception and production of phonetic detail unique to the L2 sound (Chang, 2015). This would align with the present study's findings: the phonologically dissimilar French vowel, /y/, demonstrates significant phonetic change over the RA, but not the phonologically similar French vowel, /u/.

Finally, it is important to note that findings at both phonological and phonetic levels are modulated by the preceding environment. This is because /y/ items were more target-like for preceding coronal consonants than non-coronal consonants, while the opposite held true of /u/ items, echoing previous research (e.g. Levy and Law, 2010). Again, this is likely due to the phonological constraints of GOOSE-fronting in English, whereby coronal contexts promote fronter productions compared to non-coronals (e.g. Harrington et al., 2008; Holmes-Elliott, 2021), resulting in more straight-forward L1 transfer for coronal+/y/ contexts, but a more substantial learning task for coronal+/u/ contexts. This interpretation is also supported by previous perceptual assimilation research in which /y/ is categorized as English /u/ more strongly when English GOOSE is preceded by coronal consonants (e.g. Levy, 2009).[15]

- Research question 2: Are both phonological and phonetic development determined by the quantity and quality of L2 input?

Several measures of L2 input were tested as predictor variables of phonological and phonetic variation. The LEQ measures of input were analysed first. The overall quantity of French input from the LEQ revealed that participants declaring most input over the RA were also more likely to increase the proportion of times they correctly use their back vowel in /u/ items from pre-RA to post-RA. The quality of input also determined this phonological change: greater amounts of aural input (e.g. listening to the radio, watching French TV) and visual input (e.g. reading, browsing the internet) were consistent with a steeper rise in correct usages of the back vowel in /u/ contexts.

The fact that phonological development is determined both by aural and visual input counters Flege's (2008) assumption that the latter is unlikely to affect L2 speech learning (p. 175). Potentially, the one-to-one grapheme–phoneme mappings for this specific French vowel pair increases the importance of visual input when it comes to learning this contrast: hearing words containing /u/ as well as seeing their written form containing <ou> (e.g. *bout* /bu/) may help to differentiate this phoneme from the front vowel (which has its own consistent grapheme–phoneme mapping of <u> to /y/; e.g. *bu* /by/). Research into the effects of orthography on L2 pronunciation is steadily growing (see, for example, Bassetti, 2008; Bassetti et al., 2020; Young-Scholten and Langer, 2015), and complements that of the processing domain where it has been shown that orthographic input can facilitate encoding new contrasts (e.g. Escudero et al., 2008).

While aural and visual input appear to determine phonological development in the present study, the amount of conversational French (e.g. small talk, phone calls) does not show the same effect, nor does it determine any phonetic change in the vowels. One reason why conversational input may be less likely to yield a significant result is that the measure in its current form is not sufficiently informative when it comes to the nature of these conversations. Indeed, in a residence abroad context, it is likely that French encounters will sometimes take place with other international students or colleagues (e.g. Meier and Daniels, 2013) and thus the quality of input is likely to be highly variable due to the participants' own non-native utterances and those of their non-native interlocutors. In agreement with Flege and Wayland (2019), the quality of input (rather than simply the modality) needs to be more closely scrutinised, and although the goal of the present research was to test L2 input measures from the original LEQ design, this questionnaire could be easily adapted for the purposes of analysing accented input in future research.

Several other variables were not found to be significant predictors of phonological development or phonetic learning. For example, AoA showed no effect, although it is likely that for this specific variable, the null effect is due to very limited variation in the age range tested rather than age not being an important variable itself. While the present results cannot, therefore, evaluate the hypothesis that input matters more than age (Flege, 2018), they do demonstrate that for a specific age group, the quantity and quality of L2 input on a residence abroad is important, especially for phonological learning.

Similarly, the LoR determined neither phonological development nor phonetic variation, aligning with previous research (Flege and Bohn, 2021; Flege and Liu, 2001; Flege and Wayland, 2019; Flege et al., 2006). Flege and Liu (2001) suggest that LoR may only

provide a useful estimate of the quantity of L2 input for individuals who frequently find themselves in contexts that allow and promote the use of their L2, as is discussed further in Flege and Wayland (2019). While it could be argued that the residence abroad context does not meet this threshold for enough of the participants for it to yield an effect, the LEQ input results do not suggest that participants did not engage with the L2 over the RA. Instead, findings would indicate that LoR simply is not an adequate measure of L2 input.

Another predictor of phonetic development for the back vowel was the measure LoLF: /u/ becomes more retracted over the RA for individuals who have received instructed French input for the greatest number of years. This finding suggests that the impact of L2 input on acquiring phonetic detail may not be observed instantaneously. Indeed, previous research has highlighted that there can be a delay between the processing of phonetic detail and adapting L2 pronunciations to reflect that learning (e.g. Nagle, 2018).

Furthermore, RA type may play a role. However, in the present research, different RA types appear to correspond to different types of pronunciation development. Indeed, while the phonological accuracy of /y/ improves most for students enrolled at a foreign university compared to English language assistants/teachers, the phonetic detail of /u/ becomes more nativelike for language assistants. These findings support previous research which underlines that L2 linguistic gains can occur in both contexts (e.g. Mitchell et al., 2017: 228), but more specifically, it would appear that these contexts are linked to different types of L2 speech development. While these results have implications for the study abroad literature, it remains unclear whether these differences are motivated by discrepancies in input between assistants and university students, or some other underlying factor. More careful between-group matching for factors such as French proficiency and length of learning French may help to rule these variables out as potential confounds in future research.

Finally, the previous naturalistic exposure input variable was found to significantly determine the variability in ELoF's phonetic /u/-backing over the RA. Indeed, those with previous immersive French experiences were least likely to exhibit phonetic development for /u/. This is because by the end of the RA, participants without previous naturalistic exposure start to come into line with the learners with previous immersive French experience, who had a slight advantage initially.

VII Conclusions

This study analysed the effect of L2 input measures on both the phonetic and phonological development of French /y/ and /u/ for L1 Anglophones participating in a residence abroad scheme in a French-speaking country. Significant phonetic development of /y/ was observed, and the proportion of times the learners used, what is for them, a back vowel in /u/ items also increased. The quantity and quality of input over the RA more strongly determined the latter – more phonological – change but the length of time learning French, amount of previous naturalistic exposure, and the type of RA did appear to play a role in determining phonetic development of /u/.

While the results of this research help to shed light on important questions regarding both the phonetics–phonology interface and the role of both input quantity and quality on L2 pronunciation, a number of limitations should be noted. First, the heterogeneity of the present participant sample may affect the generalizability of these results. For example, although GOOSE-fronting is extensive among all participants, certain individuals' GOOSE vowels are likely to be more overlapping with French /y/ than others. As such, the 'L1 variety' variable coded in this study may not sufficiently account for this within-group variability. Furthermore, it is unclear whether the phonetic variation of /y/ and /u/ across different varieties of French had an impact on learning, given that participants did not reside in one single location. More careful selection of RA participants in future L2 speech learning studies focusing on the interface between phonetics and phonology may help to reduce possible effects from confounding variables. Second, it has been argued that data obtained from sound perception experiments are more relevant for phonological theory (see, for example, Archibald, 2021b). Therefore, it is not clear whether production data – even if analysed categorically through clustering – have straightforward implications for abstract phonological theory. Indeed, we know that many performance factors affect realizations of L2 sounds, creating distance between speech output and the linguistic representation at its source. Nevertheless, it is generally accepted that 'production draws on the mental grammar' (Archibald, 2021b: 2) at least to some extent, and the bimodal distribution observed for F2 in both vowels in the present study suggests at least some degree of phonemic confusion can be observed in production. Finally, the present LEQ analysis treats the variation in responses between each activity as random (given that Activity is included as a random intercept). Future research might benefit from taking a more nuanced approach by conducting Item Response Theory (IRT) analysis to understand the relationship between input as a latent trait and the observed responses to items. For reasons of space, this procedure was not attempted in the present study.

Notwithstanding these limitations, the present results make a valid contribution to the field by demonstrating that phonetic and phonological development are not necessarily interdependent and that different types and quantities of input are not all equal for both types of learning. Indeed, this research highlights that L2 pronunciation can develop in various ways in line with different levels of representation being updated by L2 input. The challenge for future research is to establish how consistently both types of development are predicted by different types of input and to explore different means of capturing the complex nature of contact with a foreign or second language. For example, the present study analysed not only input quantity, but also input quality, as indexed by modality. Nevertheless, more detailed treatment of input quality (e.g. foreign-accented speech input) is likely to be an important line of enquiry moving forwards.

Although input appears to be emerging in L2 speech learning theories as one of the most determinant factors in L2 production, somewhat propelling the focus onwards from the effects of age on L2 speech, it is clear that not all input metrics are reliable. The present study demonstrates how mixed effects ordinal logistic regression modelling can be used to obtain individual scores of language engagement from the LEQ, an instrument that provides a rich overview of input and use. It is hoped that this, alongside other methodological advances presented here, such as teasing apart gradient and categorical change via k-medoid clustering, will ensure that L2 speech production development can be treated as a complex, multi-level process in the future.

Finally, given that both phonetic and phonological aspects of producing this novel L2 contrast improved, our results support the notion that residence abroad programmes are beneficial for L2 speech learning. As such, this research aligns with many previous residence abroad studies in which significant linguistic gains have been reported. Such findings should prove reassuring for both language learners who are considering residences abroad and the types of funding schemes that support these opportunities for adult language learners.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

James Turner (iD) https://orcid.org/0000-0001-6724-3056

Notes

1. All French varieties in question feature the high rounded vowel contrast under investigation. As such, RA location was not deemed grounds for exclusion.
2. As Uffmann (2021: 188) notes, however, this would mean that the feature [round] would have to become active at some point. Traditionally, [round] is considered to be inactive in English because the features [high] and [back] are sufficient to differentiate between vowels, e.g. /i/ [+high] [−back] is distinct from /u/ [+high] [+back], and /u/ is distinct from /o/ [−high] [+back]. If English GOOSE loses its phonological backness, the feature [round] must become active to ensure GOOSE remains distinct from /i/ in English.
3. As noted by one reviewer, the distribution across genders is skewed towards female. This was not by design and is likely not a skewness unique to this sample but a manifestation of the gender disparity in university language students at the institutions that were recruited from. Indeed, other residence abroad studies of native English university students of French have obtained similarly skewed samples (26 females to 3 males; see, for example, Mitchell et al., 2017: 53).
4. Due to low numbers for the last two categories, these were merged as an 'Other' (Scottish/Northern) category.
5. Given that these participants' results patterned similarly to the other native French speakers, they were included in the final sample.
6. To view the test statistics for Time of First Recording, please see supporting documentation here: https://osf.io/23xa4.

7.   This was confirmed in a questionnaire at the end of each session, in which participants did not select any of the /y/–/u/ test items as being unfamiliar.
8.   The post-RA session was conducted virtually due to Covid-19 and thus careful checks of the recordings were required to ensure there were no technical errors or background noise. No tokens had to be excluded from this manual check, however. In line with recommendations specified in Sanker et al. (2021), external microphones were used where possible (i.e. Apple Earpods + mic or other headsets/stand-alone mics) and lossless audio formats chosen in Audacity (Audacity Team, 2021). Sanker et al. (2021) report no significant difference in first and second formants between recordings from a professional solid-state recorder and an external headset. Furthermore, sound files generated from Audacity were found to perform well when it came to extracting F1 and F2 reliably (Sanker et al., 2021).
9.   Note, to avoid confusion, Naturalistic Exposure is different from the variable Time of First Recording (Pre-RA vs. Start of RA) mentioned in Section IV.2. Time of First Recording serves to analyse whether phonetic and phonological variation are affected by whether the first recording took place in the host country or in the UK before leaving for the RA. Naturalistic Exposure records whether participants have experience living abroad in a French-speaking country before the RA.
10. At this stage, the same clustering method described in this section was also performed on the /y/ and /u/ productions by native French speakers. Results indicated that both /y/ and /u/ were categorized correctly 100% of the time among the control group.
11. The latter was revealed after refitting the model with /y/ as the reference category.
12. Unlike the model for both segments where Southern British English speakers are least accurate, phonological accuracy for /y/ items specifically was significantly lowest for the L1 American English participants ($\beta=-1.14$, $SE=0.58$, $z=-1.97$, $p<.05$), and significantly higher than the grand mean for individuals with speaking Midlands (UK) English variety ($\beta=2.36$, $SE=0.91$, $t=2.59$, $p<.01$).
13. These results were obtained by refitting the same model with /y/ as the reference category.
14. Statistic for the Tukey-corrected pairwise estimated marginal mean comparison between the post-RA results for learners with no previous naturalistic exposure and the pre-RA results of the learners who had already resided in a French-speaking country.
15. This sound change may also explain why phonological accuracy is modulated by L1 variety, with certain GOOSE vowels more fronted than others.

References

Adank P, Smits R, and van Hout R (2004) A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116: 3099–107.

Aoyama K, Flege JE, Guion SG, Akahane-Yamada R, and Yamada T (2004) Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32: 233–50.

Archibald J (2005) Second language phonology as redeployment of L1 phonological knowledge. *Canadian Journal of Linguistics* 50: 285–314.

Archibald J (2021a) Ease and difficulty in L2 phonology: A mini-review. *Frontiers in Communication* 6: 1–7.

Archibald J (2021b) Phonological redeployment and the mapping problem: Cross-linguistic e-similarity is the beginning of the story, not the end. *Second Language Research* 39: 1–11.

Audacity Team (2021) *Audacity(R): Free audio editor and recorder* [computer programme]. Audacity. Available at: https://audacityteam.org (accessed November 2023).

Baayen RH, Davidson DJ, and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.

Barr DJ, Levy R, Scheepers C, and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78.

Bassetti B (2008) Orthographic input and second language phonology. In: Piske T and Young-Sholten M (eds) *Input matters in SLA*. Clevedon: Multilingual Matters, pp. 191–206.

Bassetti B, Mairano P, Cerni T, and Masterson J (2020) Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning* 70: 1218–56.

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

Bigi B and Meunier C (2018) Automatic segmentation of spontaneous speech [Segmentação automática da fala espontânea]. *Revista de Estudos da Linguagem* 26: 1489–1530.

Boersma P and Weenink D (2020) *Praat: Doing phonetics by computer: Version 6.1.28* [computer programme]. Available at: https://www.fon.hum.uva.nl/praat/ (accessed November 2023).

Brown C (2000) The interrelation between speech perception and phonological acquisition from infant to adult. In: Archibald J (ed.) *Second language acquisition and linguistic theory*. Oxford: Blackwell Publishing, pp. 4–63.

Brown CA (1998) The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14: 136–93.

Brysbaert M (2013) Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica* 53: 23–37.

Carroll SE (1999) Putting 'input' in its proper place. *Second Language Research* 15: 337–88.

Carroll SE (2017) Exposure and input in bilingual development. *Bilingualism* 20: 3–16.

Chang CB (2015) Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In: Raimy E and Cairns CE (eds) *The segment in phonetics and phonology*. Chichester: John Wiley, pp. 199–217.

Chang CB (2019) Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics* 74: 96–113.

Childers DG (1979) *Modern spectrum analysis*. Hoboken, NJ: John Wiley.

Chomsky N and Halle M (1968) *The sound pattern of English*. London: Harper and Row.

Christensen R (2019) *Ordinal: Regression models for ordinal data* [R package]. Available at: https://cran.r-project.org/package=ordinal (accessed November 2023).

Cobb T (2020) *Compleat lexical tutor* [web page]. Available at: https://www.lextutor.ca/vp/comp (accessed November 2023).

Colantoni L, Steele J, and Escudero P (2015) *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

Cromheecke O and Brysbaert M (2022) A French c-test for language assessment. *L'Année Psychologique* 122: 3–28.

DeKeyser R and Larson-Hall J (2005) What does the critical period really mean? In: Kroll JF and De Groot AM (eds) *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press, pp. 99–108.

Dell F (1980) *Generative phonology and French phonology*. Cambridge: Cambridge University Press.

Díaz-Campos M (2004) Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition* 26: 249–73.

Elvin J, Williams D, Shaw JA, Best CT, and Escudero P (2021) The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese vowels by English vs. Spanish listeners. *Languages* 6: 44.

Escudero P (2005) *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT Dissertation Series 113, Utrecht University, Utrecht, Netherlands.

Escudero P (2009) The linguistic perception of similar L2 sounds. In: Boersma P and Hamann S (eds) *Phonology in perception*. Berlin: Mouton de Gruyter, pp. 151–90.

Escudero P, Hayes-Harb R, and Mitterer H (2008) Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60.

Flege JE (1987) The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

Flege JE (1995) Second language speech learning theory, findings and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233–77.

Flege JE (2007) Language contact in bilingualism: Phonetic system interactions. In: Cole J and Hualde J (eds) *Laboratory phonology: Volume 9*. Berlin: Mouton de Gruyter, pp. 353–81.

Flege JE (2008) Give input a chance! In: Piske T and Young-Scholten M (eds) *Input matters in SLA*. Bristol: Multilingual Matters, pp. 175–90.

Flege JE (2018) It's input that matters most, not age. *Bilingualism: Language and Cognition* 21: 919–20.

Flege JE, Birdsong D, Bialystok E, et al. (2006) Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34: 153–75.

Flege JE and Bohn O-S (2021) The revised speech learning model (SLM-r). In: Wayland R (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 3–83.

Flege JE and Hillenbrand J (1984) Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76: 708–21.

Flege JE and Liu S (2001) The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–52.

Flege JE and Mackay IRA (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.

Flege JE, Munro MJ, and MacKay IRA (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–34.

Flege JE and Wayland R (2019) The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies* 2: 1–44.

Flynn N (2011) Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series* 2 11: 1–28.

Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27: 857–71.

Hamann S (2015) Phonological changes. In: Bowern C and Evans B (eds) *The Routledge handbook of historical linguistics*. London: Routledge, pp. 249–63.

Harrington J, Kleber F, and Reubold U (2008) Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123: 2825–35.

Holmes-Elliott S (2021) Calibrate to innovate: Community age vectors and the real time incrementation of language change. *Language in Society* 50: 441–74.

Iosad P (2012) Vowel reduction in Russian: No phonetics in phonology. *Journal of Linguistics* 48: 521–71.

Jansen S (2019) Change and stability in GOOSE, GOAT and FOOT: Back vowel dynamics in Carlisle English. *English Language and Linguistics* 23: 1–29.

Kenward MG and Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983–97.

Krijthe JH (2015) *Rtsne: T-distributed stochastic neighbor embedding using a Barnes–Hut implementation* [R package]. Available at: https://github.com/jkrijthe/Rtsne (accessed November 2023).

Kuznetsova A, Brockhoff PB, Christensen RHB, and Jensen SP (2020) *lmerTest: Tests in linear mixed effects models* [R package]. Available at: https://cran.r-project.org/web/packages/lmerTest (accessed November 2023).

Labov W, Ash S, and Boberg C (2006) *Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Lang B and Davidson L (2019) Effects of exposure and vowel space distribution on phonetic drift: Evidence from American English learners of French. *Language and Speech* 62: 30–60.

Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44: 325–43.

Lenth R, Buerkner P, Herve M, et al. (2020) *Emmeans* [R package]. Available at: https://www.cran.r-project.org/web/packages/emmeans (accessed November 2023).

Levy ES (2009) Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America* 125: 1138–52.

Levy ES and Law FF (2010) Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception–production relationship. *The Journal of the Acoustical Society of America* 128: 1290–305.

Linck JA and Cunnings I (2015) The utility and application of mixed-effects models in second language research. *Language Learning* 65: 185–207.

Lobanov BM (1971) Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49: 606–608.

Madhulatha TS (2011) Comparison between k-means and k-medoids clustering algorithms. *International Journal of Advanced Computing (IJAC)* 3: 65–69.

Maechler M, Rousseeuw P, Struyf A, Hubert M, and Hornik K (2021) *Cluster: Cluster analysis basics and extensions* [R package]. Available at: https://cran.r-project.org/web/packages/cluster (accessed November 2023).

Mayberry RI and Kluender R (2018) Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism* 21: 886–905.

Meier G and Daniels H (2013) 'Just not being able to make friends': Social interaction during the year abroad in modern foreign language degrees. *Research Papers in Education* 28: 212–38.

Mitchell R, Tracy-Ventura N, and McManus K (2017) *Anglophone students abroad: Identity, social relationships and language learning*. London: Routledge.

Moyer A (2008) Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In: Piske T and Young-Scholten M (eds) *Input matters in SLA*. Bristol: Multilingual Matters, pp. 159–74.

Muñoz C (2006) The effects of age on foreign language learning: The BAF project. In: Muñoz C (ed.) *Age and the rate of foreign language learning*. Bristol: Multilingual Matters, pp. 1–39.

Muñoz C and Llanes À (2014) Study abroad and changes in degree of foreign accent in children and adults. *Modern Language Journal* 98: 432–49.

Nagle CL (2018) Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning* 68: 234–70.

Oakley M (2019) Articulation of L2 French mid and high vowels. In Calhoun S., Escudero P., Tabain M., and Warren P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1724–28). Australasian Speech Science and Technology Association Inc.

Puig-Mayenco E, Chaouch-Orozco A, Liu H, and Martín-Villena F (2023) The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism* 13: 299–314.

R Core Team (2018) *R: A language and environment for statistical computing* [computer programme]. R Foundation for Statistical Computing. Available at: https://www.r-project.org (accessed November 2023).

Rankin T and Unsworth S (2016) Beyond poverty: Engaging with input in generative SLA. *Second Language Research* 32: 563–72.

Rathcke T, Stuart-Smith J, Torsney B, and Harrington J (2017) The beauty in a beast: Minimising the effects of diverse recording quality on vowel formant measurements in sociophonetic real-time studies. *Speech Communication* 86: 24–41.

Saito K (2015) Experience effects on the development of late second language learners' oral proficiency. *Language Learning* 65: 563–95.

Saito K (2019) Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English /ɹ/ pronunciation. *Second Language Research* 35: 149–72.

Sanker C, Babinski S, Burns R, et al. (2021) Supplementary materials: (Don't) try this at home! The effects of recording devices and software on phonetic analysis. *Language* 97: e360–82.

Sheldon A and Strange W (1982) The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–61.

Shinohara Y and Iverson P (2021) The effect of age on English /r/–/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics* 89: 101108.

Slabakova R (2013) Adult second language acquisition. *Linguistic Approaches to Bilingualism* 3: 48–72.

SR Research (2011) *Experiment builder: Version 1.10.165* [computer programme]. Available at: https://www.sr-research.com/experiment-builder (accessed November 2023).

Stevens JJ (2011) Vowel duration in second language Spanish vowels: Study abroad versus at-home learners. *Journal of Second Language Acquisition and Teaching* 18: 77–104.

Strange W, Weber A, Levy ES, et al. (2007) Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America* 122: 1111–29.

Strycharczuk P and Scobbie JM (2017) Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America* 142: 322–31.

Trofimovich P (2011) Language experience in L2 phonological learning: Effects of psycholinguistic and sociolinguistic variables. *IRAL – International Review of Applied Linguistics in Language Teaching* 49: 135–56.

Trubetzkoy N (1969) *Grundzüge der Phonologie [Essentials of phonology]*. Travaux du Cercle Linguistique de Prague: Volume 7. English translation by C. Baltaxe. Berkeley, CA: University of California Press.

Turner J (2022) Analysing the relationship between L2 production and different stages of L2 processing: Eye-tracking and acoustic evidence for a novel contrast. *Journal of Phonetics* 91: 1–15.

Turner J (2023) Phonetic development of an L2 vowel system and tandem drift in the L1: A residence abroad and L1 re-immersion study. *Language and Speech* 66: 756–85.

Uffmann C (2010) The non-trivialness of segmental representations. Unpublished paper presented at the Old World Conference on Phonology (OCP) 7, Nice, France.

Uffmann C (2021) London calling (or cooling?): Feature theory, phonetic variation, and phonological change. *Yearbook of the Poznan Linguistic Meeting* 7: 181–215.

Wells JC (1982) *Accents of English 1: An introduction*. Cambridge: Cambridge University Press.

Winitz H, Gillespie B, and Starcev J (1995) The development of English speech patterns of a 7-year-old Polish-speaking child. *Journal of Psycholinguistic Research* 24: 117–43.

Yang C (2018) A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 8: 665–706.

Young-Scholten M and Langer M (2015) The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics* 36: 93–114.

Appendix A. French lexical items.

Variable	Item	Word Frequency (Cobb, 2020)	Variable	Item	Word Frequency (Cobb, 2020)
/y/	su	1k	/u/	sous	1k
/y/	bu	2k	/u/	bout	1k
/y/	jus	7k	/u/	joue	1k
/y/	vue	1k	/u/	vous	1k
/y/	rue	1k	/u/	roux	6k
/y/	lu	1k	/u/	loup	5k
/y/	tu	1k	/u/	tout	1k
/y/	du	1k	/u/	doux	2k

Note. 1k = within 1,000 most frequent words, 2k = within 2,000 most frequent words, etc.

Appendix B. Language engagement activities (Mitchell et al., 2017) and coded modality.

Activity	Modality
Watch TV	Aural
Watch films	Aural
Listen to music	Aural
Listen to talk radio	Aural
Listen to lectures	Aural
Browse the internet (e.g. read news, etc.)	Visual
Use social networking sites (e.g. Facebook, Twitter/X, etc. but not including instant messaging)	Visual
Read emails	Visual
Write emails	Visual
Read literature (e.g. fiction, poetry, short stories)	Visual
Read academic texts	Visual
Read newspapers	Visual
Read magazines	Visual
Read text messages (not including instant messaging)	Visual
Write text messages (not including instant messaging)	Visual
Write reports (e.g. work, academic)	Visual
Write for leisure (e.g. journal)	Visual
Use instant messaging	Visual
Have short phone conversation (< 5 minutes)	Conversational
Have long phone conversations (> 5 minutes)	Conversational
Teach a class	Conversational
Engage in service encounters	Conversational
Engage in small talk	Conversational
Engage in long casual conversations	Conversational
Participate in organized social activities (e.g. clubs, church, sports, etc.)	Conversational
Participate in seminars / language classes	Conversational