Journal of Phonetics 91 (2022) 101134

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Analysing the relationship between L2 production and different stages of L2 processing: Eye-tracking and acoustic evidence for a novel contrast

James Turner

University of Southampton, Southampton, Hampshire SO17 1BF, United Kingdom

A R T I C L E I N F O

Article history:
Received 26 March 2021
Received in revised form 2 January 2022
Accepted 7 January 2022
Available online 2 February 2022

Keywords:
L2 speech perception
L2 speech production
Online processing
Acoustic phonetics
L2 phonology
Eye-tracking
Visual World Paradigm

A B S T R A C T

This study analyses the relationship between native English speakers' perception and production of the novel French /y/–/u/ contrast. Acoustic data were extracted from the learners' production of French minimal pairs contrasting these French vowels and compared with their processing of the same items in a Visual World eye-tracking task. Results reveal that the vowel most acoustically similar to the learners' native English /u/ vowel, French /y/, is both easier to identify at early processing stages and more acoustically similar to a native French control group in production, indicating a perception-production relationship. Furthermore, analyses of individual variation reveal that the learners who process both /y/ and /u/ more successfully at later processing stages are also more likely to mark a greater distinction between these phonemes in production. Together, these results indicate a relationship between L2 processing and L2 production at multiple levels. Implications for current L2 speech models are discussed.

1. Introduction

1.1. Background

Many of the theoretical models for second language (L2) speech learning assume a relationship between L2 processing and L2 production. For example, if native Spanish speakers learn to produce a distinction between /i/ and /ɪ/ in English, a contrast not found in their native language (L1), then these vowels are also expected to be distinguished in perception. However, while certain studies offer support for such a link (e.g. Elvin, Williams, & Escudero, 2016; Evans & Alshangiti, 2018; Flege, MacKay, & Meador, 1999), others have found little to no evidence of an L2 perception-production relationship (e.g. de Leeuw et al., 2019; Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011; Rallo Fabra & Romero, 2012).

A number of methodological concerns have been highlighted which help to explain these inconsistent findings (see e.g. Nagle & Baese-Berk, 2021, for a comprehensive review). One factor, in particular, which warrants further investigation is the level of linguistic representation targeted in the two modalities. Melnik-Leroy, Turnbull, & Peperkamp (2021), for example, analyse L1 English speakers' perception-production relationship for the French /y/–/u/ contrast and find that prelexical perception measures in an ABX categorisation task correlate with phonetic measures of pseudowords in production, but not with the production quality of real lexical items. The authors conclude that the perception-production relationship is strongest "within but not across levels" (Melnik-Leroy et al., 2021, p. 16).

The present study builds on this line of research by treating lexical processing as a dynamic event that can be tracked in real-time using eye-tracking technology. If the difficulty in L2 processing varies between the stimulus onset and the moment of the ultimate click response, these different processing stages may also correlate with L2 production patterns to different extents. By tapping both lower-level phonetic processing in critical windows closest to stimulus onset, and higher-level processing metrics in later windows, this research ensures that the relationship with production can be tested both within and between levels.

Before turning to the detail of the present study, however, the motivation for supposing a relationship between production and processing must firstly be considered. Importantly, the assumption of such a link is not unique to L2 speech learning research.

E-mail address: James.Turner@soton.ac.uk

https://doi.org/10.1016/j.wocn.2022.101134
0095-4470/Crown Copyright © 2022 Published by Elsevier Ltd.

1.2. L1 speech models and the perception-production relationship

Several theories of L1 acquisition, for example, rely on the premise that perception precedes production. Indeed, Eimas et al. (1971) demonstrate that infants are sensitive to variation in phonetic detail within and between the categories of adult speakers, before they are able to produce such sounds. Although this capability has been shown to diminish as time continues (Werker & Tees, 1984), correlations between perception and production continue to be revealed in the native phonetic inventories of adults (e.g. Beddor et al., 2018; Brunner et al., 2011; Newman, 2003). The specific nature of the link between perception and production of native speech varies between theoretical approaches, but often such approaches converge on the assumption that a shared level of representation (e.g. motor, articulatory, acoustic) guides both modalities (e.g. Diehl & Kluender, 1989; Fowler, 1986; Liberman et al., 1967).

Learning the sounds of a foreign language has, of course, several similarities with L1 acquisition, but crucially, articulatory gestures specific to the native language are already frequently utilised, unlike in initial stages of L1 acquisition. Whether processing success will play as much of a pivotal role in determining adult language learners' pronunciation is, therefore, open to debate.

1.3. L2 speech models and the perception-production relationship

The revised Speech Learning Model (*SLM-r*) (Flege & Bohn, 2021), for example, claims that perception and production in L2 speech learning "co-evolve without precedence" (Flege & Bohn, 2021, p. 29), indicating that learners who are generally more successful in L2 processing are also more likely to produce relevant phonetic L2 detail. This is in contrast to previous versions of the *SLM*, in which accurate perception is deemed a necessary step in learning to accurately produce L2 sounds (e.g. Flege, 1995, 2007).

The *SLM-r* also assumes a perception-production relationship across segments: the sounds which are most difficult to process are also most unlikely to be target-like in production. Indeed, those that are perceived as too phonetically similar to an existing L1 category are less likely to be both identified and produced accurately, while for more perceptually dissimilar L2 sounds, language learners stand a higher chance of eventually hearing and producing cross-linguistic differences in phonetic detail (Flege & Bohn, 2021, p. 33).

In contrast, processing models such as *PAM-L2* (Best & Tyler, 2007) and the *L2LP* (Escudero, 2005, 2009; van Leussen & Escudero, 2015) claim the level of perceptual difficulty is not so much due to individual L1-L2 segmental perceived similarity (c.f. the *SLM-r*), but rather, how pairs of L2 sounds are mapped to a learner's L1 phonology. Both models assume that new contrasts are most difficult to acquire, but differ with respect to whether cross-linguistic *articulatory* or *acoustic* differences determine the processing challenge. Indeed, *PAM-L2* adopts a direct-realist approach in which articulatory gestures are presumed to be perceived directly from the speech stream, a process which creates a natural link with learners' own production patterns. The *L2LP* argues, instead, that L2 representations are stored initially as *acoustic* replicas of L1 productions. As such, acoustic overlap between languages is deemed facilitative to L2 processing (Elvin et al., 2021; Escudero, Simon & Mitterer, 2012) and a perception-production relationship is yielded from comparisons between the perception of acoustic forms in the target language and those of L1 productions Elvin, Williams & Escudero, 2016).

In view of these speech models, the relationship between L2 processing and L2 production has been widely investigated (see Llisterri (1995) and Flege (1999) for reviews of early studies and Nagle and Baese-Berk (2021) for more recent works). One phonemic pair that has received much attention in both perception and production research is the French /y/–/u/ contrast for L1 English speakers.

1.4. French high rounded vowels

The French /y/–/u/ contrast is phonologically "new" for L1 English speakers. That is, phonetically high rounded vowels are contrastive along the front-back cline in French, e.g. "du" [dy] (some) and "doux" [du] (soft), while similar variation for a phonetically high rounded vowel in English does not have phonemic consequences, e.g. the word "do" might be uttered [du], [dʉ] or even [dy].

Research analysing the French /y/–/u/ contrast has often suggested that French /u/ is perceived as more phonetically similar to the English /u/ vowel which leads to it being more difficult to learn for L1 English speakers (e.g. Flege & Hillenbrand, 1984; Flege, 1987; Levy, 2009a; Strange et al., 2007).

Indeed, several production studies involving L1 American English (AmE) learners of French have found /y/ to be more target-like than /u/ (e.g. Flege & Hillenbrand, 1984; Flege, 1987; Lang & Davidson, 2019). Flege and Hillenbrand (1984) analyse the production of these vowels by two groups of L1 AmE speakers who differ in terms of their French experience. Acoustic analyses suggest that the F2 of both group's front vowel, /y/, is closer to that of native French speaker values. Similarly, when native French speakers were charged with identifying both group's productions, the inexperienced group's /u/ productions were much less well identified than /y/, although the opposite pattern is observed among the experienced group. The authors conclude that /y/ is learned more quickly due to its phonetic dissimilarity to the English /u/ of the participants' L1 dialect. A later study by Flege (1987) also analyses /y/ and /u/ production by 3 different groups of L1 AmE participants with varying French experience. The most advanced group produce /y/ no differently from the native speakers according to the second formant, while /u/ was found to have a much higher second formant for all L1 English groups, a result the author attributes once more to /y/ being perceived as phonetically "new". Finally, Lang and Davidson (2019) analyse the L2 productions of /y/ and /u/ by American English speakers who have lived in Paris for 5 years or more. Their results suggest that the group's /y/ production is most acoustically similar to the native French control group, while their /u/ is produced with significantly higher F2 values. Once again, then, the advantage for the front vowel is clear.

However, robust evidence of GOOSE (Wells, 1982) fronting has been reported in many varieties of English over recent decades, and especially in Standard Southern British English

(SSBE) (e.g. Holmes-Elliott, 2020; Strycharczuk & Scobbie, 2017). Due to this sound change, French /u/ is likely to be more acoustically dissimilar to the SSBE /u/ vowel, and potentially, therefore, also more acquirable. Rebranding /y/ as a phonetically similar vowel for English speakers builds on Levy and Law's (2010) interpretation; in that study, /y/ was deemed phonetically similar in alveolar contexts for L1 AmE speakers because this environment favours GOOSE-fronting in their L1 variety (Levy & Law, 2010, p.1304). Nevertheless, /y/ was still more target-like in production than /u/, suggesting that phonetic similarity may not necessarily impede L2 acquisition.

While production research often analyses the quality of /y/ and /u/ individually, perception research on French /y/ and /u/ generally focuses on the learners' ability to discriminate the phonemes in comparison to other contrasts, rather than the identification accuracy of each individual sound (e.g. Darcy et al., 2012; Levy & Law, 2010; Levy & Strange, 2008; Levy, 2009b). As such, it is not always clear which vowel is most easily perceived by L1 English speakers and perception-production research has often relied on comparing discrimination measures of the contrast as a whole with both the production quality of /y/ and /u/ individually, revealing no significant effect (e.g. Levy & Law, 2010).

A further issue for L2 perception-production research is whether there is a direct correspondence between success in L2 processing for phonetic detail at a sub-lexical level (accurate perception of non-words, for example) and lexical identification of real words containing the same sounds. Indeed, previous /y/–/u/ research, such as Darcy et al. (2012), has shown that intermediate L1 English learners of French struggle to lexically differentiate /y/ and /u/ items despite exhibiting relatively low error rates in an ABX categorisation task for non-words, while in advanced learners, the opposite pattern is observed: the learners can successfully differentiate lexical contrasts but error rates remain high in the phonetic categorisation task. Although no comparisons with production data are made, it is likely that such variation across processing results will yield different strengths of perception-production relationships, as has been shown previously (e.g. Melnik-Leroy et al., 2021). This demonstrates the need for L2 perception-production research to analyse different stages of lexical processing in more detail to establish whether the relationship is strongest as soon as the phonetic detail of the stimulus is heard, at later stages when this detail is mapped to phonological representations, or at the moment when an offline lexical decision is made.

1.5. The present study

This research analyses the processing and production of the French vowels /y/ and /u/ by 23 L1 English learners of French (ELoF) and 10 native French speakers (NFS). Firstly, I determine which French vowel is most acoustically dissimilar to ELoF's English /u/ and whether this dissimilar French vowel is also more nativelike in ELoF's L2 production. Secondly, I ask whether the phoneme found to be most nativelike in ELoF's L2 production is also processed more successfully in a Visual World eye-tracking task, a correspondence which would indicate a perception-production relationship. Finally, to analyse the link between L2 processing and L2 production at a more phonological level, statistical analyses determine whether individual learners with better processing measures for both vowels are more likely to mark a larger acoustic distinction between phonemes in production.

In all of these investigations, participants' processing is monitored continually from the very millisecond that the listening task commences. Given that phonetic processing is likely to precede phonological categorisation and lexical recognition (Escudero, 2009, p. 154; McQueen, 2005, p. 265), the temporally rich nature of eye-tracking data from a Visual World Paradigm (Cooper, 1974; Tanenhaus et al., 1995) allows production results to be compared with both lower and higher levels of ELoF's L2 processing. Indeed, this research compares measures of L2 processing that are more phonetic-based (early-window fixation proportions as soon as the stimulus is heard), measures that are likely to be more phonological (late-window processing fixation proportions) and lexical measures of L2 processing (proportions of accurate offline click response) with both phonetic and phonological measures of L2 production to test the relationship within and between levels.

1.6. Data collection

Data collection sessions involved English and French read-aloud tasks, a French eye-tracking Visual World task, a language background and engagement questionnaire adapted from Mitchell, Tracy-Ventura and McManus (2017), and a French LEXTALE test as an index for linguistic proficiency (Brysbaert, 2013).[1] In the proficiency test, a learner started with a score of 0.5 and was penalised or rewarded depending on their lexical recognition from a list of mixed real and nonce words. Scores could range from 0 (entirely incorrect responses) to 1 (100% real word identification and 100% non-word rejection). English tasks were always performed first and production tasks before processing ones. Participants viewed short videos in the same language of subsequent tasks in order to help them acclimatise and instructions were also given in the same language to facilitate a more monolingual language mode (Grosjean, 1998).

2. Production

2.1. Hypotheses

Due to the frontness of the English /u/ vowel across preceding phonetic contexts in SSBE (Holmes-Elliott, 2020; Strycharczuk & Scobbie, 2017), a higher degree of overlap with the French front vowel /y/ is expected, while French /u/ is hypothesised to be more dissimilar compared to the English counterpart:

- *H1: Compared to native French /y/, native French /u/ is more acoustically dissimilar to ELoF's L1 English /u/ vowel.*

Although, a universally accepted method of quantifying phonetic similarity cross-linguistically has yet to be established (Flege & Bohn, 2021, p. 33), previous research generally varies between acoustic analyses (e.g. Escudero, Simon, &

[1] Other tasks completed in the data collection session, but not reported in the current paper, include a silent film narration exercise, a sociolinguistic interview in both languages, and an English eye-tracking perception task.

Mitterer, 2012) and perceptual assimilation tests (e.g. Levy, 2009a). While the two methods do not always yield the same results (Strange, 2007), some consistencies have been observed for certain vowels (e.g. Elvin et al., 2021; Flege, Fox, & Munro, 1994).

If cross-linguistic acoustic dissimilarity is facilitative to L2 speech learning, lexical items containing the dissimilar French vowel, /u/, are predicted to be more similar to native French speaker values in production:

- *H2: ELoF's L2 French /u/ productions are acoustically more native-like than their L2 French /y/ productions.*

2.2. Methodology

2.2.1. Participants

Twenty-three native English university students of French (3 male, 19 female and 1 non-binary) with a mean age of 20.17 (range: 19–21) took part in the research. Participation was compensated with a small monetary gesture and lasted around 1.5 hours. Twenty-one of these participants had never lived in a French-speaking country for an extended period of time and had acquired French in UK classroom settings. Two other participants had resided in France before: one 6 months, the other for 2 years but over ten years before testing, and both had predominantly learned French in L1 instructed settings in similar fashion to the other participants.

The learners were either speakers of Standard Southern British English (SSBE) or had had extensive contact with this variety (due to parents' dialect or living in the south of England for school/university). They were also generally advanced instructed learners, having studied French on average 9.41 years (range: 4–15) and now enrolled in French university degrees at undergraduate level. However, given the lack of extensive naturalistic French input, the participants fit within the definition of "foreign language learners" as opposed to "second language learners" (see Best & Tyler, 2007; Piske, 2007 for more on this distinction). Any remaining variation in French proficiency was to be controlled for in statistical analyses as a continuous factor (Colantoni, Steele, & Escudero, 2015, p. 82), using vocabulary knowledge as an index (Brysbaert, 2013). Some participants spoke other languages to an advanced level, but French was by far and large the group's strongest foreign language, minimising the compound of second-to-third language transfer (Marx & Mehlhorn, 2010).

Finally, a group of ten native European French speakers was included (mean age: 21.89, range: 19–25). These participants were predominantly from the Parisian area and were relatively proficient in English, having studied it at school and university. As such, they may differ, if only slightly, from French monolinguals in terms of their productions (Chang, 2012) and performance more generally (Schmid, Gilbers, & Nota, 2014). Nevertheless, bilingual control groups have previously been encouraged to ensure that the inherent differences of bilinguals compared to monolinguals are not only observable in the experimental group (e.g. Schmid, Gilbers & Nota, 2014). As such, these speakers were deemed appropriate for the present purposes.

No participants reported impairments of sight (other than corrected vision), hearing or speech that could affect the results of this research.

2.2.2. Procedure

Monosyllabic lexical items (see Tables 1 and 2, Appendices) containing high-rounded vowels in both English and French were selected. These were CV /y/–/u/ minimal pairs in French, with high-frequency pairs preferred to increase the likelihood that ELoF would be familiar with the items, which was also confirmed by participants in a questionnaire post-experiment. In English, partially contrastive members of the GOOSE/GHOUL lexical sets (Wells, 1982) were elicited as part of a wider project analysing high rounded vowels. The data are reported here to provide a SSBE /u/ baseline for these speakers and by including both GOOSE and GHOUL lexical sets, an appropriate range of L1 phonetic variation for English /u/ was ensured, with following laterals contexts predicted to be slightly less fronted given their resistance to this sound change (Strycharczuk & Scobbie, 2017).

These words were embedded in the following carrier phrases, respectively:

- *Say <ITEM> 'cause it's the word you're looking for.*
- *C'est <ITEM> qui est le mot que tu cherches.*
  (It's <ITEM> which is the word you're looking for)

Embedding items after "*Say*" and "*C'est*" ensured the phonetic environment preceding the lexical items was similar cross-linguistically (Nearey & Rochet, 1994), and the preceding environment *within* lexical items was also controlled post-hoc in statistical analyses. The following environment was consistently PAUSE + [ki] in French because of silent word-final consonants and OBSTRUENT/LATERAL + PAUSE + [kə] in SSBE.[2] Due to vowel harmony (Nguyen and Fagyal, 2008), the French carrier phrase may promote fronter articulations due to the following [i], as pointed out by one anonymous reviewer. This concern is addressed more fully in the discussion but does not appear to affect which vowel is most acoustically similar to the English counterpart.

Stimuli phrases were presented one at a time and read aloud by the participants. ELoF participated in both the English and French production tasks, while NFS read aloud only the French phrases. All sentences were pseudo-randomised so that no two stimuli containing the same segment of interest, nor two items of the same minimal pair, would be consecutive. Distractor phrases were also employed which contained six tokens for each monophthong vowel of each language. These phrases were the same as those used in Lang and Davidson (2019) for both English and French (phrase-medial items with vowels embedded between obstruent consonants). The formants of these vowels were then used to normalise the high rounded vowels, as detailed in the following section.

2.2.3. Data preparation

Transcriptions of recordings were force-aligned using SPPAS (Bigi & Meunier, 2018) and the vowel tokens were hand-checked, with segmentation boundaries adjusted manually where necessary in *Praat* (Boersma & Weenink, 2020). The start of the vowel was said to be the onset and offset of the steady portions of the first two formants to minimise the

[2] Although participants were not instructed to leave a pause after the experimental lexical items, a natural short silence was inserted in the majority of utterances as a means of emphasising these words before reading the rest of the phrase.

effect of the phonetic environment (Ladefoged, 2003, p. 105). A *Praat* script was composed to extract these formant measurements. The standard LPC algorithm by Burg (Childers, 1979) was used with the computation employing a 25 ms window length, a maximum number of 5 formants and a pre-emphasis of 50 Hz. A maximum formant frequency of 5000 Hz was set for speakers with lower voices and 5500 Hz for higher voices. The first (F1), second (F2) and third (F3) formant readings were recorded from the midpoint of the segment. These were normalised using the Lobanov (1971) method, a technique that has been shown to correct for inter-speaker anatomical differences in vocal tract size while maintaining meaningful variation (Adank, Smits, & van Hout, 2004; Flynn, 2011).

2.2.4. Statistical analyses

Linear mixed effects regression (LMER) models (Models 1–4) were fitted using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova, Brockhoff, Christensen, & Jensen, 2020) in *R* (R Core Team, 2018). All models included random intercepts for Speaker and Word to account for random variation (Baayen, Davidson, & Bates, 2008; Linck & Cunnings, 2015) and maximal random slopes were employed in preliminary models (Barr et al., 2013) before being stepwise reduced using the 'Step' function of the *lmerTest* package (Kuznetsova et al., 2020). Likelihood ratio tests determined whether inclusion of a given variable as a random slope significantly improved model fit and AIC and BIC values were also compared manually (for structures and model outputs see https://osf.io/rv7kd/). A similar backwards stepwise reduction technique was also used to determine which fixed effects should be controlled for. Such models were built from the bottom up (with a maximal random structure) following Melnik-Leroy et al. (2021) and stepped (Kuznetsova et al., 2020). Preceding context (coded Coronal vs. Non-Coronal) and Word Frequency (see Tables 1 and 2, Appendices) were not found to significantly improve model fit and so were consistently removed. To analyse the effect of ELoF's proficiency on their L2 speech, ELoF's French data were modelled aside from the native speaker data following the step-wise process described above. No significant effect of Proficiency was observed for either French vowel, however, and so these results are not discussed further.

Categorical variables fitted as fixed effects were contrast coded using simple coding (e.g. −0.5, +0.5 for two-category factors and −0.33, +0.67 for three-category contrasts). Where "pairwise comparisons" are reported within the text, these refer to Tukey corrected estimated marginal means, which are conducted post-hoc for fixed factors using the package *Emmeans* (Lenth et al., 2020) and the Kenward-Roger degrees-of-freedom method (Kenward & Roger, 1997).

2.3. Baseline production results

To analyse the native English and native French data, three separate models were fitted using normalised F1, F2 and F3 as dependent variables and Segment (coded 3 ways: French /y/ & /u/ and English /u/) inserted as a fixed factor in each. Speaker and Word were included as random factors and while by-Speaker random slopes for Segment were included in preliminary models, these were removed in Models 2 and 3 after a likelihood ratio suggested their inclusion did not improve model fit and because including these random slopes led to model convergence issues.

Findings for F1 (Model 1) suggest that there is no significant difference between English /u/ and the French vowels in terms of vowel height.

In contrast, pairwise comparisons within Segment in the analysis of F2 (Model 2) reveal /y/ (β = 0.91, SE = 0.33, t = 2.76, p < 0.05) and /u/ (β = −1.51, SE = 0.33, t = −4.59, p < 0.001) are significantly different from the English equivalent in terms of advancement. Importantly, this difference is greater for French /u/, demonstrated by the larger estimate and its significance at the higher confidence level.[3]

Finally, results of Model 3 do not suggest that F3 varies significantly between English and French high rounded vowels.

All pairwise comparisons for the factor of Segment in models of F1, F2 and F3 are visualised in Fig. 1. In short, these results confirm that of the two French high rounded vowels, /u/ is more acoustically distinct from the English counterpart, especially along the front-back cline.[4]

2.4. L2 French production results

Descriptive plotting (Fig. 2) was firstly conducted to analyse ELoF's French productions in comparison to those of NFS. This suggested that /y/ and /u/ productions by the ELoF group are substantially more overlapping than the productions by native French speakers.

A model of the second formant (Model 4) analyses this statistically by testing an interaction between Segment (/y/ or /u/) and Group (ELoF or NFS), with Segment included as a by-Speaker random slope and Group included as a by-Item random slope. This interaction was indeed significant ($F_{(1,34.9)}$ = 32.2, p < 0.001), suggesting the effect of Segment on the second formant was not constant across groups.

Crucially, Fig. 2 also demonstrates that ELoF's productions of the acoustically similar vowel, /y/, largely overlap those of the French control group, while /u/ is produced with substantially higher F2 values on average and with greater variation. Pairwise comparisons for this interaction confirm this: ELoF's /y/ is not significantly backer than NFS' /y/ (β = −0.33, SE = 0.15, t = −2.24, p > 0.1), while ELoF's /u/ is much less retracted than the native speaker /u/ (β = 1.29, SE = 0.23, t = 5.55, p < 0.001). These findings indicate that the acoustic variation in ELoF's /y/ productions is most similar to that of native French speakers.[5]

2.5. Discussion

In this section, I investigated which phoneme of the French /y–/u/ contrast is acoustically most dissimilar to the English /u/ vowel for L1 SSBE speakers and whether this dissimilar

[3] Further analyses also revealed that GHOUL tokens were significantly more retracted than GOOSE tokens (β = −1.88, SE = 0.13, t = −15.02, p <.001), as expected.

[4] Further analyses of vowel duration (controlling for speaking rate) revealed a significant difference between English /u/ and both of the French vowels. However, given that this effect was similar for both French vowels, it does not affect our claim of which French vowel is acoustically most similar to the English /u/ of these learners.

[5] Preliminary analyses of the L2 data also revealed that from F1, F2, F3 and durational information, F2 is the only dimension ELoF used to contrast /y/ and /u/.

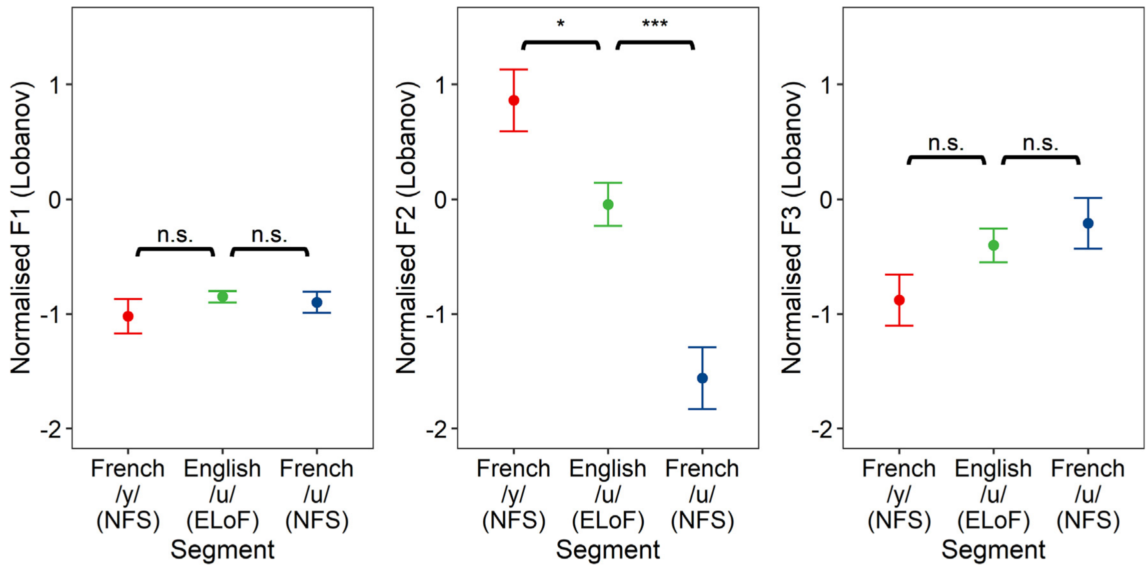

Fig. 1. Cross-linguistic acoustic comparisons of SSBE and French. *Estimated marginal means extracted and plotted from LMER models with pairwise comparisons using the kenward-roger DF and Tukey correction method* (Lenth et al., 2020).

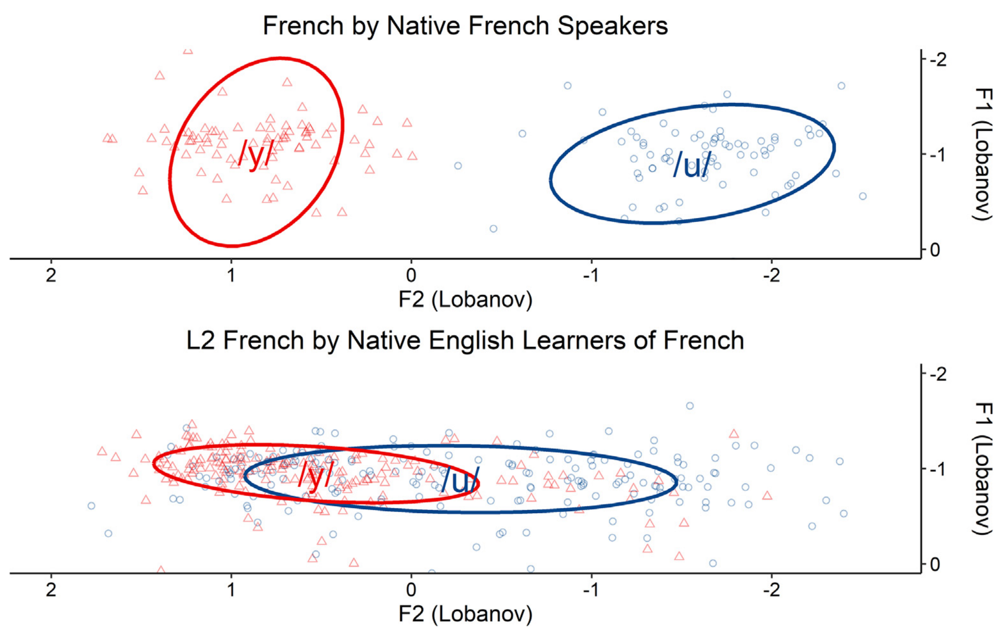

Fig. 2. Production of high rounded French vowels by NFS and ELoF. *Upper vowel space plots of normalised F1 and F2 for /y/ and /u/ by NFS (row 1) and ELoF (row 2). Ellipses assume a multivariate normal distribution and are set at 50% of the distribution with mean values marked centrally by vowel labels.*

French vowel is also produced in a more nativelike manner than the acoustically similar counterpart.

Due to GOOSE-fronting in SSBE across multiple phonetic environments, it was hypothesised that the French back vowel was most acoustically dissimilar to the English equivalent:

- *H1: Compared to native French /y/, native French /u/ is more acoustically dissimilar to ELoF's L1 English /u/ vowel.*

Results support H1, particularly along the dimension of F2 where the French vowel /u/ is most dissimilar to the high rounded English vowel. This suggests that /y/ is also most phonetically similar to the L1 English counterpart for these speakers (c.f. Flege, 1987).

Although the French carrier phrase may have induced a fronter context than the English equivalent (see Section 2.2.2), this is unlikely to affect which French vowel is most acoustically similar. Indeed, a slightly less fronted /y/ would mean that it overlaps English /u/ to a greater extent, while a more retracted French /u/ would only result in it being even more acoustically dissimilar to the English counterpart.

These results are in contrast to perceptual assimilation studies involving L1 American English (AmE) speakers in which French /u/ is mapped more consistently to the English counterpart than is French /y/ (e.g. Levy, 2009a; Strange, Levy, & Law, 2009). This is not unexpected, however, given that L1 American English speakers are less likely to have such a fronted L1 counterpart vowel, as is generally confirmed by acoustic data (e.g. Strange et al., 2007).

Due to the likelihood of ELoF having a fronted English /u/ vowel, it was hypothesised that their production of the dissimilar French /u/ vowel was most similar to native French speaker norms:

- *H2: ELoF's L2 French /u/ productions are acoustically more nativelike than their L2 French /y/ productions.*

Contrary to H2, however, this experiment reveals that ELoF's productions of /y/ lexical items are acoustically more nativelike than those containing /u/.

Although this was unexpected given that perceived phonetic dissimilarity is assumed to be facilitative in L2 speech production (Flege & Bohn, 2021; Flege, 1995, 2007), it is not uncommon for the most acoustically similar vowel cross-linguistically to be most target-like in production. For example, although in Flege (1987) /y/ is deemed perceptually "new", the acoustic distance reported for F2 is slightly smaller between English /u/ and French /y/ (416 Hz) than between English /u/ and French /u/ (450Hz), echoing the results of the present study. Similarly, Lang and Davidson (2019) find that native American English speakers living in Paris produce /y/ in more nativelike fashion. Crucially, the English /u/ of the participants was not substantially different to the /y/ of native French speakers, while compared to the native French /u/, their English /u/ had a significantly higher second formant. Presumably then, the minimal change needed from L1 articulators in order to approximate the acoustic quality of /y/ facilitates the learning task for this vowel.

This is further supported by Levy and Law (2010) in which L1 American English participants produce /y/ more accurately in alveolar contexts due to GOOSE-fronting in this phonetic environment in their L1 variety. Given the prevalence of GOOSE-fronting in both coronal and non-coronal contexts in SSBE (Holmes-Elliott, 2020), the consistent advantage for /y/ in the present results could also be explained by a high second formant for their English /u/ vowel across multiple phonetic environments.

Finally, ELoF may also produce /y/ in a more nativelike manner due to the extent of vowel dispersion in English compared to French. Iverson and Evans (2007, 2009), for example, find that speakers of languages with larger sound inventories are more likely to successfully acquire new sound systems because L2 sounds can be processed and produced through phonetic categories already present in the L1. Although English and French are relatively similar in their respective vowel densities (Meunier et al., 2003), the point stands that for /y/ acquisition, the greater extent of phonetic overlap in the high front region compared to the high back region is likely to be facilitative.

3. Processing

3.1. Hypotheses

In this section, we turn to the performance of the same learners in a processing task involving the same French /y/–/u/ lexical items. The first line of enquiry aims to establish whether the phoneme which was pronounced in the most nativelike manner in the production task is also processed more easily in a Visual World eye-tracking paradigm:

- *H3: ELoF's identification of lexical items containing /y/ is more accurate than /u/ items and eye-tracking data also indicate more successful online processing for /y/ stimuli.*

Secondly, the relationship between perception and production is analysed in terms of interspeaker/listener variation. If such a relationship exists, a learner with better processing of both French vowels may be more likely to produce both a high F2 for /y/ and a low F2 for /u/:

- *H4: L2 listeners who identify both /y/ and /u/ more easily in the processing task also produce a more substantial /y/–/u/ contrast.*

3.2. Methodology

3.2.1. Participants

Participants were the same 23 SSBE learners of French (ELoF) and 10 native French speakers (NFS) described in Section 2.2.1.

3.2.2. Procedure

A native French speaker from Paris, (male, 26 years old) was recorded uttering the French lexical items from a word list (see Table 2, Appendices). Distractor words for the participants were also elicited from the Parisian speaker and a carrier phrase for the experiment was recorded three times: *Cliquez sur l'option* (Click on the option). The first two served as practice while the third recording was imported for the experiment and inserted 50 ms before each lexical item to form an entire trial stimulus, e.g. *Cliquez sur l'option*... *Vous* (Click on the option... You).[6]

The task was designed on Experiment Builder (SR Research, 2011) and compatible with the Eyelink1000Plus eye-tracker, as used by this study. At the start of the experiment, a 13-point calibration task was completed to ensure precision in tracking. In each trial, participants would see a blank screen followed by a central fixation cross. This allowed the eye-tracker to correct for drift and for participants to self-start the trial by fixating on the cross. Throughout the experiment, the chin of each participant was rested on a mount that was positioned opposite a 24-inch CRT monitor and their dominant eye was tracked.[7]

Four lexical items were presented in the corners of the screen: the target, e.g. *Vous (you)*, the phonological competitor, e.g. *Vue (sight)*, and two unrelated distractors, e.g. *Bras* and *Lire (arm* and *reading*, respectively). This follows the Huettig and McQueen (2007) Visual World text design, allowing a greater choice of stimuli by overcoming the concern that lexical items are too abstract to represent pictorially. Although the effect of orthography is an important consideration in L2 processing (e.g. Escudero, Hayes-Harb, & Mitterer, 2008),

[6] This carrier phrase was adapted from previous research (e.g. McQueen and Viebahn, 2007) and ensured that participants were accustomed to the moment at which the target lexical item would be played.

[7] Participants' dominant eye was determined in a warm-up task. Each individual focused both eyes on an object in the room around 3 metres away through a gap between their hands. They were then instructed to shut one eye. If the object was no longer visible, the shut eye was said to be their dominant eye, while if the object remained central in the gap, this was deemed to be their sub-dominant eye (and the opposite eye was then closed instead).

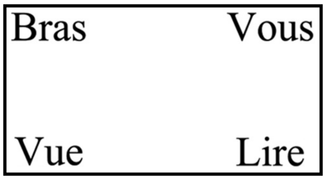

Fig. 3. Eye-tracking trial for stimulus *Vous* [vu] (you). *Experimental trial from the Visual World eye-tracking perception task based on* Huettig and McQueen (2007).

the consistent mapping between the graphemes <u> and /y/, and <ou> and /u/ in French helps to mitigate biases towards any one vowel in particular. Fig. 3 demonstrates an experimental trial for the stimulus *Vous* /vu/ (you).

Subsequently, participants would hear a stimulus phrase over Betron HD1000 headphones. The audio commenced 500 ms after the visual options were presented, with stimulus word onset occurring 1065 ms later. This allowed time for pre-scanning (c.f. McQueen & Viebahn, 2007), but in case the target had not been pre-scanned directly before the critical window, a variable labelled 'Target Fixated Pre-Stimulus' was also coded post-hoc with a binary Yes/No response and tested in statistical models.

The same four lexical options appeared for four separate trials with each item being the target only once. Furthermore, for each set of four (and, therefore, for both segments), each response (Target, Competitor, Distractor 1 and Distractor 2) appeared an equal number of times in each position to counter fixation biases towards a particular corner. The order of trials was pseudo-randomised such that stimuli of the same segment or same minimal pair were not presented consecutively.[8]

Although lexical items were considered to be commonly used in foreign language classrooms, word frequencies of all items were recorded (Cobb, 2020) (Appendices, Table 2) and tested in statistical analyses. As reported earlier, individuals' familiarity with these items was also checked post-experiment and no items needed to be excluded as a result.

3.2.3. Data analysis

The onset of each vowel in the stimuli was established using the waveform and spectrogram in *Praat* (Boersma & Weenink, 2020) and the variable Time was rescaled such that 0 was equal to vowel onset. For stimuli that had stops preceding the vowel, this was labelled as the zero crossing preceding the first reliable glottal pulsation after the burst; for a preceding voiced fricative, a sudden change to a new regular patterning was used; and for a preceding lateral, the steady state of the third formant was the criterion.

The fixation data were aggregated into 25 ms time bins using functions from the package *eyetrackingR* (Dink & Ferguson, 2015). To control for any biases towards particular visual stimuli (due to word size for example), the mean proportion of looks to each interest area was calculated for the first 150 ms of each trial to act as a baseline. The difference between the fixation proportion observed in the baseline and each 25 ms time bin was then calculated (c.f. Huettig & Altmann, 2011; Orenes et al., 2016 for similar baseline correction techniques). The start of the analysis window was set at 200 ms after segment onset to allow for the delay in planning an initial saccade in response to the audio (Matin, Shao, & Boff, 1993; Salverda, Kleinschmidt, & Tanenhaus, 2014) and the end of the analysis window was limited to 1000 ms after segment onset (c.f. Beddor et al., 2018; Ito et al., 2018; McQueen & Viebahn, 2007). Where trackloss occurred, these data were excluded from analyses (0.5% of data in a 200–1000 ms window).

3.2.4. Statistical models

For all eye-tracking analyses (Models 5–10), LMER Growth Curve models were fitted with third order orthogonal polynomial time terms to model fixation proportions as a dynamic event (Mirman, Dixon, & Magnuson, 2008). A series of models were built starting with random intercepts of Participant and Item (Baayen, Davidson, & Bates, 2008; Linck & Cunnings, 2015) and maximal random slopes (Barr et al., 2013). These were reduced using the 'Step' function of the *lmerTest* package (Kuznetsova et al., 2020) as well as manual comparisons of AIC and BIC results (again, code and model outputs made publically available: https://osf.io/rv7kd/).

To test whether further factors should be controlled for, such variables were fitted as singular fixed effects in separate models, again with maximal random structures following Melnik-Leroy et al. (2021). These variables include whether the target interest area had been directly pre-scanned before the stimulus was heard in each trial (Target Fixated Pre-Stimulus), French Proficiency (Lextale scores), Word Frequency and Preceding Phonetic Environment (coded coronal or non-coronal). Such variables never improved the model fit of the fixation data in these analyses, however, and so were excluded from final models.

Finally, a mixed effects binomial logistic regression model (Model 11) was fitted to analyse ELoF's accuracy in lexical identification. 'Target Fixated Pre-Stimulus' was controlled for after a likelihood ratio test using the 'drop1' function of the *lmerTest* package (Kuznetsova et al. Brockhoff, Christensen, & Jensen, 2020) revealed its inclusion significantly improves model fit. Proficiency, Word Frequency and Preceding Environment were not significant, however, and so were not included in the final model.

All categorical variables were contrast coded using simple coding but statistics reported in the text referring to "pairwise comparisons" are within-factor post-hoc contrasts conducted through *Emmeans* (Lenth et al., 2020). These results are Tukey method corrected and use the Kenward-Roger degrees-of-freedom method (Kenward & Roger, 1997).

3.3. L2 processing results

ELoF's eye-tracking data were first inspected visually by comparing the proportion of fixations to the target interest area, the phonological competitor, and the two distractors over time (Fig. 4).

A Growth Curve statistical model (Model 5) compared the target fixation proportions and those of the competitor from

[8] Other stimuli included in the design, but not reported here, include those of 12 minimal pairs contrasting voiced/voiceless French stops and 4 pairs of items contrasting /s/ and /ʃ/. As above, two unrelated distractor items were included per pair, constituting a total of 48 experimental trials and 48 distractors, along with a further quadrant of distractor stimuli for training purposes.

200-1000 ms after segment onset for /u/ stimuli (Fig. 4, right-hand panel). It revealed that the target is not fixated substantially more than the phonological competitor ($F_{(1,24.0)} = 0.01$, $p > 0.1$). For /y/ stimuli, however (Fig. 4, left-hand panel), the target is fixated at a consistently higher rate than the competitor ($\beta = 0.19$, SE = 0.09, $t = 2.08$, $p = .05$), according to pairwise comparisons (Model 6).[9] These results offer the first indication that ELoF may process the front vowel more easily than the back vowel.

However, the processing advantage for /y/ appears to diminish over time: the difference between the target and competitor fixation proportions decreases from 600 to 1000 ms, while for /u/ this difference increases as seen in Fig. 4. A second set of LMERs uses the difference between the competitor and target proportions as the dependent variable and to analyse its change over time, the 200–1000 ms window was divided into four smaller windows. Results confirm that the difference between ELoF's target fixation proportions and the phonological competitor fixation proportions is only greater for /y/ initially: in a 200–400 ms window (Model 7) pairwise comparisons suggest the difference was greater for /y/ items ($\beta = 0.22$, SE = 0.09, $t = 2.58$, $p < 0.05$) and to lesser extent in a 400–600 ms window ($\beta = 0.26$, SE = 0.13, $t = 1.98$, $p < 0.1$, Model 8), while no effect of Segment was found in both a 600 ms-800 ms (Model 9) window nor a 800–1000 ms (Model 10) window ($F_{(1,25.6)} = 1.76$, $p > 0.1$ and $F_{(1,29.8)} = 0.41$, $p > 0.1$, respectively), demonstrating the /y/ advantage is short-lived.[10] Further evidence for /y/ being processed more successfully in the initial stages of the listening task is the significant interaction between Segment and the linear time term in Model 7 ($F_{(1,2875.1)} = 4.2$, $p < 0.05$), indicating that in the 200–400 ms window, the difference between target and competitor fixations increases more quickly for /y/ than /u/ ($\beta = 0.16$, SE = 0.08, $t = 2.05$, $p < 0.05$). This suggests faster identification of the target and quicker rejection of the competitor.

Finally, ELoF's accuracy in their click response (lexical identification) revealed little advantage for /y/ items over /u/ items as demonstrated by Fig. 5.

A mixed effects binomial logistic regression model (Model 11) analysing these accuracy scores statistically revealed no significant effect of Segment ($\chi^2(1) = 0.89$, $p > 0.1$), indicating that ELoF did not identify either /y/ or /u/ lexical items more successfully by the end of the trial.[11]

Together, these results suggest that ELoF's processing of their most nativelike vowel in L2 production, French /y/, is also more successful than for /u/ initially, but that in later stages of processing and ultimate lexical identification, both /y/ and /u/ are confused to a similar extent.

In the following section, I explore the relationship between ELoF's L2 processing and their production of these vowels in further detail by turning to the individual variation exhibited in both modalities.

3.4. L2 processing-production relationship

3.4.1. Statistical models

The following analyses examine whether individuals with more successful processing of both sounds also tend to realise more of a distinction between phonemes in production. Firstly, the difference between the target fixation proportions and competitor fixation proportions from the eye-tracking task was mean-summarised per speaker and per vowel to provide individual processing scores. Given that the level of processing difficulty changes for each segment over time (see Models 7–10 reported earlier), the difference in fixation proportions between the two interest areas was calculated for all four time windows analysed in Section 3.3 (200–400 ms, 400–600 ms, 600–800 ms, and 800–1000 ms). Secondly, each learner's mean lexical identification accuracy for each vowel was calculated.

To test these processing results as predictors of ELoF's production data, the mean second formant was calculated for each learner's productions of both segments and used as the dependent variable. Interactions between Segment (labelled /u/, /y/, and contrast coded −0.5, +0.5) and each of the five sets of processing scores (4 × time windows and 1 × accuracy scores) were fitted in separate linear regression models. No random intercepts were found to improve model fit and while French proficiency was tested as a fixed factor, this was not found to be significant and so was removed from final models.

3.4.2. L2 processing-production results

Model 12 examines the effect of ELoF's 200–400 processing scores on their mean production values, but does not reveal a significant interaction between the processing results and Segment ($F_{(1,42)} = 1.98$, $p > 0.1$). This indicates that individuals who look to the target substantially more than the competitor in the earliest stages of processing for both /y/ and /u/ stimuli are not necessarily those who produce a fronter /y/ and a backer /u/ in French.

However, Models 13 and 14 which test the fixation means in the 400–600 ms window and the 600–800 ms window as a predictor, *do* reveal interactions with the factor Segment ($F_{(1,42)} = 7.31$, $p < 0.01$; and $F_{(1,42)} = 3.83$, $p < 0.1$, respectively). The first of these is visualised in Fig. 6 and demonstrates that individuals who are focusing on the target more than the competitor in the 400–600 ms region for audio stimuli containing both /y/ and /u/ are more likely to produce /y/ with higher F2 values and /u/ with lower F2 values.

This relationship fades by the very end of the critical window, however; the difference in target-competitor fixation means for the 800–1000 ms window (Model 15) does not interact with Segment ($F_{(1,42)} = 0.65$, $p > 0.1$).

Finally, Model 16 reveals that ELoF's identification accuracy scores in perception also interact with Segment significantly ($F_{(1,42)} = 15.66$, $p < 0.001$). The effect is visualised in Fig. 7 and demonstrates that learners who identify both /y/ and /u/ lexical items more successfully are also more likely to produce /y/ and /u/ more distinctly in terms of F2.

Overall, these results indicate a relationship between delayed measures of L2 processing and the production quality of both vowels. However, ELoF listeners with processing

[9] Modelling of the Native French Speaker eye-tracking data confirmed that the target was fixated significantly more than the competitor for both /u/ and /y/ stimuli in the same 200-1000ms window, as expected.

[10] Further modelling of the Native French Speaker eye-tracking data revealed no effect of Segment in any of these 4 time windows, as expected.

[11] Native speaker data were not analysed statistically given the 100% accuracy rates for both segments.

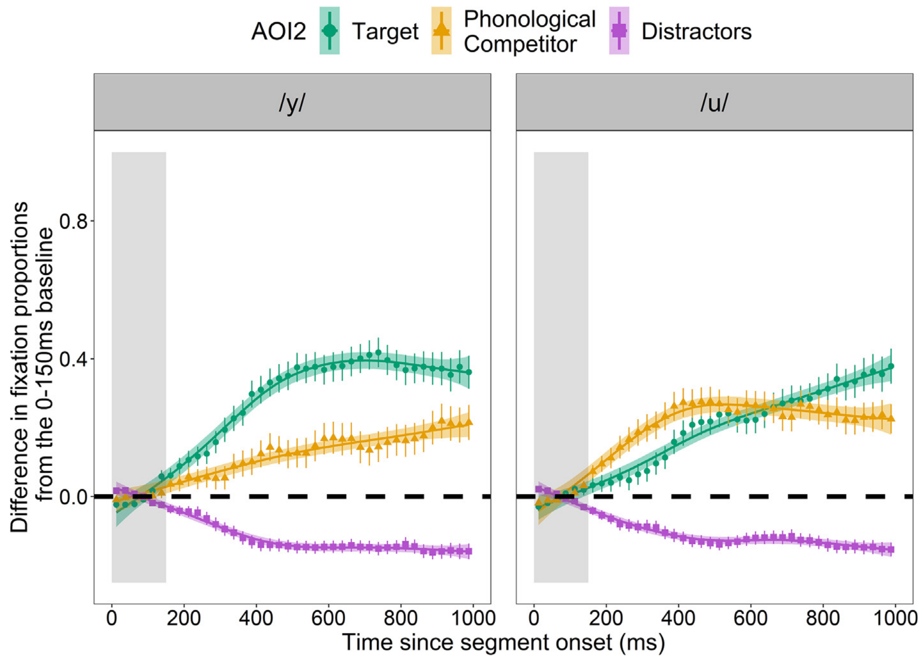

Fig. 4. ELoF's processing of /y/ and /u/ stimuli. *Mean fixation proportion for each Area of Interest (AOI) minus the mean fixation proportion for the 0–150 ms baseline.*

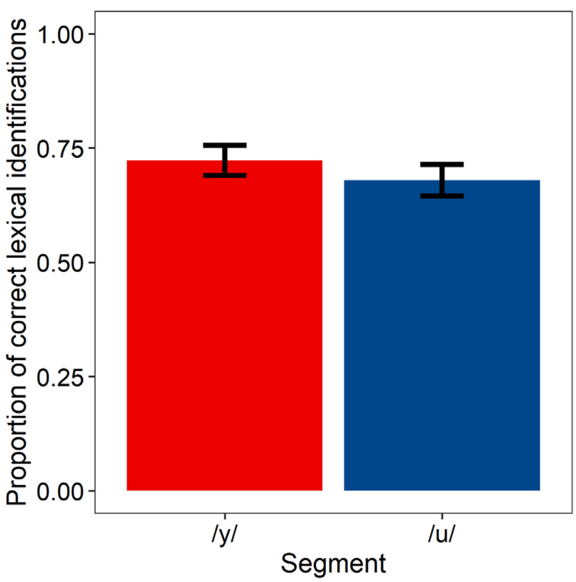

Fig. 5. ELoF's mean proportion of correct lexical identifications. *ELoF's mean accuracy when identifying French words containing /y/ and /u/ (error bars = ±SE).*

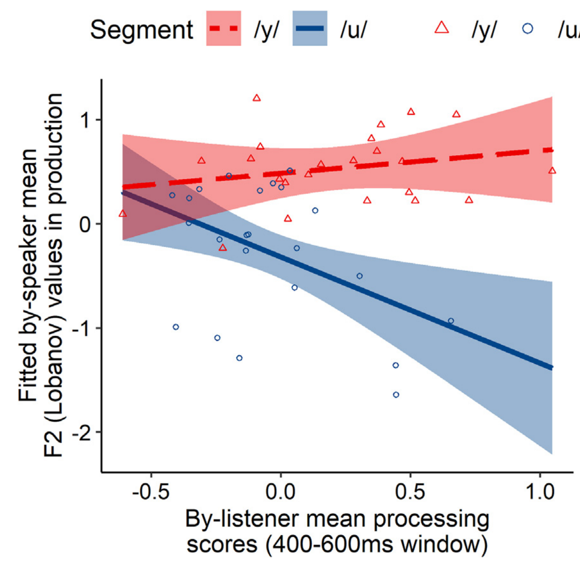

Fig. 6. Effect of ELoF's L2 processing on the acoustics of their L2 productions. *Fitted interaction between the difference in proportion of Target fixations and Competitor fixations (mean-summarised per learner and vowel) and Segment, as a predictor of each learner's mean F2 (Lobanov) in production. NB: scatterplot = raw values.*

measures most indicative of successful early-window processing for both vowels do not appear to realise the contrast any more distinctively in production.

3.5. Discussion

This experiment analyses the relationship between L2 processing and L2 production in two respects. Firstly, ELoF's processing of French /y/ and /u/ is tested to determine whether the front vowel – the phoneme found to be most nativelike in ELoF's L2 French production – is also identified more easily in a Visual World eye-tracking task:

- *H3: ELoF's identification of lexical items containing /y/ is more accurate than /u/ items and eye-tracking data also indicate more successful online processing for /y/ stimuli.*

The data offer partial support for H3: when ELoF were played lexical items containing /y/, the learners looked to the target significantly more than the competitor in the time window from 200 to 1000 ms after vowel onset, while in /u/ trials this was not the case. Importantly, the advantage for /y/ items was found to exist primarily in the first half of this processing window (from 200 ms to 400 ms after vowel onset and to a lesser extent from 400 ms to 600 ms). In these windows, the difference between target fixation proportions and competitor

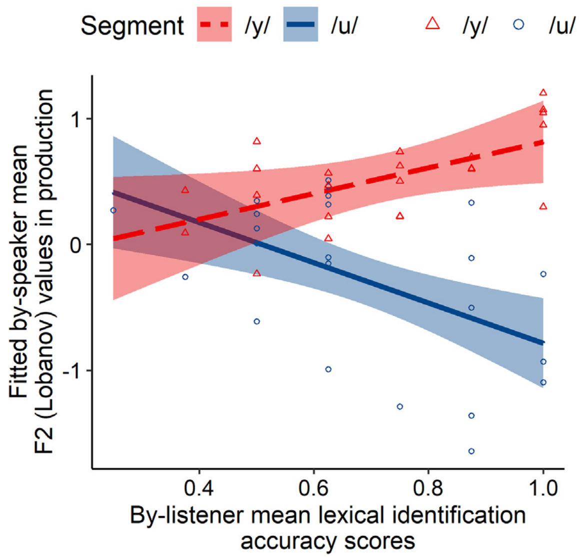

Fig. 7. Effect of ELoF's L2 lexical identification accuracy on the acoustics of their L2 productions. *Fitted interaction between lexical accuracy scores (mean-summarised per learner and vowel) and Segment, as a predictor of each learner's mean F2 (Lobanov) in production. NB: scatterplot = raw values.*

fixation proportions was larger for /y/ stimuli than /u/, and in a 200–400 ms window, this difference increased more quickly for /y/ compared to /u/, suggesting faster stimulus identification.

In opposition to H3, however, later processing windows of 600–800 ms and 800–1000 ms did not reveal /y/ items to be processed more easily. Furthermore, lexical identification accuracy scores (click response) did not suggest that ELoF have significantly higher rates for /y/ items compared to /u/ items.

The finding that ELoF process /y/ more easily, at least initially, offers support towards a perception-production relationship and echoes previous research which finds that sounds posing most challenge perceptually are also less target-like in production (e.g. Evans & Alshangiti, 2018). This is likely due, in part, to the acoustic similarity between French /y/ and the English /u/ of SSBE across many phonetic environments, as discussed for the production results. Indeed, several other eye-tracking studies have found that greater levels of acoustic similarity between the L2 sound in question and the nearest L1 sound correlate with more successful L2 processing (Cutler, Weber, & Otake, 2006; Escudero, Hayes-Harb, & Mitterer, 2008; Weber & Cutler, 2004). In contrast to these studies, however, the present results indicate that both /y/ and /u/ stimuli induce a similar level of competition between the target and the competitor at later processing stages. Why then is it only initially that /y/ is processed more easily than /u/?

Early-window processing is known for lower-level cognitive processes such as perception of phonetic detail; learners are sensitive to the acoustic detail of the input but are more likely to engage phonological knowledge, such as how this detail contributes to a phonemic distinction, at later stages (Escudero, 2009, p. 154; McQueen, 2005, p. 265). As such, although ELoF appear to be more successful at classing stimuli with a high F2 as a good match for the phonetic detail of /y/, the mapping between a high F2 and the phonological status of their English /u/ as a high rounded back vowel may become more prevalent as processing continues, promoting in turn French /u/ as an acceptable candidate. Crucially these findings demonstrate that acoustic similarity is not entirely predictive of the difficulty in L2 processing (c.f. Escudero, Simon, & Mitterer, 2012), otherwise /y/ items would be easier to process throughout the trial, rather than only initially. Instead, results align with other L2 perceptual research in which both phonetic and phonological representations appear to play a role.

For example, Chen et al. (2020) demonstrate that Mandarin learners of Thai consistently map the Thai steady tone to their L1 phonological category but remain sensitive to the cross-linguistic differences in phonetic detail, as shown by low goodness-of-fit ratings. If both phonetic and phonological representations affect L2 processing in this respect, the relationship between production and processing is unlikely to remain constant across different stages of processing. The present results support this claim: /y/ is perceived more successfully than /u/ only in initial stages of low-level processing.

The second line of enquiry in this experiment analyses the perception-production relationship using a more phonological production measure:

- *H4: L2 listeners who identify both /y/ and /u/ more easily in the processing task also produce a more substantial /y/–/u/ contrast.*

Individual learners' processing scores were calculated from their eye-tracking data for both /y/ and /u/ trials and tested as a predictor of production scores calculated from the second formant of their productions. Offline click-response accuracy scores in the processing task were also tested as a fixed effect.

Results reveal that individuals with better online processing for both vowels are also more likely to mark a greater distinction between /y/ and /u/ realisations in production. However, this effect was not observed at all stages of processing: mean fixations from the earliest window (200–400 ms after segment onset) were not significant; only those for the 400–600 ms window and the 600–800 ms window, as well as the lexical accuracy scores were found to determine the acoustics of ELoF's L2 productions.

These results once more emphasise that the different levels of linguistic representation tapped by perception and production measures are likely to modulate the strength of any relationship between the two modalities. Indeed, the learners who were most successful in early phonetic processing of the French vowels in the present study were not necessarily those producing the most distinct contrast in production, but participants with more successful phonological processing and those with higher lexical identification rates *did* produce a higher F2 for /y/ and a lower F2 for /u/.

One exception, however, is the finding that better processing scores in the latest 800–1000 ms window did not correspond to a greater amount of contrast maintained in production. Nevertheless, this is likely due to these processing scores being less reliable by this stage: in the 800–1000 ms window, the proportion of target looks may start to decrease if identification has already been successfully made, diminishing the difference between the proportion of competitor looks.

As such, a smaller difference between target and competitor fixation proportions in this window may not indicate processing difficulty as straight-forwardly as in earlier windows.

In summary, the results of this section support previous research which has found a relationship between perception and production in L2 speech learning (e.g. Evans & Alshangiti, 2018; Flege, MacKa,y & Meador, 1999). However, the present study demonstrates that combined with the many methodological considerations that modulate the relationship between these two modalities (see Nagle & Baese-Berk, 2021), one factor which must not be overlooked is the level of linguistic representation targeted in both processing and production. Indeed, results suggested that the two modalities corresponded within but not between levels, indicating that both phonetic and phonological aspects of L2 speech learning may be required to fully understand the relationship between L2 processing and L2 production. The extent to which such a notion is compatible with current L2 speech models is now explored.

4. General discussion

This research investigated both the production and processing of the novel /y/–/u/ contrast by L1 English learners of French (ELoF). The perception-production relationship was analysed in two respects: firstly I tested whether the same segment within the contrast was both produced in more nativelike fashion and processed more easily than its counterpart. Results revealed that French /y/ – the most acoustically similar vowel to ELoF's fronted SSBE /u/ vowel – is both produced in more nativelike fashion by ELoF and processed more easily than French /u/, particularly at initial stages of processing when the vowel was first heard.

Secondly, the perception-production relationship was investigated phonologically by looking at variation across individuals. Results suggested that participants with higher processing scores for both vowels at more delayed stages of processing were more likely to mark a greater acoustic distinction in their production of this contrast. Similarly, participants who identify both /y/ and /u/ lexical items more accurately via click response were also more likely to produce /y/ with higher F2 values and /u/ with lower F2 values. No such perception-production relationship was found for fixation data extracted from the very initial stages of processing, however.

These findings have broader implication for current L2 speech models. First and foremost, they support a relationship between L2 processing and L2 production as proposed by the *SLM-r* (Flege & Bohn, 2021), & arguably assumed although not explicitly stated by *PAM-L2* (Best & Tyler, 2007) and *L2LP* (Escudero, 2005; van Leussen & Escudero, 2015).

In particular, participants with better processing of the contrast are also more likely to produce an acoustic distinction, a result which is compatible with the co-evolution hypothesis of the *SLM-r* (Flege & Bohn, 2021, p. 29). That is, results do not reveal a clear dichotomous relationship, e.g. that bad processors are bad producers (c.f. Flege, 1995, 2007), but rather, that the link between L2 processing and producing is gradient in nature and is only observed when averaging over a substantial amount of variation. To provide one concrete example, three learners in the present sample who are actually more likely to look at the competitor when hearing a stimulus containing /u/ than the target in a 400–600 ms window (see bottom left of Fig. 6), are in the top 5 learners for producing /u/ with a low F2. Such findings counter the claim that successful L2 phonetic processing needs to be in place before other elements of L2 speech, such as production, can progress Elvin, Williams & Escudero, 2016, p. 8). Similarly, these results highlight the disparity between learning L2 sounds and L1 speech learning given that in the latter, perceptual targets are generally assumed to anticipate production (e.g. Eimas et al., 1971).

Instead, this correspondence between modalities lends weight to the notion that a "bi-directional connection exists between perception and production" in L2 speech learning (Flege & Bohn, 2021, p. 30), a conclusion also supported by L2 pedagogic research. For example, perceptual training interventions have been shown to translate into production gains (e.g. Bradlow et al., 1997; Inceoglu, 2016) and training in production has also been shown to facilitate perceptual processing (e.g. Carlet & De Souza, 2018; Kissling, 2015).

Nevertheless, one difficulty that does appear to be posed for the *SLM-r* is that this model assumes L2 speech acquisition is primarily a *phonetic* learning task because both native and foreign sounds are stored in "a common phonetic space" (Flege & Bohn, 2021, p. 42). In contrast, the present results demonstrate that a relationship between the two modalities is also modulated by the extent to which the *phonological* relevance of phonetic detail is encoded. As such, models which assume both phonetic categories and phonological representations in L2 speech learning appear better supported by the present findings.

One such model is *PAM-L2* (Best & Tyler, 2007). Although primarily concerned with testing the difficulty level in perceiving contrasts and therefore more phonological in nature than the *SLM-r*, *PAM-L2* predicts that language learners will remain sensitive to within-category phonetic variation (Best & Tyler, 2007; Best, 2015; Tyler, 2021). As discussed earlier, even when an L2 sound is consistently assimilated to an L1 phonological category, language learners may yet assign low "goodness-of-fit" ratings, demonstrating their sensitivity to phonetic detail (e.g. Chen, Best, & Antoniou, 2020). In this respect, the high level of competition between the vowels in L2 processing observed in the present study can be explained through multiple levels: French /u/ is an appropriate mapping for the phonological status of English /u/ as a back vowel, but is less acoustically similar and thus a bad phonetic fit, while the opposite is true of French /y/.

While the theoretical underpinnings of *PAM-L2* are appealing in this respect, it is unclear how, precisely, the model can be applied to production data, and although a relationship between L2 processing and the patterns of *native* production is assumed, without specific hypotheses for the L2 perception-production relationship, the implications of the present results for the model is limited.

Similarly, the *L2LP* (Escudero, 2005; van Leussen & Escudero, 2015) focuses on the perception domain. Nevertheless, the present results do offer a certain amount of support to the model's central claims. Firstly, this study suggests that acoustic similarity between the L1 and the L2 facilitates L2 processing, echoing previous *L2LP* research (Elvin et al., 2021; Escudero, Hayes-Harb, & Mitterer, 2008; Escudero, Simon, &

Mitterer, 2012). Such a claim may even extend to production given that ELoF's /y/ acoustics are most nativelike, an argument also made in Levy and Law (2010) after taking into account the effect of preceding phonetic environment. These results are somewhat in contrast to the *SLM-r*, however, where perceived phonetic dissimilarity is predicted to facilitate learning (Flege & Bohn, 2021). Nevertheless, perceptual assimilation measures would be needed to confirm that /y/ is assimilated to ELoF's fronted /u/ vowel in English more consistently than is the case for French /u/.

Secondly, the *L2LP* engages with multiple levels of representation in L2 processing, which increases its relevance for the present results. The model assumes that when encountering L2 input, acoustic cues are matched to language-specific phonetic knowledge which is then converted into possible phonological representations (van Leussen & Escudero, 2015, p. 4). From this level, appropriate lexical representations are recognised and the target selected through the optimal path (*ibid*). The present research demonstrates the benefits of treating the sub-lexical stages of processing for phonetic detail as separate from phonological and lexical stages by offering evidence that the perception-production relationship is not constant from the onset of phonetic processing to the ultimate offline lexical response. Indeed, measures of production which are more phonetic in nature correspond to more phonetic stages of processing: /y/ is produced in most nativelike fashion and is perceived more easily in early stages of low-level processing, but not higher levels of phonological and lexical processing. Furthermore, a relationship is observed at a phonological level: individuals who produce a more distinct contrast between /y/ and /u/ lexical items are also more likely to process both vowels with greater success in later processing windows. If the relationship is analysed between levels, however, it becomes clear that ELoF's success in processing the phonetic detail of both vowels in the earliest time window does not predict how phonemically distinct the contrast is in production.

Overall, the relationship between processing and production observed in the present study is encouraging for the premises of these three speech models. However, although current speech models provide a necessary grounding for both processing and production research, few explicit hypotheses are made regarding the relationship between the two domains. While the *SLM-r* is most unequivocal in its claims regarding this relationship, the model assumes a phonetic level of segmental learning. In contrast, the present study suggests that by considering both phonetic and phonological levels in combination, a more comprehensive account of the relationship between L2 processing and L2 production may be provided.

5. Conclusions

This research analyses the perceptual processing and production of the novel phonemic contrast /y/–/u/ by L1 SSBE university students of French. Results support a relationship between L2 processing and L2 production in two respects. Firstly, the most nativelike phoneme in production was also the easiest phoneme to identify in the earliest processing stages of the eye-tracking task. Secondly, learners who were successful in processing both vowels at later, phonological stages of the eye-tracking task were also more likely to mark a greater acoustic distinction between the vowels in production. These results support the claim that the L2 perception-production relationship is strongest within, but not between, linguistic levels (Melnik-Leroy et al., 2021).

CRediT authorship contribution statement

James Turner: Conceptualization, Methodology, Software, Validation, Data curation, Visualization, Investigation.

Acknowledgements

This research was guided by extensive feedback from Dr. Sophie Holmes-Elliott, Dr. Jaine Beswick, Lewis Baker, Christine Turner, as well as three anonymous reviewers whom I thank unreservedly. Any errors, however, remain entirely my own.

Funding

The author gratefully acknowledges financial support from the Economic and Social Research Council (ESRC) (Grant Number: ES/P000673/1), and by extension, the South Coast Doctoral Training Partnership (SCDTP).

Declarations of interest

None.

Appendix

Table 1
English lexical items.

Variable	Item	Word Frequency (Cobb, 2020)	Variable	Item	Word Frequency (Cobb, 2020)
/u/ (GOOSE)	Goose	5 k	/u/ (GHOUL)	Ghoul	11 k
/u/ (GOOSE)	Toot	12 k	/u/ (GHOUL)	Tool	2 k
/u/ (GOOSE)	Rude	2 k	/u/ (GHOUL)	Rule	1 k
/u/ (GOOSE)	Youths	3 k	/u/ (GHOUL)	You'll	1 k
/u/ (GOOSE)	Poop	1 k	/u/ (GHOUL)	Pool	2 k
/u/ (GOOSE)	Food	1 k	/u/ (GHOUL)	Fool	2 k
/u/ (GOOSE)	Stoop	6 k	/u/ (GHOUL)	Stool	5 k
/u/ (GOOSE)	Coop	9 k	/u/ (GHOUL)	Cool	1 k

NB: 1 k = within 1000 most frequent words.

Table 2
French lexical items.

Variable	Item	Word Frequency (Cobb, 2020)	Variable	Item	Word Frequency (Cobb, 2020)
/y/	Su	1 k	/u/	Sous	1 k
/y/	Bu	2 k	/u/	Bout	1 k
/y/	Jus	7 k	/u/	Joue	1 k
/y/	Vue	1 k	/u/	Vous	1 k
/y/	Rue	1 k	/u/	Roux	6 k
/y/	Lu	1 k	/u/	Loup	5 k
/y/	Tu	1 k	/u/	Tout	1 k
/y/	Du	1 k	/u/	Doux	2 k

NB: 1 k = within 1000 most frequent words.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America, 116*(5), 3099–3107. https://doi.org/10.1121/1.1795335.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48 https://doi.org/10.18637/jss.v067.i01.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language, 94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051.

Best, C. T. (2015). Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *Phonetics-Phonology Interface: Representations and Methodologies* (pp. 3–31). Amsterdam: John Benjamins Publishing Company.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing Company.

Bigi, B., & Meunier, C. (2018). Automatic Segmentation of Spontaneous Speech / Segmentação automática da fala espontânea. *Revista De Estudos Da Linguagem, 26*(4), 1489–1530 https://doi.org/10.17851/2237-2083.26.4.1489-1530.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer [computer programme]. *Version, 6*(1), 28 https://www.praat.org.

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

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., & Perkella, J. (2011). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *Journal of Speech, Language, and Hearing Research, 54*(3), 727–739. https://doi.org/10.1044/1092-4388(2010/09-0256).

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*(1), 23–37. https://doi.org/10.5334/pb-53-1-23.

Carlet, A., & De Souza, H. K. (2018). Improving L2 pronunciation inside and outside the classroom: Perception, production and autonomous learning of L2 vowels. *Ilha do Desterro, 71*(3), 99–123. https://doi.org/10.5007/2175-8026.2018v71n3p99.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007.

Chen, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics, 83*. https://doi.org/10.1016/j.wocn.2020.101013.

Childers, D. G. (1979). *Modern spectrum analysis, IEEE Press Selected Reprint Series*. Hoboken, NJ: John Wiley & Sons Inc.

Cobb, T. (2020). Compleat Lexical Tutor [web page]. https://www.lextutor.ca/vp/comp.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second Language Speech: Theory and Practice*. Cambridge, UK: Cambridge University Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. *Second Language Research, 28*(1), 5–40. https://doi.org/10.1177/0267658311423455.

Diehl, R. L., & Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology, 1*(2), 121–144. https://doi.org/10.1207/s15326969eco0102_2.

Dink, J. W., & Ferguson, B. (2015). EyetrackingR: An R Library for Eye-tracking Data Analysis [R Package]. http://www.eyetracking-r.com.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303.

Elvin, J., Williams, D., Shaw, J. A., Best, C. T., & 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*(1). https://doi.org/10.3390/languages6010044.

Elvin, J., Williams, D., & Escudero, P. (2016). The relationship between perception and production of Brazilian Portuguese vowels in European Spanish monolinguals. *Loquens, 3*(2), 2386–2637. https://doi.org/10.3989/loquens.2016.031.

Escudero, P. (2005). Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization. *LOT Dissertation series*. 113th edn.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151–190). Berlin; New York: Mouton de Gruyter.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics, 40*(2), 280–288. https://doi.org/10.1016/j.wocn.2011.11.004.

Evans, B. G., & Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics, 68*, 15–31. https://doi.org/10.1016/j.wocn.2018.01.002.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47–65. https://doi.org/10.1016/s0095-4470(19)30537-6.

Flege, J. E. (1995). Second Language Speech Learning Theory, Findings and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In *14th International Congress of Phonetic Sciences (ICPHS)* (pp. 1273–1276).

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–381). Berlin; New York: Mouton de Gruyter.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge, UK; New York, USA; Melbourne, Australia: Cambridge University Press.

Flege, J. E., Fox, R. A., & Munro, M. J. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*(6), 3623–3641. https://doi.org/10.1121/1.409931.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*(3), 708–721. https://doi.org/10.1121/1.391257.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–2987. https://doi.org/10.1121/1.428116.

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

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics, 14*(1), 3–28. https://doi.org/10.1016/s0095-4470(19)30607-2.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 131–149. https://doi.org/10.1017/s136672899800025x.

Holmes-Elliott, S. (2020). Calibrate to innovate: Community age vectors and the real time incrementation of language change. *Language in Society, 1–34*. https://doi.org/10.1017/S0047404520000834.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing 'frog': How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology, 64*(1), 122–145. https://doi.org/10.1080/17470218.2010.481474.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics, 37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America, 122*(5), 2842. https://doi.org/10.1121/1.2783198.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866–877. https://doi.org/10.1121/1.3148196.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*, 1–17. https://doi.org/10.3389/fpsyg.2014.01246.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997. https://doi.org/10.2307/2533558.

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research, 19*(3), 254–275. https://doi.org/10.1177/1362168814541735.

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA, USA; Oxford, UK; Carlton, Victoria, Australia: Blackwell Publishing.

Lang, B., & Davidson, L. (2019). Effects of Exposure and Vowel Space Distribution on Phonetic Drift: Evidence from American English Learners of French. *Language and Speech, 62*(1), 30–60. https://doi.org/10.1177/0023830917737111.

de Leeuw, E., Stockall, L., Lazaridou-Chatzigoga, D., & Gorba Masip, C. (2019). Illusory vowels in Spanish-English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research, 1–32*. https://doi.org/10.1177/0267658319886623.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., & Jensen, S.P. (2020). lmerTest: Tests in Linear Mixed Effects Models [R Package] https://cran.r-project.org/web/packages/lmerTest.

Lenth, R., Buerkner, P., Herve, M., Love, J., Riebl, H. & Singmann, H. (2020). Emmeans [R Package] https://www.cran.r-project.org/web/packages/emmeans.

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology, 6*, 1–12. https://doi.org/10.3389/fpsyg.2015.01000.

Levy, E. S. (2009a). 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*(2), 1138–1152. https://doi.org/10.1121/1.3050256.

Levy, E. S. (2009b). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America, 126* (5), 2670–2682. https://doi.org/10.1121/1.3224715.

Levy, E. S., & Law, F. F. (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*(3), 1290–1305. https://doi.org/10.1121/1.3466879.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36*(1), 141–157. https://doi.org/10.1016/j.wocn.2007.03.001.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception and the speech code. *Psychological review, 74*(6), 431–461. https://doi.org/10.1037/h0020279.

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning, 65*(S1), 185–207. https://doi.org/10.1111/lang.12117.

Llisterri, J. (1995). Relationships between Speech Production and Speech Perception in a Second Language. *13th International Congress of Phonetic Sciences (ICPHS)*, *4*, 92–99.

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America, 49*(2B), 606–608. https://doi.org/10.1121/1.1912396.

Marx, N., & Mehlhorn, G. (2010). Pushing the positive: Encouraging phonological transfer from L2 to L3. *International Journal of Multilingualism, 7*(1), 4–18. https://doi.org/10.1080/14790710902972271.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380. https://doi.org/10.3758/BF03206780.

McQueen, J. M. (2005). Speech Perception. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 255–275). London, UK: Sage Publications.

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60* (5), 661–671. https://doi.org/10.1080/17470210601183890.

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2021). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u /–/y/ contrast. *Second Language Research*. https://doi.org/10.1177/0267658320988061.

Meunier, C., Frenck-mestre, C., Lelekov-boissard, Taissia, Meunier, C., Frenck-mestre, C., Lelekov-boissard, Taissia, Pro-, M. L. B., Meunier, C., Frenck-mestre, C., Lelekov-boissard, Taïssia, & Besnerais, M. Le (2003). Production and perception of vowels : does the density of the system play a role? *15th International Congress of Phonetic Sciences (ICPHS)*, 723–726.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006.

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

Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition, 1–26*. https://doi.org/10.1017/S0272263121000371.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on vot production and perception for french and english stops. *Journal of the International Phonetic Association, 24*(1), 1–18. https://doi.org/10.1017/S0025100300004965.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America, 113*(5), 2850–2860. https://doi.org/10.1121/1.1567280.

Nguyen, N., & Fagyal, Z. (2008). Acoustic aspects of vowel harmony in French. *Journal of Phonetics, 36*(1), 1–27. https://doi.org/10.1016/j.wocn.2007.07.001.

Orenes, I., Moxey, L., Scheepers, C., & Santamaría, C. (2016). Negation in context: Evidence from the visual world paradigm. *Quarterly Journal of Experimental Psychology, 69*(6), 1082–1092. https://doi.org/10.1080/17470218.2015.1063675.

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. *Proceedings of the Annual Conference of the International Speech Communication Association*.

Piske, Thorsten (2007). Implications of James E. Flege's research for the foreign language classroom. In O.-S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 301–314). Amsterdam: John Benjamins Publishing Company.

R Core Team (2018). *R: A language and environment for statistical computing. [computer programme]*. R Foundation for Statistical Computing.

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics, 40*(3), 491–508. https://doi.org/10.1016/j.wocn.2012.01.001.

Research, S. (2011). Experiment Builder 1.10.165 [computer programme]. https://www.sr-research.com/experiment-builder.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002.

Schmid, M. S., Gilbers, S., & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch-English bilingualism. *Second Language Research, 30*(2), 129–157. https://doi.org/10.1177/0267658313505314.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America, 122*(2), 1111–1129. https://doi.org/10.1121/1.2749716.

Strange, W. (2007). Cross-language phonetic similarity of vowels. Theoretical and methodological issues. In Bohn, O.-S., & Munro, M. J. (Eds.), Language experience in second language speech learning: In honor of James Emil Flege (pp. 35–56). Amsterdam: John Benjamins Publishing Company.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naïve American listeners. *The Journal of the Acoustical Society of America, 126*(3), 1461–1476. https://doi.org/10.1121/1.3179666.

Strycharczuk, P., & Scobbie, J. M. (2017). Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America, 142*(1), 322–331. https://doi.org/10.1121/1.4991010.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 1632–1634*. https://doi.org/10.1126/science.7777863.

Tyler, M. D. (2021). Phonetic and Phonological Influences on the Discrimination of Non-native Phones. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 157–174). Cambridge, UK; New York, USA; Melbourne, Australia: Cambridge University Press.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0.

Wells, J. C. (1982). *Accents of English 1: An introduction*. Cambridge, UK; New York, USA; Melbourne, Australia: Cambridge University Press.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3.