



HAL
open science

Phonetic convergence and vowel overlap in simultaneous bilinguals, language learners, and monolinguals

Andres Felipe Lara, Claire Pillot-Loiseau

► **To cite this version:**

Andres Felipe Lara, Claire Pillot-Loiseau. Phonetic convergence and vowel overlap in simultaneous bilinguals, language learners, and monolinguals. *Radical: A journal of Phonology*, 2023, 3, pp.238-279. hal-04374971

HAL Id: hal-04374971

<https://hal.science/hal-04374971v1>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

volume 3, 2022

**Lara, Andres Felipe
Pillot-Loiseau, Claire**

**Phonetic convergence
and vowel overlap in
simultaneous
bilinguals, language
learners, and
monolinguals**

Lara, Andres Felipe & Pillot-Loiseau, Claire. 2023. "Phonetic convergence and vowel overlap in simultaneous bilinguals, language learners, and monolinguals". *Radical: A Journal of Phonology*, 3, 238-279.

Editor: Quentin Dabouis
Reviewers: Jalal Al-Tamimi, Damien Chabanal

PHONETIC CONVERGENCE AND VOWEL OVERLAP IN SIMULTANEOUS BILINGUALS, LANGUAGE LEARNERS, AND MONOLINGUALS

LARA, ANDRES FELIPE (LABORATOIRE DE PHONETIQUE ET PHONOLOGIE, UMR 7018 CNRS
SORBONNE NOUVELLE, FRANCE)

PILLOT-LOISEAU, CLAIRE (LABORATOIRE DE PHONETIQUE ET PHONOLOGIE, UMR 7018 CNRS
SORBONNE NOUVELLE, FRANCE)

This preliminary study explores segmental phonetic convergence and vowel overlap in simultaneous bilinguals, highly proficient American late learners of French, and monolinguals. A shadowing task solicited French /i e ε y ø œ a ɔ o u/ and American English /i ɪ ε æ ɑ ɔ ʊ u/ vowels in four different phonetic contexts: isolated, /bVb/, /pVp/ and /pVɁ/ for the French vowel /œ/. F1, F2, F3 and duration were extracted and analyzed for convergence towards the target values. In addition, phonological accuracy was evaluated by measuring the space between contrastive vowel classes through the use of the Pillai Bartlett trace (Pillai 1976). This study found that bilinguals and language learners showed more signs of divergence for English than French but found more vowel merging in French than in English for both speaker profiles.

phonetic convergence, language learners, simultaneous bilinguals,
shadowing task, vowel overlap

INTRODUCTION

Variability is a phenomenon often encountered in areas of research where one must rely on the phonetic-acoustic form of an utterance. It can manifest itself at both the segmental and suprasegmental level and can vary depending on age, sex, regional dialect, and a speaker's language background (Pardo et al. 2010). Regan and Bayley (2004) postulate that variability is a critical component of a speaker's linguistic competence. They go on to say that, in order for non-native speakers to master a foreign language, they must first be able to reproduce the intrinsic variation patterns found in native speakers. The question remains: to what extent can bilinguals or language learners diverge phonetically from modeled target values produced by native speakers before phonological accuracy is affected?

This preliminary study measured phonetic divergence at the segmental level in English and in French using a version of Goldinger's shadowing paradigm (Goldinger 1998). Numerous studies on phonetic speaker adaptation have implemented shadowing tasks (Delvaux and Soquet 2007; Namy et al. 2002; Nielsen 2008; Babel 2009; Brouwer et al. 2010). The methodological paradigm in which the vowels are produced is key as "laboratory speech-shadowing tasks provide a favorable context to elicit phonetic convergence" (Pardo et al. 2017:640). Phonetic convergence describes the process by which the acoustic-phonetic form of an individual assimilates that of a person or model after recent or prolonged exposure (Babel 2010; Goldinger 1998; Pardo 2006; Sancier and Fowler 1997). Given the nature of the study, we can expect cases of complete, partial and null divergence, which can equate with, null, partial and complete convergence, respectively.

We compared the shadowed productions of simultaneous bilinguals and highly proficient American late learners of French to those of monolingual speakers of both languages. Monolingual speakers have been known to converge on both vowel formants (Delvaux and Soquet 2007; Schweitzer and Lewandowski 2014; Babel 2012) and

duration (Delvaux and Soquet 2007). However, few studies have explored convergence in non-native speakers at the segmental level. The relatively small number of studies that have been done evaluate this phenomenon through perceived phonetic convergence by native speakers (Horton and Bradlow 2011) and suprasegmental measures of amplitude envelopes and intonational contours “at the word level” (Lewandowski and Dogil 2009, Lewandowski and Jilka 2019). Nevertheless, convergence seems to be the consensus and speakers with the same native language (L1) are more likely to converge than speakers that do not share the same L1 (Horton and Bradlow 2011).

It has been demonstrated that the age of first exposure to a language bears significant weight on the degree of cross-linguistic transfer (Flege 2007:467). Studies indicate that late language learners are more likely to show transfer from the L1 to the L2 and experience difficulty perceiving and producing new sounds in the L2 (Flege 1995). The revised Speech Learning Model (SLM-r; Flege and Bohn 2021) posits that, although L2 sounds might never be native-like in late language learners, this does not translate to a hindered capacity for speech learning. Instead, variation in L2 production could be due to other factors such as the amount of input in the L2 or the interaction between phonological categories in the L1 and L2 sound systems. Similarly, Best’s Perceptual Assimilation model (PAM) posits that, contrasts, which are not present in the native language inventory, may be more difficult to discriminate in the L2 (Best 1994, 1995). Both models predict that, in the cases where one phonetic category is assimilated to another, the new sound in the L2 will be represented by the closest sound (phonetically) in the L1, effectively yielding merged categories (Flege 2007). For example: front rounded vowels /y, ø, œ/ are particularly difficult for late learners of French (Gottfried 1984), especially for native speakers of the American varieties of English. The French vowel /y/ has been largely attested as being assimilated to the vowel /u/ by Americans learning French as an L2 (Flege 1987; Flege and Hillenbrand 1984; Gottfried 1984; Levy 2009).

Simultaneous bilinguals would, in theory, outperform late language learners of French.

The fact that they have been exposed to both English and French from birth would admittedly mean that the quantity of input in French is by far greater than that of the late language learners. However, the literature suggests that bilinguals undergo a process of cross-linguistic interaction that influences perception and production in both languages. The Interactional Dual System Model (Paradis 2001) posits that simultaneous bilinguals have two separate phonological systems and that those two systems exert mutual influence on each other. The bidirectional interaction between the two systems would thus result in merged vowel categories in both languages. Although recent research (Morris 2013, 2017; Watson 2007; Simonet 2011; Mayr et al. 2019, Simonet and Amengual 2020; Strandberg et al. 2021) has turned focus to cross-linguistic transfer in simultaneous bilinguals none of them compare them to late language learners. As far as convergence is concerned, the list of studies on bilinguals is short. Mayr et al. (2019) studied monophthong vowel productions in Welsh-English bilinguals and found high degrees of convergence in their realizations. Simonet and Amengual (2020) also found cross-linguistic vowel convergence in Catalan-Spanish bilinguals.

To our knowledge, no previous acoustic study has compared the shadowed productions of late language learners and simultaneous bilinguals, let alone in French. As a consequence, most studies on vowel convergence limit themselves to socio-dialectal differences and very few document segmental features of the acoustic signal. Furthermore, only one other study has analyzed F3 values for French vowels, albeit, in Belgian French and within the scope of monolingual French productions (Delvaux and Soquet 2007). F3 is of particular interest for French vowels as French has vowels with the same tongue position but with different lip positions like in the case of the front-unrounded /i, e, ε/ and rounded vowels /y, ø, œ/ respectively. (Ladefoged and Disner 2012). Similarly, most studies in second language acquisition juxtapose language learners to either sequential bilinguals (Flege et al. 2002; Abrahamsson and Hyltenstam 2009; Yeni-Komshian et al. 2000) or heritage learners (Montrul 2010; Amengual 2019; Kang and Nagy 2016; Elias et al. 2017). This preliminary study adds to the already extensive

work on bilingualism and second language acquisition by analyzing convergence patterns and vowel contrast overlap between simultaneous French-English bilinguals and American late language learners of French.

American English (AE) /i ɪ ε æ ɑ ɔ ʊ u/ and French /i e ε y ø œ a ɔ o u/ vowels were analyzed for divergence. Our subjects shadowed vowels in four different monosyllabic phonetic contexts: /V/, /bVb/, /pVp/, and /pVɁ/ for the French vowel /œ/. F1 to F3 formants were measured for all vowels. In addition to formant values, duration was also measured. The literature suggests that vowels followed by a voiced stop are significantly longer than vowels followed by a voiceless stop; this is particularly true of English (Chomsky and Halle 1968; Mack 1982). Furthermore, the tense-lax distinction of the vowels in English carries with it an intrinsic difference in duration. Tense vowels in AE /i ɑ ɔ u/ are significantly longer than lax vowels /ɪ ε ʊ/, with the exception of /æ/ often on par with most tense vowels (Peterson and Lehiste 1960). French vowels, on the other hand, do not vary in length (Gottfried and Beddor 1998; Fourgeron and Smith 1999).

The data in this preliminary study served to evaluate accuracy in monolingual, bilingual and language learner production. Accuracy is measured not just in regard to the level of convergence in comparison to the native-speaker values for vowel formants and duration. It is also represented by whether the space between phonological value contrasts is maintained or not. Measuring vowel overlap is not a common methodological choice when it comes to evaluating pronunciation in second language acquisition (Perry et al. 2019; Mairano and Santiago 2019). That being said, a recent study done by Kelley and Tucker (2020) sees great potential in measuring vowel mergers in second language acquisition research. They even go as far as evaluating and critiquing different overlap measures ultimately recommending the Pillai Barlett (Pillai 1976) trace as being theoretically preferable to other approaches; amongst those the euclidean distance, the spectral overlap assessment metric, and the vowel overlap assessment with convex hulls metric. Consequently, we chose to use the Pillai Bartlett trace to measure vowel overlap in formantic spaces F1 – F2 and F2 – F3.

Taking the literature previously discussed into account we can expect the following:

- (1) H.0 We predict that too much variation could hinder reproduction of target values for both language learners and simultaneous bilinguals. Variation would manifest itself through vowel class mergers due to cross-linguistic transfer.
 - H.1 English-speaking language learners of French are expected to deviate in French (L2) but to converge in English (L1) while simultaneous bilinguals could either diverge or converge for both English and French. Monolinguals would converge for both languages as they share the same L1 as the target values modeled in the shadowing task.
 - H.2 English-speaking language learners of French are expected to manifest merged vowel categories in French (L2) especially for the front rounded vowels /y, ø, œ/ while simultaneous bilinguals would produce merged vowel categories with perceptually linked speech sounds in English and French.

1. METHOD

1.1. SUBJECTS AND CORPUS

F1 to F3 formants values of American English /i ɪ ε æ ɑ ɔ ʊ u/ and French /i e ε y ø œ a ɔ o u/ vowels were analyzed for 5 simultaneous bilinguals (age 27, ±4) living in France having been exposed to both languages from birth, 5 American second language learners of French (age 24, ±2) having learned the target language after the age of 13 and having lived in France for more than 3 years on average and, 10 monolinguals (age 30, ±7) having only lived either in France (5) or in the USA (5). The learners' proficiency in French was self-assessed by means of a survey filled out at the beginning of the study. All learners

were considered advanced learners as per the responses on the questionnaire. Three native speakers were solicited to create the target values both in American English (AE) and in French. One participant was used for the training test, a man and a woman were used to create the sentences for the main task. With the exception of /œ/ in French, all vowels were recorded in three different phonetic contexts: isolated, /bVb/, and /pVp/. The phonetic context was chosen in order to limit the number of occurrences and the articulatory factors that could eventually cause variation. A bilabial context reduces the influence of lingual coarticulation. An exception was made for the French front rounded vowel /œ/. In order to facilitate distinction for the contrast /ø – œ/, the target for the vowel /œ/ were recorded in isolated, /bVɤ/, and /pVɤ/ contexts (see 2). The final rhotic was added to the French vowel /œ/ in order to respect the phonotactical rules of the language. Pseudo-words were used where necessary in order to allow for homogeneity in the corpus. Ten out of the sixteen tokens used in English were pseudo-words and 12 out of the twenty tokens used in French were real words. Isolated vowels were recorded in a carrier sentence containing a reference word at the beginning and at the end. Vowels in CVC contexts were recorded without a reference word. Each vowel was repeated three times. Canonical examples of AE (Hillenbrand et al. 1995) and French (Georgeton et al. 2012) vowels were chosen whenever possible. Measurements for these vowels constitute the target values shadowed in the study.

- | | | |
|-----|----------|--|
| (2) | Isolated | I said /i/ like in bee.
J'ai dit /u/ comme dans poule. |
| | /bVC/ | I said <i>beeb</i> three times.
J'ai dit <i>boub</i> trois fois. |
| | /pVC/ | I said <i>peep</i> three times.
J'ai dit <i>poup</i> trois fois. |
| | /Cœɤ/ | J'ai dit <i>beur</i> trois fois.
J'ai dit <i>peur</i> trois fois. |

1.2. PROCEDURES

All participants signed consent forms before they were allowed to join the study. Participants also filled out a questionnaire prior to beginning the experiment. Participants that did not match the profiles defined in the study were systematically excluded. Recordings were done either in a calm environment or in the anechoic chamber at the laboratory. A Zoom H6 recorder was used along with a Shure WH30XLR cardioid headband microphone. Sennheiser headphones were used in order to avoid background noise when listening to the target values. Audacity was used both to record and to segment the utterances. Target sentences were presented to the participants using the software Psychopy (Peirce 2007). Instructions appeared on the computer screen and the participants used the spacebar to move from one task to the next. Vowels were presented in the carrier sentences as recorded. The initial reference word was omitted in the case of the isolated vowels. A training period ensured that participants were familiarized with the task. Productions recorded during the training period were not included in the study. Participants were asked to listen to the stimulus and to repeat it back into the microphone three seconds after the stimulus stopped playing. Each vowel was repeated twice in each of the different contexts, once after the man and once after the woman.

1.3. DATA ANALYSIS

Praat software (Boersma and Van Heuven 2019) was used to annotate and (semi-automatically) extract formants and duration values. Formants were extracted at midpoint in order to minimize coarticulation effects. Outliers were checked manually and were not excluded in order to better represent variability amongst speaker profiles. The study included women and men, so all formant values were normalized using the ANAE method (Labov 2006). The advantages being that it is speaker-extrinsic, it allows for F3 values, and physiologically-caused differences in formant values are factored out while

retaining variability amongst speakers. As mentioned in the introduction, F3 is an important correlate of roundness in French. We conducted a Shapiro Wilk test on all formants and duration values, which revealed abnormal distributions for English and French. Consequently, a Kruskal Wallis (Ostertagova et al. 2014) test was conducted as an alternative to the ANOVA test in order to measure divergence from target values. Pairwise comparisons using Dunn's test indicated significant differences in more detail. In addition, we extracted the Pillai-Bartlett (Pillai 1976) trace to measure the overlap between vowel classes. The value ranges from 0 to 1, with values closer to 0 indicating more overlap and values closer to 1 representing complete separation (Nycz and Hall-Lew 2013). Statistical analyses were performed with R (RStudio Team 2020). A 95% confidence interval was used in order to avoid type I errors. Error bars seen in the graphs correspond to the standard deviation; values represent the means for each “group”.

2. RESULTS

We performed statistical analyses on all groups as whole. We compared vowel formants in relation to three phonetic contexts: isolated, /bVb/, and /pVp/. In English, no significant differences were found for formants F1 and F2 in each of the different contexts F1: ($H_{(2, 867)} = 1.788$; $p < .409$), F2 : ($H_{(2, 867)} = 3.483$; $p < .175$). Significant differences were found for F3 however: ($H_{(2, 867)} = 15.38$; $p < .0005$). F3 for vowels in /pVp/ was significantly lower than F3 in both isolated vowels and vowels in the /bVb/ context. Seeing as no significant differences were found for F1 and F2, we decided to group all contexts. F3 was also grouped since English vowels can be identified based solely on F1 and F2 (Ladefoged 2006). In French, a significant difference was found for F1: ($H_{(2, 1082)} = 10.646$; $p < .005$) and F3: ($H_{(2, 1082)} = 11.834$; $p < .003$). The post-hoc test (4) showed that F1 for vowels in /pVp/ context was significantly higher than F1 in isolated vowels. F3 in /pVp/ context was significantly higher than F3 in isolated vowels and vowels in /bVb/ context. No significant differences were found for F2. Following the procedure

established for English vowels, all phonetic contexts were grouped for duration analysis. Duration for English and French vowels will be discussed at the end of each section.

2.1. ENGLISH

A Kruskal Wallis test detected significant differences between the target values and the three speaker profiles for all three vowel formants as well as duration (Table 3). F3 shows an extremely significant amount of variance. Speakers diverged the least in terms of F2.

- (3) KRUSKAL WALLIS TEST FOR THE DIFFERENCE IN FORMANTS F1 TO F3 AND DURATION FOR ENGLISH VOWELS BETWEEN TARGET VALUES AND ALL SPEAKER PROFILES.

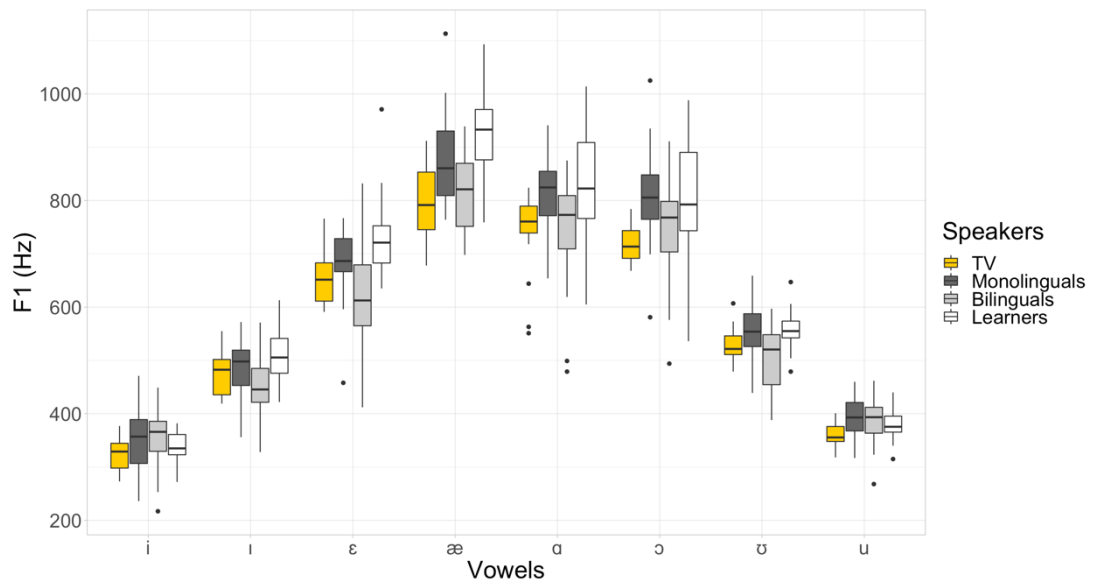
FORMANT	N	H	DF	P
F1	867	12.364	3	0.00624 **
F2	867	9.989	3	0.0187 *
F3	867	56.776	3	2.87e-12 ****
Duration	867	12.184	3	0.00678 **

Table 4 shows the results from the *post-hoc* test. Language learners diverged significantly from the target values in terms of F1. Both bilinguals and monolinguals converged for F1. Bilinguals diverged significantly for F2. Monolinguals and language learners converged towards F2 target values. All speaker profiles diverged significantly from the F3 target values.

(4) DUNN TEST COMPARING TARGET VALUES TO EACH SPEAKER PROFILE WITH BONFERRONI CORRECTION AND P ADJUSTED SIGNIFICANT VALUES

FORMANT	SPEAKERS	N1	N2	H	P ADJ
F1	Bilinguals	240	145	-0.652	0.100
	Learners	240	145	-2.843	0.026 *
	Monolinguals	240	145	-2.419	0.093
F2	Bilinguals	240	145	-2.786	0.049 *
	Learners	240	145	-0.700	0.592
	Monolinguals	240	145	-1.829	0.250
F3	Bilinguals	240	145	-7.314	1.55e-12 ****
	Learners	240	145	-5.969	1.43e-08 ****
	Monolinguals	240	145	-4.954	4.36e-06 ****
Duration	Bilinguals	240	145	2.683	0.044 *
	Learners	240	145	0.440	1.000
	Monolinguals	240	145	2.339	0.116

(5) F1 ENGLISH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).



- (6) F2 ENGLISH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).

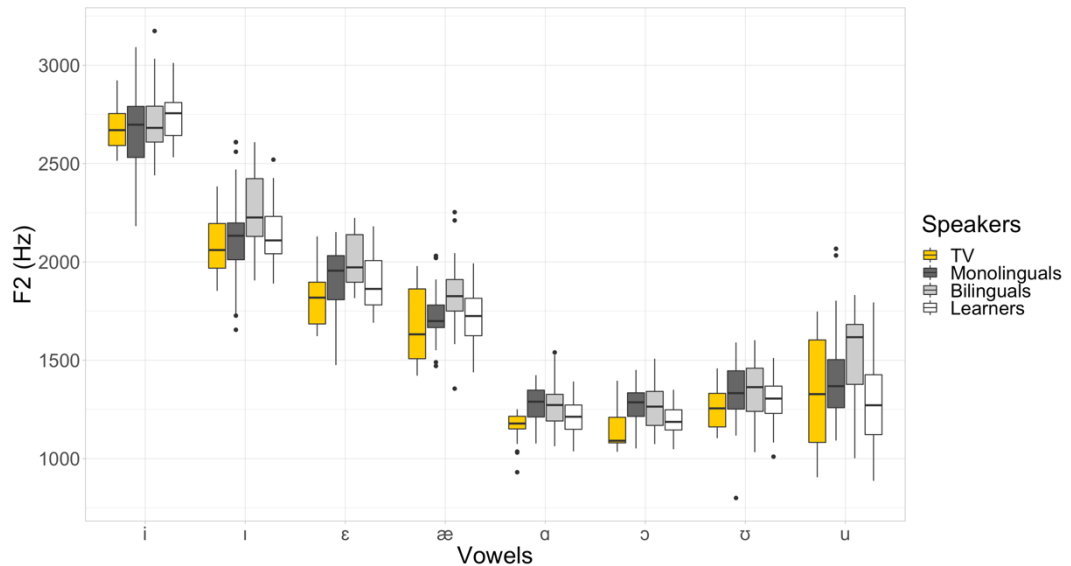
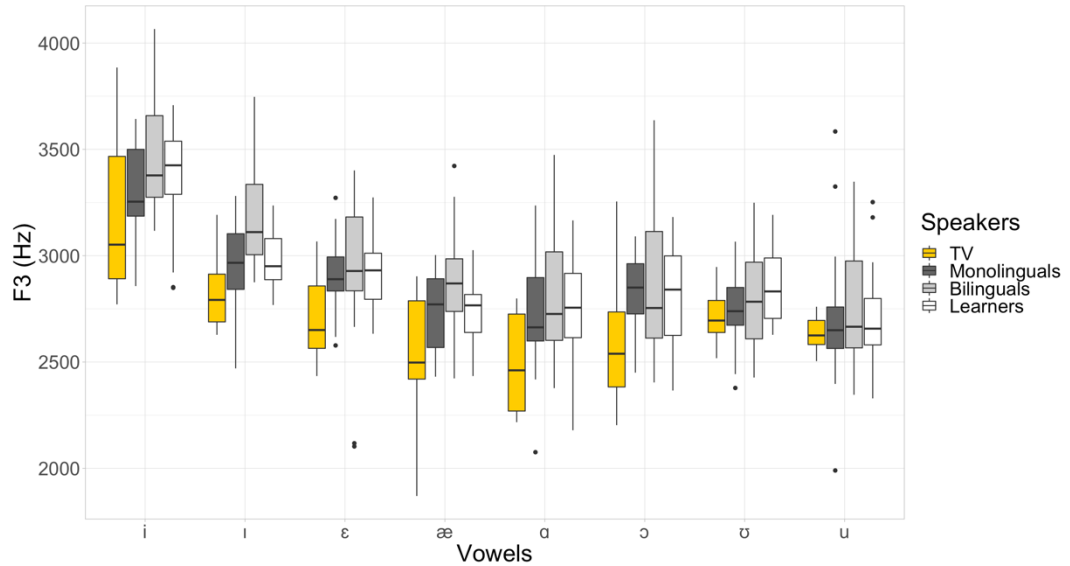


Figure 5 plots the mean F1 values for each speaker profile. Language learners and monolinguals tend to diverge from the target values when it comes to open-mid / ϵ æ ɔ / and open / ɑ / vowels. In terms of articulation, open vowels produced by language learners and the monolinguals appear more open than those modeled by the target values. F2 corresponds to the anterior/posterior movement of the tongue. F2 was higher for bilinguals (see high [ɪ u], mid [ϵ], and low [æ] vowels in figure 6). Variation is most prominent in terms of F3. This is particularly true for target values and bilingual productions. F3 in target values was lower than F3 in the rest of the speakers. Target values vowels were (with the exception of / ʊ / and / u /) more rounded than the vowels produced by the other speaker profiles (7).

- (7) F3 ENGLISH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).



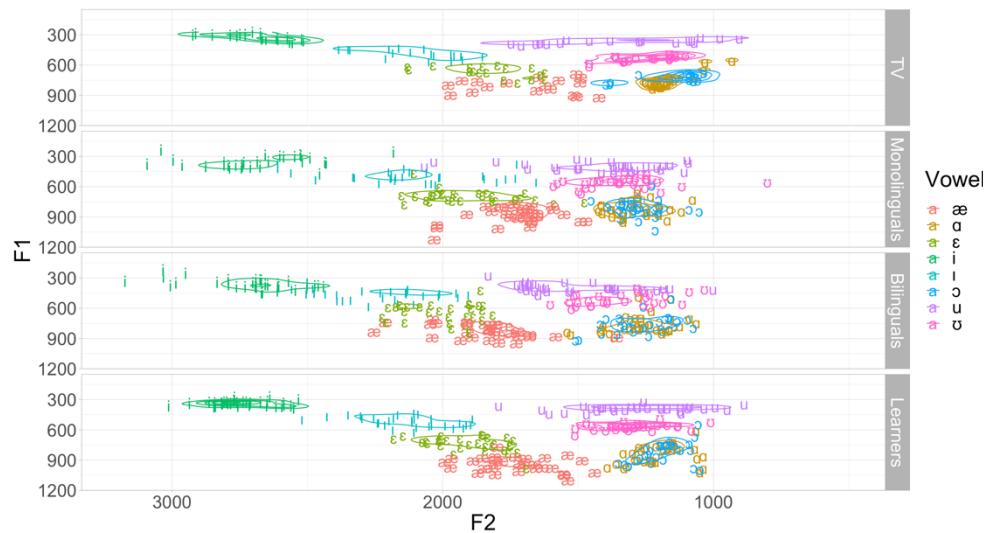
In order to better ascertain the maintenance of vowel contrasts by the different speaker profiles, pillai scores were obtained for all minimal pairs (8). /a - ɔ/ is the only contrast that shows significant merging for all speaker profiles except the target values. All other contrasts were retained albeit at varying degrees, especially for the F2 – F3 space.

- (8) F1 – F2 AND F2 – F3 PILLAI SCORES FOR ENGLISH VOWEL CONTRASTS. COLOR INTENSIFIES AS IT GETS CLOSER TO 0 SUGGESTING VOWEL OVERLAP. SIGNIFICANT VALUES IN RED.

SPEAKER	FORMANTS	/i - ɪ/	/ɛ - æ/	/a - ɔ/	/ʊ - u/	/ɛ - ɑ/
Target Values	F1 – F2	0.847	0.571	0.062	0.945	0.889
	F2 – F3	0.882	0.161	0.025	0.136	0.923
Monolinguals	F1 – F2	0.748	0.683	0.005	0.773	0.849
	F2 – F3	0.653	0.35	0.063	0.044	0.842
Learners	F1 – F2	0.882	0.68	0.02	0.902	0.902
	F2 – F3	0.831	0.375	0.019	0.164	0.901
Bilinguals	F1 – F2	0.628	0.605	0.001	0.607	0.912
	F2 – F3	0.61	0.233	0.009	0.22	0.901

Vowel production density is plotted in (9). The plot confirms the results from the Pillai score table (8). All speaker profiles show little differentiation for the contrast /a - ɔ/. There appear to be slight difficulties in the distinction of the contrast /ɛ - æ/ for all speaker profiles. The /u - ʊ/ and /i - ɪ/ categories seem less well defined for Bilinguals. Monolinguals show variation in the production of /ɪ/ and /i/.

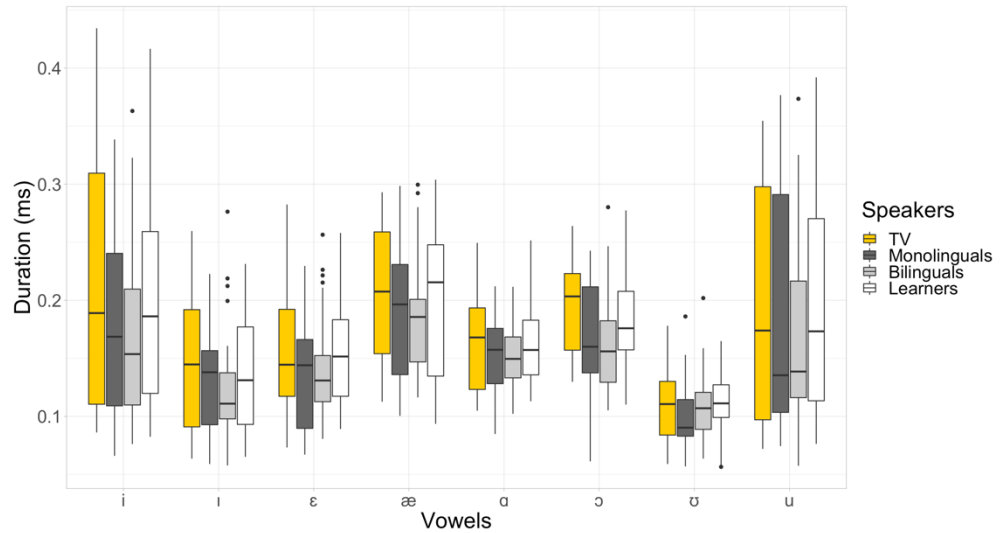
(9) F1 AND F2 VALUES FOR ALL ENGLISH VOWEL PRODUCTIONS IN TERMS OF DENSITY. DIAGRAMS SEPARATED BY SPEAKER PROFILE. TV = TARGET VALUES.



Vowel duration varied significantly depending on the context ($H_{(2, 867)} = 409.625$; $p < 1.13e-89$). Isolated vowels were significantly longer than vowels in /bVb/ context ($H_{(1, 290)} = 8.202$; $p < 7.05e-16$), which were in turn longer than vowels in /pVp/ ($H_{(1, 290)} = -11.922$; $p < 2.72e-32$). A Kruskal Wallis test found significant differences in duration between speaker profiles (Table 3). A *post-hoc* test showed that bilinguals diverged significantly from the target values (Table 4). (10) shows that both bilinguals and monolinguals produce shorter vowels as a whole. Language learners converge with slightly shorter vowels than the target values. The intrinsic duration-related

distinction between tense and lax vowels was well maintained by all speakers.

- (10) ENGLISH VOWEL DURATION IN MS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES (OUTLIERS NOT EXCLUDED).



2.2. FRENCH

The Kruskal Wallis test conducted (11) on all formant values between the target values and each of the speaker profiles did not reveal any significant differences for F1 or F2. It did however find significant differences in F3 values as well as duration.

- (11) KRUSKAL WALLIS TEST FOR THE DIFFERENCE IN FORMANTS F1 TO F3 AND DURATION FOR FRENCH VOWELS BETWEEN TARGET VALUES AND ALL SPEAKER PROFILES.

FORMANT	N	H	DF	P
F1	1082	7.322	3	0.062
F2	1082	4.700	3	0.195
F3	1082	10.624	3	0.014 *
Duration	1082	21.168	3	9.72e-05 ****

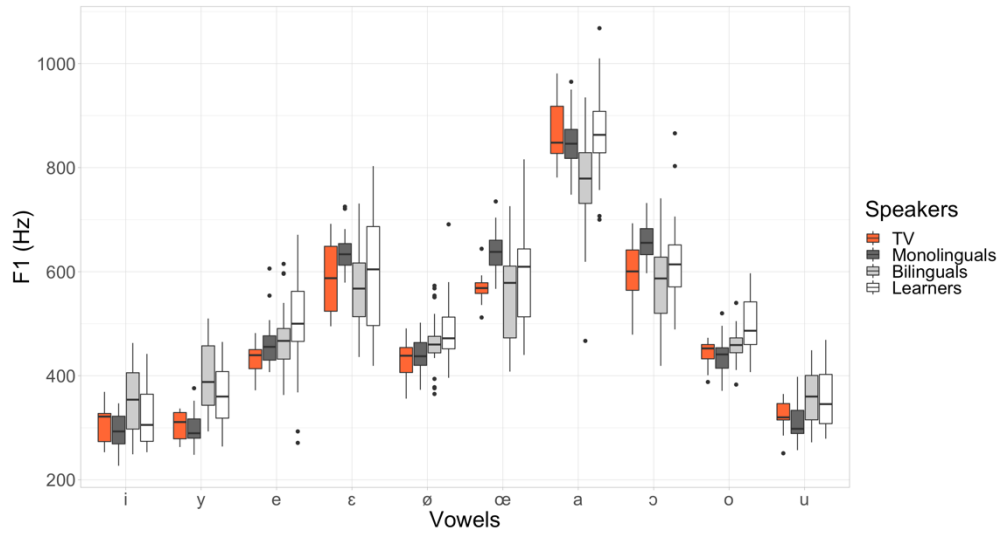
A Dunn test with Bonferroni correction (12) showed that Language learners converge the least out of all the speaker profiles in terms of F1 (though not significantly). None of the groups diverge significantly for F2 but Bilinguals converge the least out of all three groups. No significant differences found for F3.

(12) DUNN TEST COMPARING TARGET VALUES TO EACH SPEAKER PROFILE WITH BONFERRONI CORRECTION AND P ADJUSTED SIGNIFICANCE VALUES

FORMANT	SPEAKERS	N1	N2	STATISTIC	P ADJ
F1	Bilinguals	300	182	-1.612	0.642
	Learners	300	182	-2.619	0.053
	Monolinguals	300	182	-1.132	1.000
F2	Bilinguals	300	182	-1.965	0.296
	Learners	300	182	-0.536	1.000
	Monolinguals	300	182	-1.150	1.000
F3	Bilinguals	300	182	-1.876	0.364
	Monolinguals	300	182	-1.582	0.683
	Learners	300	182	0.562	1.000
Duration	Bilinguals	300	182	1.255	1.000
	Learners	300	182	-1.922	0.328
	Monolinguals	300	182	1.745	0.486

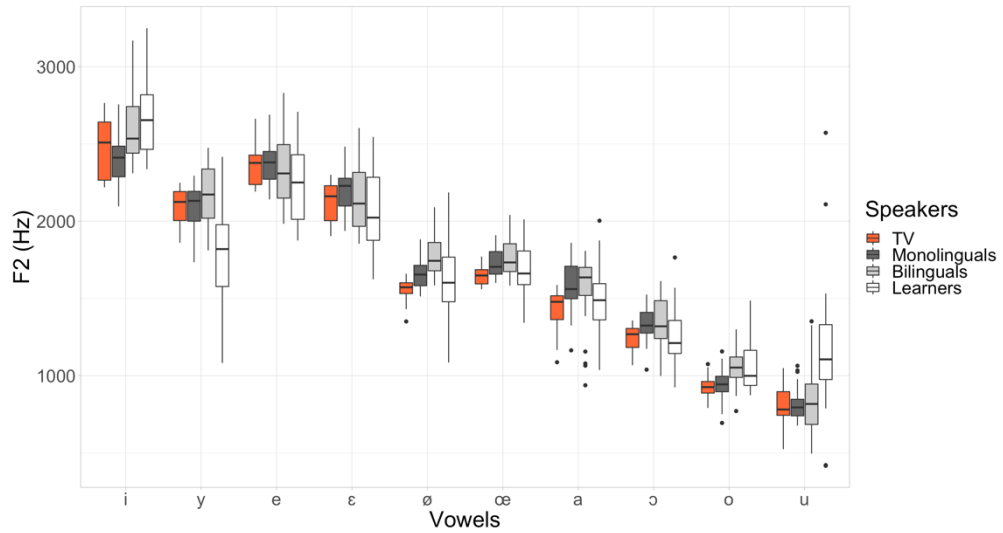
(13) plots F1 for the different vowels and the different groups. Bilinguals diverge from target values for vowels /i, y, e, a, u/. F1 for /y/ was higher in bilinguals (392Hz) and language learners (361Hz) than it was for both monolinguals (298Hz) and the target values (304Hz). Additionally, F1 values for the low vowel [a] in bilinguals (773Hz) were lower than the target values (864Hz). Moreover, variability is generally more prominent in bilinguals and language learners than it is for monolinguals. Language learners diverge in most vowels with the exception of /i, ε, a, ə/. They have a tendency to produce vowels with a higher F1 frequency and therefore more open than the target values. Monolingual open-mid vowels /ε œ ə/ are more open than those of the other groups.

- (13) F1 FRENCH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).

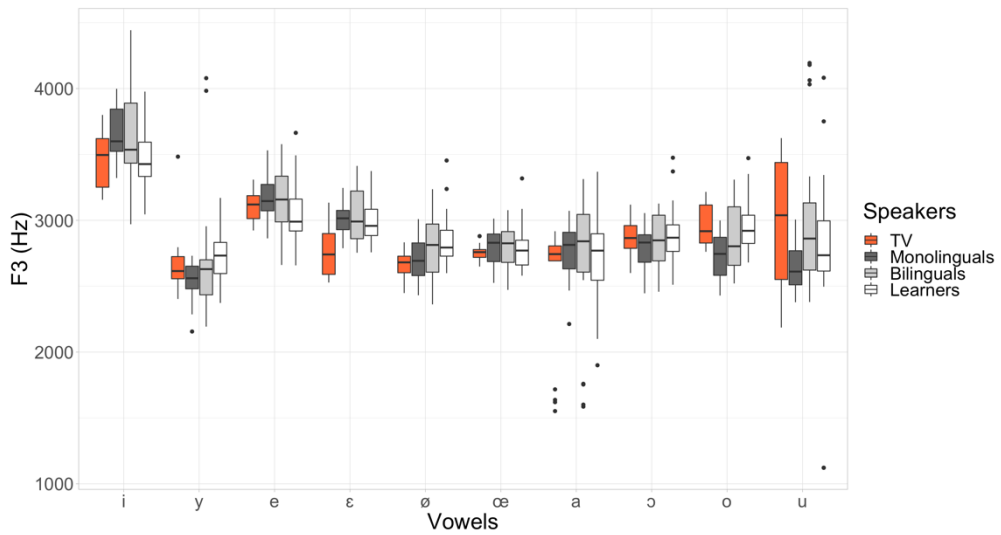


(14) compares target values to the different speaker profiles by plotting the F2 values for each of the French vowels. Language learners diverge slightly on most vowels with the exception of / ϵ , \emptyset , œ , ɔ /. Two vowels stand out, /u/ and /y/, with noticeable divergence from the target values. Furthermore, the F2 distance between /i/ and /y/ is much more marked in language learners than in the other groups. Bilinguals diverge on vowels / \emptyset , œ , a, ɔ /. (15) plots the F3 values for all vowels and for the different groups. Language learners appeared to merge vowel [y] with vowel [u] in terms of F3 as well. Bilinguals and language learners did not mark the contrast / \emptyset – œ / in terms of the F3.

- (14) F2 FRENCH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).



- (15) F3 FRENCH VOWELS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES. ERROR BARS SHOW STANDARD DEVIATION (OUTLIERS NOT EXCLUDED).



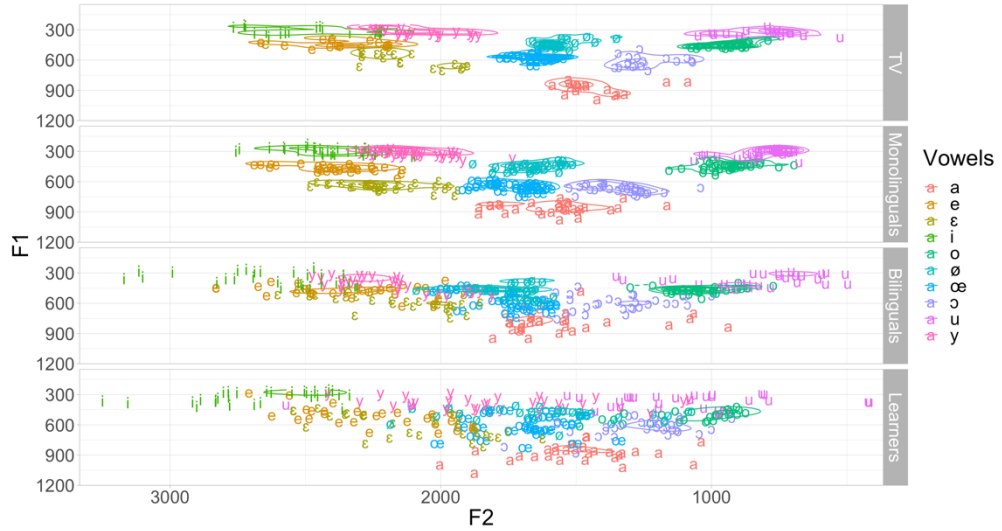
Pillai scores were obtained for all minimal pairs in French as well (16). There aren't any significant vowel mergers apart for the vowel contrast /ø – œ/ in the formantic space F2 – F3. It is interesting to note that most of the heavily shaded area corresponding to the Pillai scores that are closest to 0 and therefore indicating a tendency towards merging corresponds to bilingual and learner productions. This is especially true for contrasts /e - ε/ and /ø – œ/, but the same can be said for /ɔ - o/, /ε - œ/, and /e - ø/. Bilinguals however seem to have a slight advantage over the language learners. Language learners experience difficulty maintaining the vowel contrast /u – y/.

(16) F1 – F2 AND F2 – F3 PILLAI SCORES FOR FRENCH VOWEL CONTRASTS. COLOR INTENSIFIES AS IT GETS CLOSER TO 0 SUGGESTING VOWEL OVERLAP. SIGNIFICANT VALUES IN RED.

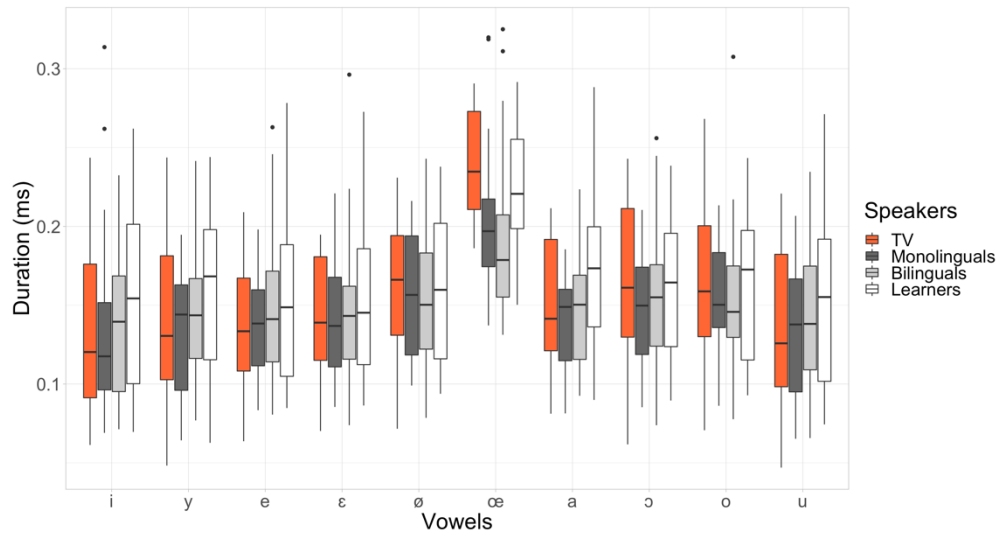
SPEAKER	FORMANTS	/i - y/	/e - ø/	/ε - œ/	/ɔ - o/	/u - y/	/e - ε/	/ø – œ/
Target Values	F1 – F2	0.654	0.929	0.902	0.848	0.964	0.702	0.834
	F2 – F3	0.777	0.934	0.848	0.777	0.962	0.632	0.342
Monolinguals	F1 – F2	0.596	0.91	0.811	0.936	0.969	0.853	0.891
	F2 – F3	0.916	0.912	0.787	0.809	0.971	0.313	0.254
Learners	F1 – F2	0.718	0.658	0.519	0.439	0.399	0.176	0.336
	F2 – F3	0.829	0.59	0.509	0.294	0.485	0.081	0.05
Bilinguals	F1 – F2	0.523	0.702	0.672	0.665	0.923	0.344	0.331
	F2 – F3	0.624	0.7	0.566	0.527	0.922	0.146	0.005

Vowel production density is plotted in (17). The plot confirms the results from the Pillai score table (16). Variability is much more prominent in bilingual and language learner productions. Vowel classes appear more stable for monolinguals. Vowel /y/ is backed in some of the language learner productions causing an overlap with the vowel /u/. Bilinguals and language learners seem to struggle with maintaining the space between vowel classes /e/ - /ε/ and /ø/ - /œ/.

- (17) F1 AND F2 VALUES FOR ALL FRENCH VOWEL PRODUCTIONS IN TERMS OF DENSITY. DIAGRAMS SEPARATED BY SPEAKER PROFILE. TV = TARGET VALUES.



- (18) FRENCH VOWEL DURATION IN MS: COMPARISON BETWEEN TV (TARGET VALUES) AND ALL SPEAKER PROFILES (OUTLIERS NOT EXCLUDED).



Vowel duration varied significantly depending on the context ($H_{(2, 1082)} = 503.805$; $p < 3.98e-11$). Just like in English, isolated vowels were significantly longer than vowels in /bVb/ context ($H_{(1, 360)} = 8.226$; $p < 5.79e-16$), which were in turn longer than vowels in /pVp/ ($H_{(1, 360)} = -13.95872$; $p < 8.352e-44$). A Kruskal Wallis test found significant differences in duration between the speaker profiles (11). A post-hoc test showed that none of the groups deviated significantly from the target values (12). However, significant differences were found between the language learners and both bilinguals and monolinguals. (18) shows that both bilinguals and monolinguals produce shorter vowels as a whole.

3. DISCUSSION

Analyses showed a tendency for speakers to diverge more in AE than in French. This contradicts part of our hypothesis H.1 and the literature (Horton and Bradlow 2011), which claims that speakers tend to diverge when they don't share the same L1. A particularly interesting finding in this study was the way in which speakers diverged in both AE and French. In AE, convergence analysis revealed a differentiation in production strategies for both bilinguals and language learners. Bilinguals diverge the most in terms of F2 while language learners diverge in terms of F1. It has been suggested that the acoustic cues used by bilinguals differ from those used by monolinguals and that it is a strategy used to either highlight perceptual differences or simply due to patterns of articulatory economy (Sundara et al. 2006). Watson (2007: 1536: 112) argues that the factor of language exposure should be incorporated into models that attempt to describe the process of phonological acquisition in simultaneous bilingual. However, this study found more convergence in French but the results from the Pillai Bartlett trace found more overlap in French than in English. Various median Pillai score values were also attested. This points towards a continuum between vowel-contrast convergence and divergence, and the very probable existence of a third category. Thus, inherent varying levels of cross-

linguistic transfer among simultaneous bilinguals and learners could explain the variability present in the study (Detey et al. 2005). Bilinguals in this study had spent the greater part of their life in France and were living in France at the time of the study. Most had only lived in France with regular (short-stay) visits to the United States. A questionnaire aimed at quantifying the amount of time each language was used on a daily basis showed that most bilingual speakers (4 out of 5) split the difference in time of use/exposure for both AE and French right down the middle (50/50). We thus excluded the notion of language dominance (Grosjean 1997) as a possible explanation. In French, statistical analyses did not show signs of significant divergence from the target values for any of the speaker profiles. That said, based on the varying levels of convergence we were able to determine that language learners diverged the most in terms of F1 and bilinguals in terms of F2 and F3. Monolinguals were consistently convergent in both English and French. This suggests that age of first exposure might not be the only predictive factor for native-like production in an L2 or a second L1. Kuhl et al.'s (2008) NLM-e (Native Language Magnet) model suggests that experience, not just time of exposure, is a “critical factor driving phonetic learning and perception of a second language”. Additional data is still necessary to fully comprehend bilingual phonetic development.

In English, significant differences in vowel duration depending on the context were expected. Isolated vowels were longer than vowels in /bVb/ phonetic context, which were longer than vowels in /pVp/ phonetic context. This confirms the extant literature on the subject which posits that an intrinsic context-duration relationship exists in AE (Chomsky and Halle 1968, Mack 1982). The tense-lax duration-related distinction between vowels was well maintained by all speakers, which might suggest that duration might not be affected by cross-linguistic influence in simultaneous bilinguals. In French, like in English, context-based differences were found in relation to vowel length. Isolated vowels were significantly longer than vowels in /bVb/ context. Vowels in /bVb/ context were significantly longer than vowels in /pVp/ phonetic context. This would suggest that

even in French, a language with relatively low vowel-length effect (Mack 1982), the differences are detectable and even significant. Surprisingly, duration was constant for most vowels in all speakers, even language learners. This would suggest that cross-linguistic transfer in the language learners does not influence duration in the L2. This could be due to their high level of proficiency in the language.

The Pillai Bartlett trace was instrumental in measuring the phonological space between vowel category contrasts. In English, the contrast /ɑ - ə/ was particularly difficult to maintain for all speakers. This is not surprising as the literature suggests that this vowel contrast could be undergoing neutralization in younger generations and in certain regional dialects of AE (Ladefoged and Disner 2012). Language learners seem to have a slight advantage over bilinguals when it comes to English. Moreover, variability seemed more present in bilingual productions than in language learner productions, which could suggest that too much variability could in fact be a hinder to language mastery and maybe even intelligibility (H.0). In French, apart from the formantic space F2 – F3 for the vowel contrast /ø – œ/ produced by language learners and bilingual speakers, no significant merging took place. However, the Pillai scores closest to zero were, for the most part, tied to bilingual and language learner productions. Bilinguals and language learners were in disadvantage for most vowel contrasts in comparison to monolingual productions. The statistical analyses revealed differences between French vowels that were context-specific. Differences concentrated on F1 and F3 values which would suggest that certain phonetic contexts could be more difficult for either learners or bilinguals; or both.

The PAM model (Best 1994, 1995) claims that the phonological and articulatory-phonetic similarities in non-native contrasts are a good predictor of discrimination patterns for language learners. Similarly, the SLM (Flege 1995) and the SLM-r (Flege and Bohn 2021) predict phonological competence in the L2 based on acoustic-phonetic similarity. Therefore, it is not surprising that contrasts containing front rounded vowels seem more difficult to reproduce in the case of bilinguals and language learners. As predicted in our hypothesis H.2, the contrast /y – u/ was particularly difficult for language

learners. The Pillai scores obtained for bilinguals show attenuated distinction for vowel contrasts /i - ɪ/ and /u - ʊ/. Overall, the relatively high levels of vowel overlap in French attested in language learners and in both languages for simultaneous bilinguals are indicative of cross-linguistic transfer. In the case of the simultaneous bilinguals, this transfer seems to be bilateral as initially hypothesized in H.2 and as modeled by Paradis (2001).

CONCLUSION

The results of this study suggest varying levels of cross-linguistic transfer in both simultaneous bilinguals and language learners, testifying to the existence of provisional and unstable phonological systems in both groups (Detey et al. 2005). Language learners and bilinguals tend to deviate from English but converge for French. With the exception of two tense-lax vowel contrasts, the Pillai Bartlett trace for the vowel contrasts produced by bilinguals and language learners seemed relatively stable in English. The distinction between vowel categories seemed less accurate in French for both bilinguals and language learners. The contradictory results indicate that convergence alone is not a reliable measure for accuracy. Other measures such as the Pillai Bartlett trace are necessary in order to evaluate phonological competence. Additionally, we argue that variation should be explicitly defined in either phonetic or phonological terms. Articulatory variation could in fact be overlooked if the vowel contrasts are maintained and intelligibility is not affected. Such seems to be the case for English, where divergence was significant but the different acoustic cues reproduced by language learners and bilinguals did not seem to affect vowel contrast distinction. In French however, no significant divergences were found but vowel overlap in language learners and bilinguals was much more prominent. The results of this study show that phonetic divergence does not necessarily result in more vowel mergers. Vowel contrasts can undergo acoustic differences without losing inherent discriminability. This entails that a separate study that looks at phonological variation specifically is also necessary in order to accurately describe and characterize

phonological system(s) interaction in bilinguals and language learners. Further research will include more participants and individual linguistic profiles will be discussed in relation to acoustic and statistical analyses. This next step would also involve individual evaluation for each speaker and thus a mixed-effects modeling would be implemented to account for individual variability. This is necessary to reinforce the trends found in this preliminary study. In addition, perception and articulatory components will be added to the study so as to identify the nature of the difference in production strategies implemented by both bilinguals and language learners. We are also aware of a possible lexical bias introduced by the indiscriminate use of real words and pseudo-words in this study. A search on [lexique.org](https://www.lexique.org) revealed that not all words had the same frequency, neither in French (pipe: 16.38, peur: 557, pupe: .01) nor in English (peep: 226, bib: 25, pep: 178). Research on language acquisition has found that world-level information can affect the way in which phonetic contrasts are treated (Feldman et al. 2013). Pseudo-words and words with lower frequency would provide little to no information consequently affecting vowel contrast distinction. In the future, a more detailed discriminating corpus that separates words from pseudo-words would be necessary in order to avoid this phenomenon altogether. The context-specific influences found in the French corpus and the rhotic consonant after /œ/ would also merit further study in order to ascertain possible influences on vowel quality and the possibility that certain contexts could prove more difficult for some speakers. In conclusion, this preliminary study highlights the importance of research that seeks to investigate the interfaces between variation, phonological space organization, and the effects of age of L1/L2 exposure in simultaneous bilinguals and late language learners.

REFERENCES

Amengual, M. (2019). Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism*, 23(5), pp. 954-970.

- Babel, M. (2009). Selective vowel imitation in spontaneous phonetic accommodation. *UC Berkeley PhonLab Annual Report*, 5(5), pp. 163-194.
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, pp. 437-456.
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), pp. 177-189.
- Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee- Collier & L. P. Lipsitt (Eds.), *Advances in infancy research*. Norwood, NJ: Ablex, Vol. 9, pp. 217- 304.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross- language research*. Baltimore: York Press, pp. 171-204.
- Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5(9/10), pp. 341-347.
- Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment, *The Journal of the Acoustical Society of America*, 128(1), pp. 32-37.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York : Harper and Row.
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2-3), pp. 145-173.
- Detey, S., Durand, J., & Nespoulous, J. L. (2005). Interphonologie et représentations orthographiques. Le cas des catégories /b/et/v/chez des apprenants japonais de Français Langue Etrangère. *Revue Parole*, 34, pp. 140-185.

- Elias, V., McKinnon, S., & Milla-Muñoz, Á. (2017). The effects of code-switching and lexical stress on vowel quality and duration of heritage speakers of Spanish. *Languages*, 2(4), pp. 29.
- Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological review*, 120(4), 751.
- Flege, J. E. (1987). A critical period for learning to pronounce foreign languages?. *Applied linguistics*, 8(2), pp. 162-177.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92, pp. 233-277.
- Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. *Laboratory phonology*, 9, pp. 353-381.
- Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). *second language speech learning: theoretical and empirical progress*, pp. 3-83.
- Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America*, 76(3), pp. 708-721.
- Flege, J. E., MacKay, I. R., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23(4), pp. 567-598.
- Fourgeron, C., & Smith, C. L. (1999). French. *Handbook of the International Phonetic Association*, pp. 78–81.
- Georgeton, L., Paillereau, N., Landron, S., Gao, J., & Kamiyama, T. (2012, June). Analyse formantique des voyelles orales du français en contexte isolé : à la recherche

- d'une référence pour les apprenants de FLE. In *Conférence conjointe JEP-TALN-RECITAL 2012*, pp. 145-152.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105(2), pp. 251-279.
- Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, 12(2), pp. 91-114.
- Gottfried, T. L. and Beddor, P. S. (1998). Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, pp. 57-75.
- Grosjean, F. (1997). The bilingual individual. *Interpreting*, 2(1-2), pp. 163-187.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97(5), pp. 3099-3111.
- Horton, Kim, M. & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory phonology*, 2(1), pp. 125-156.
- Kang, Y., & Nagy, N. (2016). VOT merger in Heritage Korean in Toronto. *Language Variation and Change*, 28(2), pp. 249-272.
- Kelley, M. C., & Tucker, B. V. (2020). A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America*, 147(1), pp. 137-145.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), pp. 979-1000.

- Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of phonetics*, 34(4), pp. 500-515.
- Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Thomson, Wadsworth.
- Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. Hoboken: John Wiley & Sons.
- Lewandowski, N., & Dogil, G. (2009). Perception-production loop in native–non-native dialogs: Phonetic convergence. *The Journal of the Acoustical Society of America*, 125(4), pp. 2768-2768.
- Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, 4: 18.
- Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, 125(2), 1138-1152.
- Mairano, P., & Santiago, F. (2019, November). On the link between L2 learner's vocabulary knowledge and pronunciation accuracy: a corpus-based study. In *10èmes Journées Internationales de Linguistique de Corpus*, 26-28 novembre 2019, Grenoble.
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America*, 71(1), pp. 173-178.
- Mayr, R., López-Bueno, L., Fernández, M. V., & Lourido, G. T. (2019). The role of early experience and continued language use in bilingual speech production: A study of Galician and Spanish mid vowels by Galician-Spanish bilinguals. *Journal of Phonetics*, 72, pp. 1-16.

- Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, pp. 3-23.
- Morris, J. (2013). *Sociolinguistic variation and regional minority language bilingualism: an investigation of Welsh-English bilinguals in North Wales*. United Kingdom : The University of Manchester.
- Morris, J. (2017). Sociophonetic variation in a long-term language contact situation:/l/-darkening in Welsh-English bilingual speech. *Journal of Sociolinguistics*, 21(2), pp. 183-207.
- Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4), pp. 422-432.
- Nielsen, K. Y. (2008). *Word-level and Feature-level Effects in Phonetic Imitation*. Los Angeles: University of California.
- Nycz, J., & Hall-Lew, L. (2013, December). Best practices in measuring vowel merger. In *Proceedings of Meetings on Acoustics 166ASA*. Acoustical Society of America, 20(1), pp. 060008.
- Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. In *Applied Mechanics and Materials*. Trans Tech Publications Ltd, Vol. 611, pp. 115-120.
- Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International journal of bilingualism*, 5(1), pp. 19-38.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), pp. 2382-2393.

- Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79(2), pp. 637-659.
- Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, 72(8), pp. 2254-2264.
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, 162(1-2), pp. 8-13.
- Perry, S. J., Holko, G., Kelley, M. C., & Tucker, B. V. Consonants, vowels, and lexical representations: evidence from auditory lexical decision? In the 2021 annual conference of the Canadian Linguistic Association, pp. 1-12.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, 32(6), pp. 693-703.
- Pillai, K. C. S. (1976). Distributions of characteristic roots in multivariate analysis Part I. null distributions. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pp. 157-183.
- Regan, V., & Bayley, R. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of sociolinguistics*, 8(3), pp. 323-338.
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of phonetics*, 25(4), pp. 421-436.
- Schweitzer, A., & Lewandowski, N. (2014). Social factors in convergence of f1 and f2 in spontaneous speech. In *Proceedings of the 10th international seminar on speech production, Cologne*, pp. 391-394.

- Simonet, M. (2011). Production of a Catalan-specific vowel contrast by early Spanish-Catalan bilinguals. *Phonetica*, 68(1-2), pp. 88-110.
- Simonet, M., & Amengual, M. (2020). Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism*, 24(2), 208-221.
- Strandberg, J. A., Gooskens, C., & Schüppert, A. (2021). Simultaneous bilingualism and speech style as predictors of variation in allophone production: Evidence from Finland-Swedish. *Journal of Phonetics*, 88, 101095.
- Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. *Cognition*, 100(2), pp. 369-388.
- Watson, I. (2007). Simultaneous bilinguals and Flege's Speech Learning Model. In Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 6–10 August, 2007, pp. 1533–1536.
- Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and cognition*, 3(2), pp. 131-149.

DISCUSSION WITH JALAL, AL-TAMIMI

(UNIVERSITÉ PARIS CITÉ, LABORATOIRE DE LINGUISTIQUE FORMELLE, CNRS, UMR-7110)

Al-Tamimi, Jalal. 2023. discussion in: Lara, Andres Felipe & Pillot-Loiseau, Claire (auth.) “Phonetic convergence and vowel overlap”. *Radical: A Journal of Phonology*, 3, 270-279.

COMMENTS

This research highlights some important, yet less documented, differences between various language producers’ groups. The researchers aimed to systematically compare the notions of divergence and convergence between four speakers’ profiles, producing French or American English (AE) vowels. The researchers defined the speakers’ profiles as a) native speakers target vowels; b) Monolinguals; c) Simultaneous bilinguals and d) AE learners of French as a second language. The variation in these groups allowed the researchers to systematically compare the speakers’ profiles to allow to further understand how cross-linguistic variation operates and whether there is any systematicity alongside the convergence-divergence continuum. Using a shadowing experience for speakers in groups b, c, and d, allowed the researchers to assess this type of variation, with a relatively small sample size, albeit generally appropriate for this type of preliminary investigation.

The major results obtained in this study showed that even monolingual speakers’ vowels deviated from those from the first group (native speakers’ target vowels) and alongside the convergence-divergence continuum. Depending on the dimension investigated (F1, F2, F3, duration, or the pillai scores between F1-F2 or F2-F3 dimensions), the researchers found that the speakers’ profiles varied. Learners of the language were either similar or dissimilar to native speakers’ Target Vowels, to monolinguals and/or to bilinguals. The same applies to the other groups. Another major

result was that speakers seemed to diverge more in AE than in French, whereby bilinguals diverged more on the F2 dimension, but learners diverged more on the F1 dimension (pp. 257).

The researchers used two approaches to assess the degree of convergence or divergence.

The first approach was based on statistical inferences; a statistically significant difference was used as an indication of divergence and a non-statistical difference (even if a tendency) was used as an indication of convergence. A major criticism to this approach is the non-inclusion of random effects structure, which increases Type I error and favours more false positives. It is not clear how homogenous the speakers' profiles are; a point not discussed in the current research. It is likely that some of the observed deviations are either caused by inherent group differences, or the type of material used. For instance, the TVs are likely to have been produced in isolation, whereas the comparison for both languages was done on the averaged results across the three production styles: isolation, /bVb/ or /pVp/, even when there were clear differences in English for F3 and in French for F1 and F3 (pp. 246). It is likely that this is the reason behind finding differences between the vowels using in the Target Vowels in comparison with the monolinguals. The expectations is that there should not be any differences between the two, and any observed differences are due to either contextual differences or sociophonetic/socio-linguistic difference between the two groups. In addition, it is well-known that vowels in isolation are produced as more extreme in comparison with those in contexts (Lindblom, 1963, 1990, 1996); these differences can be further enhanced when using pseudo-words; these will lead to an increase in processing time and will most likely be produced in an extreme fashion. This makes the comparison problematic at first site. However, the researchers assessed whether the three contexts differed, and their results showed a mixture of patterns, albeit using a non-mixed effects approach. The observed inherent variation in the speakers' profiles and material used should have been explored further in this study, by employing at least linear mixed effects modelling on the

raw data, to allow for this variation to be considered, even with the small sample size.

The second approach used relied on the pillai scores. This statistical approach has been widely used in the literature to assess vowel mergers. The researchers here used it over the Euclidean distances, spectral overlap assessment metric, and the vowel overlap assessment with convex hulls, as they highlight its strengths over these classical approaches (pp. 242). This is of course an excellent choice as it allows to compare groups within specific dimensions. By looking at the extremes of the Pillai scores: 1 indicating complete divergence (separation) and 0 indicating complete convergence (or overlap). The Pillai scores presented in table 8 (for AE) and 16 (for French) indicate that the picture is more complex than what was initially advocated in this paper. The results point to a gradient rather than a categorical difference, which can be quantified alongside a convergence-divergence continuum, as none of the scores reported in this paper show a categorical 0 or 1. This is of course expected due to the nature of these vowel comparisons. A reanalysis of the Pillai scores is shown in the figures 1 and 2 below. The two figures show somehow an opposite interpretation of the results to those advanced in this study: speakers diverge less in AE than in French! When taking the 0.5 as the threshold to decide on the categorical shift, English vowel pairs (figure 1) seem to show an increase in the divergence (separation) in 4 out of 5 pairs (on the F1-F2 dimension), and in 2 out of the 3 pairs (on the F2-F3 dimension); French vowel pairs (figure 2) seem to show an increase in the divergence (separation) in 3 out of 7 pairs on either dimension. The conclusion that can be made here is that “speakers [tend to] diverge more in AE than in French” (pp. 258).

In more detail, if we look further at the learners’ group, they show more variation and a tendency towards convergence (overlap) in their productions of French vowels as they seem to deviate the most from all other groups. Bilinguals also show deviations that are more marked than those of the Target Vowels or of the monolinguals; they deviate more in their productions of French rather than in English. This seems to follow what the researchers point to in their discussion that bilinguals and learners are more variable in their productions, and points to the notions of language dominance; in fact, even though

the bilinguals lived in France for almost all their life, they seem to diverge mostly in their productions of French. Can this be quantified solely by “language dominance”? Do bilinguals internalise their English language which seems to influence their French vowels the most? For learners, there is a clear indication that even at an advanced level, influence from their L1 is present as they show most of the deviations on either dimension when looking at their productions of the French vowels.

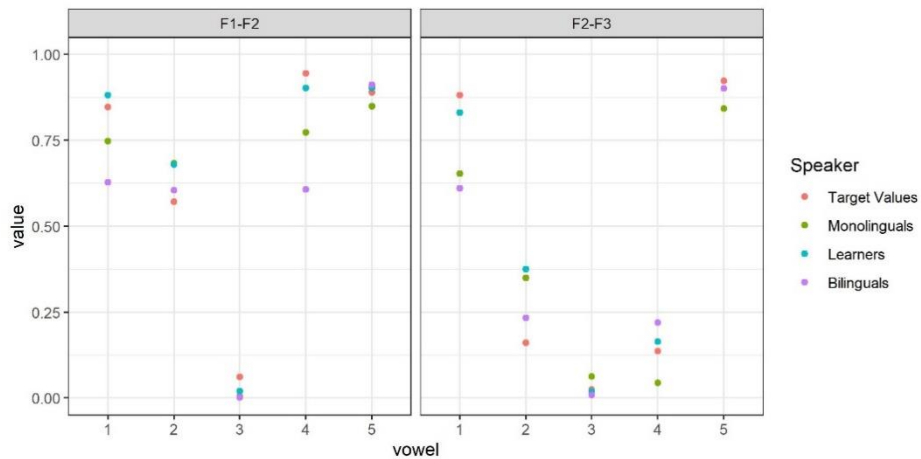
Of course, our discussion so far used the 0.5 threshold to decide on separating the vowel pairs alongside the convergence-divergence continuum. What happens if we change our threshold? If one chooses another threshold, say 0.375, the picture will change, and speakers in French will diverge in 5 out of the 7 pairs on the F1-F2 dimension! This will lead to an opposite conclusion from the one highlighted in the paper.

This threshold is problematic and depends on our view of how this convergence-divergence continuum is. It is clear from the two figures below that there is more variation alongside the convergence-divergence continuum in the productions of the French vowels (figure 2) than those seen in English. How is this variation considered? How likely are they influencing our decisions in considering vowel divergence or convergence? Are speakers systematic in how they signal differences between vowels?

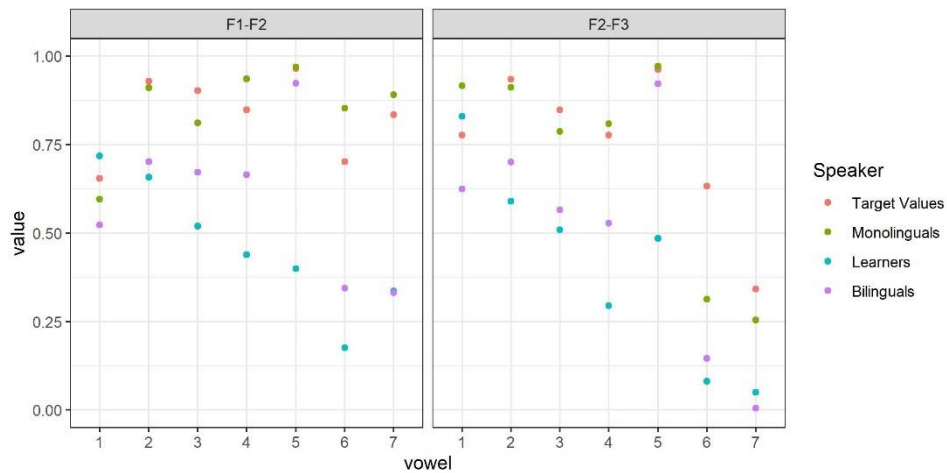
It is likely that speakers of these two languages use specific mechanisms that allow them to differentiate these vowel pairs. For instance, this research looked at individual formants as indicative of vowel separation, and used the pillai scores on the F1-F2 dimension or F2-F3 dimension. The assumption here is that individual formants are a meaningful way to identify vowel pairs. It is in fact very well-known that the distance between formants, rather than the formants themselves offer a more direct way to correlate production with perception (Fahey et al., 1996; Traunmüller, 1981, 1984; Syrdal and Gopal, 1986). This is likely to be the case here and likely that speakers of the various groups use specific metrics to signal differences between these vowel pairs. Unfortunately, a reanalysis of the current research is not possible due to the raw data not being available, however, it is likely that speakers of both languages will use different mechanisms to

differentiate their vowels that will rely on the Z1-Z0, Z2-Z1 and Z3-Z2 dimensions (Z = Critical band value), given that Z1-Z0 correlates best with the [+open] (or [+low]) vowels; Z2-Z1 correlates best with [+low, +back] and Z3-Z2 correlates best with [+back] vowel (Al-Tamimi, 2017; Syrdal and Gopal, 1986). These dimensions show how close (or far) the formants are and how they allow to separate the vowels. It is likely that these auditory dimensions allow for a better account of the convergence-divergence continuum and the within-group variation will disappear to show clearer near-categorical differences; a point to consider further in future research.

- (1) ENGLISH PILLAI SCORES PER FORMANT COMPARISON (F1-F2 OR F2-F3), VOWEL TYPE (1 = /i - ɪ/, 2 = /ɛ - æ/, 3 = /ɑ - ɔ/, 4 = /ʊ - u/, 5 = /ɛ - ɑ/) and speakers' profile



- (2) FRENCH PILLAI SCORES PER FORMANT COMPARISON (F1-F2 OR F2-F3), VOWEL TYPE (1 = /i - y/, 2 = /e - ø/, 3 = /ɛ - œ/, 4 = /ɔ - o/, 5 = /u - y/, 6 = /e - ɛ/, 7 = /ø - œ/) and speakers' profile



Finally, the researchers used predictions from the two well-known theories in L2 phonology, namely, the SLM and PAM. These are indeed the most influential approaches to explaining the behaviour of learners in L2 phonology. Using predictions from the Native Language Magnet Effect, discussed towards the end offers an opening towards understanding the behaviour of the speakers in this study. Using the predictions from the L2LP model (Escudero, 2005) can offer an additional account of the current data as the model uses predictions that come directly from acoustic analyses of the various speech sounds rather than relying on gestures (as is the case in PAM).

Due to the dynamic nature of vowels and of acoustics in general, combining various approaches advocated in my discussion will help shed light into how bilinguals differ from monolinguals on the one hand, and how they differ from (advanced) learners. What makes this variation important to deal with and how it is normalised by the listener?

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 1–40. <https://doi.org/10.5334/labphon.19>. **Escudero, P.** (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT. **Fahey, R. P., Diehl, R. L., & Traunmüller, H.** (1996). Perception of back vowels: Effects of varying F1-F0 Bark distance. *Journal of the Acoustical Society of America*, 99(4), 2350–2357. <https://doi.org/10.1121/1.415422>. **Lindblom, B.** (1963). *On vowel reduction*. Quarterly Progress Status Report 29, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden. **Lindblom, B.** (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16. **Lindblom, B.** (1996). Role of Articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America*, 99(3), 1683–1692. **Syrdal, A. K., & Gopal, H. S.** (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79(4), 1086–1100. <https://doi.org/10.1121/1.393381>. **Traunmüller, H.** (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69(5), 1465–1475. <https://doi.org/10.1121/1.385780>. **Traunmüller, H.** (1984). Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication*, 3(1), 49–61. [https://doi.org/10.1016/0167-6393\(84\)90008-6](https://doi.org/10.1016/0167-6393(84)90008-6).

REPLY

We thank Jalal Al-Tamimi for his comments and for the opportunity to expand on this work.

The reviewer has perceptively zeroed in on the conscious choice for both method and participants which allowed us to investigate cross-linguistic variation in both bilinguals and late-learners with a relatively small sample size. This study defines deviation as the degree to which the productions of the monolinguals, bilinguals, and late learners differentiate from those shadowed (i.e., the target values). This measure was drawn from statistical inferences. We agree with Jalal Al-Tamimi that mixed linear models would be more appropriate in measuring the divergence convergence continuum as they account for random effects. That said, fixed effects models are a viable choice if there is reason to assume that there is one common effect (Tufanaru et al., 2015): in this case, time of exposure to both languages. The reader is reminded that this case study has a relatively small sample size of relatively homogenous speakers and generalizations would require

a larger number of speakers per group. We also discuss the need for individual evaluations for learners and the inclusion of mixed linear models in future studies (page 262).

All speakers diverged from the target values; monolinguals included. We would like to point out that target values were not produced in isolation. They were produced in the same phonetic context as those produced by the different groups of speakers. Thus, all three phonetic contexts were taken into account when evaluating differences across all groups. Consonantal context effects were beyond the scope of this study but we see the merit of exploring group and individual differences based on interlanguage syllable structure. Studies show that consonantal context may influence perception of native and non-native vowels (Gottfried, 1984; Trofimovich, Baker, & Mack, 2001). Regarding homogeneity, it is clear that this presents a problem for all studies that seek to quantify time and quality of exposure in bilinguals and language learners. We tried, to the best of our ability, to compose homogenous groups based on the factors mentioned above (page 259). Acoustic data was normalized to account for anatomical differences based on sex.

Differences between monolingual productions and target values raise a very important question: how much inter-speaker variability should be allowed before the concept of a monolingual native speaker as a unitary phenomenon is considered not viable? (Mack, 1997) Understanding how native and non-native speakers organize natural variation into meaningful perceptual units is of utmost importance if we are to understand inter-speaker variability in native and non-native groups (McAllister Byun and Tiede, 2017; Park et al., 2019). Moreover, it is important to ascertain patterns of cross-linguistic influence among bilinguals and late learners. Language learning models like PAM (Best, 1995), NLM (Kuhl, 1995), SLM (Flege, 1995), SLM-r (Flege & Bohn, 2021), and as Jalal Al-Tamimi suggested, may be able to explain variability in late learners of French as a result of the influence the L1 has on the L2. In the case of the L2LP model (Escudero, 2005) explains variation through the notion of language modes. Much like Grosjean's proposal (2001), the L2LP model predicts that any intermediate perception is "the result of parallel activation during online speech perception" (Escudero, 2005). Thus, we recommend that

future studies be more deliberate in providing further insight into the perception-production relationship.

Pillai scores were used to measure the vowel merging in the different groups. It is worth noting that definitive conclusions of overall divergence and convergence are based on impressionistic conclusions translating to more or less merging in one group or another. We address the contradictory results from both statistical tests in the discussion (page 258) and in the conclusion (page 261). The statistical test used in the first part of the study found that speakers converged more towards target values in French. The Pillai Barlett trace test found more vowel overlap in French. The two do not mutually entail each other but we do question the reliability of categorical claims of divergence and convergence between groups based on binary statistical inferences whenever evaluating phonological competence. The article discusses the possibility of “varying levels of cross-linguistic transfer amongst simultaneous bilinguals and learners” which would also explain the variability in the study (Detey et al., 2005).

Pillai scores were drawn using absolute formant values. The reviewer is correct in citing formant-ratio theory as it is well established that intervals between resonances play an integral role in vowel perception and they serve to eliminate talker differences (Peterson, 1961). However, we believe that absolute frequencies must also play a role. Furthermore, major shortcomings of formant-ratio theory have been expressed in the following questions which remain unanswered even today. Why do some differently perceived vowels have similar formant ratios? How is it that different tokens of the same vowel can have very different formant ratios? The auditory-perceptual theory (Miller, 1989) expands on formant-ratio and calculates a perceptual path from the sensory input. In the auditory-perceptual theory, a perceptual target zone is activated to issue a category code corresponding to its associated allophone whenever conditions for segmentation are met. Such conditions can be reasonably estimated using formant values (Miller, 1989).

The use of the second approach also raises a very important issue. How do we determine an acceptable threshold when evaluating vowel mergers using the Pillai

Bartlett test? One possible reason this is so problematic is that there are no agreed-upon methods to measure the size of the auditory target at the representation level (Cheng et al., 2021). We thus relied on an ad hoc threshold to interpret vowel merging in our speakers where >40 percent overlap represents merged categories. However, as Jalal Al-Tamimi suggests, relying on provisional thresholds is problematic because sample size is rarely comparable across studies or between speakers (Stanley et Sneller, 2023). Accordingly, future studies should take sample sizes into account when determining a reasonable threshold as indication of merging.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, 171-204. **Cheng**, H. S., **Niziolek**, C. A., **Buchwald**, A., & **McAllister**, T. (2021). Examining the relationship between speech perception, production distinctness, and production variability. *Frontiers in Human Neuroscience*, 15, 660948. **Detey**, S., **Durand**, J., & **Nespoulous**, J. L. (2005). Interphonologie et représentations orthographiques. Le cas des catégories/b/et/v/chez des apprenants japonais de Français Langue Etrangère. *Revue Parole*, 34, pp. 140-185. **Escudero**, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT. **Flege**, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92, 233-277. **Flege**, J. E., & **Bohn**, O. S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, 10(9781108886901.002). **Gottfried**, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, 12(2), 91-114. **Kuhl**, P. K., & **Iverson**, P. (1995). Linguistic experience and the “perceptual magnet effect”. *Speech perception and linguistic experience: Issues in cross-language research*, 121-154. **McAllister Byun**, T., & **Tiede**, M. (2017). Perception-production relations in later development of American English rhotics. *PloS one*, 12(2), e0172022. **Mack**, M. (1997). The monolingual native speaker: Not a norm, but still a necessity. *Studies in the Linguistic Sciences*, 27(2), 113-146. **Miller**, J. D. (1989). Auditory-perceptual interpretation of the vowel. *The journal of the Acoustical society of America*, 85(5), 2114-2134. **Park**, Y., **Perkell**, J. S., **Matthies**, M. L., & **Stepp**, C. E. (2019). Categorization in the perception of breathy voice quality and its relation to voice production in healthy speakers. *Journal of Speech, Language, and Hearing Research*, 62(10), 3655-3666. **Peterson**, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, 4(1), 10-29. **Stanley**, J. A., & **Sneller**, B. (2023). Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America*, 153(1), 54-67. **Trofimovich**, P., **Baker**, W., & **Mack**, M. (2001). Context-and experience-based effects on the learning of vowels in a second language. *Studies in the linguistic Sciences*, 31(2), 167-186. **Tufanaru**, C., **Munn**, Z., **Stephenson**, M., & **Aromataris**, E. (2015). Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *JBI Evidence Implementation*, 13(3), 196-207.