Sky Onosson's Writing Samples

This document contains the following publications compiled into a single .pdf file, with master page numbers shown in the lower right in large red font.

Please note that, while the papers listed here are all collaborative efforts, as first author I was responsible for approximately 70-90% of the analysis and writing for each paper.

PaPE 2019 Special Issue

Language and Speech

The Effects of Language Contact on Non-Native Vowel Sequences in Lexical Borrowings: The Case of Media Lengua

Language *and* Speech
1–30

DOI: 10.1177/00238309211014911
journals.sagepub.com/home/las
SAGE

Sky Onosson
University of Manitoba, Canada

Jesse Stewart
University of Saskatchewan, Canada

Abstract
Media Lengua (ML), a mixed language derived from Quichua and Spanish, exhibits a phonological system that largely conforms to that of Quichua acoustically. Yet, it incorporates a large number of vowel sequences from Spanish which do not occur in the Quichua system. This includes the use of mid-vowels, which are phonetically realized in ML as largely overlapping with the high-vowels in acoustic space. We analyze and compare production of vowel sequences by speakers of ML, Quichua, and Spanish through the use of generalized additive mixed models to determine statistically significant differences between vowel formant trajectories. Our results indicate that Spanish-derived ML vowel sequences frequently differ significantly from their Spanish counterparts, largely occupying a more central region of the vowel space and frequently exhibiting markedly reduced trajectories over time. In contrast, we find only one case where an ML vowel sequence differs significantly from its Quichua counterpart—and even in this case the difference from Spanish is substantially greater. Our findings show how the vowel system of ML successfully integrates novel vowel sequence patterns from Spanish into what is essentially Quichua phonology by markedly adapting their production, while still maintaining contrasts which are not expressed in Quichua.

Keywords
Media Lengua, Quichua, mixed language, generalized additive mixed models, vowels

Corresponding author:
Sky Onosson, Department of Linguistics, University of Manitoba, 534 Fletcher Argue Building, 15 Chancellor's Circle, Winnipeg, Manitoba, R3T 5V5, Canada.
Email: sky@onosson.com

1 Introduction

Mixed languages are considered a special case of contact language, unique in that instead of arising out of communicative necessity, much like a pidgin, creole, or trade language, they are created for expressive purposes (Stewart & Meakins, accepted). This is the result of the originators already being proficient bilinguals in the source languages; leaving no communicative void for the new language to fill. Because of this bilingual knowledge, mixed languages show little in the way of lexical and morphosyntactic simplification as is typical in most creole-based or pidgin-based languages.

At the same time, mixed languages tend to show highly systematic splits between the elements originating from each source language; for example, lexical–functional splits, noun phrase and verb phrase splits, or lexical–structural splits (Meakins, 2013; Stewart & Meakins, accepted). Therefore, the study of mixed languages provides linguists with a rare platform to test everything from language genesis to structural compatibility. Each new study also provides further insights to improve our understanding of cognitive mechanisms that allow humans to take two unrelated, fully functional languages, split them apart and create a new fully functional language.

Prior to the 2010s, the majority of mixed language research focused on the morphosyntactic divisions among the source languages due to their unique arrangements. However, the past decade has benefited from a surge of research in mixed language phonology that centers on the question: "What is the result of competing source language phonologies in a new mixed language?" To answer this question, phonemic conflict sites (i.e., conflicting areas of phonological convergence in the source languages' phonologies) are compared acoustically and quantitative analyses provide important details into how sounds are treated in the mixed language. Interestingly, unlike the clear splits observed in the morphosyntax, results from acoustic studies (see e.g., Buchan, 2012; Bundgaard-Nielsen & O'Shannessy, 2019; Hendy, 2019; Jones & Meakins, 2013; Jones et al., 2011, 2012; Meakins & Stewart, accepted; Rosen, 2006, 2007; Rosen et al., 2020; Stewart, 2014, 2015a, 2015b, 2020; Stewart & Meakins, accepted; Stewart et al., 2018, 2020b) show that there is a propensity for phonological material to assimilate to the phonology of the ancestral language. In other words, "the language, which was acquired[1] originally as an L2 [second language] . . . essentially conforms to the L1 [first language] phonological system of the ancestral language" (Stewart & Meakins, accepted). However, evidence also shows that the introduced language does not completely relinquish its influence on the mixed language's phonology and instead supplies material that appears to contribute to the new sound system.

While the resulting phonological system may appear straightforward at first glance, the arrangements of the introduced language's phonology in the mixed language are anything but. For instance, the resulting system does not always reflect predictions made by models of adaptive dispersion, which hypothesize that newly established categories disperse in acoustic space to avoid crowding neighboring sounds to maintain contrasts (see e.g., Flege, 2007; Johnson, 2000; Liljencrants & Lindblom, 1972; Lindblom, 1986, 1990; Livijn, 2000). Instead, the phonologies of mixed languages are fraught with "near-mergers, overlapping categories, categorical assimilation, categorical maintenance, and overshoot of target categories at the segmental level, in addition to prosodic assimilation, possible preservations of archaic patterns, and innovation at the suprasegmental level" (Stewart & Meakins, accepted).

2 Media Lengua (ML)

One of the most studied mixed languages, from a phonetic standpoint, is ML, a mixed language spoken in the northern Ecuadorian province of Imbabura. ML is a lexical–grammar mixed

Figure 1. Communities and cities of Imbabura, Ecuador where data for this study were collected: Pijal for Media Lengua; Cashaloma and Chirihuasi for Quichua; and Ibarra for Spanish. This map is freely licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF).

language (Meakins, 2013; Meakins & Stewart, accepted; Muysken, 1997), which combines Quichua's[2] agglutinating morphology and word order with the Rural Andean Spanish lexicon that has replaced over 90% of Quichua content words primarily through the process of relexification (Deibel, 2017, 2019, 2020; Gómez-Rendón, 2005; Lipski, 2016; Muysken, 1981, 1997; Stewart, 2011, 2015b). ML appears to have formed in the early 20th century and is still currently spoken today by approximately 2000 people in several communities near Lago San Pablo. The data from this study come from the community of Pijal, where the language appears to have originated in Imbabura (see map in Figure 1).

Example (1)[3] provides a sample of ML; the elements in boldface type are of Spanish origin. The first line contains the ML orthography, the second a broad International Phonetic Alphabet transcription with morpheme boundaries, the third the interlinear glosses, and the fourth and fifth lines provide translations in Imbabura Quichua and Spanish, respectively.

(1) **Mio** **ciudad puente**ca **derrumba**rircami.
mio siudad puente-ka dez̢umba-ɾi-ɾka-mi
1.POSS city bridge-TOP collapse-REF-PST-VAL
Ñuka llakta chacaka tuniririrkami. (Imbabura Quichua)

> *El puente de mi cuidad se derrumbó*. (Spanish)
> "My city's bridge collapsed."

This paper adds to this growing literature with a quantitative acoustic analysis of Spanish origin and Quichua origin vowel sequences in ML. Moreover, our study provides important insights into the functionality of an exceptional vowel system that opposes predicted arrangements in acoustic space.

2.1 The ML vowel system

2.1.1 Monophthongs. The ML's vowel system is a combination of Quichua's three-vowel system, consisting of /i, u, a/, and Spanish's five-vowel system, consisting of /i, u, e, o, a/. Stewart (2014) describes this system as a complex case of stratification based on the language of origin of the lexeme where a given vowel is found. His analysis of formant 1 (F1) and formant 2 (F2) measurements involved 2515 tokens extracted from ML elicited phrases based on the speech of 10 participants. Results revealed small, albeit significant differences, between Spanish origin and Quichua origin corner vowels (i.e., /i, u, a/) where Spanish origin /i, u, a/ were on average just 13 Hz less centralized in acoustic space compared to Quichua origin /i, u, a/. While at first glance, this difference may appear trivial, the dispersion patterns of the F1 in all the Spanish origin corner vowels fall in line with the directions predicted by models of adaptive dispersion (i.e., the F1 values of the Spanish origin vowels extend further into acoustic space compared to those of Quichua origin), but not with the degree of dispersion expected for contrastive purposes (see e.g., Flege, 2007; Johnson, 2000; Liljencrants & Lindblom, 1972; Lindblom, 1986, 1990; Livijn, 2000, for details on adaptive dispersion theories). This means that at a "micro" level, ML speakers produce consistent differences in lexically non-contrastive vowels of the same quality (similar to near-mergers or covert contrasts) based solely on the language of origin.

Additionally, Stewart (2014) also shows that ML speakers produce Spanish origin mid-vowels in a significantly higher F1 range (i.e., with a lower tongue body position) when compared to Spanish origin and Quichua origin high-vowels (Figure 2A). However, what makes this part of the system interesting is that, like the corner vowels described above, the mid-vowel and high-vowel categories also exist with substantial overlap; yet with just enough distance (0.36 bark, 41 Hz) to meet the threshold of 0.3 bark, established by Kewley-Port (2001), for possible formant discrimination for values between 200 and 3000 Hz. In fact, it was later confirmed by Stewart (2018) that ML speakers are able to consistently identify differences between Spanish origin mid- and high-vowels in a two-alternative forced choice identification task experiment using minimal pairs (Figure 2B).

The results of these studies paint a picture of a system that may operate up to eight vowels,[4] yet it is arranged in an acoustic space that reflects that of a three-vowel system such as Quichua (see Figure 3). Stewart and Meakins (accepted) hypothesize that this unconventional system could have arisen during the genesis of the language given that the originators of ML most likely acquired Spanish as late bilinguals and thus spoke a Quichua-accented Spanish. However, instead of having the Spanish mid-vowels fully assimilate to Quichua high-vowels, enough distance was allotted to maintain some degree of contrast to benefit the newly relexified vocabulary, which brought in contrasting mid- and high-vowels with high phonological function loads. For instance, the maintenance of some Spanish phonemes avoids the realization of possible ambiguous words if complete assimilation to Quichua phonology were instead the case, for example, we might observe the word *pisu* or *pizu*, which could mean *beso* "kiss," *peso* "weight," *piso* "floor," *bezo* "part of the lip," etc. This overlapping vowel system would have then become nativized, and thus normalized, in subsequent generations of speakers.

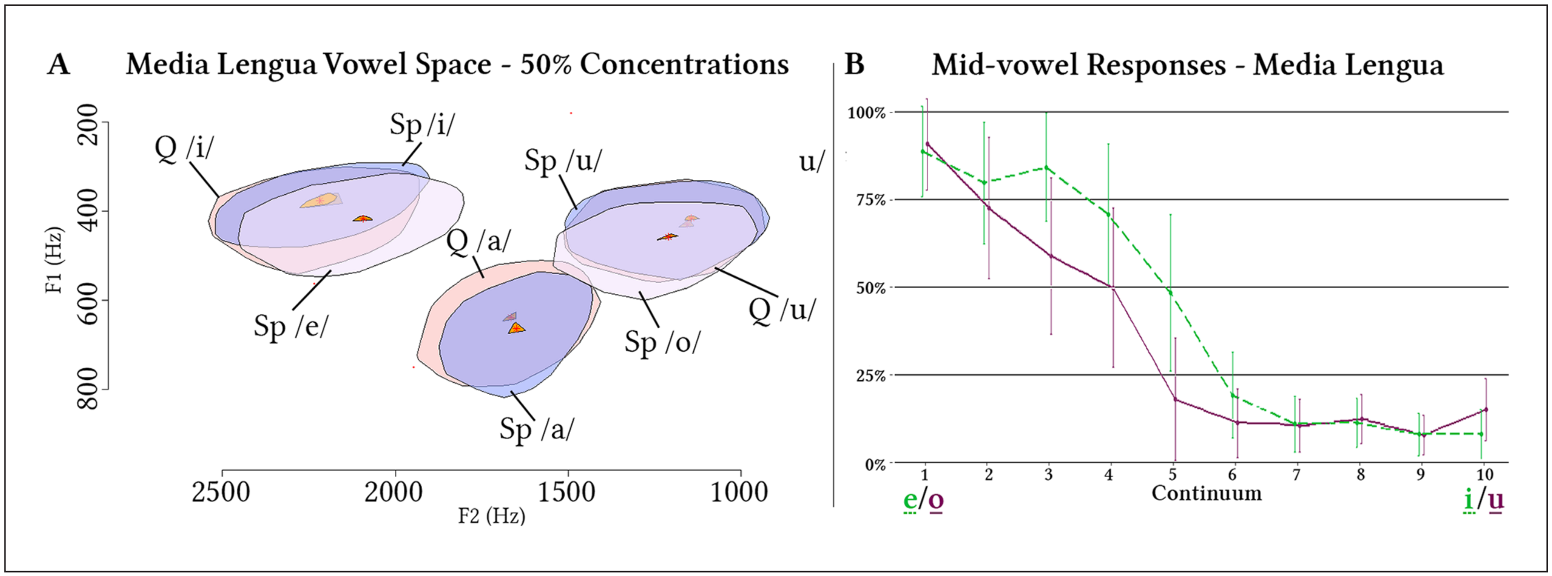


Figure 2. Image A—Media Lengua (ML) vowel space based on 50% concentrations of F1 and F2 (Hz) measurements of statistically different vowel clusters based on Stewart (2014); and Image B—ML results from a two-alternative forced choice identification task experiment with minimal pair stimuli modified along a 10-step continuum between prototypical ML mid-vowels to prototypical ML high-vowels based on Stewart (2018).

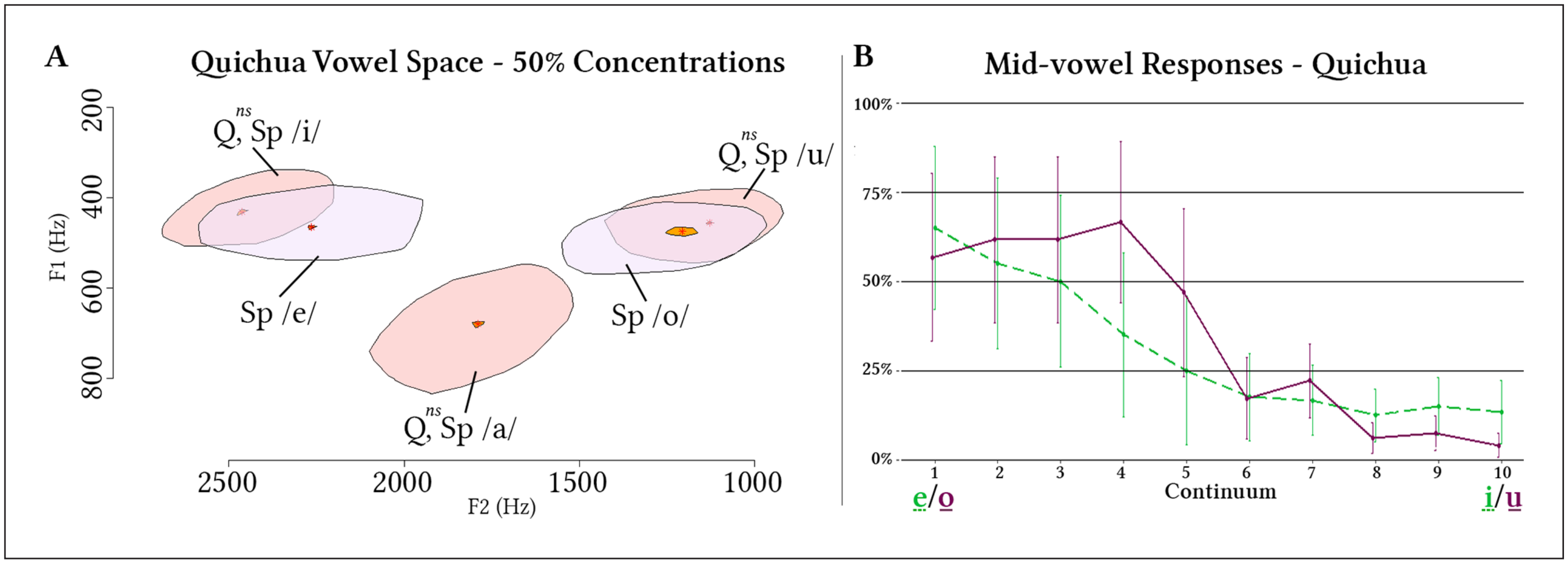


Figure 3. Image A—Quichua vowel space based on 50% concentrations of F1 and F2 (Hz) measurements of statistically different vowel clusters based on Stewart, 2014; and Image B—Quichua results from a two-alternative forced choice identification task experiment with minimal pair stimuli modified along a 10-step continuum between prototypical Media Lengua (ML) mid-vowels to prototypical ML high-vowels (nearly identical to Quichua prototypical values) based on Stewart (2018).

Evidence to support this claim also appears in the Quichua vowel system (Stewart, 2014) which has adopted approximately 40% of the Spanish lexicon (see Bakker & Muysken, 1994) over the last 500 years of intense contact. However, unlike ML, Spanish origin corner vowels are shown to be non-significant from native Quichua corner values (implying assimilation, and thus merged in Figure 3A). Moreover, while Spanish origin mid-vowels are significantly different from native Quichua high-vowels and Spanish origin high-vowels, like in ML (see Figure 3A), in Quichua they do not reach the threshold of 0.3 bark, established by Kewley-Port (2001), for possible formant discrimination for values between 200 and 3000 Hz. Indeed, with a difference of just 0.23 bark (26 Hz), Stewart (2018) shows that Quichua speakers do not consistently identify Spanish origin mid-vowels but, as would be expected, consistently identify high-vowels of Spanish origin, which are non-significantly different from native Quichua high-vowels (see Figure 3B).

2.1.2 Vowel sequences. This study focuses on what we term "vowel sequences," by which we mean vowel-like articulations containing two target positions/qualities. The term "vowel sequences" is deliberately imprecise, and intended to be taken agnostically, such that its referents may include what are in fact two different sets, from a phonological point of view: (a) sequences where both members are distinct phonological vowels, sometimes referred to as a hiatus; and (b) sequences where one member is a vowel and one member is a glide—that is, a canonical diphthong. This distinction is, however, less clear than it might initially appear. Similarities between glides and high-vowels in particular have led some phonologists to describe segmental pairs such as /i, j/ as being featurally identical aside from consonantal status (Padgett, 2008). Sánchez Miret (1998) argues that monophthongs, diphthongs, and hiatuses form a cline where the distinction between adjacent categories is one of degree rather than kind. Therefore, "vowel sequences" as units in and of themselves are of primary interest as changes in vowel quality (i.e., a "true" diphthong), component vowels affecting each other's realization, or even the presence of adjacent vowels in a syllable nucleus (i.e., hiatus), can all be considered useful contrastive structures in the phonology of a language.

However, intense language contact can call into question the functionality of previously contrastive units in lexical borrowings through processes such as assimilation, elision, or reinterpretation. Therefore, acoustic analyses and cross-language comparisons of phonetic correlates such as formant trajectories of borrowed vowel sequences in ML and Quichua will lend to a better understanding of how such units are borrowed, interpreted, and produced. This not only allows us to identify phonological processes such as assimilation, but also gradient modifications which reveal how borrowed sounds conform to the phonological constraints of a new language. Through the phonetic analysis of a large number of vowel sequences we can also identify consistent and variable articulatory gestures (in F1 × F2 space), which can correlate to the importance of a given vowel sequence. For example, variable articulatory gestures might suggest that the sequence carries a low functional load (i.e., contrastive importance is minor) whereas consistent articulatory gestures might suggest that the sequence is more important for discerning meaningful units. Understanding the precise phonological status of each vowel sequence, however, remains an area of future research, and is not something which we intend to explore within the confines of this study.

For ML and Quichua, Spanish origin vowel sequences were organized based on their original Spanish pronunciation. For example, in ML, the word *tierra* "earth, ground, dirt" might be pronounced as [ˈtie̝z̨a] or [ˈtiz̨a], yet the vowel sequence would be considered as /i͜e/, since its prelexified production was that of /i͜e/ as in Andean Spanish *tierra* [ti͜er̝a] "earth, ground, dirt." Where such individual pronunciation differences between ML versus Spanish occur, our analysis method is designed to determine whether they reach the level of statistical significance within the aggregated dataset.

Table 1. Potential Media Lengua vowel sequences. Sequences in boldface type are uniquely of Spanish origin.

a-	e-	i-	o-	u-
ae	**ea**	ia	**oa**	ua
ai	**ei**	**ie**	**oe**	**ue**
ao	**eo**	**io**	**oi**	ui
au	**eu**	iu	**ou**	**uo**

Based on its five-vowel system, ML has the potential to contain 20 distinct non-identical vowel sequences, which are listed in Table 1. Considering ML's lexical source languages, six of these vowel sequences are shared between both Spanish and Quichua, while 14 occur only in Spanish (boldface type in Table 1).

All 20 identified sequences were found throughout ML's documented lexicon (see Stewart et al., 2020a), with the exception of /ou/. The lack of this particular sequence is unsurprising given that /ou/ is exceptionally rare and only present in low frequency words in Spanish, often of foreign origin. Moreover, when found in Spanish orthography, the <ou> sequence is produced as one of /u, o, au/ depending on the word, for example, *boutique* [buˈtik] "boutique." Similarly, the frequency of some vowel sequences is less common than others. Because of this, our study excludes three existing but low-frequency ML vowel sequences: /iu/ with only two lexemes in our data (*viuda* "widow" and *ciudad* "city"); /oi/ which only appeared in four words (*hoy* "today," *proibido* "prohibited," *oina* "listen," and in the name *Zoila*); and, /oe/ which was only found in one word (*oeste* "west").

3 Method

The aim of this study is to investigate the hypothesized integration and acoustic arrangement of Quichua and Spanish origin vowel sequences in ML. To do so, we conduct an instrumental study of formant trajectories in and across all three languages.

3.1 Materials

The number of vowel sequence tokens under analysis in this study totals 1096 across all three languages. ML vowel sequence data were gathered from three sources, totaling 406 tokens. The first source (corpus 1) comes via a 2010–2012 corpus of elicited phrases and contains 150 tokens (37%). The second source (corpus 2) comes via a 2010–2014 corpus of wordlist/sentence list data and contains 195 (48%) tokens. In this corpus collection, participants were informed that they would be asked to read several short sentences (for both the Quichua and ML groups) or words (for the Spanish groups) on a computer. The ML and Quichua data were presented in sentences to avoid possible "switches" in language mode as individual lexical items in isolation may be ambiguous as to their source (e.g., *escuela* "school" is the same word in all three languages). Spelling for all three languages was based on Spanish orthography as the unified Quichua system was still unfamiliar to many Quichua and ML speakers. If a participant could not read (a rare occurrence), the author or the assistant read the sentence and asked the participant to repeat it twice. The second utterance was used for analysis. If a participant struggled with reading, they would be asked to repeat the sentence from memory allowing for a more naturalistic speech sample.

Table 2. Speech corpora token quantities.

Corpus	Elicitation type	Media Lengua	Quichua	Spanish
Corpus 1	Phrases	150	—	—
Corpus 2	Word/sentence lists	195	166	524
Corpus 3	Natural speech	61	—	—

The last source (corpus 3) comes via a 2015–2019 corpus of natural speech data and contains 61 tokens (15%). The Spanish and Quichua data come solely via corpus 2, containing wordlist/sentence list data: 166 (100%) tokens for Quichua; and 524 (100%) tokens for Spanish. Quichua has a lower token count as it contains roughly one-third of the number of vowel sequence combinations as Spanish and ML. Token quantities in our data are summarized by language and corpus in Table 2.

During the elicitation sessions for corpus 1, participants were asked to give their best oral translation of a sentence into ML. The participants' oral translations were recorded on a TASCAM DR-1 portable digital recorder using TASCAM's compatible TM-ST1 MS stereo microphone set to 90° stereo width. Elicitations were recorded in 16-bit Waveform Audio File (WAV) format with a sample rate of 44.1 kHz. During the wordlist and sentence list sessions for corpus 2, participants from each language were asked to read words or sentences in their native language from a computer screen. For the Spanish participants there was a mix of sentence list data and wordlist data. However, ML and Quichua required sentences to prime the target language; for example, seeing the word *caña* "cane/stick" instead of *cañahuanmi* "cane-INST-VAL" might elicit a more Spanish-like pronunciation since ML and Quichua speakers are bilingual. The participants were recorded on a TASCAM DR-1 portable digital recorder using a NEXXTECH unidirectional dynamic microphone (50–13,000 Hz response). The wordlist was recorded in 16-bit WAV format with a sample rate of 44.1 kHz. Finally, corpus 3 was created by two assistants from Pijal who recorded the data. Participants were asked to gather in pairs and converse about a topic of their choosing. These data were recorded on a Zoom H6 Handy Recorder with its internal microphone (unidirectional XYH-6 capsule). See Stewart (2015b) for more details on the data collection procedures for corpora 1 and 2.

3.2 Participants

For ML, 23 trilingual speakers (Quichua, ML and L2 Spanish) participated in this study. This group consisted of fourteen women and nine men. All participants were from the community of Pijal Bajo and acquired Quichua and ML simultaneously from birth and began learning Spanish upon entering primary school, typically at 6–7 years of age.[5]

Ten Quichua-speaking participants also participated in this study, all of whom were bilinguals (Quichua and L2 Spanish) ranging from low to high proficiency in Spanish.[6] This group consisted of six women and four men. Four women had a rudimentary level of Spanish, one man and one woman were simultaneous bilinguals, and one man acquired Spanish at the age of 18, while the rest acquired Spanish upon entering primary school, typically at 6–7 years of age. Participants were born, raised, and lived in the neighboring communities of Chirihuasi and Cashaloma at the time of recording. These slightly more distant communities were chosen to gather Quichua data over Pijal and neighboring communities to avoid any influence from ML on the Quichua speech of these participants.

From the Spanish group 14 participants took part in this study and all were monolinguals except two with late English L2 acquisition. Eight participants were female and six were male. All participants had graduated from college and were born, raised, and lived in urban centers (11 from Quito and three from Ibarra) at the time of recording. More demographic information pertaining to all participants is listed in the Appendix.

3.3 Procedures

The procedure followed in this study has been successfully implemented in several contexts, including: the comparison of diphthong productions across different phonetic contexts (Onosson, 2018); the cross-dialect comparison of diphthong productions (Onosson, 2018); and, the investigation of vowel-glide sequence production differences between L1 and L2 speakers (Onosson & Bird, 2019). This study potentially represents the first application of this particular methodology to production differences across different languages, and to lexical items deriving from different source languages, and is therefore informative not only for the specific results pertaining to ML, but also with regards to the methodology itself and its potential for application to a diversity of linguistic contexts.

3.3.1 Vowel extraction. Each target vowel sequence token was manually segmented in Praat (v6.1.04; Boersma & Weenink, 2019) and acoustic data were extracted using a Praat script (Xu & Gao, 2018) which takes discrete formant measurements at specified intervals. For this study, we retrieved F1, F2, and F3 measurements at 5% intervals across vowel duration, yielding 20 discrete measurements per formant per token. This method permits high-fidelity reconstruction of the original formant trajectories versus more coarse methods which utilize fewer (e.g., 2, 3, or 5) such measurements.

3.3.2 Generalized additive mixed models (GAMMs). Following formant extraction, the data were analyzed in R (v3.6.1; R Core Team, 2019) using RStudio (v1.2.5001; RStudio Team, 2019). Our primary method of analysis uses generalized additive models (GAMs; Hastie & Tibshirani, 1990; Wood, 2017). GAMs are designed to analyze and compare dynamic non-linear data, such as the formant trajectories of diphthongs or vowel sequences. A GAM computes a "smooth" across the dynamic component of the data under investigation, in this case being the temporal dimension, which is to say, vowel duration. A pair of smooths can then be directly compared, allowing the computation of a "difference smooth" which indicates temporal regions where the two smooths differ at a given confidence level, typically made at 95%.

The addition of mixed-modeling to GAMs is referred to as generalized additive *mixed* models or GAMMs and allows the inclusion of random effects to cover such factors as unconstrained inter-speaker variance. Our GAMMs analysis was implemented with the bam function of the mgcv package (Wood, 2011) using the formula in Example (2), modeled after Sóskuthy (2017):

(2) GAMMs formula code:

```
Formant ~ Language +
s(Time, bs = "cr") +
s(Time, by = Language, bs = "cr") +
s(Time, Speaker, bs = "fs", m = 1) +
s(Time, Word, bs = "fs", m = 1)
```

The main effect tested in each GAMMs analysis is the relationship between *Formant* as the dependent variable and *Language* as the independent variable. This correlation was conducted across both F1 and F2 for all available languages; wherever possible, all three languages were compared against each other, but as a number of vowel sequences do not occur in Quichua (see subsection 2.1.2), only the relevant ML and Spanish data could be compared for those sequences. Following the main effect, the GAMMs formula includes a series of smooths, each prefaced by the letter s in the formula in Example (2). The first smooth considers only the *Time* value at each formant measurement, taken as a percentage of overall sequence token duration. This method normalizes across individual token durations and thereby accommodates durational variance between different vowel + vowel (V+V) sequences. The second smooth additionally computes separate *Time* smooths per language, which is intended to further accommodate systematic differences that may exist between languages. Finally, two random smooths are also computed: one for *Speaker*; and one for *Word*. The specific coding here, which includes the use of bs = "cr" to specify the use of a *cubic regression spline* as the basis function for fixed effects smooths and *factor smooth interactions* (bs = "fs") for random effects smooths, follows recommendations by Sóskuthy (2017). The addition of a random smooth for *Speaker* is commonplace in mixed methods, as it allows the existence of individual speaker variation to be incorporated into the model. The random smooth for *Word* is included as a proxy for differing phonetic contexts between V+V tokens. The dataset for our study derives from elicited forms which were selected specifically for the presence of V+V sequences, and do not systematically cover all the possible phonetic or phonological contexts in which a V+V sequence could potentially occur in for every language. In fact, because the phonotactic restrictions on V+V sequences differ between the three languages, and because of differences in the range of possible phonetic environments, which are far greater in Spanish versus Quichua, it would be impossible to build a comprehensive set of elicited forms that covered all possible phonetic contexts, and which were equivalent between all three languages. Because of these inherent limitations, we elected to provide *Word* as a random effect as a means of allowing the specific phonetic context of each elicited form to be included in the model, albeit without necessarily being able to comprehensively compare like contexts between different languages in a more robust and systematic way. In any case, we do not expect phonetic context to exhibit a strong effect on V+V formant production except at the onset and offset points—the central phase of the sequence's formant trajectories would not be expected to be overly impacted by adjacent segment context, in the majority of cases. We do, however, acknowledge this as a limitation of our methodology, one which could be addressed in a future study with more comprehensive coverage of phonetic context within the elicitation materials.

Our GAMMs comparison method involves the following. For each formant of each V+V sequence, two models are computed: a model which incorporates the main effect of *Language* (M1); and a null model that excludes it (M0). M1 uses precisely the formula listed above, while the formula for M0 removes the sections containing the main effect of *Language* and its associated smooth. The compareML function of the itsadug R package (van Rij et al., 2020) is then used to conduct a comparison between M0 and M1. This function compares each model's maximum likelihood and degrees of freedom (*df*) results using a Chi-squared (χ^2) test and reports a *p*-value for the overall comparison, which indicates whether the χ^2 result is deemed significant. With respect to the difference between the two models, which centers around the inclusion versus exclusion of *Language* as a main effect, we interpret the *p*-value from this comparison as an "indicator of surprise" rather than in its traditional use as an indicator of statistical significance, reflective of the exploratory nature of this study (see Baayen et al., 2017, on the evaluation of the GAMMs comparison methods and Roettger et al., 2019, on significance-testing in confirmatory versus exploratory studies).[7] Our model comparison is fundamentally focused on one question: *Do the set of formant trajectories derived from tokens of a particular V+V sequence differ according to language?* As such, we have endeavored to keep our model structure relatively simple, such that success or failure of the model should be clearly interpretable as pertaining to the inclusion/exclusion of the effect of *Language*. Following the indication by the χ^2 test result that *Language* is a significant effect, we then analyze the GAMMs difference smooth, or smooths in the case of multiple language-to-language comparisons. This allows identification of specific regions across the vowel sequence duration where the formant trajectories of each of the two (or three) languages differ robustly from each other. An illustration of the application of this method follows below.

To illustrate, we present a comparison of the F1 trajectories of /au/ across the three languages in our study using the procedure outlined above. The M0 versus M1 model comparison for /au/ F1 indicates that M1, which includes the main effect of *Language*, is a better fit than M0, $\chi^2 = 10.142$, $df = 6$, $p = 0.002$. Within M1, the difference between ML versus Quichua is non-significant, $p = 0.0167$, while the difference between ML versus Spanish is significant, $F = 4.002$, $p = 0.00133$. Figure 4 shows the difference smooth comparisons from M1. In Figures 4A and 4B, the ML formant trajectory is represented by the thin, solid black curved line which indicates the computed GAMMs smooth for that trajectory. The horizontal zero-line on the y-axis represents the reference level for each comparison—in Figure 4A this is the Quichua vowel sequence formant trajectory, and in Figure 4B it is the Spanish trajectory. The plotted smooth for ML thus indicates where (i.e., *when*) and in which direction the ML trajectory diverges from the comparison reference trajectory; the shaded gray regions in Figures 4A and 4B represent 95% confidence intervals for the difference smooth. In Figure 4A, the ML confidence interval fully overlaps the reference level zero-line, which indicates that the ML F1 trajectory does not differ substantially from Quichua at any point across the duration of the vowel sequence, a reflection of the non-significant difference between the two languages reported for M1 under the χ^2 test. The regions where the confidence intervals depart from the zero-line indicate regions of particularly robust difference between the two trajectories under comparison. This region is highlighted in the difference smooth by vertical dashed lines at the edges of the region of difference, accompanied by a red bar along the x-axis. Such a region appears in Figure 4B from the beginning of the sequence until approximately 70% of duration, indicating that the ML F1 trajectory differs most substantially from that of Spanish during this portion of production of the sequence.

Finally, a visual comparison across multiple smooths can be achieved via a comparison plot which overlays each smooth and its associated confidence intervals. Figure 4C shows the comparison of the /au/ F1 smooths across the three languages, where it can be readily observed that the ML and Quichua productions follow a very similar trajectory, while Spanish diverges quite substantially, occurring largely during the initial and medial portions of the trajectory.

4 Results

4.1 Language-internal GAMMs comparisons

The initial set of GAMMs comparisons were run on vowel sequences occurring within the ML data to investigate potential differences between Quichua origin versus Spanish origin lexical items containing the same vowel sequence. Only six distinct vowel sequences occur in the native Quichua lexicon (/ai, au, ia, iu, ua, ui/), and the Quichua origin words in ML do not fully represent this set. As a result, we were able to identify only three sequences occurring across both Quichua and Spanish origin words in ML: /ai, au, ui/.

The GAMMs comparisons were made between F1 and F2 trajectories across the constituent different-source-language lexical items for each sequence, such that, for example, tokens of /au/ in ***tau**ca* [ˈtau.ka] "a lot, much, many" (Quichua origin) were compared to tokens of *b**au**tismo* [bau.ˈtis.mo] "baptism" (Spanish origin).[8] Main-effect models (M1) were built for each

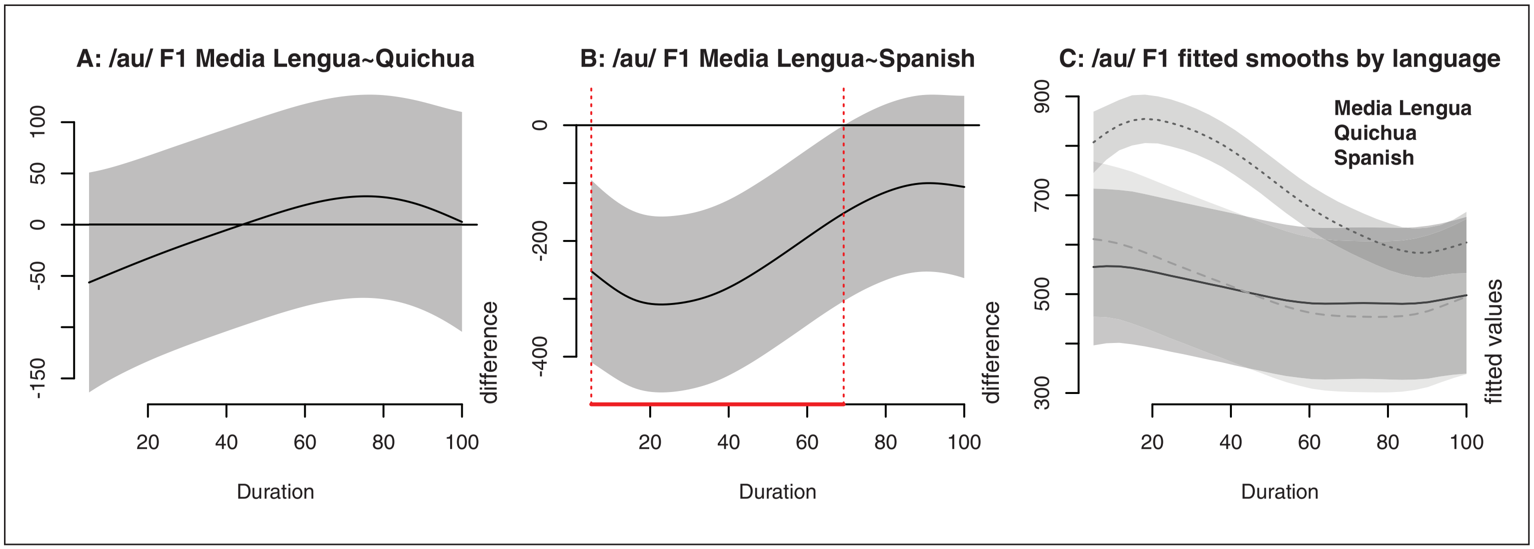


Figure 4. /au/ F1 (Hz) generalized additive mixed models' three-language comparison. In 4C, the solid line indicates the smooth for Media Lengua; the dashed line the smooth for Quichua; and the dotted line the smooth for Spanish.

sequence-formant with *Source Language* as main effect, and an analysis of variance (ANOVA) compared that model's output with a null model (M0) for the same vowel sequence formant. None of these comparisons indicated any significant differences according to *Source Language*. In other words, ML vowel sequence productions do not differ significantly between lexical items from different source languages, but instead represent a single phonological system with similar productions across all similar vowel sequences, regardless of language of origin.

Because ML represents the only dataset in our study derived from all three corpora listed in Table 2, and thus containing a variety of recording types, we ran an additional set of GAMMs models (M1) with *Recording Type* as main effect. ANOVA comparisons of the per-sequence M1 models versus parallel M0 models did not find any significant differences in this case, either, indicating that recording type (and presumably speech register or style) is not correlated with production differences for these vowel sequences.

4.2 Cross-language GAMMs comparisons

In Table 1 we identified 20 vowel sequences which exist across ML and Spanish (six of which are also present in Quichua). Of these, /iu, oe, oi, ou/ were excluded from further analysis due to non-occurrence or low token counts in our data (see subsection 2.1.2), leaving the set of sequences listed in Example (3):

1. (3) /ae, ai, ao, au, ea, ei, eo, eu, ia, ie, io, oa, ua, ue, ui, uo/

The GAMMs models including *Language* as main effect were computed for each formant, F1 and F2, of each sequence in Example (3), along with a parallel null model which excluded the effect of *Language*. Comparisons of each null versus non-null model pair (see subsection 3.3.2 for discussion of this methodology) indicated that the *Language* model was superior to the null model (where the comparison *p*-value was below 0.05), for either one or both formants, in eight of the 16 sequences investigated; these are listed in Table 3. Where the ANOVA indicated a significant difference between models, thus confirming that the addition of *Language* as a main effect improved the GAMMs model, we investigated the GAMMs output to identify which languages differ from each other, where more than two languages are involved in the comparison, and critically which specific portions of the formant trajectories differ.

The upper half of Table 3 lists sequences for which Quichua data were available, and thus for which GAMMs comparisons were carried out across all three languages.[9] The lower half of Table 3 lists sequences for which only ML and Spanish data were available for comparison. These two groupings are discussed separately in subsection 4.2.2 and subsection 4.2.3. Before that, we present a descriptive account of the formant trajectories for these eight sequences, in subsection 4.2.1.

4.2.1 Vowel sequence trajectories. In subsection 4.2 we established via GAMMs analysis significant differences between ML and ML's two lexical source languages, Spanish and Quichua, in the formant trajectories of eight distinct vowel sequences (see Table 3). Before examining the GAMMs results in more detail, in this subsection we first take a closer look at this set of sequences by plotting their trajectories in F1 × F2 space.

In Figure 5, the trajectories of the eight vowel sequences listed in Table 3 are plotted in F1 × F2 space along with the five ML monophthongs in the background to provide context. For the monophthong data, we use formant measurements extracted at 50% of duration; the positions of the monophthong vowel labels indicate the median F1 and F2 values at this point, and the ellipses represent one standard deviation. For the vowel sequence trajectories, the text label (e.g., "ia")

Table 3. Cross-language vowel sequence generalized additive models' comparisons where the model including *Language* as the main effect is superior to the null model at $p < 0.05$.

Languages compared	Sequence	F1	F2
Media Lengua (ML), Quichua, Spanish	/au/	$\chi^2 = 10.142$, $df = 6$, $p = 0.002$	—
	/eo/	$\chi^2 = 6.337$, $df = 6$, $p = 0.049$	—
	/ia/	$\chi^2 = 11.417$, $df = 6$, $p < 0.001$	$\chi^2 = 12.376$, $df = 6$, $p < 0.001$
	/ie/	—	$\chi^2 = 18.153$, $df = 6$, $p < 0.001$
	/ue/	—	$\chi^2 = 14.024$, $df = 6$, $p < 0.001$
ML, Spanish	/ei/	$\chi^2 = 5.401$, $df = 3$, $p = 0.013$	$\chi^2 = 8.735$, $df = 3$, $p < 0.001$
	/io/	$\chi^2 = 13.795$, $df = 3$, $p < 0.001$	$\chi^2 = 4.771$, $df = 3$, $p = 0.023$
	/uo/	$\chi^2 = 7.574$, $df = 3$, $p = 0.002$	—

marks 20% of sequence duration, and the arrowhead marks 80% of duration; we excluded the initial and final 20% portions of the sequences, as coarticulatory effects are more apparent in these regions. Both ML and Spanish vowel sequence trajectories are included in each plot in Figure 5 and distinguished by grayscale-shading as shown in the figure, with ML appearing darker than Spanish; the Quichua /ie/ sequence from Spanish origin words is also included in Figure 5E (in the lightest shading), as this was the only sequence to exhibit a significant difference (restricted to F2, equivalent to articulatory front–back position) between Quichua and ML

Several differences in acoustic target positions between ML and Spanish can be seen in Figure 5. For /a/, which occurs in /au/ (plot A) and /ia/ (D), the ML target is substantially higher than Spanish, most especially so when it is the initial target in the sequence in /au/. For /e/ the overall pattern is less clear-cut. In /ei/ (B) the ML /e/ target is more retracted than in Spanish, but this is not the case for /eo/ (C) or /ue/ (G) where the targets appear quite close. And in /ie/ (E) the ML /e/ is higher and slightly more advanced than Spanish. ML /i/ is similarly complex in its realizations relative to Spanish. It is more retracted and higher in /ia/ (D) and /ie/ (E), similarly retracted but slightly lower in /ei/ (B), and nearly identical in /io/ (F). ML /o/ exhibits more variation, being substantially higher than Spanish in both /io/ (F) and /uo/ (H), but somewhat retracted and lower in /eo/ (C). Finally, Spanish /u/ exhibits a split pattern. It is fairly similar to ML /u/ in both /au/ (A) and /ue/ (G), but massively more retracted in /uo/ (H).

Regarding intra-language observations of ML, nearly all the vowel sequences in Figure 5 show a noticeable shift in vowel quality or some degree of diphthongization, in the phonological sense (e.g., /au/-A, /ei/-B, /eo/-C, /ia/-D, /io/-F, and /ue/-G). However, there is little in the way of formant movement in /ie/ (E) and /uo/ (H), suggesting that these sequences may be functioning more like monophthongs in ML.

Figure 6 is identical to Figure 5 except that it includes the Spanish monophthongs in the background as context. Cross-comparisons of Figure 5 and Figure 6 suggest that the relationship between Spanish and ML vowel sequences are not necessarily parallel in structure even when considering differences in shape and size of the vowel spaces. For example, there are differences in proportionality with respect to the angular difference (measured in degrees)[10] between monophthong averages and the initial and end point averages of V+V sequences. For example, the ML cline is four times as steep in /ei/ compared to its average cline between monophthongs /i/ and /e/ (a 30° trajectory in /ei/ vs. a 7.4° average difference between monophthongs /e/ and /i/), while the Spanish /ei/ cline is only two times as steep compared to the average cline between Spanish monophthongs /i/ and /e/ (a 22.2° trajectory in /ei/ vs. a 42.2° between /e/ and /i/). Within this sequence, there are also small proportional differences in the average distances

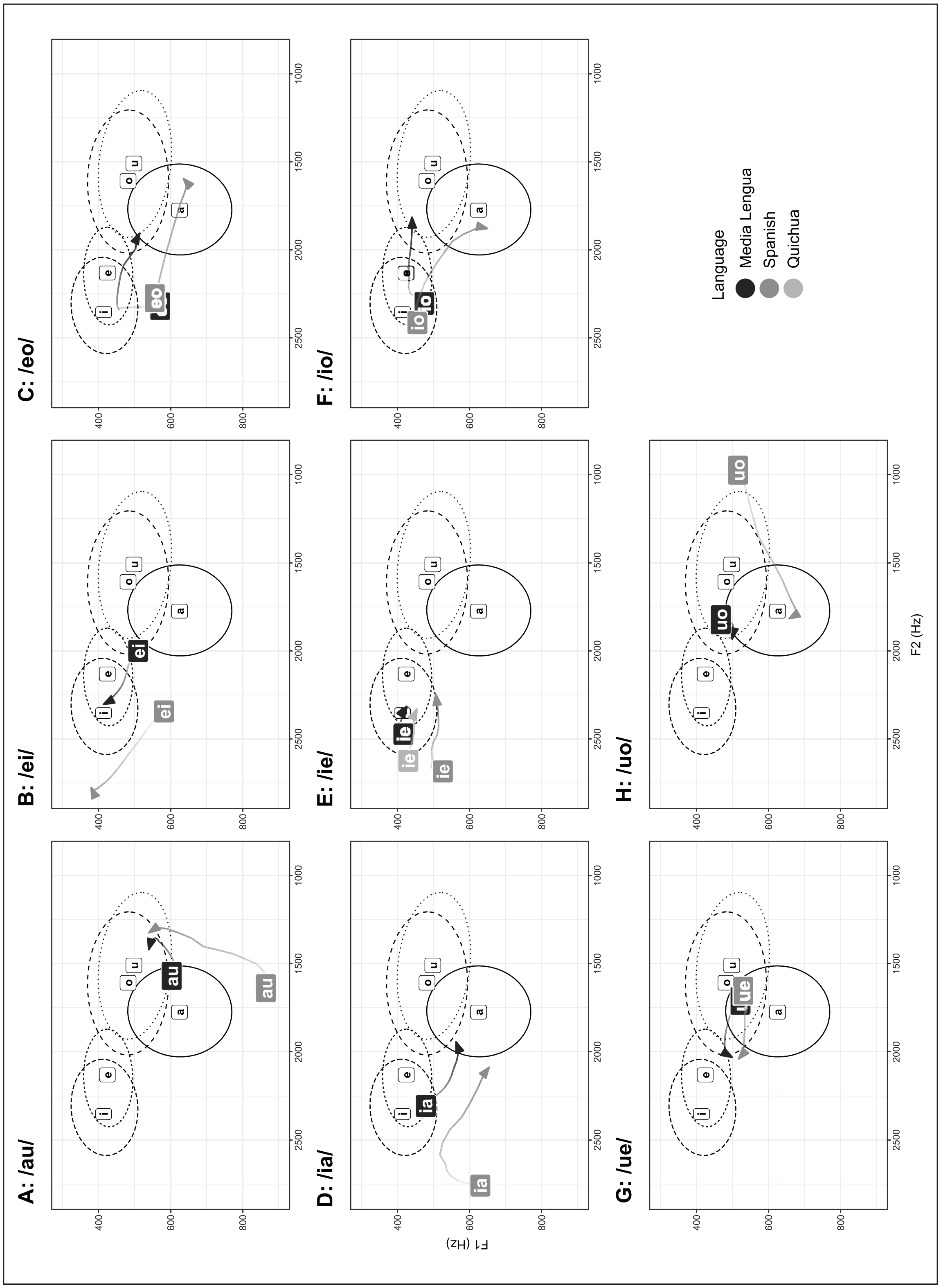


Figure 5. Vowel sequence 20%–80% trajectories overlaid on the Media Lengua vowel space; background ellipses indicate one standard deviation.

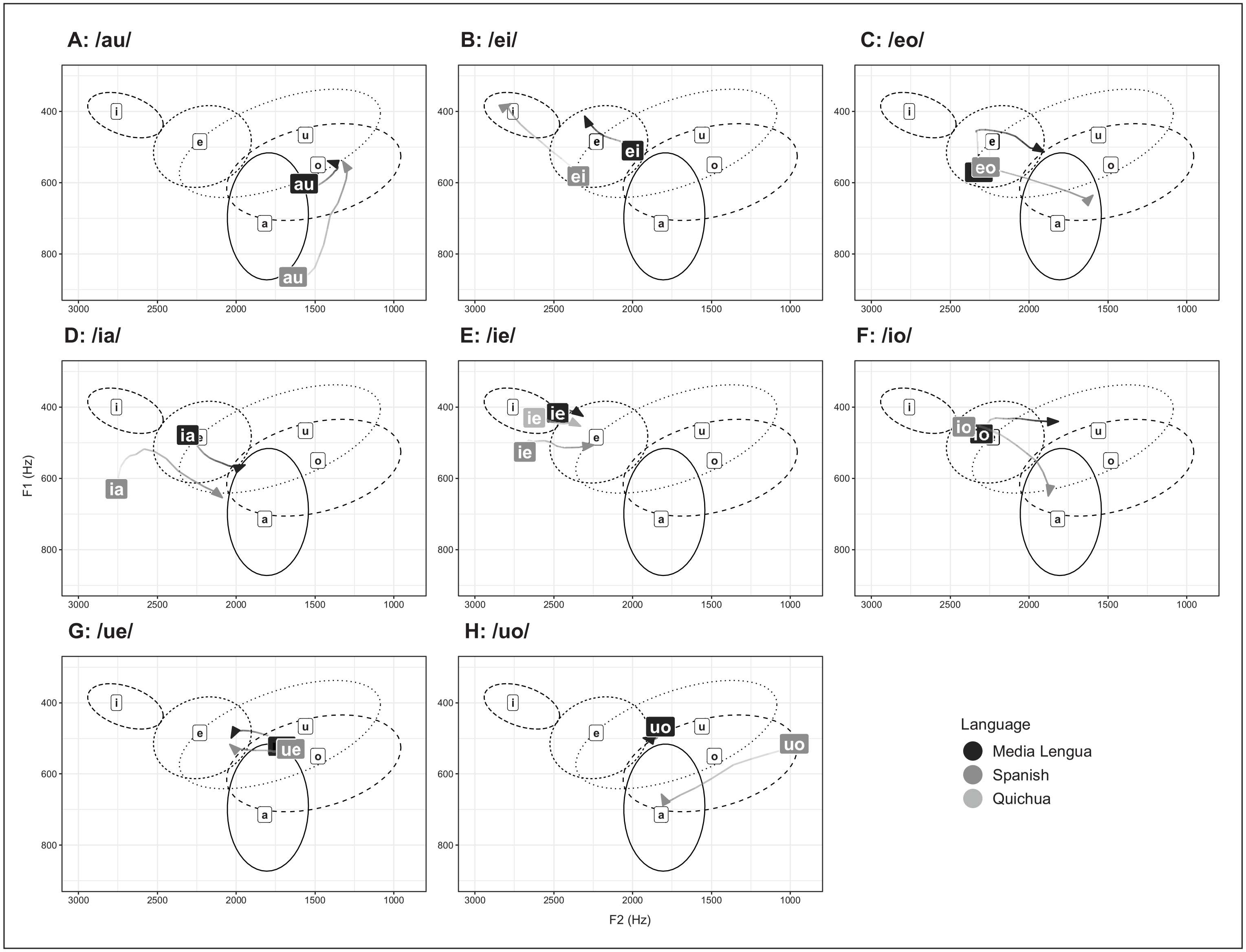


Figure 6. Vowel sequence 20%–80% trajectories overlaid on the Spanish vowel space; background ellipses indicate one standard deviation.

between monophthongs and V+V sequences in acoustic space. For instance, the distance between the average /e/ quality and the average /i/ quality in the /ei/ vowel sequence was 1.5 times more distant than the average distance between monophthongs /e/ and /i/, while Spanish showed a difference of just 1.25 times in the same analysis. These indicators may suggest that ML speakers reshaped the /ei/ trajectory to emphasize the change in vowel quality even though, when compared proportionally to Spanish, it is not immediately apparent. Similar proportional differences with respect to trajectory angles can be observed in /eo/ and /eu/ as well. However, two Spanish origin vowel sequences, /ie/ and /uo/, appear to have collapsed into monophthongs in ML (/i/ and either /o/ or /u/, respectively), as there is little in the way of dispersion (angular or distal) from the initial and end qualities in these sequences. This supports the hypothesis that Spanish and ML vowel sequences are not parallel in structure. Rather, ML has reshaped and reanalyzed Spanish origin vowel sequences to fit its phonological demands.

4.2.2 GAMMs comparisons across three languages. Five of the vowel sequences which indicated significant differences by *Language* in formant trajectories include data from Quichua as well as ML and Spanish, and are therefore most revealing of the relationship between ML and each of its contributory languages; those sequences are: /au, eo, ia, ie, ue/. Following computation of the GAMMs difference smooths between ML and each of Spanish and Quichua for F1 and F2 of each sequence, there was only a single instance of a significant difference in the M1 GAMM between ML and Quichua, F2 of /ie/; this will be discussed separately at the end of this subsection. The difference smooths between ML and Spanish, for these sequences according to which formants exhibited significant differences under GAMMs analysis, are shown in Figure 7. Note that F1 is negatively correlated with vowel height, such that the *lower* ML F1 productions in Figures 7A, B, and C are indicative of *higher* vowel articulations in terms of tongue position. And, F2 is positively correlated with vowel advancement, such that *lower* ML F2 productions (seen in Figures 7D and E) represent a more *retracted* articulation, while *higher* F2 productions (Figure 7F) represent a more *advanced* or *fronted* articulation. In general, ML vowel sequence productions show greater reduction, more centralization, and less aperture than their Spanish counterparts, which reflects the relatively smaller range of ML's overall vowel space relative to Spanish.

As mentioned above, F2 of /ie/ represents the only example in our data where a significant difference was found between ML and Quichua formant trajectories, illustrated in Figure 8. The difference smooth between ML and Quichua (as reference level) is shown in Figure 8A while the difference between ML and Spanish (as reference level) is repeated in Figure 8B (originally Figure 7E) for comparison. While ML F2 of /ie/ is lower (farther back, i.e., more retracted) than Quichua by between 100 and 200 Hz during the initial 63.5% of duration, this is a notably smaller degree of difference in comparison to that observed between ML and Spanish (Figure 8B), where the difference well exceeds 200 Hz for half of the sequence duration. These differences are further illustrated by layering the fitted per-language smooths into a single plot in Figure 8C, where the closer proximity between the ML and Quichua trajectories is readily apparent. In comparison, Spanish is substantially higher at onset and has a much steeper gradient during the transition towards offset. It is also worth reiterating that /ie/ in Quichua is non-native in origin, occurring only in Spanish origin words.

4.2.3 GAMMs comparisons across two languages. The other set of sequences exhibiting significant cross-linguistic differences in formant trajectories are: /ei, eu, io, uo/. As no Quichua data were available for these sequences due to low occurrence in any Quichua lexical items of Spanish origin, all differences reported in this subsection concern only ML versus Spanish. Recall that F1 is

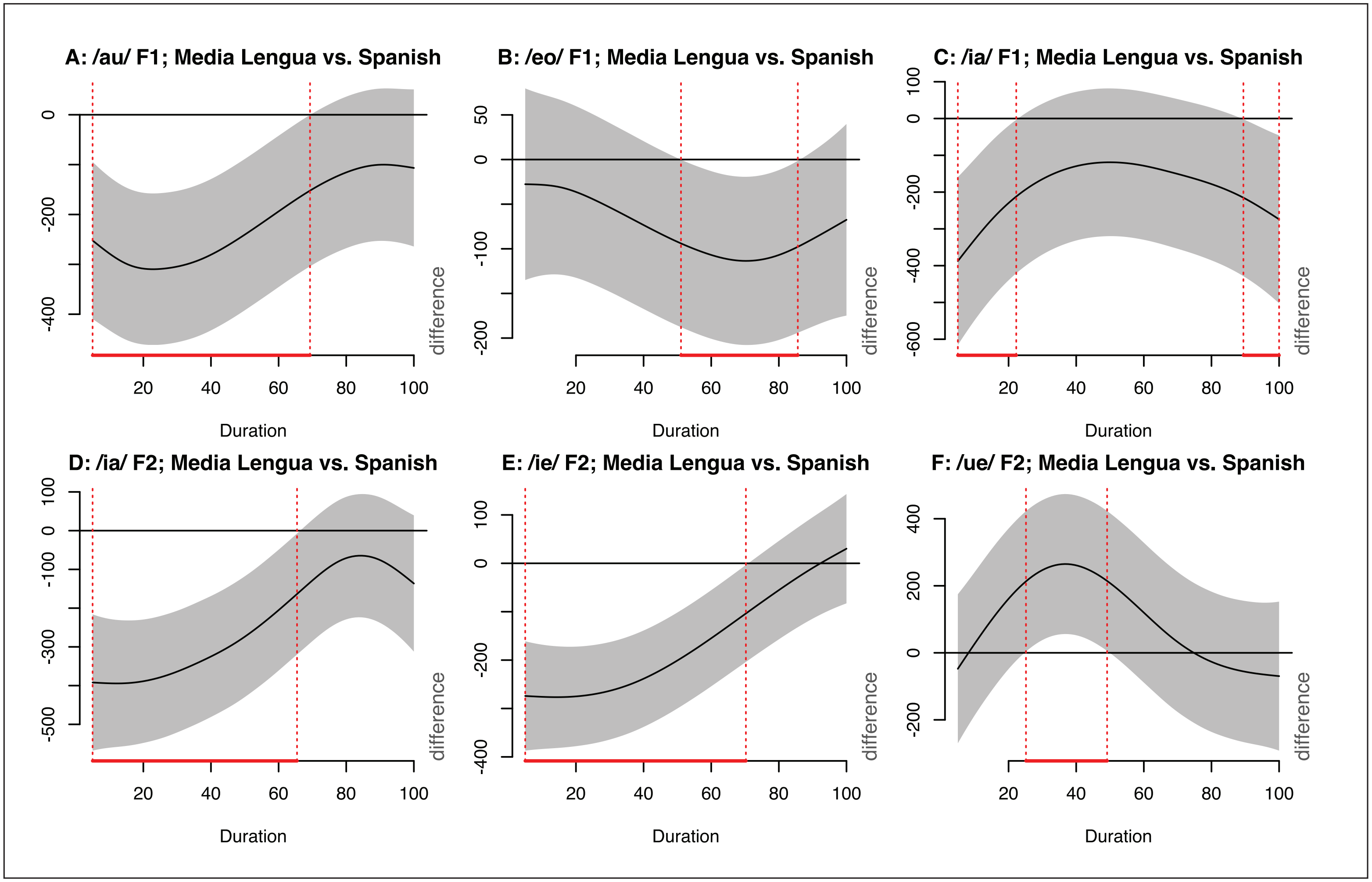


Figure 7. Formant (Hz) difference smooths: Media Lengua (ML) versus Spanish (three-language comparisons).

(A) F1 of /au/ (Figure 7A): The generalized additive mixed models (GAMMs) output indicates that ML F1 is lower than Spanish from onset until 69.3% of duration. This indicates that the ML articulation of at least the /a/ portion of /au/ is significantly *raised* in comparison with Spanish production;

(Continued)

Figure 7. (Continued)

(B) F1 of /eo/ (Figure 7B): The difference between ML F1, which is lower versus Spanish, with the GAMMs indicating the greatest degree of difference occurring between 51% and 85.6% of duration. This suggests that the difference is centered around the /o/ portion of /eo/, with ML having a *raised* articulation here relative to Spanish.

(C) F1 of /ia/ (Figure 7C): Some differences between ML versus Spanish production are most strongly indicated during onset (prior to 27.3% of duration) and just at offset (after 89.4% duration); due to their limited scope at the edges of the sequence, these differences might be assumed to be due to coarticulatory effects with adjacent consonants (see discussion in subsection 3.3.2 regarding the inclusion of *Word* as a random effect).

(D) F2 of /ia/ (Figure 7D): For this formant there is a more substantial difference from Spanish, with ML production being lower by several hundred Hz for the initial 65.5% of duration, indicative of a more *retracted* articulation centered around the /i/ portion of /ia/ and extending into the /a/ portion as well.

(E) F2 of /ie/ (Figure 7E): ML production is substantially lower than Spanish for the initial 70.3% of duration, indicative of a more *retracted* articulation during the /i/ portion of /ie/ and perhaps extending to some degree into the /e/ portion as well.

(F) F2 of /ue/ (Figure 7F): There is only a small window of robustly higher F2 production in ML versus Spanish for a brief period between 25.2% and 49.1% of duration, indicative of a more *advanced* articulation centered around the /u/ portion of /ue/.

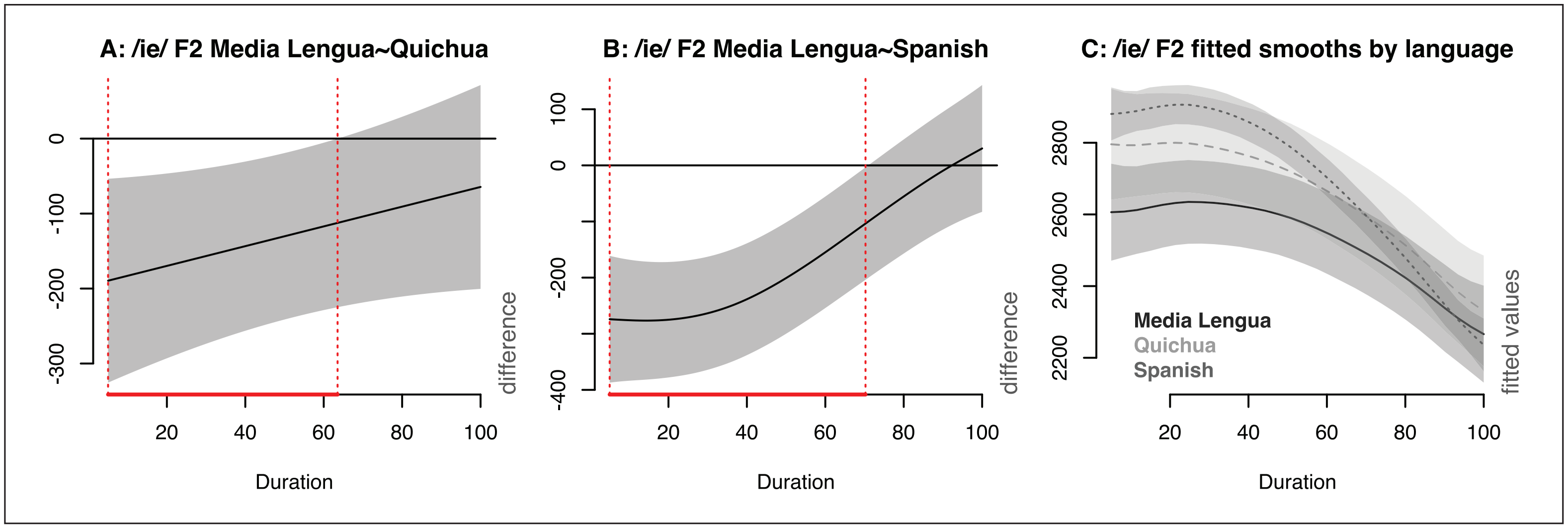


Figure 8. /ie/ F2 (Hz) generalized additive mixed models' three-language comparison. In 8C, the solid line indicates the smooth for Media Lengua; the dashed line the smooth for Quichua; and the dotted line the smooth for Spanish.

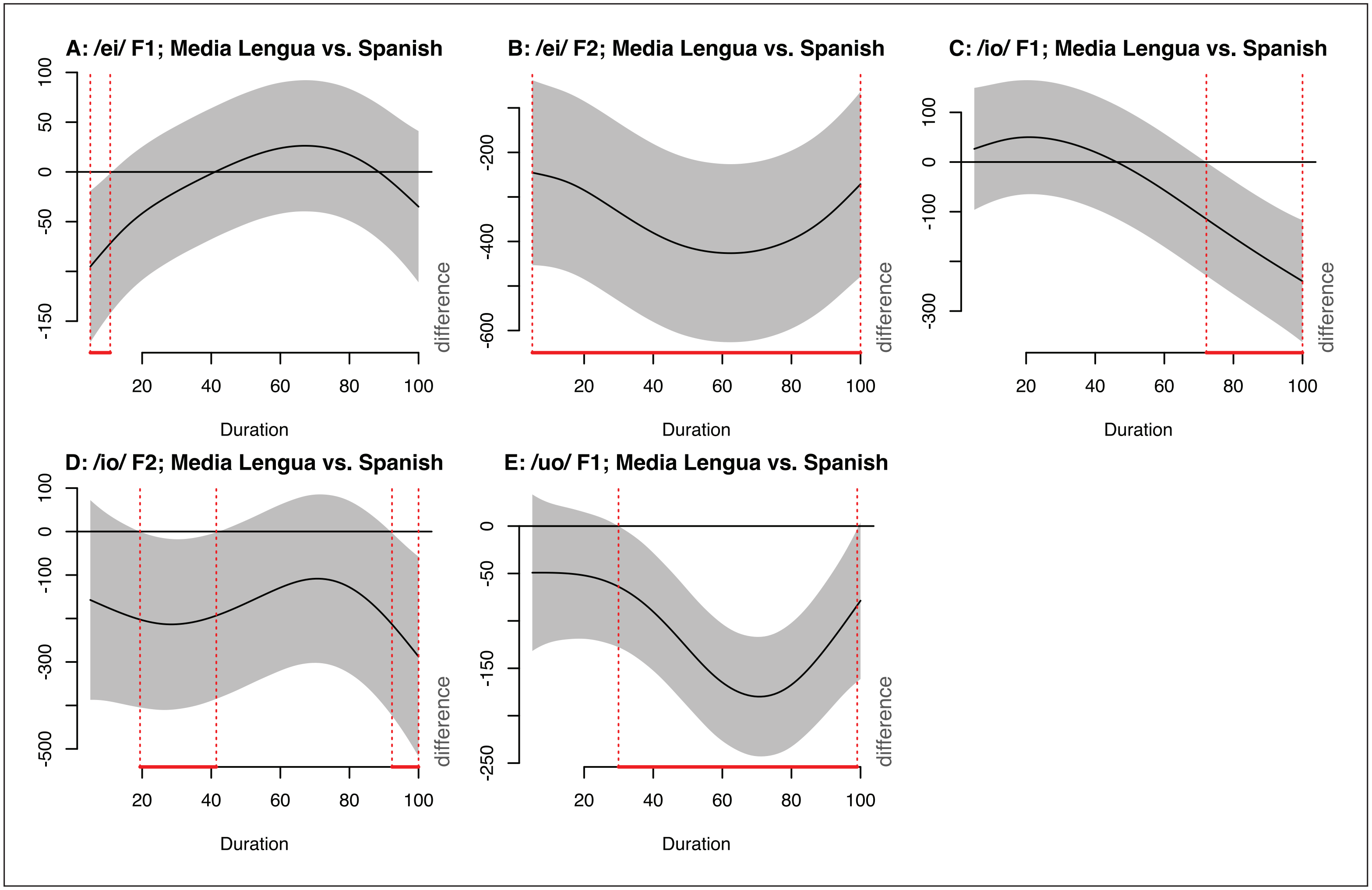


Figure 9. Formant (Hz) difference smooths: Media Lengua (ML) versus Spanish (two-language comparisons).
(A) F1 of /ei/ (Figure 9A): This difference is very short, occurring only at the onset until 10.8% of duration, and can probably thus be ignored.

(Continued)

Figure 8. (Continued)

(B) F2 of /ei/ (Figure 9B): This difference covers 100% of duration, and is quite substantial, with ML being as much as 400 Hz lower than Spanish close to the mid-point, indicative of a more *retracted* ML articulation relative to Spanish. Note that, due to this difference extending across the full duration, the reference level indicating the Spanish F2 production is omitted from this plot—it is located above the plotted region, as indicated by the negative values shown on the y-axis. Although this difference occurs for the entire sequence, it reaches its greatest magnitude during the latter half of the trajectory, centered around the /i/ portion of /ei/.

(C) F1 of /io/ (Figure 9C): The sequence /io/ exhibits some complex cross-linguistic patterning. For F1 the most robust region of difference occurs subsequent to 72.2% of duration, during which ML F1 progressively descends to as much as 200 Hz lower than Spanish, indicative of a more *raised* ML articulation relative to Spanish.

(D) F2 of /io/ (Figure 9D): ML is lower than Spanish, with the greatest region of difference occurring during the early portion of the trajectory between 19.4% and 41.5% of duration, and again at the tail end of the sequence subsequent to 92.3% duration, with ML having a lower F2 during both periods, indicating a more *retracted* articulation in ML.

(E) F1 of /uo/ (Figure 9E): The difference here occurs during the majority of the trajectory, being most robust beginning from 29.9% of duration; ML F1 is lower by a substantial amount over this period, reaching nearly 200 Hz around the 70% mark, indicating a relatively *raised* ML articulation which is centered around the /o/ portion of /uo/.

negatively correlated with vowel height, such that the *lower* ML F1 productions in Figures 9A, C, and E, are indicative of *higher* vowel articulations in terms of tongue position, and that F2 is positively correlated with vowel advancement, such that *lower* ML F2 productions (seen in Figures 9B and D) represent a more *retracted* articulation.

5 Discussion

To the best of our knowledge, this is the first study to explore the acoustic production of vowel sequences in a mixed language along with its source languages. The results are novel in that we observe a vowel system that formed essentially like that of the original ancestral source language (Quichua, the original L1) yet has incorporated non-native vowel sequences from the introduced source language (Spanish, the original L2). Interestingly, the trajectories of these "borrowed" vowel sequences are modified to occupy a more central (Quichua-like) region of the vowel space and frequently exhibit markedly reduced trajectories compared to Spanish. At the same time, we find only one case, the sequence ie/, where a ML vowel sequence differs significantly from its Quichua counterpart—and even in this case the difference from Spanish is substantially greater in comparison. Our findings show how the vowel system of ML successfully integrates nearly all novel (originally L2) vowel sequence patterns from Spanish into what is essentially Quichua phonology by markedly adapting their production while still maintaining contrasts which are not expressed in Quichua. The only two sequences, based on the plots in Figures 5 and 6, that do not appear diphthongal in ML are /ie/ and /uo/ (two phonologically parallel sequences). As an anonymous reviewer pointed out, these pairs look simply like the monophthongs /i/ and /o/ (or possibly /u/) and there appear to be some novel innovations in the behavior of /e/-initial V+V sequences, in that /e/ begins lower than one would expect based on the phonetics of the monophthong /e/. It seems that some interesting changes occurred when ML adapted these sequences from Spanish, and they are no longer easily analyzable as simple sequences of phonologically independent vowels.

The ML vowel space is also interesting in that its contrastive mid- and high-vowels do not conform to models of adaptive dispersion, which predict greater distances between categories in acoustic space to allow for optimal contrastability. Therefore, the shorter trajectories in ML vowel sequences (revealed in this study) and the overlapping mid-vowel and high-vowel categories (revealed in Stewart, 2014) suggest we are looking at a system that was created with unequal influences from its source languages. In fact, ML mid-vowels are only minimally distant enough from their high-vowel counterparts to be contrastive perceptually. This begs the question, why did the mid-vowels not just assimilate to the high-vowels or, opposingly, undergo greater dispersion? As per Stewart and Meakins (accepted), the ML vowel system reflects that of a late bilingual where interference from their L1 (Quichua) impedes native-like production in their L2 (Spanish). Therefore, the arrangement and shape of the vowel categories in acoustic space would suggest that ML was created by late bilinguals instead of early or simultaneous bilinguals. However, they hypothesize that the phonological "stresses" from relexification (e.g., high functional loads of contrastive non-native phonemes) from the mid-vowels may have been driving forces for maintaining/creating contrasts with the high-vowels in the predominately Quichua system. Subsequently, the vowel system could have been nativized with the overlaps "frozen" in place. Evidence from Stewart, et al. (2020a) supports the functional load hypothesis. From a sample of 1415 ML words transcribed by native speakers, approximately 80% of the lexicon contains one of /e, o/ either individually or as part of a vowel sequence. Thus, assimilating every /e/ → /i/ and /o/ → /u/ could have strained the usage of the Spanish lexicon which was optimized for high-vowel versus mid-vowel contrasts, creating unwarranted cognitive processing loads in the process.

Given that ML has indeed adopted contrastive Spanish origin mid-vowels (both productively and perceptually), there is little reason to believe that non-native vowel sequences would not also follow as the categories were already established. However, as the ML vowel categories are more centralized in acoustic space, like that of Quichua, the trajectory of the vowel sequences would also need to adapt to these regions. This can be seen in the ML formant trajectories of the vowel sequences which reflect Quichua vowel sequences in shape, size, and degree of variation. This might suggest why ML vowel sequences act more like those one might expect from Quichua, if such vowel sequences existed in the language, rather than like native Spanish vowel sequences.

One interesting contrast that appears among the vowel sequences which differ between ML versus Spanish (see Figures 5 and 6) concerns the trajectories of /ei/ and /ie/, two sequences which, if they are composed of simple monophthong sequences in hiatus, should be expected to be something like mirror images of each other. In fact, this is true for neither the ML nor the Spanish forms. In both languages, the rising-and-fronting sequence /ei/ shows formant trajectories which reflect this description, moving upwards and forwards in the vowel space. For /ie/, what might be predicted to be a lowering-and-retracting sequence occurrence instead shows in Spanish merely as a retracting movement. While this is also true for ML, this motion is so slight that, as noted above, it could easily be categorized as monophthongal. Ideally, we could compare the behavior of /ie, ei/ with the contrastive pair /uo, ou/ across Spanish and ML. Unfortunately (see subsection 2.1.2) /ou/ does not occur in ML or is not produced as [uo] in Spanish. We do note, however, that ML /uo/ appears, in parallel fashion to /ie/, as a simple monophthong with respect to its formant trajectory, while the same cannot be said of Spanish /uo/ which follows a clear lowering-and-fronting path, marking this development in ML as innovative.

6 Conclusion

This study adds to the mixed language literature with the first acoustic analysis of vowel sequences. It supports findings from other mixed languages studies suggesting that phonological arrangements are not clear-cut cases of stratification or assimilation. Instead, phonological arrangements are more complex and suggest that age of acquisition of the source languages, cognitive factors, and functional load play a much more substantial role in determining the mixed language phonological system than divisions in the language's morphosyntax.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sky Onosson https://orcid.org/0000-0001-9308-7202

Jesse Stewart https://orcid.org/0000-0001-8678-7884

Notes

1. The term "ancestral" is used in the chronological sense referring to the original homeland, pre-contact language as opposed to "introduced" which refers to the language introduced to this group, either through trade, colonization, or their own migration (i.e., post-contact).
2. The Quechuan languages spoken in Ecuador are broadly known as Quichua (or Kichwa) with internal

variants prefaced by the province where they are spoken; for example, Imbabura Quichua or Cotopaxi Quichua.

3. In this example, Media Lengua (ML) orthography is based on the spelling system developed in Pijal and used in the ML dictionary (Stewart et al., 2020a). The Quichua orthography is based on the unified system adapted to Imbabura.
4. As noted by an anonymous reviewer, the corner vowels are not contrastive vowel categories in the traditional sense. We agree and only mention them to remain true to the original analysis of the Media Lengua vowel system in Stewart (2014). This study does not differentiate between Quichua-origin and Spanish-origin corner vowels due to their negligible differences.
5. Possible variation from each individual participant was considered in the statistical models, which made use of *participant* as a random effect.
6. Since providing a standardized assessment test/written questionnaire would be culturally insensitive, Spanish proficiency for both groups was judged using less intrusive methods: (a) feedback from our Quichua-speaking assistant, who was familiar with the participants; (b) our own interactions in Spanish with the participants after the experiment; and (c) informal oral self-assessment loosely based on a number of questions from the Bilingual Language Profile (BLP) (Birdsong et al., 2012).
7. We thank an anonymous reviewer for bringing these papers to our attention.
8. It should be noted that lexical stress in Spanish (e.g., tonic vs. atonic) was not considered as changes in formant trajectories have not been readily identified as a primary correlate in Spanish stress/accent. Instead, most studies focus on pitch, intensity, and duration as acoustic correlates of stress in most Spanish dialects (see e.g., Contreras, 1964; Navarro-Tomás, 1964). In addition, when Ortega-Llebaria and Prieto (2007) analyzed formant frequency as a correlate of stress, they only found a slight tendency towards centralization and only with unstressed [o]; a very similar finding from a forthcoming study on Media Lengua vowel correlates shows that unstressed vowels were only slightly retracted (55 Hz on F2) (Onosson & Stewart, accepted).
9. Although Quichua data were included in the comparisons of /ie/ and /ue/, these sequences are non-native to Quichua and exclusively occur in borrowings from Spanish.
10. Differences in F1 × F2 averages were measured with a protractor using the initial vowel quality of a sequence as the starting point, for example, [e] in the [ei] vowel sequence.

References

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Bakker, P., & Muysken, P. (1994). Mixed languages and language intertwining. In J. Arends, P. Muysken, & N. Smith (Eds.), *Pidgins and creoles: An introduction* (pp. 41–52). John Benjamins.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer. *Computer application*. http://www.fon.hum.uva.nl/praat/

Buchan, H. (2012). *Phonetic variation in Gurindji Kriol and Northern Australian English: A longitudinal study of fricatives in maternal speech* (Doctoral dissertation, University of Wollongong). https://ro.uow.edu.au/theses/3789/

Bundgaard-Nielsen, R., & O'Shannessy, C. (2019). Voice onset time and constriction duration in Warlpiri stops (Australia). In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3612–3616). Australasian Speech Science and Technology Association Inc. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_3661.pdf

Contreras, H. (1964). ¿Tiene el español un acento de intensidad? [Does Spanish have an accent of intensity?] *Boletín Del Instituto de Filología de La Universidad de Chile*, *16*, 237–239. [In Spanish.]

Deibel, I. (2017). *VO vs. OV: What conditions word order variation in Media Lengua?* Colloquium on Mixed Languages, Bremen, Germany.

Deibel, I. (2019). Adpositions in Media Lengua: Quichua or Spanish? – Evidence of a lexical–functional split. *Journal of Language Contact*, *12* (2), 404–439. https://doi.org/10.1163/19552629- 01202006

Deibel, I. (2020). *Language representations in the presences of a lexical–functional split: An experimental approach targeting the Quichua–Media Lengua–Spanish interface* (Doctoral dissertation, The Pennsylvania State University).

Flege, J. (2007). Language contact in bilingualism: Phonetic system interactions. In J. I. Hualde & J. S. Cole (Eds.), *Laboratory phonology 9* (pp. 353–380). Mouton de Gruyter.

Gómez-Rendón, J. (2005). La Media Lengua de Imbabura [The Media Language of Imbabura.] In H. Olbertz & P. Muysken (Eds.), *Encuentros y conflictos: Bilingüismo y contacto de lenguas en el mundo Andino* [Encounters and conflicts: Bilingualism and language contact in the Andean world] (pp. 39–58). Iberoamericana. [In Spanish.]

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall.

Hendy, C. R. (2019). *The distribution and acoustic properties of fricatives in Light Warlpiri* (B.A. Honours Thesis, Australian National University). https://openresearch-repository.anu.edu.au/handle/1885/200483

Johnson, K. (2000). Adaptive dispersion in vowel perception. *Phonetica, 57*(2–4), 181–188. https://doi.org/10.1159/000028471

Jones, C., & Meakins, F. (2013). Variation in voice onset time in stops in Gurindji Kriol: Picture naming and conversational speech. *Australian Journal of Linguistics*, *33*(2), 196–220. https://doi.org/10.1080/07268602.2013.814525

Jones, C., Meakins, F., & Buchan, H. (2011). Comparing vowels in Gurindji Kriol and Katherine English: Citation speech data. *Australian Journal of Linguistics*, *31*(3), 305–326. https://doi.org/10.1080/07268602.2011.598629

Jones, C., Meakins, F., & Mauwiyath, S. (2012). Learning vowel categories from maternal speech in Gurindji Kriol. *Language Learning*, *62*(4), 1052–1078. https://doi.org/1010.1111/j.1467-9922.2012.00725.x

Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America*, *110*(4), 2141–2155. https://doi.org/10.1121/1.1400737

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–862. https://doi.org/1010.2307/411991

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer Academic.

Lipski, J. (2016). Language switching constraints: More than syntax? Data from Media Lengua. *Bilingualism: Language and Cognition*, *1*(4), 1–25. https://doi.org/10.1017/S1366728916000468

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories – hot spots or adaptive dispersion? In *Proceedings of the XIIIth Swedish Phonetics Conference (FONETIK 2000), Skóvde, Sweden*, *May 24–26, 2000* (pp. 93–96). https://www2.ling.su.se/fon/perilus/2000_11.pdf

Meakins, F. (2013). Mixed languages. In Y. Matras & P. Bakker (Eds.), *Contact languages: A comprehensive guide* (pp. 159–228). Mouton de Gruyter.

Meakins, F., & Stewart, J. (accepted). Mixed languages. In S. Mufwene & A. M. Escobar (Eds.), *Cambridge handbook of language contact*. Cambridge University Press.

Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In A. R. Highfield (Ed.), *Historicity and variation in Creole studies* (pp. 57–78). Karoma Publishers.

Muysken, P. (1997). Media Lengua. In S. Thomason (Ed.), *Contact languages: A wider perspective* (pp. 365–426). John Benjamins.

Navarro-Tomás, T. (1964). La medida de la intensidad [The measure of intensity]. *Boletín Del Instituto de Filología de La Universidad de Chile*, *16*, 231–235. https://boletinfilologia.uchile.cl/index.php/BDF/article/view/49444/51905 [In Spanish.]

Onosson, S. (2018). *An acoustic study of Canadian raising in three dialects of North American English*. (Doctoral dissertation, University of Victoria). https://dspace.library.uvic.ca/handle/1828/9274

Onosson, S., & Bird, S. (2019). Differences in vowel-glide production between L1 and L2 speakers of Hul'q'umi'num'. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 984–988). Australasian

Speech Science and Technology Association Inc. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_1033.pdf

Onosson, S., & Stewart, J. (accepted). A multi-method approach to correlate identification in acoustic data: The case of Media Lengua. *Laboratory Phonology*.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 155–176). John Benjamins.

Padgett, J. (2008). Glides, vowels, and features. *Lingua*, *118*(12), 1937–1955. https://doi.org/10.1016/j.lingua.2007.10.002

R Core Team. (2019). R: A language and environment for statistical computing. Programming language. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.r-project.org/

Rosen, N. (2006). Language contact and Michif stress assignment. *Language Typology and Universals*, *59*, 170–190. https://doi.org/10.1524/stuf.2006.59.2.170

Rosen, N. (2007). *Domains in Michif phonology*. (Doctoral dissertation, University of Toronto).

Rosen, N., Stewart, J., & Sammons, O. (2020). How 'mixed' is mixed language phonology – An acoustic analysis of the Michif vowel system. *Journal of the Acoustical Society of America*, *147*(4), 2989–2999. https://doi.org/doi.org/10.1121/10.0001009

Roettger, T. B., Winter, B., & Baayen, H. (2019). Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics*, *73*, 1–7. https://doi.org/10.1016/j.wocn.2018.12.001

RStudio Team. (2019). RStudio: Integrated Development for R. Computer application. RStudio, Inc., Boston, MA. https://www.rstudio.com/

Sánchez Miret, F. (1998). Some reflections on the notion of diphthong. In J. Fisiak (Ed.), *Papers and Studies in Contrastive Linguistics* (Vol. 34, pp. 27–51). Adam Mickiewicz University.

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. http://eprints.whiterose.ac.uk/113858/2/1703_05339v1.pdf

Stewart, J. (2011). *A brief descriptive grammar of Pijal Media Lengua and an acoustic vowel space analysis of Pijal Media Lengua and Imbabura Quichua*. (Master's Thesis, University of Manitoba). http://hdl.handle.net/1993/4882

Stewart, J. (2014). A comparative analysis of Media Lengua and Quichua vowel production. *Phonetica*, *71*(3), 159–182. https://doi.org/10.1159/000369629

Stewart, J. (2015a). Intonation patterns in Pijal Media Lengua. *Journal of Language Contact*, *8*(2), 223–262. https://doi.org/10.1163/19552629-00802003

Stewart, J. (2015b). *Production and perception of stop consonants in Spanish, Quichua, and Media Lengua*. (Doctoral dissertation, University of Manitoba).

Stewart, J. (2018). Vowel perception by native Media Lengua, Quichua, and Spanish speakers. *Journal of Phonetics*, *71*, 177–193. https://doi.org/10.1016/j.wocn.2018.08.005

Stewart, J. (2020). A preliminary, descriptive survey of rhotic and approximant fricativization in Northern Ecuadorian Andean Spanish varieties, Quichua, and Media Lengua. In R. Rao (Ed.), *Spanish phonetics and phonology in contact: Studies from Africa, the Americas, and Spain* (103–140). John Benjamins.

Stewart, J., & Meakins, F. (2021). Advances in mixed language phonology: An overview of three case studies. In M. Mazzoli & E. Sippola (Eds.), *New perspectives on mixed languages: From core to fringe. language contact and bilingualism*. De Gruyter Mouton.

Stewart, J., Ayala, G. P., & Gonza Inlago, L. (2020a). Media Lengua dictionary. *Dictionaria*, *12*, 1–3216. https://dictionaria.clld.org/contributions/medialengua

Stewart, J., Meakins, F., Algy, C., Ennever, T., & Joshua, A. (2020b). Fickle fricatives: Obstruent perception in Gurindji Kriol and Roper Kriol. *Journal of the Acoustical Society of America*, *147*, 2766. https://doi.org/10.1121/10.0000991

Stewart, J., Meakins, F., Algy, C., & Joshua, A. (2018). The development of phonological stratification: Evidence from stop voicing perception in Gurindji Kriol and Roper Kriol. *Journal of Language Contact*, *11*(1), 71–112. https://doi.org/10.1163/19552629-01101003

van Rij, J, Wieling, M, Baayen, R, & van Rijn, H (2020). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.4. https://rdrr.io/cran/itsadug/

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Taylor and Francis Group.

Xu, Y., & Gao, H. (2018). FormantPro as a tool for speech analysis and segmentation. *Revista De Estudos Da Linguagem*, *26*(4), 1435–1454. https://doi.org/10.17851/2237-2083.26.4.1435-1454

Appendix: Participants

Tables A1 through A3 provide information pertaining to the participants who took part in the production portion of this study. The data include: speaker code; age at the time of recording; gender; formal education; level of Spanish; Quichua and Media Lengua; frequency of language use; recording type; and place of residence.

Table A1. This table provides demographic information about the Media Lengua (ML) participants including: age at the time of recording; gender; level of formal education; level of Spanish; ML; and Quichua; frequency of ML usage; recording type; and place of residency. It also includes the type of recording used to gather the data.

ML participants								
Code	Recording type	Gender	Age	Formal education	Spanish level	ML / Quichua level	ML Usage	Place of residence
41	Conversations	M	65	Primary	High	Native	Intermittently	Pijal
42	Wordlist	F	50	None	High	Native	Daily	Pijal
43	Elicitations, wordlist	F	39	Secondary	High	Native	Daily	Pijal
44	Elicitations, wordlist	F	39	Primary	High	Native	Intermittently	Pijal
48	Wordlist	F	50	Primary	High	Native	Intermittently	Pijal
49	Wordlist	F	50	Primary	High	Native	Intermittently	Pijal
50	Elicitations, wordlist	F	42	Secondary	Mid	Native	Intermittently	Pijal
51	Elicitations, wordlist	F	60	None	Mid	Native	Infrequently	Pijal
52	Wordlist	M	40	Primary	High	Native	Infrequently	Pijal
53	Conversations	M	58	Primary	High	Native	Intermittently	Pijal
56	Conversations, wordlist	F	64	Primary	High	Native	Intermittently	Pijal
58	Wordlist	M	33	University	Native	Native	Infrequently	Pijal
59	Wordlist	M	38	University	Native	Native	Infrequently	Pijal
62	Wordlist	M	54	Primary	High	Native	Intermittently	Pijal
79	Elicitations, wordlist	M	45	Primary	High	Native	Intermittently	Pijal
S011b	Conversations	F	60+	Primary	High	Native	Intermittently	Pijal
S01a	Conversations	F	60+	Primary	High	Native	Intermittently	Pijal
S02a	Conversations	F	45+	Primary	High	Native	Intermittently	Pijal
S03a	Conversations	F	60+	Primary	Mid	Native	Intermittently	Pijal
S03b	Conversations	F	60+	Primary	Mid	Native	Intermittently	Pijal
S04b	Conversations	M	60+	Primary	High	Native	Intermittently	Pijal
S07b	Conversations	M	60+	Primary	High	Native	Intermittently	Pijal
Z013c	Conversations	F	60+	Primary	Mid	Native	Intermittently	Pijal
Z016b	Conversations	F	60+	Primary	High	Native	Intermittently	Pijal

Table A2. This table provides demographic information about the Quichua speaking participants including: age at the time of recording; gender; level of formal education; level of Spanish and Quichua; frequency of Quichua usage; recording type; and place of residency.

Quichua participants								
Speaker code	Age	Gender	Formal education	Spanish level	Quichua level	Quichua usage	Recording type	Place of residence
63	66/68	M	Primary	Mid/High	Native	Daily	Elicitation/wordlist	Chirihuasi
64	62	F	None	Low	Native	Daily	Elicitation	Chirihuasi
65	45	F	None	Low	Native	Daily	Elicitation	Chirihuasi
69	29	F	None	Low	Native	Daily	Elicitation	Chirihuasi
70	21	F	None	Low	Native	Daily	Elicitation	Chirihuasi
72	42	M	Not applicable (NA)	Mid/High	Native	Daily	Elicitation	Chirihuasi
73	70	M	University	Mid/High	Native	Daily	Elicitation	Chirihuasi
75	52	M	Secondary	Native	Native	Daily	Elicitation	Chirihuasi
76	55	F	NA	Mid/High	Native	Daily	Elicitation	Chirihuasi
77	49	M	Secondary	High	Native	Daily	Elicitation	Chirihuasi

Table A3. This table provides demographic information about the Spanish speaking participants including: age at the time of recording; gender; level of formal education; level of Spanish and Quichua; recording type; and place of residency.

Spanish participants							
Speaker code	Age	Gender	Formal education	Spanish level	Quichua level	Recording type	Place of residence
68	27	Male (M)	University	Native	None	Sentence	Ibarra
73	34	Female (F)	University	Native	None	Sentence	Ibarra
71	21	M	University	Native	None	Sentence	Ibarra
97	22	F	University	Native	None	Wordlist	Quito
98	35	F	University	Native	None	Wordlist	Quito
100	38	M	University	Native	None	Wordlist	Quito
101	33	M	University	Native	None	Wordlist	Quito
102	35	M	University	Native	None	Wordlist	Quito
103	57	M	University	Native	None	Wordlist	Quito
104	34	F	University	Native	None	Wordlist	Quito
105	30	F	University	Native	None	Wordlist	Quito
106	24	F	University	Native	None	Wordlist	Quito
107	38	F	University	Native	None	Wordlist	Quito
108	24	F	University	Native	None	Wordlist	Quito

labphon Laboratory Phonology Journal of the Association for Laboratory Phonology

Onosson, S., & Stewart, J. 2021 A multi-method approach to correlate identification in acoustic data: The case of Media Lengua. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 13, pp. 1–30. DOI: https://doi.org/10.5334/labphon.291

JOURNAL ARTICLE

A multi-method approach to correlate identification in acoustic data: The case of Media Lengua

Sky Onosson[1] and Jesse Stewart[2]

[1] Department of Linguistics, University of Manitoba, Winnipeg, CA

[2] Department of Linguistics, University of Saskatchewan, Saskatoon, CA

Corresponding authors: Sky Onosson (Sky@Onosson.com); Jesse Stewart (Stewart.Jesse@usask.ca)

This study of Media Lengua examines production differences between mid and high vowels in order to identify the major correlates that distinguish these vowel types. The Media Lengua vowel system is unusual in that it incorporates lexical items originating in Spanish's five-vowel system into a three-vowel system inherited from Quichua, resulting in high degrees of overlap between the front versus back, mid and high vowel pairs /e, i/ and /o, u/ in F1xF2 space. As Media Lengua speakers utilize and differentiate between all five vowels despite the large degree of acoustic overlap between mid and high vowels, this raises the question of what other correlates beyond F1 and F2 might be involved. To address this, our study looks at a range of variables, both acoustic and qualitative, in a multi-method approach using both *factor analysis for mixed data* and *linear mixed effects regression modelling*. Each method provides a unique view on the correlates of vowel differentiation in Media Lengua. Taken together, our results indicate that Media Lengua speakers rely on both social and linguistic contextual cues to distinguish mid from high vowels, which overlap in acoustic space (F1 and F2).

Keywords: Media Lengua; mixed languages; vowels; factor analysis for mixed data; linear mixed effects regression; mixed methods

1. Introduction

Contact languages provide an ideal platform for exploring and testing the atypical arrangements of linguistic elements, given that such languages do not form through ordinary processes of evolutionary change. Instead, contact language formation is frequently tumultuous, rapid and, more often than not, forceful as groups with no common language must communicate. However, in the case of mixed languages, a sub-category of contact languages, the process is systematic, metalinguistic, and expressive as the originators are already proficient bilinguals in both source languages. This allows mixed languages to form through expressive means (e.g., ethnic or cultural identification) rather than through communicative necessity (Meakins & Stewart, in press). Because of this, mixed languages are often used internally within a speech community and show clear categorical source language divisions in their lexicon and/or grammar. For instance, Michif, a mixed language spoken sparsely throughout central Canada and in the northern U.S., integrates Plains Cree verb phrases and Métis French noun phrases; while Media Lengua, a mixed language spoken in the Ecuadorian Andes, integrates Quichua morphosyntax and Spanish lexicon (see Section 2).

However, while the lexicon and/or morphosyntax of these languages show clear divisions, acoustic studies on mixed language phonologies (see e.g., Buchan, 2012; Bundgaard-Nielsen & O'Shannessy, 2019; Hendy, 2019; Jones & Meakins, 2013; Jones, Meakins, & Buchan, 2011; Jones, Meakins, & Mauwiyath, 2012; Meakins & Stewart, in press; Onosson & Stewart, in press; Rosen, 2006, 2007; Rosen, Stewart, Pesch-Johnson, & Sammons, 2019; Rosen, Stewart, & Sammons, 2020; Stewart, 2014, 2015a, 2015b, 2018a, 2018b, 2020; Stewart & Meakins, 2021; Stewart, Meakins, Algy, Ennever, & Joshua, 2020; Stewart, Meakins, Algy, & Joshua, 2018) suggest a heavy influence from the original L1 of the speech community, which may be related to late acquisition of the L2 by the originators (see Stewart & Meakins, 2021, for details on this hypothesis). This influence, however, is not absolute and phonological elements from the L2 source language are present in most mixed languages, though in unexpected ways. Stewart and Meakins (2021) state that mixed language phonologies often abound in "near-mergers, overlapping categories, categorical assimilation, categorical maintenance, and overshoot of target categories at the segmental level, in addition to prosodic assimilation, possible preservations of archaic patterns, and innovation at the suprasegmental level." Media Lengua (see Section 2.2) is a mixed language of particular interest given that the primary acoustic correlates for production and perception remain unclear even after having been extensively documented (see Stewart 2014, 2018b).

Acoustic studies of vowels typically involve measuring the first two formants (F1 and F2) as a correlate for tongue body position (F1 for height and F2 for frontedness). Additionally, the third formant (F3) is used in the analysis of rhotic (r-colored) vowels, rounding (Ladefoged & Maddieson, 1996), pharyngeal constriction (Fant, 1968), and possibly to differentiate between non-low front vowels (see e.g., Maurer, Cool, Landis, & D'Heureuse, 1991). However, other vowel systems with prosodic contrasts require additional measurements to be fully described. For example, languages with a long-short quantity distinction require a length measurement (e.g., Cree /i/ versus /iː/; see Muehlbauer, 2012); tonal languages require pitch measurement (e.g., Vietnamese /a˧˨/ versus /a˨˩˦/; see Nguyễn, 1997); languages with oral versus nasal vowels (e.g., Guaraní /a/ versus /ã/; see Walker, 1999) are analyzed using a variety of methods (see e.g., Chen, 1996; Stewart & Kohlberger, 2017; Styler, 2015); languages with modal voice versus creaky voice distinction require the analysis of harmonics, pitch (f0), formant amplitudes among other correlates (e.g., Mazatec with a three way contrast between laryngealized, (/a̰/) model (/a/), and breathy phonation (/a̤/); see Garellek & Keating, 2011), etc.

Empirical evidence from Ian Maddieson (1984) shows that while a system may make use of multiple correlates for production purposes, languages typically rely on only one primary correlate (sometimes two) for contrastive purposes. For example, the primary distinguishing factor between /i/ and /ɪ/ in English is typically described as a tense-lax contrast; however, tense-lax contrasts also carry a phonetic duration distinction where tense vowels are often longer than lax vowels. Additionally, intrinsic pitch frequency differences have also been documented between high and low vowels (see e.g., Crandall, 1925; Fant, 1960; Lehiste & Peterson, 1961; Ohala & Eukel, 1987). Similarly, intrinsic vowel duration has been shown to be directly correlated to the size of the articulatory gesture (lower vowels being longer than high vowels) or the nature of the following consonant (see e.g., House & Fairbanks, 1953; Lehiste, 1970; Lehiste & Peterson, 1961; Sharf, 1962). Likewise, Lehiste and Peterson (1959) show that vowel amplitude is also an intrinsic characteristic of vowel production, lower vowels being produced with greater intensity than higher vowels.

Nonetheless, given the unique nature of the development of Media Lengua's vowel system as a contact language derived from two languages with differently-sized inventories—and specifically developed by speakers with a smaller L1 inventory who have incorporated vocabulary from an L2 system with a wider variety of vowel phonemes—it should not be presumed which correlates serve as the primary means of distinction between

Media Lengua vowels. To sufficiently allow for the determination of the most important correlates in this regard requires a suitably unbiased methodology, one which does not a priori assign certain factors an elevated rank over others. The present study seeks to accomplish this through the application of statistical methods for dimensional reduction to a dataset composed of a multiplicity of acoustic and other variables related to Media Lengua vowel production. Dimensional reduction methods construct unobserved variables by identifying patterns within the observed variables in a dataset, attributing variation in the most concise means possible through the reduction of the number of dimensions needed to describe the primary patterns of variability, and relating the original empirically-measured variables to the constructed dimensions so as to rank their relative influence as primary correlates of variation. Whether or not such methods indicate that the primary correlates of acoustic vowel production in Media Lengua pattern similarly to or differently from those of non-contact languages, we believe that this approach is worthwhile, as it is inherently less biased than other methods. Moreover, where dimensional reduction methods are able to determine unexpected patterns, such approaches may offer something of value to linguistic research in a more specific sense.

2. Media Lengua

2.1. About the Media Lengua language

Media Lengua is described as a lexical-grammar mixed language (Meakins, 2013; Meakins & Stewart, in press; Muysken, 1997), which is spoken in the northern highlands of Ecuador in the province of Imbabura. Media Lengua combines Quichua's[1] suffixing morphological system and word order with the lexicon of the Rural Spanish dialect spoken in the northern Andes. The Spanish lexicon has replaced over 90% of Quichua's vocabulary primarily through relexification (Deibel, 2017, 2019, 2020; Gómez-Rendón, 2005; Lipski, 2016; Muysken, 1981, 1997; Stewart, 2011, 2015b). Media Lengua is spoken by an estimated 2,000 people in a handful of communities near Lago San Pablo. Muysken (1997) and Stewart (2011, 2015b) both suggest that the language formed in the early 20th century. Data from this study come from the community of Pijal (see **Figure 1**).

A sample of Media Lengua is proved in example 1; the first line contains orthography with the bold elements being of Spanish origin, the second line contains a broad IPA transcription with morpheme boundaries, the third line contains the interlinear glosses, and the fourth and fifth lines provide translations in Quichua and Spanish (respectively) for cross comparison.

1.	**Mas**	**bueno**my	**trilla**ngapa	**caballo**huan.
	mas	bueno-mi	triʒa-ngapa	kabaʒo-wan
	more	good-VAL	thresh-PURP	horse-INST

Ashtahuan alymi aisangapa bishtiahuan. (Quichua)
Es mejor trillar con un caballo. (Spanish)
'It's better to thresh using a horse.'

2.2. The Media Lengua vowel system

This section describes Media Lengua's unconventional vowel system based on acoustic and experimental data; for impressionistic descriptions of Media Lengua vowels, see Muysken (1997) for Cotopaxi Media Lengua, and Gómez-Rendón (2005) for Imbabura Media Lengua.

[1] The Quechuan languages spoken in Ecuador are broadly known as Quichua (or Kichwa) with internal variants often prefaced by the province where they are spoken e.g., Imbabura Quichua or Cotopaxi Quichua.

Figure 1: Map of Imbabura province, Ecuador. Data for this study come from the community of Pijal. Map source: https://freevectormaps.com/.

Media Lengua vowels are an integrated system comprised of Quichua's three-vowel system, (/i, u, a/), and Spanish's five-vowel system (/i, u, e, o, a/). The Media Lengua system is described in detail in Stewart (2014) based on F1 and F2 measurements from 2,515 vowel tokens taken from elicited phrases produced by 10 participants. Stewart's results show small, yet statistically significant, differences between Spanish origin and Quichua origin /i/, /u/, and /a/, where the Spanish origin vowels are only 13 Hz less centralized in acoustic space, on average, compared to the Quichua origin vowels (2A). In most cases, a difference of 13 Hz could simply be chalked up to a type I statistical error attributed to the quantity of tokens tested. However, the dispersion patterns of F1 in each Spanish origin corner vowel correspond to the directions predicted by adaptive dispersion models. In other words, vowel systems with more vowels, such as Spanish's five, are predicted to occupy more extreme acoustic spaces, e.g., lower F1 values for high vowels and higher F1 values for low vowels, compared to systems with fewer vowels, such as Quichua's three. However, the degree of dispersion expected for contrastive purposes is not met with an average difference of only 13 Hz (see e.g., Flege, 2007; Johnson, 2000; Liljencrants & Lindblom, 1972; Lindblom, 1986, 1990; Livijn, 2000, for details on adaptive dispersion theories). Therefore Stewart's (2014) results suggest that Media Lengua speakers produce imperceptible, yet consistent differences in vowels of the same quality, similar to near-mergers or covert contrasts, in a complex case of stratification based solely on the language of origin of the morpheme.

Stewart (2014) also describes the differences between Spanish origin mid-vowels and Quichua/Spanish origin high vowels (**Figure 2A**). His results suggest that Media Lengua speakers produce statistically significant F1 differences between both groups with /e/ and /o/ having lower tongue body positions, as indicated by formant differences of 41 Hz on average. When converted to 0.36 bark, this is just enough distance to surpass Kewley-Port's (2001) threshold of 0.3 bark for possible formant discrimination (for values between 200–3000 Hz). However, like the corner vowels previously described, the mid-vowel and high-vowel categories also exhibit substantial overlap. It was later confirmed by Stewart (2018a) that Media Lengua speakers take advantage of this small albeit important distance between categories to identify differences between Spanish origin mid and high vowels. Stewart's results (**Figure 2B**) were based on a 10-step, two-alternative force choice (2AFC) identification task experiment using minimal pairs as stimuli, which contained modified F1, F2, F3, pitch, duration, and intensity values. Although this line of research has successfully described the arrangement of Media Lengua's different-origin vowels and identified their role in perception, it is still not determined which correlate or correlates are primarily responsible for consistently identifying contrasts in the minimal pairs.

Onosson and Stewart (in press) have recently shown that Media Lengua speakers have also successfully integrated Spanish vowel-sequences (i.e., diphthongs) into the overlapping mid and high vowel space by reducing the overall range of the formant trajectories compared to equivalent sequences produced by Spanish speakers. This was even true for vowel-sequences, which only differ by mid and high vowels in the same articulatory region. For example, e.g., /ie/ and /ei/ consistently had opposing initial and final targets even within the tightly overlapping spaces (see 2A).

The results of these studies taken together describe a dense vowel system with the capacity to operate up to eight monophthong vowels and 19 vowel-sequences which are arranged in an acoustic space that reflects the original Quichua three-vowel system. Stewart and Meakins (2021) suggest that such a system may be a result of Media Lengua's originators having acquired Spanish as late bilinguals, and who spoke a Quichua-accented Spanish. However instead of Spanish origin /e/ assimilating to Quichua /i/ and Spanish origin /o/ assimilating to Quichua /u/, enough acoustic distance was maintained to handle the newly relexified vocabulary of the emerging mixed language, which brought with it a substantial mid versus high vowel phonological functional load.

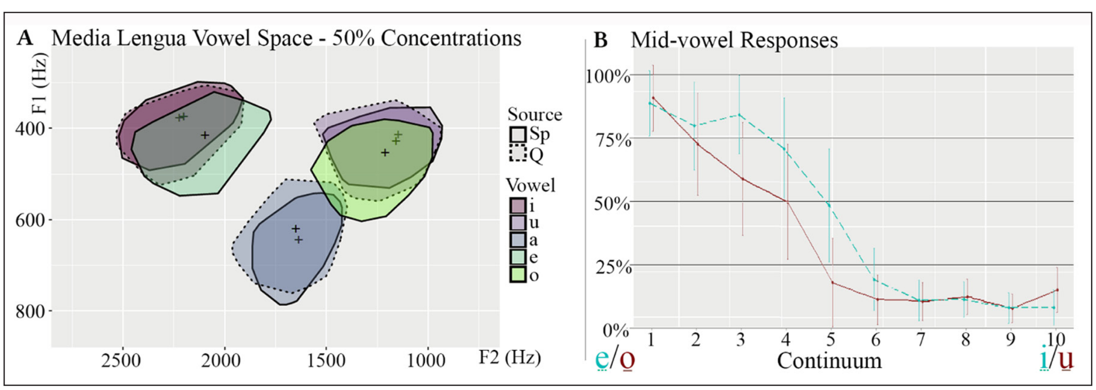


Figure 2: A: Media Lengua vowel space 50% concentrations of statistically different vowel clusters (based on Stewart, 2014); **B:** Media Lengua results from a 2AFC identification task experiment with minimal pair stimuli modified along a 10-step continuum between Media Lengua mid-vowels to Media Lengua high-vowels (based on Stewart, 2018b).

Given that Media Lengua's vowel space has highly overlapping mid and high vowels in the F1 and F2 dimensions (Stewart, 2014), and that Stewart's (2018b) perception experiment incorporated multiple correlates into the stimuli (F1, F2, F3, pitch, duration, intensity), it is not fully clear if the limited differences in F1 and F2 are the primary correlates used in contrasting the pairs. Therefore, this study implements a novel, unbiased approach to better identify potential acoustic correlates involved in producing and perceiving differences in Media Lengua's unconventional vowel system; a system which opposes predicted arrangements in standard dispersion models, and yet still shows clear contrastibility even with overlapping clusters in acoustic space.

3. Methods

The aim of this study is to identify the acoustic and other correlates involved in differentiating highly overlapping mid and high vowel clusters in Media Lengua. To do so, we make use of the exploratory tool *factor analysis of mixed data* (FAMD) which is used to categorize multiple quantitative and qualitative variables. FAMDs, and particularly their quantitative component Principal Component Analysis (PCA), are a common exploratory technique used in natural sciences (e.g., biology, chemistry); however, FAMDs have not been readily applied to date in linguistics and acoustic phonetics. We also make use of a well-established statistical method used in acoustic phonetics, linear mixed effects regression modelling (MEM), as a second layer of analysis to corroborate the usefulness of the application of FAMDs to phonetic or other linguistic data.

3.1. Materials

For this study, $n = 1{,}202$ vowel tokens were analyzed. The vowel data were gathered from two sources.[2] The first source (corpus 1) comes via a 2010–2014 corpus of wordlist/sentence list data and contains 242 (20%) tokens. The second source (corpus 2) comes via a 2015–2019 corpus of natural speech data and contains 960 (80%) tokens. Token quantities based on the corpora are summarized in **Table 1**.

During the wordlist and sentence list sessions (corpus 1), participants were asked to read words or sentences in Media Lengua from a computer screen. Since Media Lengua speakers are multilingual (Media Lengua, Quichua, and Spanish) Media Lengua inflectional morphology was added to each word to prime the target language; e.g., seeing the word *casa* 'house' instead of *casamanmi* 'house-DIR-VAL' might elicit a more

Table 1: Speech corpora and token counts.

Vowel	Corpus 1	Corpus 2	Total
	Wordlist/sentences	Natural speech	
a	92	232	324
e	25	188	213
i	29	176	205
o	82	170	252
u	14	194	208
Total	242	960	1,202

[2] It should be noted that neither of these corpora were used in Stewart (2014), and that even though some of the same speakers participated in both studies, the corpora for the present study contain data from 17 additional speakers. Data gathering methods also differ between the two studies: elicited oral translations for Stewart (2014), and a combination of wordlist/sentence readings and natural speech for these more recent corpora.

Spanish-like pronunciation. The participants were recorded on a TASCAM DR-1 portable digital recorder using a NEXXTECH unidirectional dynamic microphone (50–13,000 Hz response) in 16-bit WAV format with a sample rate of 44.1 kHz; see Stewart (2015b) for more details on the data collection procedures for corpus 1.

Data from corpus 2 were recorded by two assistants from Pijal working in the community. Participants were asked to converse in pairs about a topic of their choosing. These data were recorded in 16-bit WAV format with a sample rate of 44.1 kHz on a Zoom H6 Handy Recorder with its internal microphone (unidirectional XYH-6 capsule), in 16-bit WAV format with a sample rate of 44.1 kHz.

3.2. Participants

The participants in this study included 26 trilingual speakers (L1 Media Lengua, L1 Quichua, and L2 Spanish). This group consisted of 15 women and 11 men. All participants were from the community of Pijal Bajo and had acquired Quichua and Media Lengua simultaneously from birth. Upon entering primary school, typically at six–seven years of age, they learned Spanish. Each language has its own niche with Media Lengua only used within the community and typically among speakers aged 40 and above. Spanish is now the dominant language in Pijal and most Media Lengua speakers use Spanish to communicate with their children and those younger than 40; in fact some of the consultants' children had no idea they even spoke Media Lengua and were in awe during the first recording session (see Stewart, 2011). Quichua, while occasionally used alongside Media Lengua, is more commonly used outside Pijal when conversing with people from other Indigenous communities. Given that Media Lengua is typically used as an internal language, and speakers are fluent in both Quichua and Spanish, many people outside of Pijal are not even aware of Media Lengua. However, those that are familiar with it will typically state that people in Pijal "can't speak Quichua well" or "it doesn't make sense," when prompted. As the language is not used with others outside the community, there is little in the way of external linguistic discrimination. Internally, it is a different story; Media Lengua is often referred to as *yanga shimi* 'a nothing language,' and speaker attitudes range from, "it's fun to speak Media Lengua as it has different intonation and it's expressive" to nostalgia, with some speakers remembering how their parents used to speak, to disdain with one consultant once stating, "this language is stupid and should be forgotten; we'd be better off learning English." Pijaleños in their late 20s and 30s typically have a passive knowledge of Media Lengua and may be able to carry on a basic conversation. However, Pijaleños below this age range are typically Spanish monolingual, though they may have some knowledge of Unified Quichua from classes taught at school (for more information on the social context or language ideologies see Jarrín Paredes, 2014; Stewart, 2011).

3.3. Procedures

This section discusses the procedures followed in this study, beginning with methods of vowel extraction. This is followed by a detailed description of the FAMD approach, including prior examples of the implementation of FAMD (and related methods) in phonetic research, as this method is likely to be largely unfamiliar to a number of researchers. Finally, we briefly discuss the MEM method and our rationale for employing it along with FAMD; we omit a lengthy overview of MEM methodology as it is a relatively familiar statistical tool for many phoneticians, while providing some important references for those who are less familiar with it. All statistical analysis and plotting was carried out in R (v3.5.2, R Core Team, 2020) with extensive use of the tidyverse package suite (Wickham et al., 2019).

3.3.1. Vowel extraction

Each target vowel token was manually segmented in Praat (v6.1.04, Boersma & Weenink, 2020), and acoustic data was extracted using a Praat script written by the authors, which takes F1, F2, F3, pitch, and intensity measurements from a boundary point placed at a steady formant state typically located near of the centre of the vowel. Duration was extracted from interval boundaries that demarcated the vowel based on several criteria (appearance of or changes in glottal pulse and formant patterns, changes in the wave form, intensity, etc.). The script also adjusted the ceiling of the formant search range to 5000 Hz for men and 5500 Hz for women as suggested in the Praat manual (Boersma & Weenink, 1996).

The vowel data were also marked for stress (levels: stressed and unstressed) as it has been known since at least the 1950s that stress can affect vowel formats and vowel duration (see e.g., Fry, 1955, 1965). Media Lengua, like Quichua, has fixed stress on the penultimate syllable, which 'shifts' to the new penultimate syllable when its suffixing agglutinating morphology is added e.g., [ˈgato] 'cat' ⇒ [gaˈtota] 'cat-ACC' ⇒ [gatoˈtami] 'cat-ACC-VAL.' Similarly, the vowel data were also marked for syllable type (levels: open [–CODA] versus closed [+CODA]) as vowel production can be affected based on the presence or absence of a coda, especially duration (see e.g., Maddieson, 1985). The majority of syllables in both Media Lengua follow one of four patterns: V (*abil* [ˈa.bil] 'skillful'), CV (*comini* [ko.ˈmi.ni] 'I eat'), VC (*alcalde* [al.ˈkal.de] 'mayor'), and CVC (*costal* [ˈkos.tal] 'sack'). However, both languages may have up to two consonants in onset (CCV in *creana* [kɾeˈana] 'raise') and in rare cases up to two in coda positions when coda /ɾ/ is produced as [ɾʂ] (VCC in *ayer* [ˈajeɾʂ] 'yesterday'). These cases were treated simply as open and closed, respectively.

3.3.2. Factor analysis for mixed data

Factor analysis for mixed data (FAMD; Pagès, 2004) belongs to a family of multivariate statistical methods for the dimensional reduction of data. Dimensional reduction methods seek to determine internal correlations among dependent variables within a set of observations through the construction of derived, unobserved variables (variously termed factors, dimensions, or components depending on the technique), and thereby reducing the number of overall variables needed to describe relationships within the data. Certain methods, such as *principal component analysis* (PCA; Abdi & Williams, 2010; Michailidis, 2007) are restricted to the analysis of continuous, quantitative measurements, such as measures of formant frequency or acoustic intensity, while other methods, such as *(multiple) correspondence analysis* (MCA; Abdi & Valentin, 2007) are applicable to data represented by categorical, qualitative variables, such as speaker sex or elicitation style. FAMD incorporates both PCA and MCA methods together into a single analysis, making it ideally suited for the analysis of sociophonetic data which may include a mixture of quantitative acoustic variables along with qualitative social or other categorical variables.

In most dimensional reduction methods, the derived variables are arranged according to the total amount of variance in the dataset described by that variable, such that Dimension 1 is correlated with the greatest amount of data variance among all dimensions, Dimension 2 a lesser amount, and so on.[3] Each derived dimension has a relationship to the original variables in the dataset which is described via ranked *loadings* which indicate how much of each dimension is composed of the various original variables. In this way, the relative influence or importance of the original variables can be determined, as well as

[3] The number of calculated dimensions is maximally the number of dependent variables in the dataset but can be smaller than this—hence the term dimensional *reduction* techniques. Reducing the number of dimensions involved in describing dataset-internal variation is the key feature of such methods.

the relationships between these variables, as variables which have loadings on the same dimension are necessarily correlated with each other to some degree.

PCA, the quantitative method used within FAMD, has been previously used within acoustic phonetic research in relation to the study of vowel acoustics, beginning with a series of studies on Dutch in the 1960s and 70s (Klein, Plomp, & Pols, 1970; Plomp, Pols, & van de Geer, 1967; Pols, Tromp, & Plomp, 1973; Pols, van der Kamp, & Plomp, 1969; van Nierop, Pols, & Plomp, 1973).[4] The approach taken in these studies and subsequent work has been to measure vowel spectra via a series of band-passed filters, which measure the acoustic intensity within a small frequency band. Results from those early studies, and confirmed by later research following similar methods (Jacobi, Pols, & Stroop, 2005; Leinonen, 2009), has confirmed that PC1 and PC2 tend to correlate strongly with F1 and F2 respectively, to the extent that plots of vowel spectra-derived PC1 and PC2 strongly approximate traditional F1xF2 vowel plots. Implementation of PCA analysis with speech data has been typically restricted to the band-passed spectral filter method described above. Other dimensional reduction methods such as Factor Analysis (FA) have been successfully used (e.g., Clopper & Paolillo, 2006) to examine direct measurements of vowel formants as well as vowel duration, and to further consider cross-dialectal production differences (Clopper & Paolillo, 2006; Leinonen, 2009).

In this study we utilize PCA and MCA, implemented together within an overall FAMD model, in an innovative approach to the analysis of vowel acoustics. Rather than applying the band-passed filtered spectral measurement method typical in PCA application to acoustic vowel data, we take direct formant measurements of F1-F3 at vocalic mid-point as well as intensity and pitch (F0), along with overall vowel duration.[5] Unlike other notable studies of vowel acoustics which have utilized PCA, our study considers a broad range of diverse acoustic variables rather than solely relying on band-pass-filtered spectra. Our aim in adopting this method is to maximally reduce bias in the dimensional analysis input by incorporating a variety of acoustic measures, and thereby avoid prejudicing it too much in favour of high-intensity spectral frequency bands (i.e., acoustic formants). Although there is no question that cross-linguistically formants are a primary acoustic characteristic of vowels, in this study we thought it important to also consider as many other potentially relevant acoustic qualities as possible, due to the unusual nature of Media Lengua's vowel system, and to allow dimensional reduction techniques to determine which qualities were most strongly correlated with each other, as well as with Media Lengua's various vowels. Furthermore, the use of MCA within the larger FAMD allows us to include a range of other variables such as speaker sex, elicitation style, syllable stress, etc., covering a host of non-acoustic factors potentially involved in vowel production variation.

3.3.3. Linear mixed effects regression

Given that FAMD has not been readily used as an analytical technique in the field of acoustic phonetics, we also evaluate each dependent quantitative variable using linear mixed effects regression models (a.k.a. 'mixed effects models' or MEMs). This provides us with a basis for comparing and interpreting FAMD results with a well-established analytical tool common in the field. This also allows us to observe relationships in the data which might not be apparent under a single method. Linear regressions are apt for this type

[4] PCA has also been fruitfully applied to the study of articulation (Mokhtari, Kitamura, Takemoto, & Honda, 2007; Mooshammer, 2007), prosody (Hadjipantelis, 2012; Kim, Matachana, Nyman, & Yu, 2020; Tupper, Leung, Wang, Jongman, & Sereno, 2020), and fricative acoustics (Zhao, 2010) among others (we would like to thank an anonymous reviewer for bringing several of these studies to our attention).

[5] It should be noted that these raw quantitative measures are thereafter normalized as part of the FAMD algorithm as per Kassambara, 2017 which states, "Quantitative and qualitative variables are normalized during the [FAMD] analysis in order to balance the influence of each set of variables" (p. 108).

of data as they allow for the analysis of a continuous dependent variable (e.g., formant frequencies, duration, pitch, and intensity) along with multiple independent variables that include an entire population (e.g., Sex, Vowel Stress, Syllable Type, etc.), while testing for possible interactions. The mixed effects version of a linear regression incorporates an additional layer of analysis by including variables whose populations cannot easily be exhausted e.g., speaker and word (where it is impossible to test every word in a language and its variation each time it is uttered). These models help answer two basic questions: (1) do the independent variables in question have a significant effect on the production of the continuous variable? And (2) how large is this effect? For additional information on the MEM technique, Bates et al. (2015) discuss the computational methods behind its application and Gries (2015) discusses the use of MEM specifically in linguistics; see also Kirby and Sonderegger (2018) on the topic of experimental design.

4. Results

4.1. Media Lengua vowels

Following the methods described in Section 3.3.1, per-token vowel formant values were extracted from the audio recordings. **Figure 3** shows sex-differentiated bag plots of the Media Lengua vowel formants within F1xF2 space. Bag plots (Marwick, 2018; Rousseeuw, Ruts, & Tukey, 1999) are two-dimensional analogs of box-and-whiskers plots; the inner polygonal 'bag' (analogous to the 'box') encompasses 50% of tokens and the larger, outer polygonal 'loop' (analogous to the 'whiskers') covers an area three times as large. A median point is also demarcated within the central area of the bag. See Appendix A for a listing of mean vowel formants.

4.2. Factor analysis for mixed data

A single FAMD model was built for the Media Lengua dataset, comprised of six quantitative acoustic variables (F1, F2, F3, pitch, duration, and intensity) and four qualitative variables (speaker sex, vowel stress, syllable type, and elicitation style). Note that vowel is not a variable within the FAMD model, an important point to which we return later in this section. The model was built using the FAMD function provided by the FactoMineR package (Lê, Josse, & Husson, 2008; see Appendix B for full model results). Because the acoustic observations involve different unit types (Hertz, milliseconds, and decibels) and variances, all units are automatically scaled according to within-variable internal variance following the formula (Kassambara, 2017): $\frac{x_i - mean(x)}{sd(x)}$.

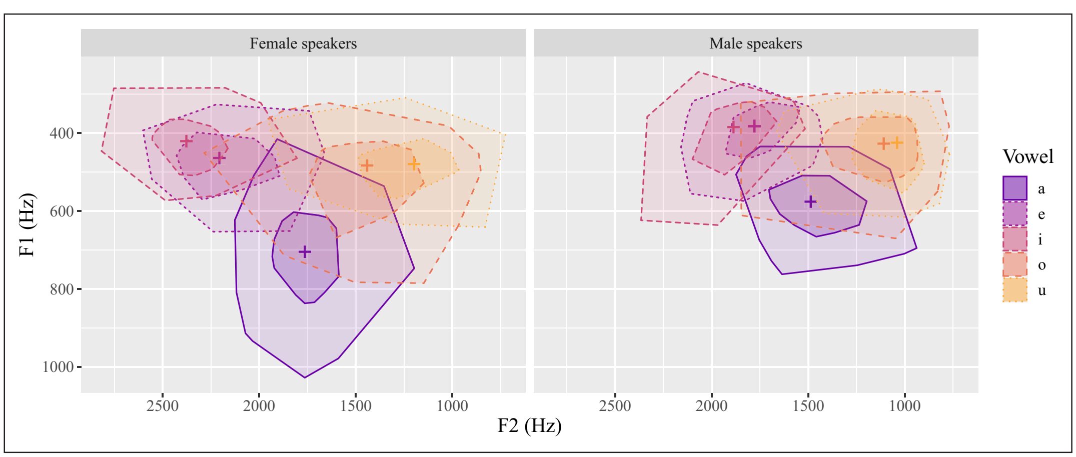


Figure 3: Bag-and-loop plots of vowel tokens according to speaker sex in F1xF2 space.

Figure 4A shows a scree plot, which is a plot type common to several dimensional analysis methods. The scree plot displays the amount of variance within the dataset across the first six computed dimensions from the FAMD. As can be seen, there are diminishing returns to the inclusion of larger numbers of dimensions, as each additional dimension explains a progressively decreasing amount of variation. Moreover, lower dimensions tend to become dominated by single variables and so do not increase explanatory power. **Figure 4B** through **G** show the ranked contributions of the individual quantitative and qualitative variables to Dimensions 1 through 6, respectively. As can be seen, Dimension 5 contains only two variables (syllable type and intensity) with contributions greater than the expected level for that dimension (which serves as a threshold when considering the relative influence of variables) while over 60% of the variation in Dimension 6 is contributed by just a single variable (syllable type). For these reasons, we focus on just the first four FAMD dimensions—together these account for 55% of the variation in the data (see Appendix B), and each contains a multiplicity (i.e., four or more) of variables

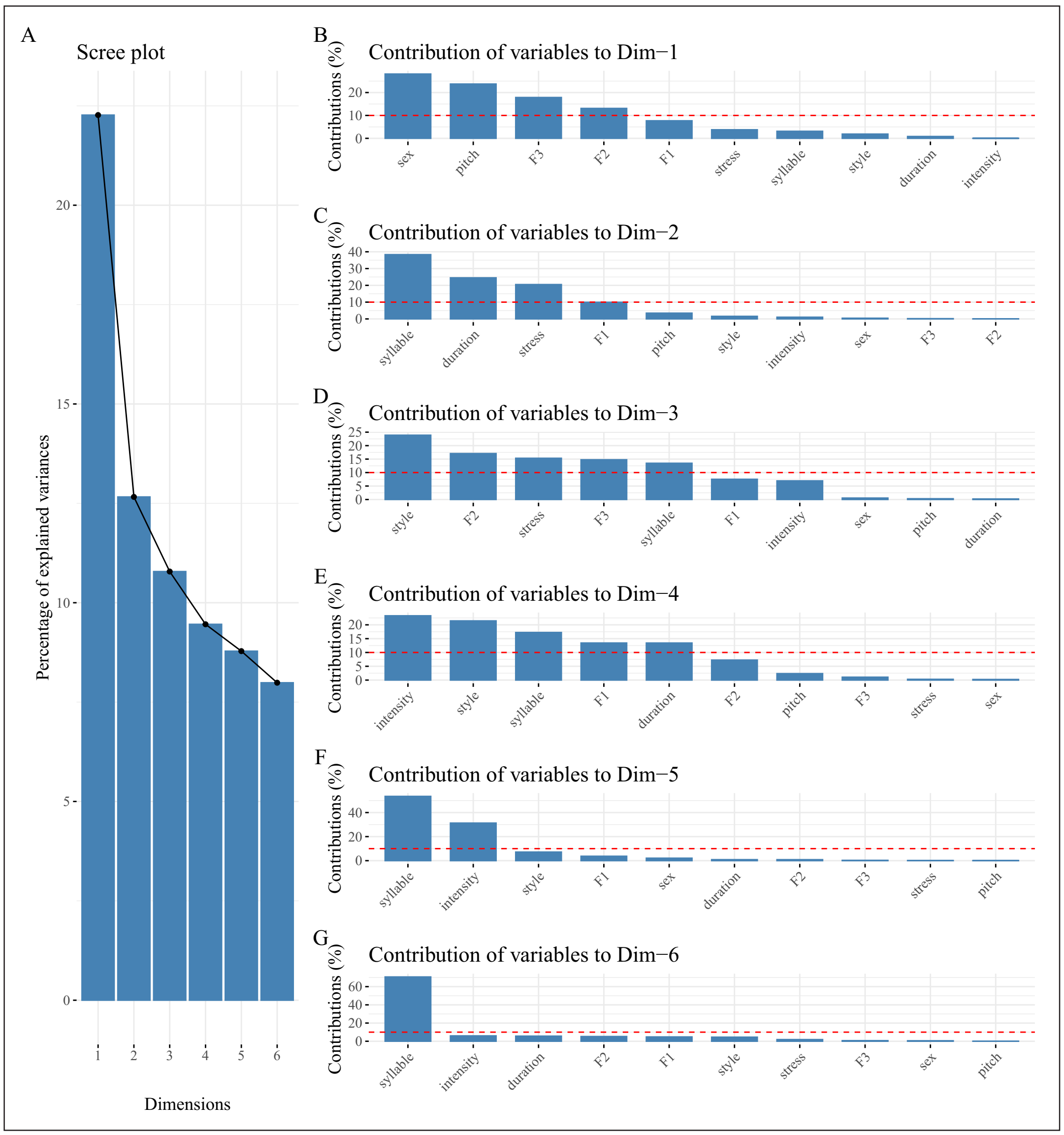


Figure 4: A: Scree plot of FAMD dimensions; **B-G:** Contributions of variables to Dimensions 1–6, red dashed lines indicate average expected contribution level.

which meet or exceed the respective expected contribution level (as indicated by the red dashed line) for that dimension.

Dimension 1 is most strongly correlated with speaker sex which accounts for more than 25% of variation along that dimension, followed by pitch, F3, and F2 in decreasing order of contribution. The remaining variables fall below the expected average contribution threshold (shown by the red dashed line), i.e., they are not very strongly correlated with Dimension 1, although F1 stands out among these as being only just below the average contribution level. The largest contributors to Dimension 2 are syllable type (nearly 40%), followed by duration and stress, and F1 which again falls just below the threshold. Dimension 3 is dominated by elicitation style as the largest contributor, but there are several other large contributors including F2, stress, F3, and syllable type. Finally, Dimension 4 is dominated by intensity and elicitation style, with syllable type, F1, and duration also making large contributions.

In interpreting which external factors the FAMD dimensions might represent according to their relative contributions, we think it is sensible to take Dimension 1 as generally reflecting most substantially the overall influence of speaker sex, but also having large contributions from vowel quality. Sex differences are well-known to correlate strongly with both pitch and vocal tract length, the latter of which further impacts upper vowel formants such as F2 and, notably, F3. Based on this interpretation, F3 appears to have an unexpectedly large role to play within Media Lengua vowel variation in comparison to the lesser-contributing F1 and F2, formants which are more traditionally associated with the acoustic maintenance of vowel quality differentiation, and which form the basis of the most commonly-used interpretation of vowel production as derived from acoustic data, i.e., the standard F1xF2 vowel plot. Dimension 2 is correlated with aspects of prosodic/syllabic structure, but importantly does not include pitch and intensity. Dimension 2 also has an important secondary contribution from F1. Taken together with the low-contribution level of F1 to Dimension 1, we can interpret this as an indication that vowel height generally is of lesser importance within the Media Lengua vowel system relative to vowel front-back position. Dimension 3 is correlated with a large range of variables, with elicitation style standing out as the most important but not by a great amount, and Dimension 4 is even less easy to characterize in a straightforward way. We will return to these interpretations of the FAMD dimensions in the Discussion.

Figure 5 shows the results of the PCA sub-component of the FAMD performed on the quantitative acoustic variables of F1, F2, F3, pitch, duration, and intensity, plotted against

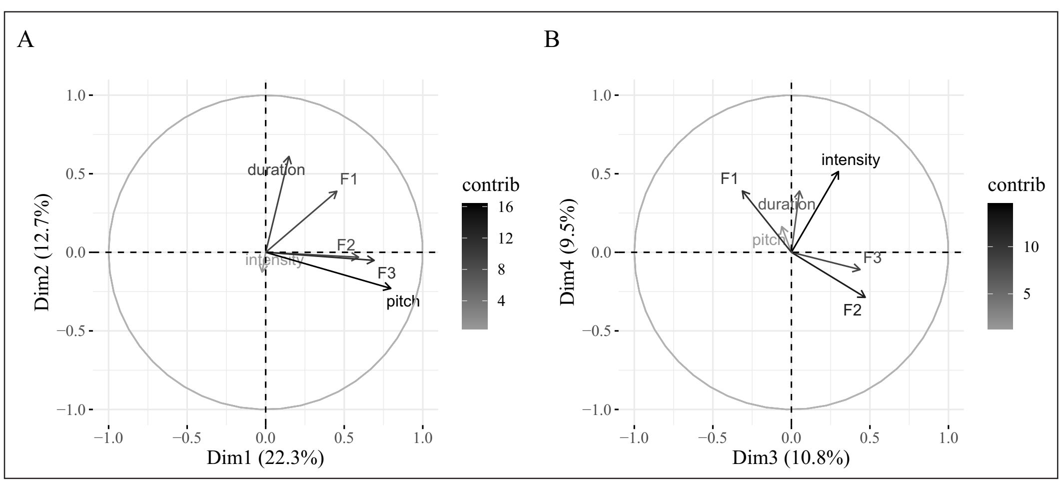


Figure 5: PCA of quantitative variables in **(A)** Dimensions 1–2 and **(B)** Dimensions 3–4.

Dimensions 1–2 in **Figure 5A** and Dimensions 3–4 in **Figure 5B**. The length and relative darkness of the shading of the text and arrows associated with each variable indicate the relative degree of their contributions, and the cross-correlation between the plotted dimensions is indicated via the angles of arrows between the horizontal and vertical axes. For example, in **Figure 5A** the length and sharply horizontal angles of the arrows for F2 and F3 indicates both the strength of their correlation with Dimension 1 (F3's longer and more darkly-shaded arrow indicating its larger contribution) and their corresponding weak-to-nonexistent correlation with Dimension 2. In contrast, the intermediary (roughly 45°) angle of F1's arrow indicates its relatively equal contribution to both dimensions, indicative of its secondary but still important role in the system. In **Figure 5B** we see that F1 and F2 are diametrically opposed, standing at a near 180° angle relative to each other, and correlated with both Dimensions 3 and 4. This opposition indicates that the two are negatively correlated. A negative correlation between F1 and F2 is generally not expected for most vowel systems, as there is typically not a very strong relationship between vowel height and frontedness, which makes this relationship somewhat notable—although it only occurs in two of the lower dimensions which together account for only about 20.3% of variation, and so should not be taken to indicate categorical opposition between these two factors.

Figure 6 shows the results of the MCA sub-component of the FAMD performed on the qualitative variables of speaker sex (2 levels: female and male), syllable type (4 levels: open, closed, open-final, and closed-final), syllable stress (2 levels: stressed and unstressed), and elicitation style (2 levels: wordlist and speech); these variables are plotted according to their contributions to Dimensions 1–2 in **Figure 5A**, and according to their contributions to Dimensions 3–4 in **Figure 6B**. In **Figure 6A**, the diametrically-opposed positions of "m" (male) and "f" (female) at great distance along Dimension 1 reflects the FAMD finding that speaker sex is the single largest contributor to variance in the dataset (see **Figure 4B**). The various levels of syllable type are mostly arranged vertically along Dimension 2, with the exception of open-final syllables which are positioned horizontally away from the other types along Dimension 1, indicating a primary division between open-final versus other syllable types. Closed-final syllables exhibit the next-largest distance from the other types at this level of the dimensional analysis. Stress shows involvement of both Dimensions 1 and 2 in differentiating stressed from unstressed vowels. And, speech style (speech versus wordlist) is mostly differentiated along Dimension 1, but with far less distance between the two levels than among the levels of the other qualitative variables.

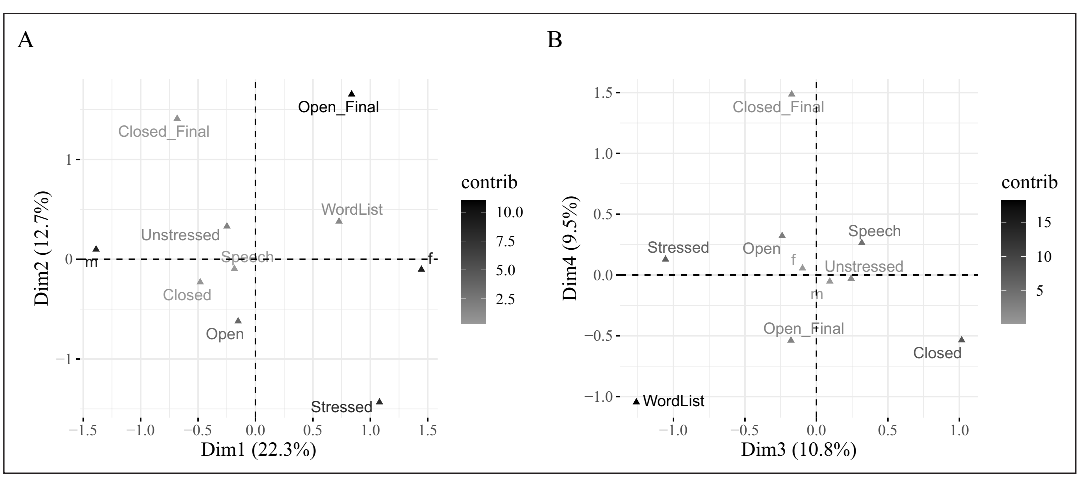


Figure 6: MCA of qualitative variables in **(A)** Dimensions 1–2 and **(B)** Dimensions 3–4.

Figure 6B covers Dimensions 3 and 4 of the MCA. Here, syllable types are again arranged mostly vertically (Dimension 4) except for closed syllables which are differentiated from other types along Dimension 3, indicating a division between them versus other types at this level. Stress shows only differentiation along Dimension 3, while speech style involves both Dimension 3 and 4, indicating a degree more of variation between levels of that variable. In contrast, speaker sex is almost completely non-differentiated in these dimensions. It is especially notable that variation related to sex, while being the largest contributor to Dimension 1, is almost entirely absent from the other dimensions. This suggests that while cross-sex variation is substantial within the Media Lengua vowel system, it does not interact with many of the other aspects of variation aside from (per the PCA) pitch and vowel quality.

Figure 7 plots the combined FAMD (that is, combining both the PCA and MCA) contributions of the original variables in the dataset along Dimensions 1 through 4. Note that **Figure 7A** and **B** have been scaled so that their axis scales are relatively proportional; the result is that **Figure 7B**, which covers a much smaller range of variance, appears much smaller visually. This is intentional and illustrates the fact that the contributions in **Figure 7A** are much more substantial in terms of their role within the overall variance of the dataset. In **Figure 7A**, the general patterning of variables exhibits a strong dichotomy between contributions to Dimensions 1 and 2, with a majority of variables being only strongly correlated with one dimension or the other, the exception being F1 which shows similar contributions to both. Some degree of clustering of variables is also evident: Sex and pitch are the largest contributors to Dimension 1, with F3 and F2 following in close succession; in Dimension 2, after the large contribution from syllable type, duration and stress are closely aligned with each other, suggesting a relationship between the two. In **Figure 7B**, Dimensions 3 and 4 are far less segregated in terms of contributing variables, with both elicitation style and syllable type accounting for relatively large proportions of both dimensions. It is again notable that sex, while strongly correlated with Dimension 1, is almost entirely relegated to that dimension and shows no meaningful contribution to Dimensions 2, 3, or 4.

Lastly, we relate the vowel phoneme categories to the FAMD model. **Figure 8** plots the positions of the 1,202 vowel token observations in the dataset within Dimensions 1 and 2 (in panel A), and within Dimensions 3 and 4 (panel B). Each vowel phoneme is identified by the colour of the individual points and outlined by an irregular polygon hull which captures 100% of the observations associated with that category (per-group mean positions are also present as somewhat larger and more darkly-shaded points among the individual observations, but we do not focus on these). It is important to emphasize, as was mentioned at the beginning of this section, that the vowel categories were not directly included in and did not inform the FAMD analysis in any way—the association between the individual observations and vowel groupings is post hoc, being made subsequent to the analyzing the structure of the dataset, and does not in any way influence the prior output of the model.

Regarding the relative positions of the vowel groups, their arrangement in **Figure 8A** only loosely corresponds to a typical vowel space plot, even if it were rotated around the central meeting point of the axes. This reflects the lesser (although not absent) importance of F1 and F2 (which form typical vowel plot axes) most especially in Dimension 1, relative to other factors such as speaker sex, pitch, and F3 (see **Figure 7**). The vowels in **Figure 8A** can be observed to fall into three broad groups, with varying degrees of overlap within and between them. Falling mostly on the negative side of Dimension 2 (i.e., below the x-axis) we find the front vowels /e/ and /i/ which show a large amount of overlap with

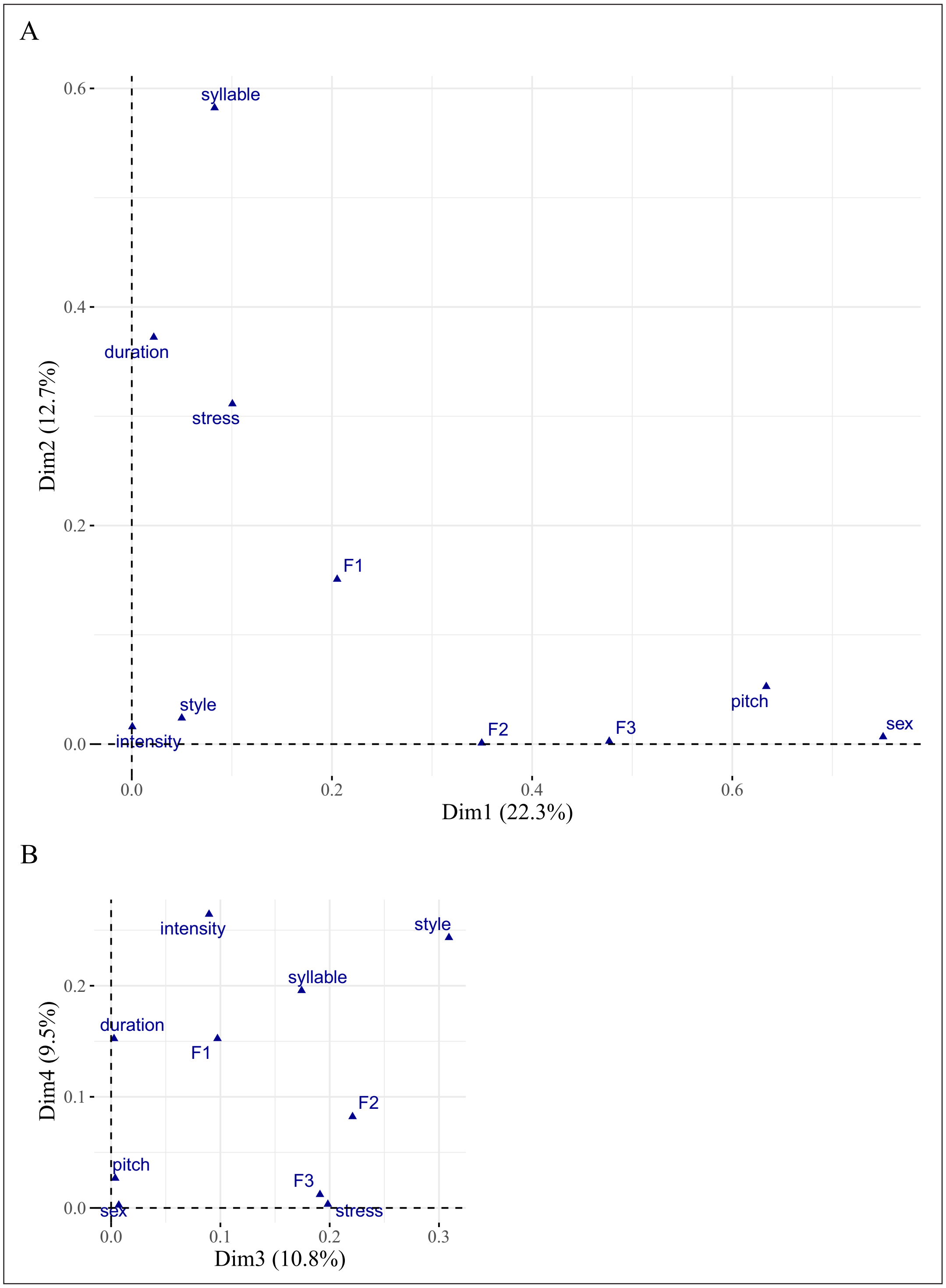


Figure 7: FAMD contributions of variables in **(A)** Dimensions 1–2 and **(B)** Dimensions 3–4.

each other but very little with the other vowels. Clustered near the central 0–0 point and skewing slightly towards the negative side of Dimension 1 (to the left of the y-axis) and the positive side of Dimension 2 (above the x-axis) are the back vowels /o/ and /u/. Like the front vowels, these show large overlap with each other. Finally, the vowel /a/ is distinguished on two counts. It shows a fair degree of overlap with the back vowels, but

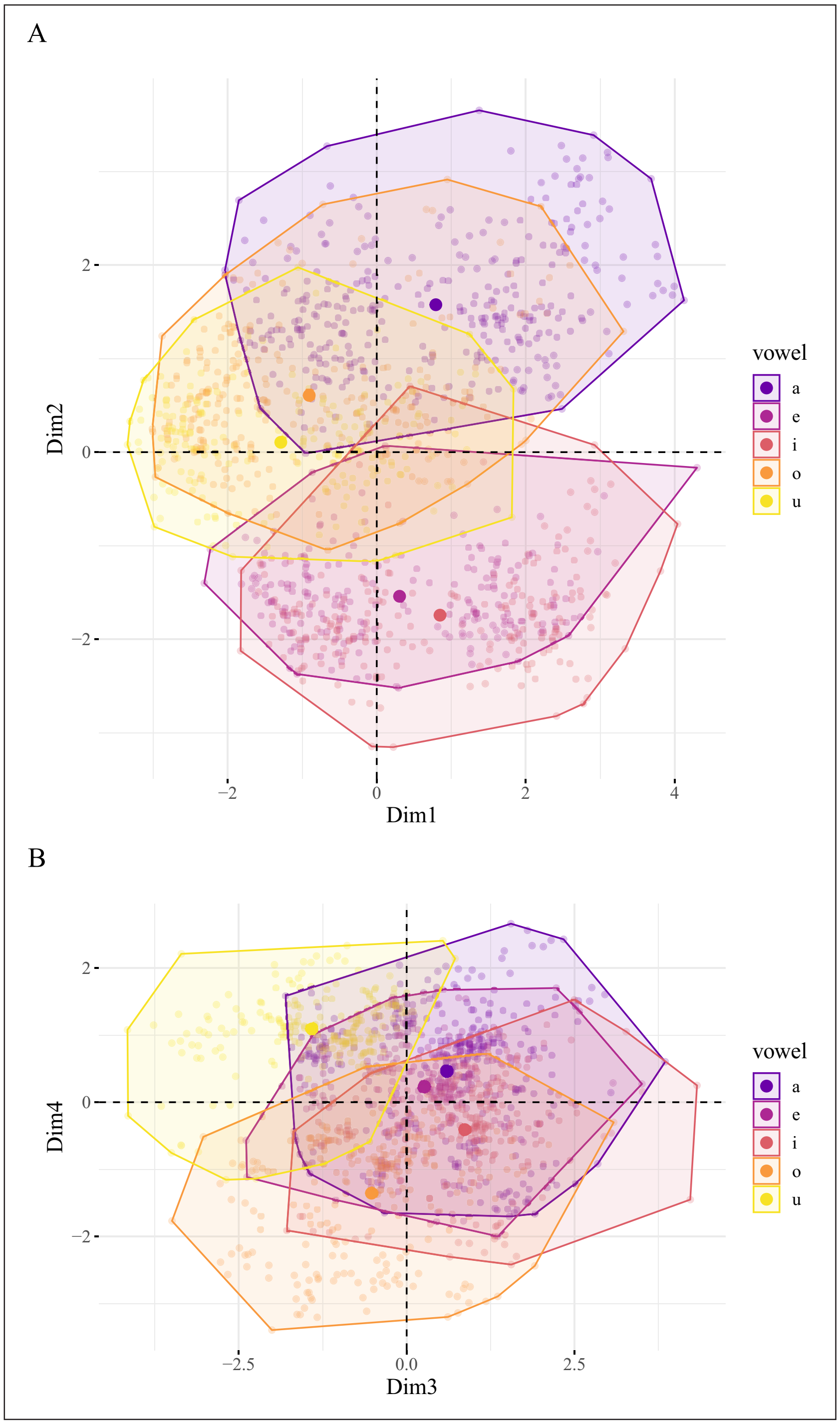


Figure 8: FAMD structure of individual observations according to vowel type in **(A)** Dimensions 1–2 and **(B)** Dimensions 3–4.

virtually none with the front vowels. It also has a relatively large part of its distribution which is more isolated from the other vowels, mostly in terms of Dimension 2, and is the only vowel to do so. Summarizing, in terms of Dimension 1 the clearest distinction is between the back vowels /o, u/ versus the others; in terms of Dimension 2 the clearest distinction is between the front vowels /e, i/ versus the others, and secondarily between /a/ versus most substantially the front vowels, but also the back vowels. **Figure 8B** provides another view on the distribution of vowel variation, showing the vowel tokens and hulls plotted against FAMD Dimensions 3 and 4. Unlike in **Figure 8A**, there is very clear separation between /o/ and /u/ here, while the other three vowels are massively overlapped. Taken as a whole, the degree of cross-vowel overlap observed in both plots suggests that /a/ is the most independent vowel, with the high-and-mid vowels less clearly distinguished in both the front and back regions of the vowel space, and that secondarily the back vowels /o/ and /u/ are more distinguished from each other than the front vowels /i/ and /e/ are from each other. Separation between front versus back vowels is also fairly clear at all levels, but also not especially revealing.

4.3. Linear mixed effects regression

Six separate MEMs were built to test each one of the acoustic correlates explored in the FAMD analysis (F1, F2, F3, pitch, duration, and intensity). Each model tests whether the correlate under analysis differs significantly across each vowel category and other independent predictors: Sex (female/male), Syllable Type (open/closed/open-final/closed-final), Stress (stressed/unstressed syllable), and Style (wordlist/conversational speech). Given that we are primarily interested in /i/'s relationship to /e/ and /u/'s relationship to /o/, we rotate /i/ and /u/ through the model intercept.[6] We did not normalize the vowel data as we are only interested in intra-speaker comparisons. Given that each speaker receives their own intercept in a MEM (using speaker as a random effect), the normalization of unequaled variances is unnecessary (see Drager & Hay, 2012, for a comprehensive overview).

MEMs were created with the *lmer* function of the *lme4* package (Bates et al., 2015), and confidence intervals (see Appendix C) were calculated using the TAB_MODEL function from the SJPLOT package (Lüdecke, 2020). P-values were estimated using the LMERTEST package (Kuznetsova, Brockhoff, & Christensen, 2017). Each model includes speaker and word as random effects. All models were kept maximal (i.e., all predictors were included, significant or not) to maintain similar environments across each model. **Table 2** provides a summary of significant coefficient estimate[7] results of each model (see Appendix C for the full model results); non-significant results are marked as n.s.

Results from the models presented in **Table 2** reveal that formants F1, F2, and F3 play a significant role in differentiating /i/ from /e/, and /u/ from /o/ (in addition to /a/ from /i/, /e/, /u/, and /o/). The F1 model shows that /i/ differs significantly from /e/ by on average 29 Hz, and /u/ from /o/ by on average 35 Hz. These results are remarkably similar to those identified in Stewart (2014), which showed a difference of 41 Hz between /i/ and /e/, and 39 Hz between /u/ and /o/; it is worth noting that his F1 and F2 analysis of Media Lengua vowel formants was based on a different dataset containing elicited speech data from 10 speakers. The only other significant effect identified in the F1 model was sex, with men producing the overall F1 vowel frequencies on average 67 Hz less than women. Interactions between sex and vowel were also tested but non-significant results were revealed for the target vowel pairs. However, men produce the /a/ vowel significantly lower in Hertz frequency, equating to a higher tongue body position compared to women. Men also produce /o/ significantly lower in Hz frequency compared

[6] Viewing each vowel from the intercept (i.e., re-parameterization) simply changes the perceptive of the model (not the model itself), allowing one to view the results from the standpoint of each vowel (see Millar, 2011 for an overview of re-parameterization in regression modeling). As the model itself does not change, any variation attributed to fixed or random effects remains constant.

[7] This is a conservative estimate of the average measurement under analysis between the Intercept and the other predictors under analysis.

Table 2: Significant coefficient estimate results from each model with intercepts rotated between /i/ and /u/ for cross-vowel comparisons.

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**		**Model 6**	
Predictors	F1 (Hz)		F2 (Hz)		F3 (Hz)		Pitch (Hz)		Duration (ms)		Intensity (dB)	
Intercept	i	u	i	u	i	u	i	u	i	u	i	u
Vowel Estimate	438	481	2360	1294	3073	2706	249	245	88	71	72	71
Vowel												
e	29	*n.s.*	–150	915	–120	246	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*
o	79	35	–912	154	–335	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*
i	—	–43	—	1066	—	366	—	*n.s.*	—	17	—	*n.s.*
u	43	—	–1066	—	–366	—	*n.s.*	—	–16	—	*n.s.*	—
a	281	238	–596	470	–242	124	*n.s.*	–11	*n.s.*	29	*n.s.*	*n.s.*
Other factors												
Sex: *Male*	–62	–71	–368	–140	–404	–397	–127	–136	–10	*n.s.*	*n.s.*	*n.s.*
Syllable: *Open*	*n.s.*		*n.s.*		*n.s.*		*n.s*		*n.s.*		*1*	
OpenFinal	*n.s.*		*n.s.*		*n.s*		–6		24		*n.s.*	
ClosedFinal	*n.s.*		–119		*n.s.*		–16		*n.s.*		*n.s.*	
Unstressed	*n.s.*		–41		*n.s.*		–9		*n.s.*		*n.s.*	
Wordlist	*n.s.*		*n.s.*		*n.s.*		19		–16		*n.s.*	
Interactions												
e * Male	*n.s.*	*n.s.*	*n.s.*	–187	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*
o * Male	–38	*n.s.*	150	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*
i * Male	—	*n.s.*	—	228	—	*n.s.*	—	*n.s.*	—	*n.s.*	—	*n.s.*
u * Male	*n.s.*	—	–228	—	*n.s.*	—	*n.s.*	—	*n.s.*	—	*n.s.*	—
a * Male	–90	–82	123	–105	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*

to women in comparison with /i/ (–38 Hz); this is likely a result of the more compact vowel space observed in the speech of men.

The F2 model shows /i/ differs significantly from /e/ by, on average 150 Hz and /u/ differs significantly from /o/ by, on average, 154 Hz. These results equate to a more centralized space for /e/ and /o/ compared to /i/ and /u/, respectively. This also corresponds to the results of Stewart's (2014) F2 analysis of /i/ and /e/, which had an average difference of 125 Hz; the F2 of /u/ and /o/ was non-significant in his study. Additionally, stressed vowels were shown to be significantly fronted by 41 Hz, while vowels in a closed final syllable in ultimate position were significantly retracted by 119 Hz. The latter may be a coarticulation effect caused by anticipating the coda targets. Interactions between sex and vowel were also tested but no significant results were revealed for the target vowel pairs. However, a number of other interactions were significant, which all point to a more retracted vowel space for men compared to women.

The F3 model shows /i/ differs significantly from /e/ by, on average 120 Hz; however, differences between /u/ and /o/ were non-significant. F3 was not analyzed in Stewart (2014). The only additional fixed factor that reached significance was sex, with men showing significantly lower Hertz values by, on average, 400 Hz. Non-significant interactions between sex and vowel were calculated with respect to F3.

Contrarily, pitch, duration, and intensity were all shown to be non-significant factors in differentiating /i/ from /e/ and /u/ from /o/. As expected, however, pitch is correlated with sex, with male speakers' pitch values being an average of 131.5 Hz lower than female speakers. Pitch also plays a significant role in stress, with stressed syllables being, on average, 9 Hz higher in pitch than unstressed syllables. Additionally, vowels in the ultimate syllable are lower in pitch than vowels in non-final syllable position. This is likely because the final syllable comes directly after the stressed syllable causing some degree of deaccentuation. Evidence from Stewart (2015a) supports these findings, which describes a bitonal pitch accent ending a high target (L + H*) on the penultimate syllable. Additionally, there is a tendency for the ultimate syllable to be lower in f0 Hz frequency compared to the baseline f0 in the contours illustrated in Stewart (2015a); especially in boundary tones that end on a low (L% BT). Moreover, the wordlist factor appears to have some influence on pitch, with an overall higher f0 of 19 Hz; an effect possibly caused by a greater tendency for careful speech produced during this data elicitation method.

The duration of a vowel was shown to be longer by, on average, 24 ms when found at the end of a word (final open syllable); a common cross-linguistic tendency corresponding to utterance-final lengthening. However, in the wordlist data, this effect was reduced by 16 ms to only 8 ms, likely due to the fact that speakers were not producing the words within an utterance in that corpus. Moreover, /a/ was shown to be longer in duration by, on average, 29 ms compared to /u/, supporting long-standing evidence that, in general, higher F1 frequencies correspond to longer vowel duration (see Toivonen et al., 2015, who recently revisited the correlation between height and duration in vowel production). The model results for duration also suggest that /i/ is marginally longer than /u/ by 17 ms. Non-significant interactions between sex and vowel were revealed with respect to pitch.

Lastly, the intensity model revealed non-significant results across the board with the exception of a negligible 1 dB difference between open and closed syllables. The lack of a strong correlation between stress and intensity may suggest that the term 'pitch accent' more aptly reflects the differences in syllable quality between penultimate position and non-penultimate position than 'stress,' which is typically used in the Quechuan literature.

Given that non-significant correlations were found across the target mid and high vowel pairs (/i/ & /e/ and /u/ & /o/) for pitch and duration, it can be inferred that qualities which can affect vowels in other languages, such as tone and/or length, are likely not responsible for differentiating the overlapping mid and high vowel clusters in Media Lengua, based on these models. From the preceding analysis, then, we are left with formants, including F3 in the case of /i/ versus /e/, as the primary acoustic correlates for differentiating mid from high vowels.

5. Discussion

5.1. Summary and comparison of results

As this study employs two distinct analytic methods, factor analysis for mixed data (FAMD) and linear mixed effects regression modelling (MEM), the first pertinent consideration is to compare the results forthcoming from each method. While both methods investigate the combined effects of several quantitative and qualitative/categorical variables on vowel production, they do so in quite different ways, each with certain strengths and weaknesses. The FAMD model incorporates all the variables within a coherent analysis; however, it does not relate the variables to the vowel categories themselves or offer direct comparisons between vowel types. In contrast, MEM does directly relate the variables to different vowel categories and compare variation between vowels—however, it does not offer a single coherent analysis of all variables across the full dataset, because separate models are built for each acoustic correlate. We proceed here with a brief summary of the results from each analysis, and then compare these to each other, within the limitations inherent to each methodology.

The FAMD analysis revealed that the 22.3% of variation was attributable to the derived Dimension 1, whose main contributing variables included speaker sex and the acoustic variables of pitch, and formants F3, F2, and F1, in relative order according to their individual degrees of contribution. A further 12.7% of variation was attributable to the derived Dimension 2, composed primarily of the variables syllable type, acoustic duration, and stress. Another 10.8% of overall variation was attributable to Dimension 3, with large contributions from style, F2, F3, stress, and syllable type, while 9.5% of variation was attributable to Dimension 4, also having strong contributions from style and syllable, as well as intensity, duration, and F1. Interpreting these various groupings of major contributing variables for each dimension, Dimension 1 appears strongly related to speaker sex (interpreted broadly and subsuming some variation in related acoustic variables),[8] F3 (discussed below), and vowel positional quality—with front-back position superseding height. Dimension 2 is associated with syllable type and factors which may be linked to it phonologically (duration and stress). While a full examination of Media Lengua stress patterning is beyond the scope of this paper, we can note that stress and syllable type are indeed strongly correlated ($\chi^2 = 100.54(3)$, $p < .001$), with stressed syllables being almost completely excluded from closed-final (0 tokens) and open-final (3 tokens) syllables (note too that syllable type makes an important contribution in Dimension 4, and dominates Dimensions 5 and especially 6; see **Figure 4**).[9] Dimensions 3 and 4 together share large contributions from style which suggests a broad, but much subordinated role for that factor. Using these interpretations, the greatest correlates of variation in Media Lengua vowel production are found to be: speaker sex, F3, vowel quality, syllable type, and speech style, roughly in that order.

Although concurring in several areas, the MEM results are not always easily compared with the FAMD findings. In terms of acoustic qualities, the MEMs show that the vowels in each mid-high pair, /e/ versus /i/ and /o/ versus /u/, are significantly different from each other and are most strongly differentiated according to vowel formants, specifically F1 and F2, rather than other acoustic factors. It is also notable that F3 serves to differentiate /e/ from /i/, but not /o/ from /u/, which seems to be a novel finding in comparison with previous studies. While these results cannot be directly compared to the FAMD output, it is notable that within the MEMs for F1 and F2 the magnitude of difference was consistently greater between /o, u/ than between /e, i/. This is congruent with the FAMD results at the level of Dimensions 3 and 4, as shown in **Figure 8B**, where /o/ and /u/ are more clearly differentiated from each other as compared with /i, e/; and even at the higher Dimensions 1 and 2 (**Figure 8A**) it is arguable that there is a similar albeit smaller difference between the two vowel pairs. This trend is actually the opposite in Stewart (2014) where the front series showed more acoustic distance compared to the back series. One possible explanation for this is that the /o/ measurements from this study show greater fronting and an overall larger vowel space compared to Stewart's (2014) measurements (compare

[8] The effect of sex on vocal pitch is well-documented in the literature and is not discussed here: See Coleman (1971, 1976), Boersma and Weenink (1996), inter alia. The relationship between F3 and sex is dealt with in the subsequent subsection.

[9] Media Lengua stress/pitch-accent is normally carried on the penultimate syllable, so this general pattern is actually expected and the three stressed, open-final tokens are exceptional.

Figure 2 with **Figure 3**), which might be attributed to the increased number of participants (over double) in this study, which allowed this effect to be captured.

Qualitative variable results are even less easily compared across methods, because within MEM analysis their role is always contextualized within a given model (**Table 2**). For example, none of the quantitative variables exhibit a significant effect in relation to F1 within the model structured around that variable, whereas the majority do so in relation to F2. The FAMD results are not discrete in this way, such that the quantitative and qualitative variables are integrated within the larger analysis. Comparing across the various MEMs, we can see that syllable type, speaker sex, and stress each have significant effects within a minimum of three distinct MEMs—and these are also the three largest qualitative contributors within the FAMD. However, while it is difficult to say more than this from the MEM results in terms of the relative magnitude of effect from each of these variables within the system as a whole, the FAMD analysis does provide such a ranking: sex > syllable > stress.

Taken as a whole, then, the FAMD and MEM results both support and complement each other. Some divergence between the /e, i/ versus /o, u/ pairs emerges in both analyses, as does the relative importance of the categorical variables of speaker sex, syllable type, and stress. The MEM provides direct comparison across vowel types, and greater detail regarding the significance of differences in variation therein. FAMD provides a full-system analysis of variables of both types, quantitative and qualitative/categorical. The two methods, at least within the present study, function in a compatible way where their combined results are greater in both quantity and quality than they are when taken independently of each other.

5.2. The role of F3

One of the more intriguing results emerging from the combined analysis in this study concerns the role of F3. Under the FAMD, F3 emerges as the third-largest contributor to Dimension 1 (and the second-largest quantitative contributor) after speaker sex and pitch. As pitch is a known correlate of sex, and vowel formants are also known to vary systematically between female versus male speakers, we have therefore interpreted Dimension 1 as mostly reflecting speaker sex taken broadly, with variation in the acoustic factors contributing to Dimension 1 (in relative order: pitch, F3, F2, F1) all being correlated with sex to some degree. Under the MEM model for F3, sex was the most substantial determinant, differing between male versus female speakers to a larger degree than for any other categorical variable. Although the MEMs for both F2 and pitch also indicate a significant and substantial role for sex, it is difficult to compare the different MEM results to each other as they are scaled differently—although F2, F3, and pitch are all measured in Hz, the variance across this unit differs widely between the three variables. In contrast, the FAMD explicitly ranks these variables in accordance with each dimension. As noted, F3 is the third-ranked contributor to Dimension 1—however, it shows negligible levels of contribution to Dimensions 2 and 4, and only a slightly above average contribution to Dimension 3.

A comparison between the MEMs for F2 and F3 in relation to vowel categories is also informative. While both models produce similar results in terms of the occurrence of significant differences (with the exception that F3 is non-significantly different between /u/ versus /o/), the magnitude of difference is larger for F2 in every case. That is, F2 and F3 results pertaining to vowel types are nearly always compatible, and F2 is moreover a better indicator of significant difference between types. The general, and perhaps unexpected conclusion regarding F3 variation, then, is that it is *primarily* related to speaker sex, and more secondarily to vowel type.

The relationship between F3 and speaker sex is not heretofore unknown, although it is typically subsumed within a broader investigation of formant-frequency differences (Skuk & Schweinberger, 2014), and/or often goes unmentioned next to F1 and F2 (Simpson, 2009). However, some research has shown that F3 provides a superior function to either F1 or F2 in discrimination tasks of speaker sex (Hillenbrand & Clark, 2009). One of the most interesting findings related to our study is Whiteside (2001), who examined production differences between sexes across a wide age-range of speakers. Whiteside found that F3 exhibited sex-based differentiation even in pre-pubescent speakers, and that the relationship between F3 and pitch values showed sex-specific developmental patterns. While the present study is not focused on sex differences in vowel production, nor does it consider production differences over the lifespan, these other studies demonstrate that our finding regarding the strong connection between F3 production and speaker sex is not entirely novel, although it may not be as widely known as e.g., the relationship between sex and vocal pitch.

5.3. Contextualizing the results

What does this mean for Media Lengua? Results from these analyses suggest that F1 and F2 may not be the most salient cues used by speakers to differentiate mid and high vowels. Given the extensive literature on vowel formants as a primary cue for vowel production, the fact that other cues may be more salient may, at first glance, seem peculiar. However, after considering that mid and high vowels are highly overlapping in their respective regions of acoustic space (**Figure 2A**), it becomes clear that additional cues may be an important factor for extrapolating lexical meaning of a given word. Our results suggest that speakers may be relying more on the structure of the utterance (syllable type, stress), who is producing the utterance (sex), and other factors such as context, when parsing the speech stream. If distinguishing mid and high vowels are a low priority for Media Lengua speakers, who instead rely on the overall phonological shape and context of a word, inferring meaning based solely on a mid and high vowel contrast may only be important in limited contexts.

A brief analysis of the recently published descriptive Media Lengua dictionary (Stewart, Prado Ayala, & Gonza Inlago, 2020) shows that while speakers use mid vowels in approximately 71% of the lexicon, there is a high degree of spelling variation (24%) where mid vowels are interchangeable with high vowels. Additionally, only 0.67% of the lexicon were identified as minimal pairs differing by mid and high vowels; all of which were either from different parts of speech or different semantic categories, making their meaning easily identifiable based on context. A context-based rationale is even more justifiable when approximately four times the number of lexical entries in Media Lengua are homonyms. Nevertheless, while the contrastive functional load for mid and high vowels may be low, it is not zero; there are cases where speakers will reject replacing a high vowel for a mid vowel and vice versa e.g., *vos* ‘2nd person pronoun’ is rejected as **vus* and *enseñana* ‘teach’ is rejected as **inseñana* or **insiñana*, though *ensiñana* is an accepted variation (and even though the rejected variants of both words were still understandable). Additionally, vowel-sequences (i.e., diphthongs), which are found in approximately 28% of the lexicon, are rarely accepted (2%) as monophthongs in spelling variations and never found with the target vowels in the same acoustic region switched e.g., /ei/ for /ie/. These observations are in line with our results suggesting that formants still play a role in vowel production and lexical interpretation with respect to the corner vowels, even if they are not found to be the most salient cue for high and mid vowel contrasts. This may motivate the maintenance of the small, albeit significant, separation of mid and high vowel categories observed in acoustic space (**Figure 2A**).

In summary, in most cases, if a monophthong is produced in its expected trisect of acoustic space (front, back, low), little more is needed in a vowel slot to extrapolate meaning, as long as the rest of the pre- and post-sequences are recognizable in phonological shape and the context is unambiguous.

5.4. Conclusion

Our study has compared the role of a range of acoustic and categorical variables in relation to Media Lengua vowel phonemes using two different statistical methods, factor analysis for mixed data (FAMD) and linear mixed effects regression modelling (MEM). The aggregated results indicate that the largest influence on variation is speaker sex, and related acoustic correlates including pitch (a well-known correlate) and F3 (less well-known but documented in the literature), with syllable type and, to some extent, stress also being major contributing factors. We interpret these results to indicate that Media Lengua speakers rely on a variety of contextual cues beyond the first two vowel formants including syllable structure and prosodic elements, or social factors such as speaker sex when differentiating mid versus high vowels. This interpretation is supported by prior research on Media Lengua perception (Stewart, 2018b) and spelling variation (Stewart, Prado Ayala, & Gonza Inlago, 2020), which show a high degree of overlap between mid and high vowels. However, an important role for F1 and F2 is still maintained amidst these other cues, also confirmed via recent work on Media Lengua diphthong production (Onosson & Stewart, in press). The implications of the results from our study point the way towards future research explicitly focused on areas such as: the role of F3 and its relation to speaker sex, in Media Lengua or indeed other languages as it is certainly under-explored relative to F1 and F2; syllable- and stress-patterning in Media Lengua; and the impact of speech style, a lesser-contributing factor in our analysis but nonetheless one with some apparent effect on production.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** PDF file with the descriptive statistics. DOI: https://doi.org/10.5334/labphon.291.s1
- **Appendix B.** PDF file with the factor analysis for mixed data. DOI: https://doi.org/10.5334/labphon.291.s2
- **Appendix C.** PDF file with the mixed effects models. DOI: https://doi.org/10.5334/labphon.291.s3

Acknowledgements

We would also like to thank Lucía Gonza Inlago, Mercedes Tobango, Antonio Maldonado, and the consultants for taking part in this investigation.

Funding Information

Data collection for this research was supported in part by funding from the Social Sciences and Humanities Research Council of Canada (IDG 430-2018-00032).

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Jesse Stewart carried out and/or coordinated fieldwork for this study and conducted the linear mixed-effects regression analysis. Sky Onosson conducted the factor analysis for mixed data.

References

Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics*. DOI: https://doi.org/10.4324/9781315516257-3

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 417–520. DOI: https://doi.org/10.1002/wics.101

Bates, D., Mächler, M., Bolker, B., Walker, S., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer, version 3.4* (Report 132). Institute of Phonetic Sciences of the University of Amsterdam.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer. Computer application, version 6.1.16. Retrieved from http://www.fon.hum.uva.nl/praat/

Buchan, H. (2012). *Phonetic variation in Gurindji Kriol and Northern Australian English: A longitudinal study of fricatives in maternal speech* (Doctoral dissertation, University of Wollongong). Retrieved from https://ro.uow.edu.au/theses/3789/

Bundgaard-Nielsen, R., & O'Shannessy, C. (2019). Voice onset time and constriction duration in Warlpiri stops (Australia). In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3612–3616). Australasian Speech Science and Technology Association Inc. Retrieved from https://assta.org/proceedings/ICPhS2019/papers/ICPhS_3661.pdf

Chen, M. Y. (1996). *Acoustic Correlates of Nasality in Speech* (Doctoral dissertation, MIT). Retrieved from http://hdl.handle.net/1721.1/11180

Clopper, C. G., & Paolillo, J. C. (2006). North American English vowels: A factor-analytic perspective. *Literary and Linguistic Computing*, *21*(4), 445–462. DOI: https://doi.org/10.1093/llc/fql039

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, *14*(3), 565–577. DOI: https://doi.org/10.1044/jshr.1403.565

Coleman, R. O. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, *19*(1), 168–180. DOI: https://doi.org/10.1044/jshr.1901.168

Crandall, I. B. (1925). The sounds of speech. *Bell Systems Technical Journal*, *4*, 586–626. DOI: https://doi.org/10.1002/j.1538-7305.1925.tb03969.x

Deibel, I. (2017). *VO vs. OV: What conditions word order variation in Media Lengua?* Bremen, Germany: Colloquium on Mixed Languages.

Deibel, I. (2019). Adpositions in Media Lengua: Quichua or Spanish? – Evidence of a lexicalfunctional split. *Journal of Language Contact*, *12*(2), 404–439. DOI: https://doi.org/10.1163/19552629-01202006

Deibel, I. (2020). The contribution of grammar and lexicon to language switching costs: Examining contact-induced languages and their implications for theories of language representation. *Bilingualism: Language and Cognition*, 1–16. DOI: https://doi.org/10.1017/S1366728919000865

Drager, K., & Hay, J. (2012). Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change*, *24*(1), 59–78. DOI: https://doi.org/10.1017/S0954394512000014

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 171–272). Amsterdam: North Holland Publishing Company.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. I. Hualde & J. S. Cole (Eds.), *Laboratory phonology 9* (pp. 353–380). Mouton de Gruyter.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*, 765–768. DOI: https://doi.org/10.1121/1.1908022

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. In E. Zwirner & W. Bethge (Eds.), *Phonetic Sciences. 5th International Congress, Münster, August 1964* (pp. 306–311). DOI: https://doi.org/10.1159/000426965

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*(2), 185–205. DOI: https://doi.org/10.1017/S0025100311000193

Gómez-Rendón, J. (2005). La Media Lengua de Imbabura. In H. Olbertz & P. Muysken (Eds.), *Encuentros y Conflictos: Bilingüismo y Contacto de Lenguas en el Mundo Andino* (pp. 39–58). Madrid: Iberoamericana. DOI: https://doi.org/10.31819/9783865278968-003

Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, *10*(1), 95–125. DOI: https://doi.org/10.3366/cor.2015.0068

Hadjipantelis, P. Z. (2012). Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models. *The Journal of the Acoustical Society of America*, *131*, 4651. DOI: https://doi.org/10.1121/1.4714345

Hendy, C. R. (2019). *The Distribution and Acoustic Properties of Fricatives in Light Warlpiri*. (BA Honours Thesis, Australian National University). Retrieved from https://openresearch-repository.anu.edu.au/handle/1885/200483

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166. DOI: https://doi.org/10.3758/APP.71.5.1150

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*, 105–113. DOI: https://doi.org/10.1121/1.1906982

Jacobi, I., Pols, L. C. W., & Stroop, J. (2005). Polder Dutch: aspects of the /ei/-lowering in standard Dutch. *Interspeech*, 2877–2880. Retrieved from http://cf.hum.uva.nl/poldernederlands/pdf/is05_ja1.pdf

Jarrín Paredes, G. (2014). *Estereotipos Lingüísticos En Relación al Kichwa y a La Media Lengua En Las Comunidades de Angla, Casco Valenzuela, El Topo y Ucsha de La Parroquia San Pablo Del Lago* (Doctoral dissertation, Pontificia Universidad Católica Del Ecuador, Quito). Retrieved from http://repositorio.puce.edu.ec/handle/22000/8234

Johnson, K. (2000). Adaptive disperion in vowel perception. *Phonetica*, *57*, 181–188. DOI: https://doi.org/10.1159/000028471

Jones, C., & Meakins, F. (2013). Variation in voice onset time in stops in Gurindji Kriol: Picture naming and conversational speech. *Australian Journal of Linguistics*, *33*, 196–220. DOI: https://doi.org/10.1080/07268602.2013.814525

Jones, C., Meakins, F., & Buchan, H. (2011). Comparing vowels in Gurindji Kriol and Katherine English: Citation speech data. *Australian Journal of Linguistics*, *31*, 305–326. DOI: https://doi.org/10.1080/07268602.2011.598629

Jones, C., Meakins, F., & Mauwiyath, S. (2012). Learning vowel categories from maternal speech in Gurindji Kriol. *Language Learning*, *62*(4), 1052–1078. DOI: https://doi.org/10.1111/j.1467-9922.2012.00725.x

Kassambara, A. (2017). *Practical Guide To Principal Component Methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra*. STHDA.

Kewley-Port, D. (2001). Vowel formant discriminaton II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America*, *85*, 1726–1740. DOI: https://doi.org/10.1121/1.1400737

Kim, S., Matachana, C., Nyman, A., & Yu, K. M. (2020). Creak in the phonetic space of low tones in Beijing Mandarin, Cantonese, and White Hmong. In *10th International Conference on Speech Prosody 2020, Tokyo* (pp. 523–527). Retrieved from https://www.isca-speech.org/archive/SpeechProsody_2020/pdfs/252.pdf. DOI: https://doi.org/10.21437/SpeechProsody.2020-107

Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, *70*, 70–85. DOI: https://doi.org/10.1016/j.wocn.2018.05.005

Klein, W., Plomp, R., & Pols, L. C. W. (1970). Vowel Spectra, Vowel Spaces, and Vowel Identification. *The Journal of the Acoustical Society of America*, *48*(4B), 999–1009. DOI: https://doi.org/10.1121/1.1912239

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Ladefoged, P., & Maddieson, I. [Ian]. (1996). *Sounds of the world's languages*. Blackwell.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. DOI: https://doi.org/10.18637/jss.v025.i01

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America*, *31*(4), 428–435. DOI: https://doi.org/10.1121/1.1907729

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, *33*, 419–425. DOI: https://doi.org/10.1121/1.1908681

Leinonen, T. (2009). Factor analysis of vowel pronunciation in Swedish dialects. *Computing and Language Variation: A Special Issue of International Journal of Humanities and Arts Computing*, *2*(October 2008), 189–204. DOI: https://doi.org/10.3366/E175385480900038X

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862. DOI: https://doi.org/10.2307/411991

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Orlando: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403–439). DOI: https://doi.org/10.1007/978-94-009-2037-8_16

Lipski, J. (2016). Language switching constraints: More than syntax? Data from Media Lengua. *Bilingualism: Language and Cognition*, *25*. DOI: https://doi.org/10.1017/S1366728916000468

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories – hot spots or adaptive dispersion? In *Proceedings of the XIIIth Swedish Phonetics Conference (FONETIK 2000),* Skövde, Sweden, May 24–26, 2000 (pp. 93–96).

Lüdecke, D. (2020). sjPlot: Data Visualization for Statistics in Social Science. R package, version 2.8.4. Retrieved from https://cran.r-project.org/package=sjPlot

Maddieson, I. [I.]. (1985). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic linguistics: essays in honor of Peter Ladefoged* (pp. 203–221). Academic Press.

Maddieson, I. [Ian]. (1984). *Patterns of Sounds (Cambridge Studies in Speech Science and Communication)*. Cambridge: Cambridge University Press.

Marwick, B. (2018). 000_geom_bag.r. R package. Retrieved from https://gist.github.com/benmarwick/00772ccea2dd0b0f1745

Maurer, D., Cool, N., Landis, T., & D'Heureuse, C. (1991). Are measured differences between the formants of men, women and children due to F0 differences? *Journal of the International Phonetic Association*, *21*(2), 66–79. DOI: https://doi.org/10.1017/S0025100300004412

Meakins, F. (2013). Gurindji Kriol. In S. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *The survey of pidgin and creole languages: Vol. III* (pp. 131–139). Oxford University Press.

Meakins, F., & Stewart, J. (in press). Cambridge Handbook of Language Contact. In S. Mufwene & A. M. Escobar (Eds.), *The Cambridge Handbook of Language Contact in Population Structure*. Cambridge University Press.

Michailidis, G. (2007). Principal component analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: With examples in R, SAS and ADMB*. John Wiley and Sons. DOI: https://doi.org/10.1002/9780470094846

Mokhtari, P., Kitamura, T., Takemoto, H., & Honda, K. (2007). Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients. *Journal of Phonetics*, *35*(1), 20–39. DOI: https://doi.org/10.1016/j.wocn.2006.01.001

Mooshammer, C. (2007). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *The Journal of the Acoustical Society of America*, *127*, 1047. DOI: https://doi.org/10.1121/1.3277160

Muehlbauer, J. (2012). Vowel spaces in Plains Cree. *Journal of the International Phonetic Association*, *42*(1), 91–105. DOI: https://doi.org/10.1017/S0025100311000302

Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In A. R. Highfield (Ed.), *Historicity and variation in Creole studies (Vols 57–78)*. Karoma Publishers.

Muysken, P. (1997). Media Lengua. In S. Thomason (Ed.), *Contact languages: A Wider Perspective* (pp. 365–426). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cll.17.13muy

Nguyễn, Đ.-H. (1997). *Vietnamese: Tiếng Việt không son phấn*. John Benjamins.

Ohala, J. J., & Eukel, B. W. (1987). Explaining the Intrinsic Pitch of Vowels. In R. Channon & L. Shockey (Eds.), *In Honor of Ilse Lehiste* (pp. 207–215). Dordrecht/Providence: Foris Publications. DOI: https://doi.org/10.1515/9783110886078.207

Onosson, S., & Stewart, J. (in press). The effects of language contact on non-native vowel sequences in lexical borrowings: The case of Media Lengua. *Language and Speech*. DOI: https://doi.org/10.1177/00238309211014911

Pagès, J. (2004). Analyse Factorielle de Données Mixtes: Principe et Exemple d'Application. *Revue Statistique Appliquee*, *4*, 93–111.

Plomp, R., Pols, L. C. W., & van de Geer, J. P. (1967). Dimensional Analysis of Vowel Spectra. *The Journal of the Acoustical Society of America*, *41*(3), 707–712. DOI: https://doi.org/10.1121/1.1910398

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America*, *53*(4), 1093–1101. DOI: https://doi.org/10.1121/1.1913429

Pols, L. C. W., van der Kamp, L. J. T., & Plomp, R. (1969). Perceptual and Physical Space of Vowel Sounds. *The Journal of the Acoustical Society of America*, *46*(2B), 458–467. DOI: https://doi.org/10.1121/1.1911711

R Core Team. (2020). R: A language and environment for statistical computing. Programming language, version 4.0.2. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rosen, N. (2006). Language contact and Michif stress assignment. *Sprachtypol. Univ. Forsch (STUF)*, *59*, 170–190. DOI: https://doi.org/10.1524/stuf.2006.59.2.170

Rosen, N. (2007). *Domains in Michif Phonology* (Doctoral dissertation, University of Toronto). Retrieved from https://twpl.library.utoronto.ca/index.php/twpl/article/view/6495

Rosen, N., Stewart, J., Pesch-Johnson, M., & Sammons, O. (2019). Michif VOT. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences,* Melbourne, Australia 2019 (pp. 1372–1376). Canberra, Australia: Australasian Speech Science and Technology Association Inc. Retrieved from https://assta.org/proceedings/ICPhS2019/papers/ICPhS_1421.pdf

Rosen, N., Stewart, J., & Sammons, O. (2020). How "mixed" is mixed language phonology? An acoustic analysis of the Michif vowel system. *Journal of the Acoustical Society of America*, *147*(4), 2989–2999. DOI: https://doi.org/10.1121/10.0001009

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, *53*(4), 382–387. DOI: https://doi.org/10.1080/00031305.1999.10474494

Sharf, D. (1962). Duration of the post-stress intervocalic stops and preceding vowels. *Language and Speech*, *5*(1), 26–30. DOI: https://doi.org/10.1177/002383096200500103

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Linguistics and Language Compass*, *3*(2), 621–640. DOI: https://doi.org/10.1111/j.1749-818X.2009.00125.x

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language and Hearing Research*, *57*, 285–296. DOI: https://doi.org/10.1044/1092-4388(2013/12-0314)

Stewart, J. (2011). *A Brief Descriptive Grammar of Pijal Media Lengua and an Acoustic Vowel Space Analysis of Pijal Media Lengua and Imbabura Quichua*. (Master's Thesis, University of Manitoba). Retrieved from http://hdl.handle.net/1993/4882

Stewart, J. (2014). A comparative analysis of Media Lengua and Quichua vowel production. *Phonetica*, *71*, 159–182. DOI: https://doi.org/10.1159/000369629

Stewart, J. (2015a). Intonation patterns in Pijal Media Lengua. *Journal of Language Contact*, *8*, 223–262. DOI: https://doi.org/10.1163/19552629-00802003

Stewart, J. (2015b). *Production and perception of stop consonants in Spanish, Quichua, and Media Lengua* (Doctoral dissertation, University of Manitoba). DOI: https://doi.org/10.1159/000369629

Stewart, J. (2018a). Voice onset time production in Spanish, Quichua, and Media Lengua. *Journal of the International Phonetic Association*, *48*(2), 173–197. DOI: https://doi.org/10.1017/S002510031700024X

Stewart, J. (2018b). Vowel perception by native Media Lengua, Quichua, and Spanish speakers. *Journal of Phonetics*, *71*, 177–193. DOI: https://doi.org/10.1016/j.wocn.2018.08.005

Stewart, J. (2020). A preliminary, descriptive survey of rhotic and approximant fricativization in Northern Ecuadorian Andean Spanish varieties, Quichua, and Media

Lengua. In R. Rao (Ed.), *Spanish Phonetics and Phonology in Contact: Studies from Africa, the Americas, and Spain*. DOI: https://doi.org/10.1075/ihll.28.05ste
Stewart, J., & Kohlberger, M. (2017). Earbuds: A method for measuring nasality in the field. *Language Documentation and Conservation*, *11*, 49–80. Retrieved from http://hdl.handle.net/10125/24724
Stewart, J., & Meakins, F. (2021). Advances in mixed language phonology: An overview of three case studies. In M. Mazzoli & E. Sippola (Eds.), *New Perspectives on Mixed Languages: From Core to Fringe. Language Contact and Bilingualism 18* (pp. 57–92). De Gruyter Mouton. DOI: https://doi.org/10.1515/9781501511257-003
Stewart, J., Meakins, F., Algy, C., Ennever, T., & Joshua, A. (2020). Fickle fricatives: Obstruent perception in Gurindji Kriol and Roper Kriol. *Journal of the Acoustical Society of America*, *147*(4), 2766–2778. DOI: https://doi.org/10.1121/10.0000991
Stewart, J., Meakins, F., Algy, C., & Joshua, A. (2018). The Development of Phonological Stratification: Evidence from Stop Voicing Perception in Gurindji Kriol and Roper Kriol. *Journal of Language Contact*, *11*(1), 71–112. DOI: https://doi.org/10.1163/19552629-01101003
Stewart, J., Prado Ayala, G., & Gonza Inlago, L. (2020). Media Lengua Dictionary. *Dictionaria*, *13*, 1–3179. DOI: https://doi.org/10.5281/zenodo.4147099
Styler, W. (2015). *On the Acoustical and Perceptual Features of Vowel Nasality* (Doctoral dissertation, University of Colorado). Retrieved from https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/0g354f20t
Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Logan, J., Ramlakhan, N., & Stone, A. (2015). Vowel height and duration. In U. Steindl, T. Borer, H. Fang, A. P. García, P. Guekguezian, B. Hsu, … I. C. Ouyang (Eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 64–71). Cascadilla Proceedings Project.
Tupper, P., Leung, K., Wang, Y., Jongman, A., & Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America*, *147*, 2570. DOI: https://doi.org/10.1121/10.0001024
van Nierop, D. J., Pols, L. C. W., & Plomp, R. (1973). Frequency Analysis of Dutch Vowels From 25 Female Speakers. *Acustica*, *29*(2), 110–118.
Walker, R. (1999). Guaraní voiceless stops in oral versus nasal contexts: An acoustical study. *Journal of the International Phonetic Association*, *29*, 63–94. DOI: https://doi.org/10.1017/S0025100300006423
Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a crosssectional study. *The Journal of the Acoustical Society of America*, *110*(1), 464–478. DOI: https://doi.org/10.1121/1.1379087
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. DOI: https://doi.org/10.21105/joss.01686
Zhao, S. W. (2010). Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America*, *128*, 2009. DOI: https://doi.org/10.1121/1.3478856

How to cite this article: Onosson, S., & Stewart, J. 2021 A multi-method approach to correlate identification in acoustic data: The case of Media Lengua. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 13, pp. 1–30. DOI: https://doi.org/10.5334/labphon.291

Submitted: 12 August 2020 **Accepted:** 25 May 2021 **Published:** 16 July 2021

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS

ETHNOLINGUISTIC DIFFERENTIATION AND THE CANADIAN SHIFT

Sky Onosson[1], Nicole Rosen[1], and Lanlan Li[1]

[1] University of Manitoba
Sky@Onosson.com, Nicole.Rosen@umanitoba.ca, Lanlan.Li@umanitoba.ca

ABSTRACT

The acoustic qualities of the Canadian Shift vowels /æ, ɛ, ɪ/ were examined among two ethnically distinct populations in Winnipeg, Canada: speakers of Filipino ancestry, and non-Filipino "white" Canadians. Results indicate that ethnic Filipinos participate more strongly in the Canadian Shift, indicated by greater retraction and/or lowering of the relevant vowels, in contrast with white speakers who only just meet the criteria for Canadian Shift. Canadian-born L1 English-speaking children of Filipino immigrants exhibited the greatest degree of retraction and/or lowering, while English L2 Filipino immigrants were intermediary in this respect. These results support the hypothesis that a new ethnolect may be forming within Winnipeg's large Filipino ethnic community. The divergence of Canadian Shift patterns between ethnic groups in Winnipeg may be viewed within an "Emergent linguistic marketplace" model [7], with implications for further research on Filipino communities in other Canadian locales, and for ethnolinguistic variation more broadly.

Keywords: vowels, Canadian Shift, sociophonetics, ethnolinguistics, Canadian English

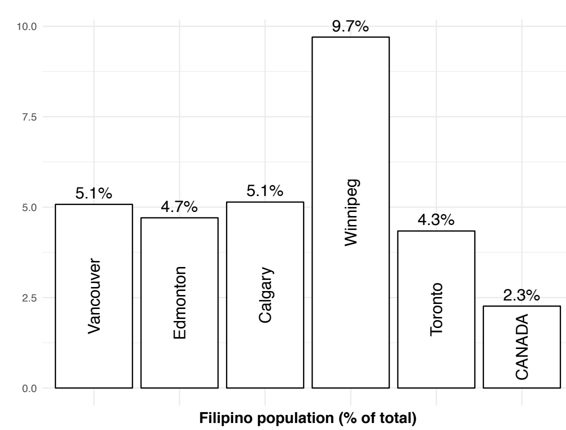
1. INTRODUCTION

Research on ethnolects in Canada is currently in its infancy. While some studies have documented distinctive features associated with particular ethnic groups in Montreal [2] and Toronto [9], in most of Canada's smaller cities the state of ethnolectal variation is largely unknown. For the present study, we have chosen to investigate one of the most important ethnic populations in Canada in recent decades, Filipino-Canadians.

Filipino migration to Canada began in earnest in the late 1960s [22] and continues to be strong today. Since that time, the largest per capita concentration of ethnic Filipinos in Canada has come to be located in the western Canadian city of Winnipeg, where ethnic Filipinos comprise nearly 10% of the city's population [21], roughly twice the proportion found in other comparable cities (Fig. 1). Furthermore, this population is most concentrated in the north-west quadrant of the city, creating distinct Filipino neighbourhoods.

Given such a substantial and concentrated population, we hypothesized that the Filipino-Winnipegger population should have a strong potential to develop a distinctive ethnolect. To investigate this hypothesis, we selected a hallmark feature of Canadian English, known as the Canadian Shift, a well-known chain shift involving the front lax vowels /æ, ɛ, ɪ/ (similar to the California Shift, although the relationship between the two is currently uncertain [6; 9]). In most varieties of Canadian English, the vowel /ɔ/ has merged with /ɑ/. The gap in the lower part of the vowel system resulting from the so-called *cot-caught* merger has been argued to be the impetus for /æ/ to retract in Canadian English [4], with /ɛ/ subsequently lowering and retracting in turn, later followed by a similar trajectory in the position of /ɪ/ in some regional dialects. The Canadian Shift is an ongoing change, with various regions of the country exhibiting subtle differences regarding the degree of retraction vs. lowering of each of the various vowels involved [5; 6; 7; 15].

Figure 1: The relative proportions of the ethnic Filipino population in Canada's four major western cities and Toronto, Canada's largest city.



The investigation of ethnic differences in Canadian Shift production has produced mixed results: in Montreal no significant production differences were found across several groups [2], while in Toronto significant differences were found which largely differentiated one ethnic group

(Chinese) from several others [9]. The present study aims to add to our knowledge of ethnolectal variation of Canadian Shift by broadening the scope of the investigation both geographically and ethnically through focus on the Filipino community in Winnipeg.

2. METHODS

2.1. Speakers

50 Winnipeg residents participated in the study. Of these, 21 were ethnically non-differentiated “white”, referred to hereafter by the moniker “Anglo”. The Anglo participants included 10 males and 11 females, all Canadian-born native speakers of English. The 29 remaining participants were ethnically Filipino, 14 male and 15 female, including 17 Canadian-born native speakers of English and 12 immigrants from the Philippines speaking English as a second language. Among the second language speakers, 6 were native speakers of Tagalog [19; 20], 4 of Kapampangan [12, 13], and 2 of Ilocano [23]; each of these languages has a similar, symmetrical 5-vowel system. All participants were recruited and recorded between 2014–2016.

2.2. Materials

A list of words was compiled for the purpose of eliciting vowel tokens comprising the full vowel system of Canadian English; the present study will only discuss those tokens relevant to Canadian Shift. Note that while /ɑ/ is not considered to be one of the vowels participating in Canadian Shift per se, as its merger with /ɔ/ has been argued to have provided the initial impetus for the shift, it is included in this study along with the canonical Canadian Shift vowels /æ, ɛ, ɪ/.

2.3. Procedure

Each participant was recorded reading aloud from the prepared word list, presented in a semi-randomized fashion (adjusted to prevent multiple successive tokens of the same vowel). Each ethnic group was recorded by a different in-group field researcher, with distinct recording technology. Filipino participants were recorded with a Sennheiser EK 100G2 wireless lavalier microphone connected to a Zoom handheld recorder. Anglo participants were recorded with an Apogee One microphone connected to a MacBook Pro computer. All recordings were made as 44.1 kHz, 16-bit uncompressed .wav files. In total, 4,452 tokens were obtained from Filipino participants, and 2,697 from Anglo participants, for a combined total of n = 7,150 tokens.

2.4. Analysis

Vowel tokens were manually segmented and transcribed in Praat [3] by two trained phoneticians (Authors 1 & 3). Waveforms and transcriptions were uploaded to FAVE [18] which measures formant values at one-third of vowel duration, and normalized under FAVE’s default Lobanov [11] normalization method. The resulting output was exported to R [14] which was used to perform statistical analysis, including mixed effects linear modelling using the lme4 package [1] for R. For this purpose, speaker profiles were coded with demographic information including speaker ethnicity, sex, and first language.

3. RESULTS

3.1. Canadian Shift vowel thresholds

The Atlas of North American English [10] offers several benchmarks for determining the occurrence of Canadian Shift, listed in Table 1 alongside the corresponding vowel means for the Anglo and Filipino groups. Note that /ɪ/ was not addressed by [10] and so does not have a well-defined benchmark.

Table 1: Canadian Shift vowel benchmarks and means (bold indicates that benchmark is met).

Benchmark	Anglo	Filipino
/ɛ/ F1 > 650 Hz	**655 Hz**	**689 Hz**
/æ/ F2 < 1825 Hz	**1804 Hz**	**1801 Hz**
/ɑ/ F2 < 1275 Hz	1307 Hz	1387 Hz

As indicated in Table 1, although /ɑ/ does not meet the indicated benchmark among either Winnipeg ethnic group, the benchmarks for the Canadian Shift vowels /ɛ, æ/ are both met within each group. We take this to indicate that both groups do participate in Canadian Shift, with Anglo speakers lagging notably behind Filipino speakers in terms of the degree of retraction or lowering.

3.2. Linear mixed effects models

Mixed effects models indicated that speaker sex was not strongly correlated with differences in formant values; mean differences between the two sexes ranged only from 1–12 Hz for F1 and 8–43 Hz for F2, depending on the particular vowel. In contrast, speaker ethnicity was often correlated with substantial differences in mean formant values. Mean F1 values differed between ethnic groups (Filipino vs. Anglo), with Filipino speakers having F1 values approximately 50 Hz higher on average (lower vowel position) for all vowels. For /ɛ/ and /ɪ/, the two most recent vowels to be affected by Canadian Shift, Filipino speakers had lower F2 values (more retracted vowel position) by > 200 Hz; /ɑ/ and /æ/ did not exhibit large F2 differences between ethnicities.

As the Filipino cohort contained both English L1 and L2 speakers, we investigated differences correlated with English language status internal to this cohort, finding that differences in mean formant values were quite variable. The largest mean differences occurred between English L1 and Ilocano L1 speakers, and the smallest between English L1 and Tagalog L1 speakers. Given the small size of the English L2 sub-groups (between 2 to 6 individuals) we did not investigate formant differences between the English L2 sub-groups any further.

3.2. Vowel token distributions

Following the mixed effects model results, we focused on exploring the differences in vowel formant output between the two ethnicities, with the Filipino cohort split into two sub-groups, English L1 and (aggregated) English L2 speakers; we refer to the latter two groups as Filipino L1 and Filipino L2, respectively. Per-vowel differences in F1xF2 token distributions between each group are illustrated in Figs. 2–5. Ellipses indicate 50% and 90% confidence intervals for mean token distribution, and shaded circles indicate the mean F1xF2 positions.

For /ɑ/ (Fig. 2), both Filipino groups exhibit distributions which are somewhat more centralized and lowered than Anglo speakers, but nonetheless there is a great degree of overlap for the central 50% of each group's tokens, indicated by the smallest, inner ellipses.

For /æ/ (Fig. 3), the degree of overlap between distributions is similar to /ɑ/. The main observable difference concerns the relatively wide spread of F2 values for Anglo speakers, extending in both directions beyond the range of variation attributed to either Filipino group.

For both /ɛ/ and /ɪ/ (Figs. 4 & 5), a similar pattern emerges: Anglo speakers have distributions which are relatively high and front, Filipino L2 speakers have somewhat lowered and retracted distributions in comparison, while Filipino L1 speakers extend this trend even further, exhibiting the most retracted (centralized) and lowered distributions of all. This difference is most striking for /ɪ/, as the central 50% of distributions for Anglo and Filipino L1 speakers barely overlap at all.

Mean F1xF2 positions of each Canadian Shift vowel per group are plotted in Fig. 6, where the Filipino vs. Anglo lowering and retracting trend for the vowels /ɛ/ and /ɪ/ can be clearly observed.

Figure 2: /ɑ/ token distributions by ethnolinguistic group.

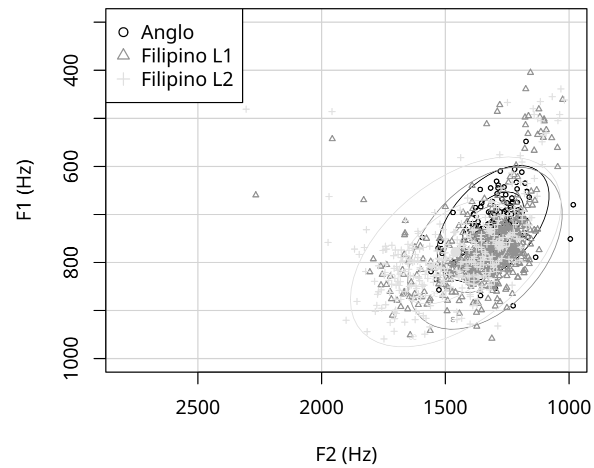


Figure 3: /æ/ token distributions by ethnolinguistic group.

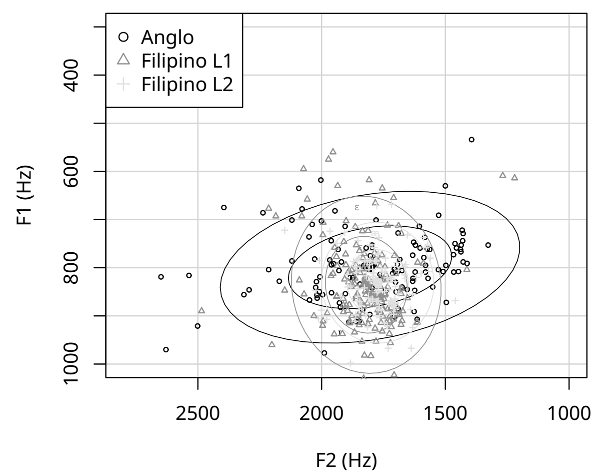


Figure 4: /ɛ/ token distributions by ethnolinguistic group.

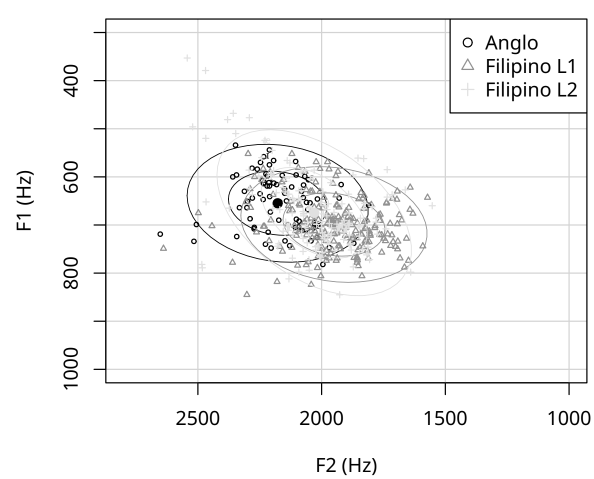


Figure 5: /ɪ/ token distributions by ethnolinguistic group.

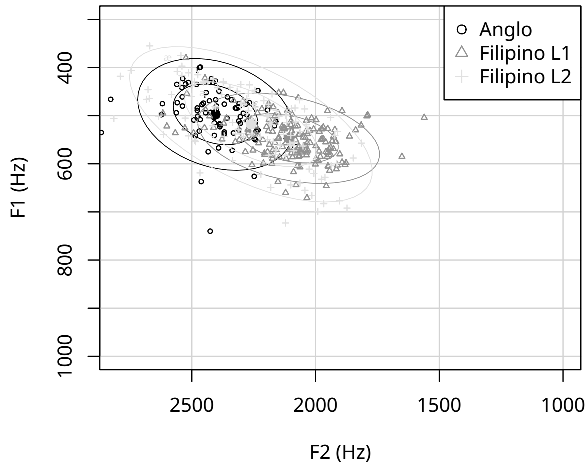
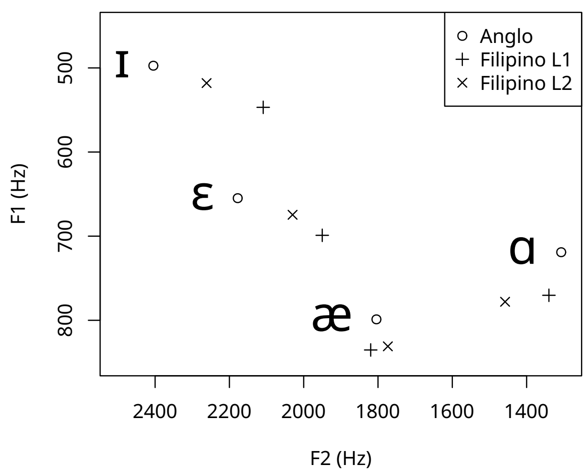


Figure 6: Canadian Shift vowels in F1xF2 space.



4. DISCUSSION

The differences in position we have observed for Canadian Shift vowels among Filipino vs. Anglo speakers in Winnipeg have two main implications, one concerning the Filipino ethnic community specifically, and the other concerning the progress of Canadian Shift itself.

For Filipino speakers, our findings indicate that a distinctive ethnolect may be developing within Winnipeg. The phonetic characteristics of this ethnolect, at least in terms of the Canadian Shift vowels, can be attributed to two sources: the local realizations of the relevant vowels among the L2 predecessors to the L1 Filipinos, and the realizations of these vowels in other parts of Canada.

While Filipino L2 group has realizations which are more advanced in terms of Canadian Shift retraction/lowering than Anglo speakers in Winnipeg, this does not necessarily indicate that the L2 realizations served as the most direct model for the subsequent generations of Filipino L1 speakers. However, this becomes more significant when we consider that for Winnipeg generally (that is, undifferentiated for ethnicity), Canadian Shift realizations have tended to lag behind the rest of the country. Previous study of Canadian Shift in Winnipeg, while limited [6], has indicated that (Anglo) Winnipeggers generally did not participate in the shift. While our more recent data shows that the situation has changed somewhat, our Anglo participants only just meet the benchmark criteria for Canadian Shift (see Table 1). Both L1 and L2 Filipinos have stronger indications of participation in the shift, and it can therefore be argued that the Filipino community in Winnipeg more closely follows the national, rather than local trends.

In this respect, Filipino-Winnipeggers may be more cognizant of and adherent to extra-local trends exemplified by communities such as e.g. Toronto where the shift has progressed further. This type of ethnically-differentiated variation may be seen as the exchange of “symbolic capital” on the “Emergent linguistic marketplace” [7]. In other words, because Winnipeg is smaller Canadian city, we hypothesize that there is little social capital to be gained by adhering to its local conservative trends versus the more innovative national pattern of larger Canadian cities with more cultural influence, such as Toronto or Vancouver. This reflects similar findings among Asian communities of San Francisco [7].

Regarding the general progress of Canadian Shift, the differences between the various vowels across ethnolinguistic lines is particularly relevant. While significant ethnolinguistic differences are largely relegated to the vowels /ɛ/ and /ɪ/, these are the most recent vowels to participate in the shift. Differences in position for /æ/ and /ɑ/ are much less substantial across ethnicities, indicating that these vowels are more stable. This may be taken as an indication that the shift is no longer in progress for /æ/ in Winnipeg generally (/ɑ/ is not argued to actually participate in the shift, being its origin). If retraction and lowering of /æ/ was in fact the initial phase of the shift, then it is expected that it would be the first vowel to have reached a stable position, while the other lax front vowels remain somewhat in flux. This places Winnipeg generally further behind in the progress of Canadian Shift in comparison to more eastern Canadian communities such as Toronto [16], Thunder Bay [15], or Montreal [2], but ahead of e.g. Victoria [17] (in western Canada), where /æ/ retraction is still active.

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

[2] Boberg, C. 2005. The Canadian Shift in Montreal. *Language Variation and Change* 17, 133–154.

[3] Boersma, P., Weenink, D. 2015. *Praat: doing phonetics by computer*. (Version 5.4.19). http://www.praat.org/

[4] Clarke, S., Ford E., Youssef, A. 1995. The third dialect of English: Some Canadian evidence. *Language Variation and Change* 7, 209–228.

[5] Esling, J., Warkentyne, H. 1993. Retracting of /æ/ in Vancouver English. In: Clarke, S. (ed.), *Focus on Canada*. Amsterdam: John Benjamins, 229–246.

[6] Hagiwara, R. 2006. Vowel production in Winnipeg. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 51(2-3), 127–141.

[7] Hall-Lew, L. (2009). *Ethnicity and Phonetic Variation in a San Francisco Neighborhood*. Ph.D. dissertation. Stanford University.

[8] Hoffman, M. F., Walker, J. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, 22(1), 37–67.

[9] Kennedy, R., Grama, J. 2012. Chain Shifting and Centralization in California Vowels: An Acoustic Analysis. *American Speech*, 87(1), 39–56.

[10] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonology, Phonetics, and Sound Change. A Multimedia Reference Tool.* Berlin: Mouton de Gruyter.

[11] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different listeners. *J. Acoust. Soc. Am*., 49, 606–08.

[12] Natividad, P. E. 1967. *A taxonomic phonological analysis of Tagalog and Pampango*. Master's thesis. University of British Columbia.

[13] Pangilinan, M. R. M. 2009. Kapampangan Lexical Borrowing from Tagalog: Endangerment rather than Enrichment. *11th International Conference on Austronesian Linguistics.*

[14] R Core Team. 2015. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

[15] Roeder, R. 2012. The Canadian Shift in two Ontario cities. *World Englishes*, 31(4), 478–492.

[16] Roeder, R., Jarmasz, L.-G. 2010. The Canadian Shift in Toronto. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, 55(3), 387-404.

[17] Roeder, R., Onosson, S., D'Arcy, A. 2018. Joining the Western Region: Sociophonetic Shift in Victoria. *Journal of English Linguistics*, 46(2), 87–112.

[18] Rosenfelder, I., Fruehwald, J., Evanini, K., Yuan J. 2011. *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. http://fave.ling.upenn.edu.

[19] Schachter, P., Otanes, F. T. 1983. *Tagalog Reference Grammar*. University of California Press.

[20] Schachter, P., Reid, L. A. 2009. Tagalog. In: Comrie, B. (ed.), *The World's Major Languages.* London and Sydney: Croom Helm, 833–855.

[21] Statistics Canada. 2016. *Focus on Geography Series, 2016 Census*. Statistics Canada Catalogue no. 98-404-X2016001. Ottawa, Ontario. Data products, 2016 Census. https://www12.statcan.gc.ca/

[22] Vachon, M., Toews, W. 2008. A geography of the Filipino migration to Winnipeg. *Canadian Journal of Urban Research*, 17(1), 107–129.

[23] Yamamoto, K. 2017. A phonological sketch of Ilocano. *Kyoto University Linguistic Research*, 36, 21–49.