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

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

Abstract

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 a high degrees of overlap between the front vs. 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 & 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 though 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

1

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 though out 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 Andeans, 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 & 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/ vs. /iː/; see Muehlbauer, 2012); tonal languages require pitch measurement (e.g., Vietnamese /a↓/ vs. /a↘/; see Nguyễn, 1997); languages with oral vs. nasal vowels (e.g., Guaraní /a/ vs. /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 vs. 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 (sometimes two) primary correlate 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 to 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.

	Mas	**bueno**my	**trilla**ngapa	**caballo**huan.
1.	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.'

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

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.

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 (2014)'s 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

A Media Lengua Vowel Space - 50% Concentrations
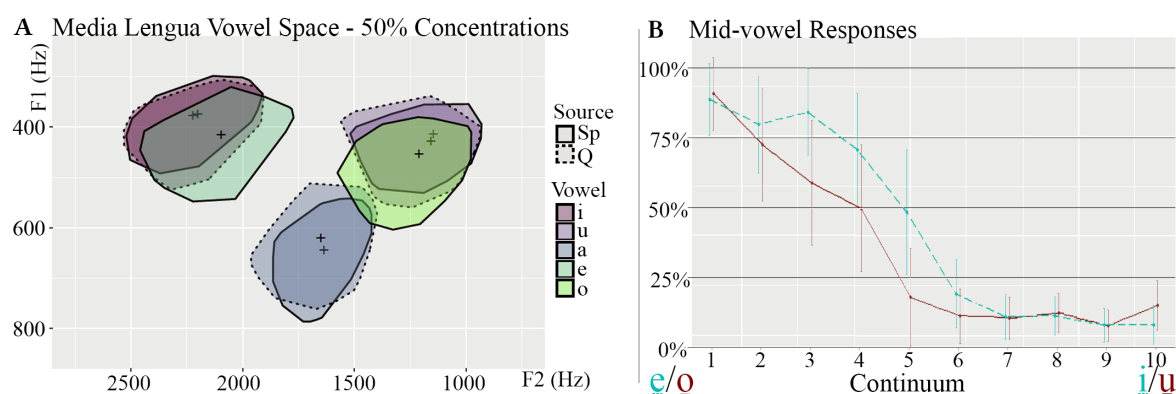
B Mid-vowel Responses

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)

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 (2001)'s 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/ & /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 8 monophthong vowels and 19 vowel-sequences which are arranged in an acoustic space that reflects the original Quichua 3-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 vs. high vowel phonological functional load.

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 (2018b)'s perception experiment incorporated multiple correlates into the stimuli (F1, F2, F3, pitch, duration, inten-

sity), 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, they 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, & 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 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.

Table 1. Speech corpora and token counts.

Vowel	Corpus 1 Wordlist/sentences	Corpus 2 Natural speech	Total
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

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

26 trilingual speakers (L1 Media Lengua, L1 Quichua & L2 Spanish) participated in this study. This group consisted of fifteen women and eleven 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 6–7 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 along side 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 use 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* & *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] vs. *closed* [+CODA]) as vowel production can be affected based on the presence or absence of a coda, especially duration (see e.g., I. 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 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 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.
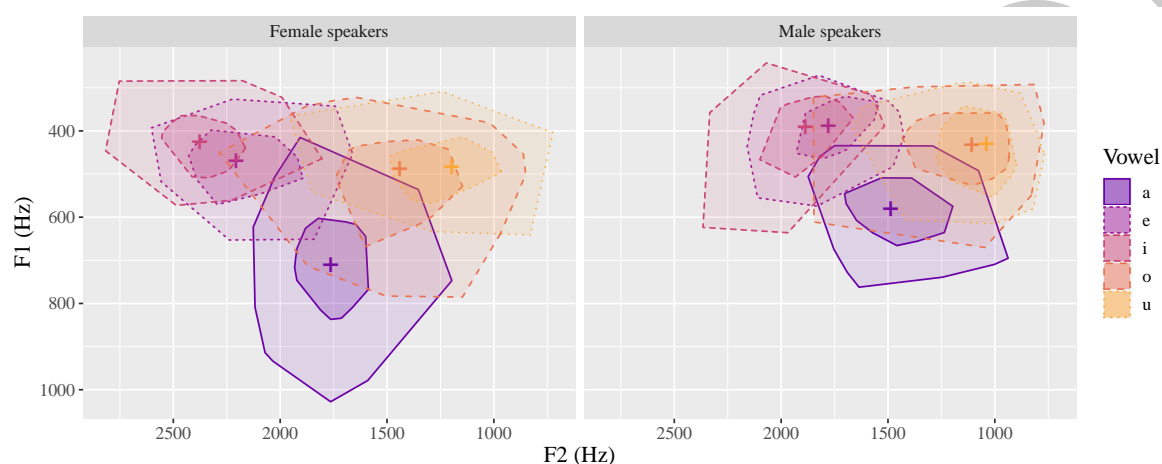

Figure 3. Bag-and-loop plots of vowel tokens according to speaker sex in F1xF2 space

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,* & *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, & 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 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. Figures 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 & intensity) with contributions greater than the expected level for that dimension (which serves as a threshold when considering the relative influence of variables)
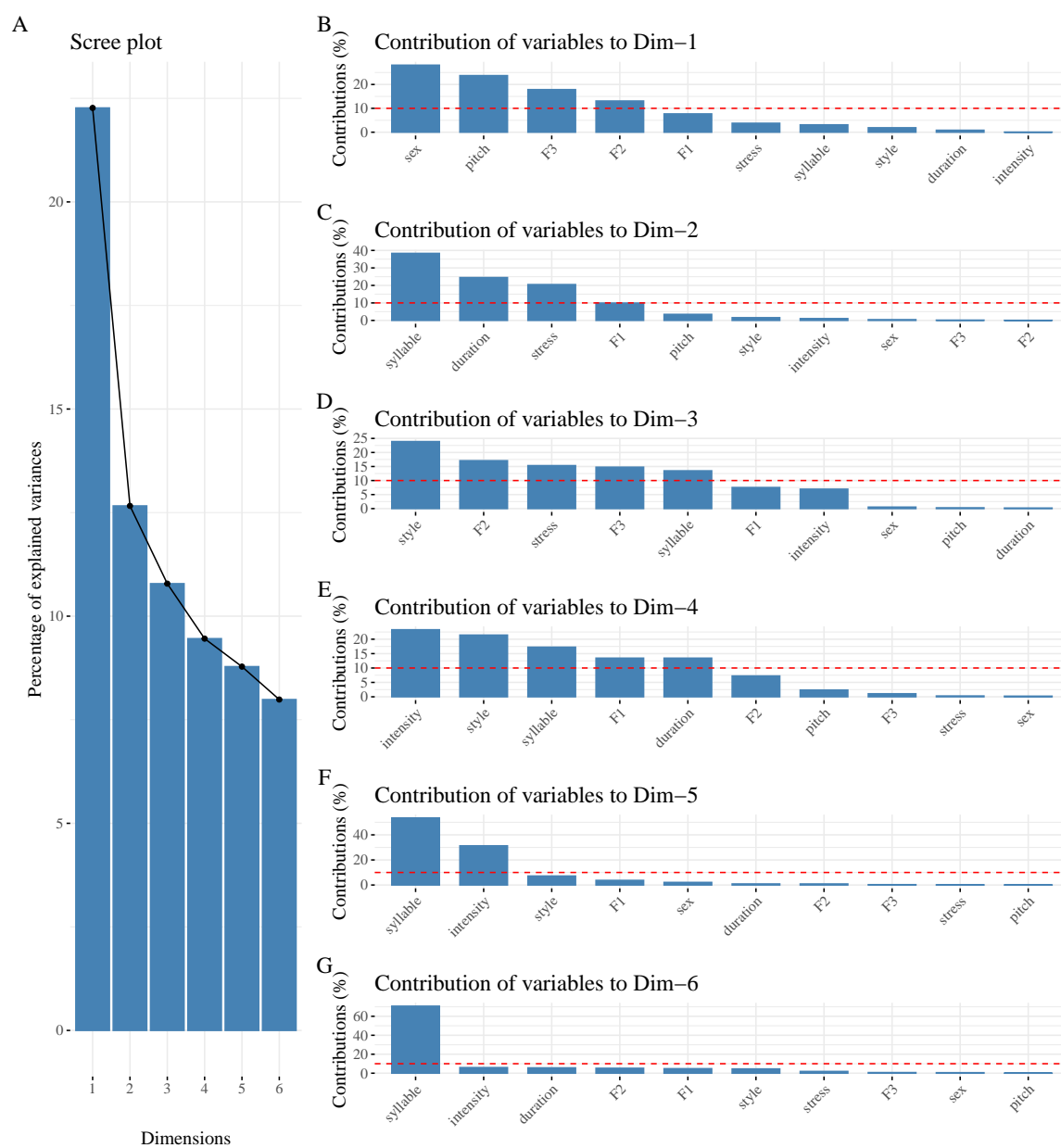
Figure 4. **A:** Scree plot of FAMD dimensions; **B-G:** Contributions of variables to Dimensions 1-6, red dashed lines indicate average expected contribution level

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 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 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,

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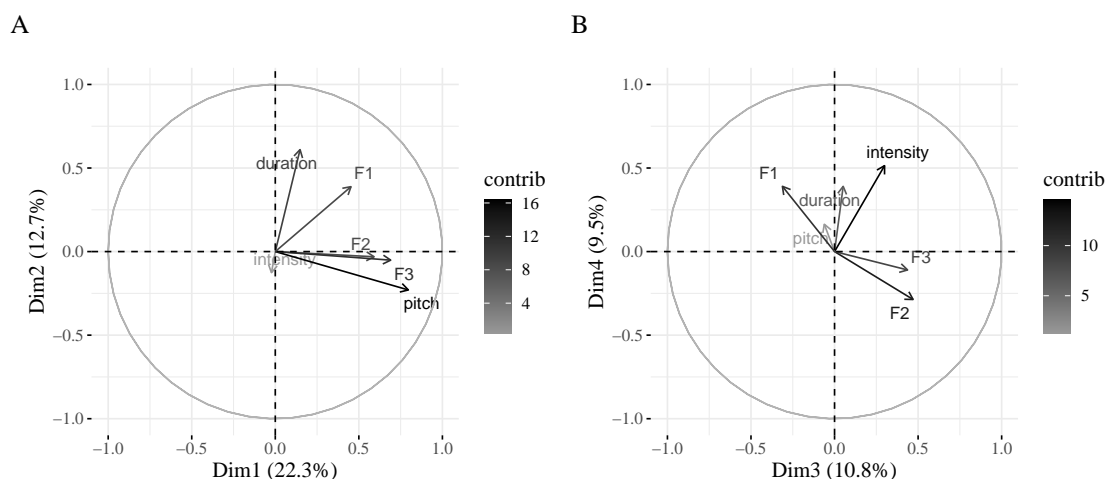
Figure 5. PCA of quantitative variables in **(A)** Dimensions 1-2 and **(B)** Dimensions 3-4

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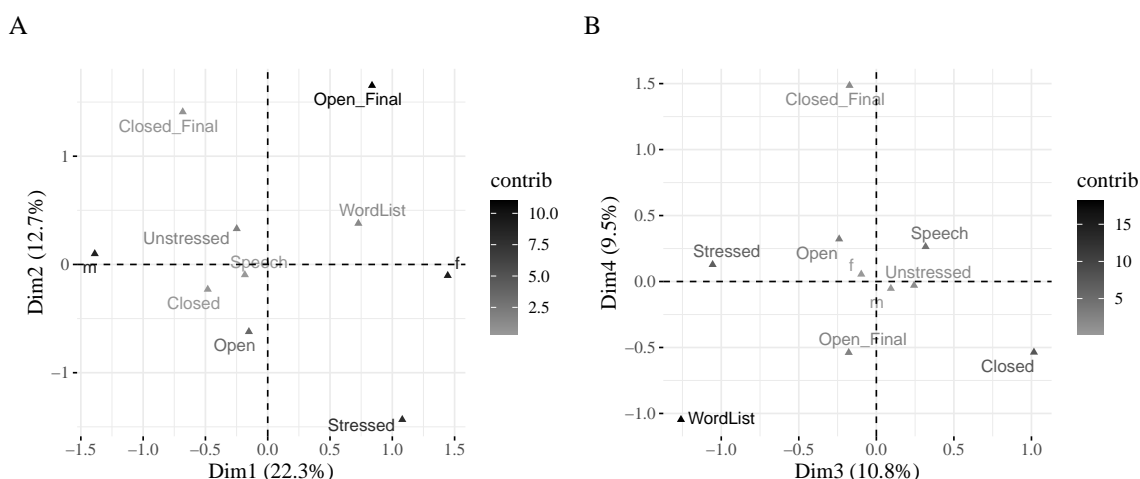
Figure 6. MCA of qualitative variables in **(A)** Dimensions 1-2 and **(B)** Dimensions 3-4

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 & male), syllable type (4 levels: open, closed, open-final, & closed-final), syllable stress (2 levels: stressed & unstressed), and elicitation style (2 levels: wordlist & 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 contributors 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 vs. 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 vs. 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 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 vs. 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 Figures 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 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
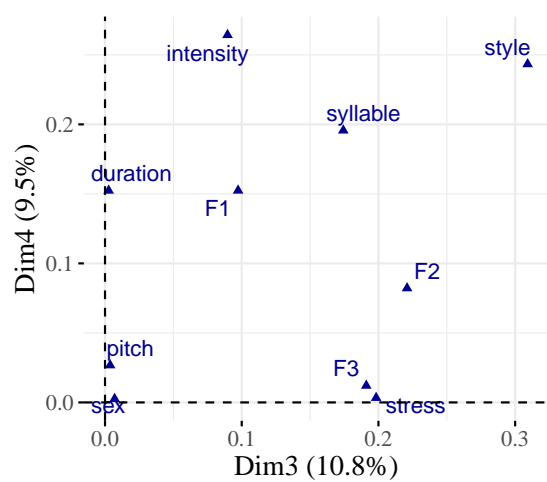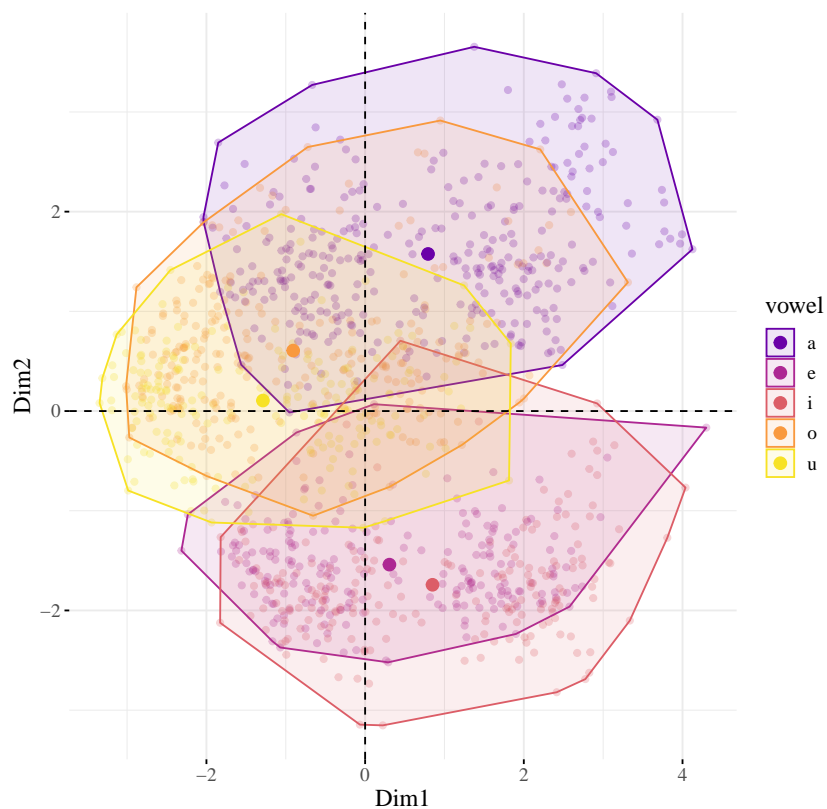
A
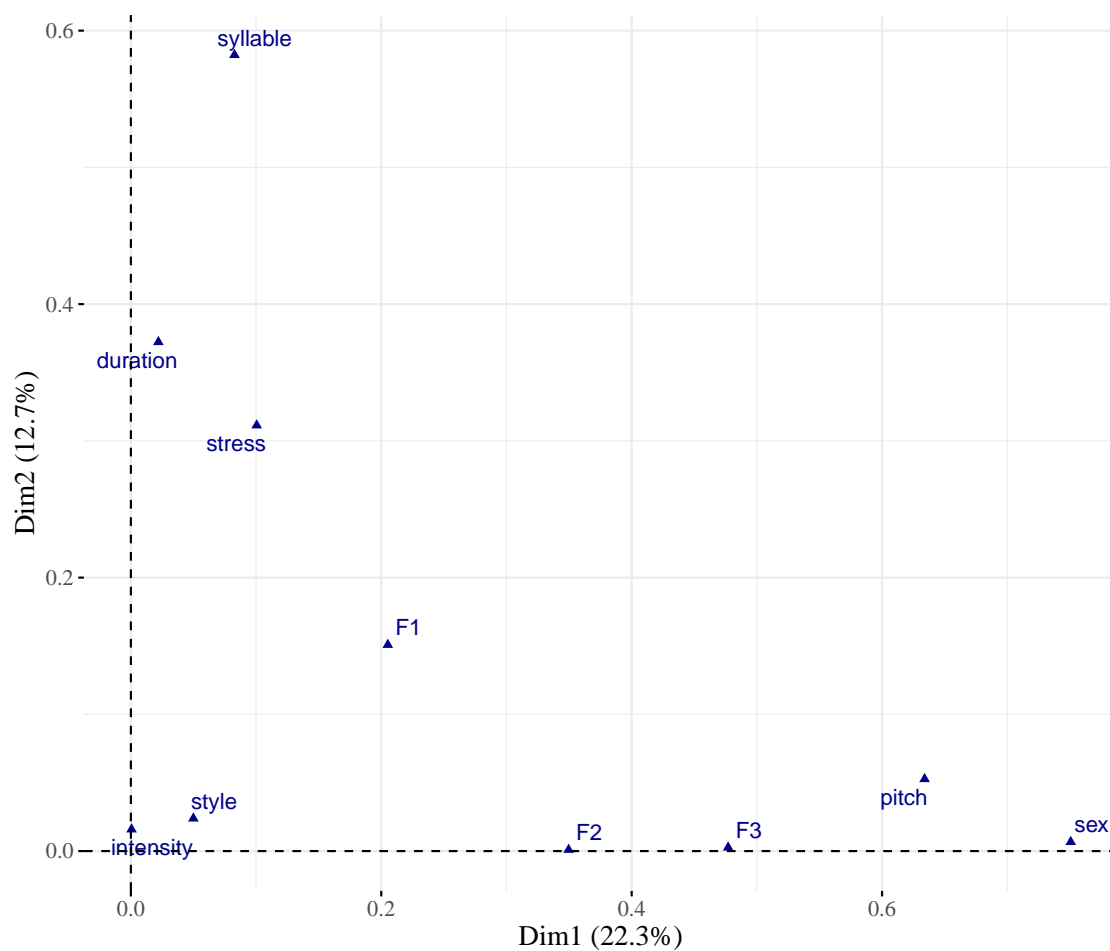
B

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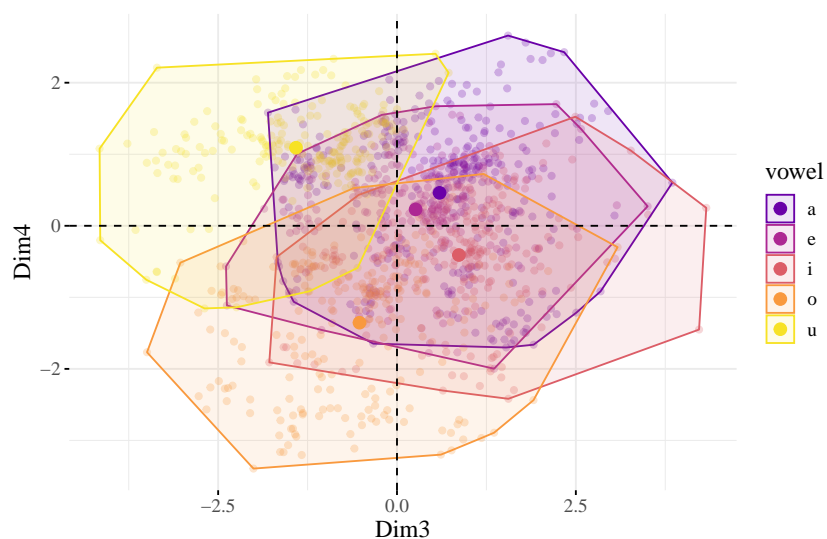
A

B

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

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/ shows a fair degree of overlap with the back vowels, but virtually none with the front vowels. /a/ 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/ vs. the others; in terms of Dimension 2 the clearest distinction is between the front vowels /e, i/ vs. the others, and secondarily between /a/ vs. 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 vs. 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,* & *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/, & /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 to women in comparison with /i/ (-38 Hz); this is likely a

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

Predictors	Model 1 F1 (Hz) i	u	Model 2 F2 (Hz) i	u	Model 3 F3 (Hz) i	u	Model 4 Pitch (Hz) i	u	Model 5 Duration (ms) i	u	Model 6 Intensity (dB) i	u
Intercept 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.*

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 (2014)'s 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 deaccen-

tuation. 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 & 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/ vs. /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 & 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/ vs. /i/ and /o/ vs. /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 (2014)'s measurements (compare 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/ vs. /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 vs. 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 vs. 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/ vs. /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 vs. 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.

Acknowledgements

Funding Information

Competing Interests

The authors have no competing interests to declare.

Author's 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.

Notes

[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.

[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.

[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.

[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).

[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.

[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.

References

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

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 417–520. doi: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: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: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: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:10.1044/jshr.1901.168

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

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. doi: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: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: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: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: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: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.

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: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: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: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: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: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: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:10.1080/07268602.2011.598629

Jones, C., Meakins, F., & Mauwiyath, S. (2012). Learning vowel categories from maternal speech in Gurindji Kriol. *Language Learning2*, *62*(4), 1052–1078. doi: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: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

Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics, 70*, 70–85. doi: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: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: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: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: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: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:10.3366/e175385480900038x

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839–862. doi: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: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:10.1017/S1366728916000468P

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: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.

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: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:10.1121/1.3277160

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

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.

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: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: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: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: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: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: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: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.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Linguistics and Language Compass, 3*(2), 621–640. doi: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: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:10.1159/000369629

Stewart, J. (2015a). Intonation patterns in Pijal Media Lengua. *Journal of Language Contact, 8*, 223–262. doi: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:10.1159/000369629

Stewart, J. (2018a). Voice onset time production in Spanish, Quichua, and Media Lengua. *Journal of the International Phonetic Association2, 48*(2), 173–197. doi:10.1017/S002510031700024X

Stewart, J. (2018b). Vowel perception by native Media Lengua, Quichua, and Spanish speakers. *Journal of Phonetics, 71*, 177–193. doi: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: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.

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: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:10.1163/19552629-01101003

Stewart, J., Prado Ayala, G., & Gonza Inlago, L. (2020). Media Lengua Dictionary. *Dictionaria, 13*, 1–3179. doi: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: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:10.1017/S0025100300006423

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America, 110*(1), 464–478. doi: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: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:10.1121/1.3478856

Appendix A   Descriptive Statistics

Table 3. Vowel means

	Vowel	F1 (Hz)	s.d.	F2 (Hz)	s.d.
All speakers					
	a	657.4	115.6	1627.9	230.7
	e	433.5	78.3	1973.3	276.4
	i	417.1	77.5	2131.7	291.6
	o	470.1	102.1	1276.1	286.8
	u	463.3	79.2	1158.5	249.4
Female speakers only					
	a	716.8	110.2	1749.0	172.4
	e	474.7	70.3	2192.8	203.6
	i	429.6	62.3	2367.6	172.6
	o	515.5	114.4	1443.0	267.5
	u	484.0	71.7	1232.9	276.4
Male speakers only					
	a	582.2	69.2	1474.7	202.5
	e	397.8	66.6	1782.7	168.8
	i	405.2	88.3	1907.1	184.3
	o	439.8	80.1	1164.5	242.2
	u	440.8	81.3	1078.1	187.0

Appendix B   Factor Analysis for Mixed Data

Table 4. Factor analysis for mixed data: dimensions

Dimensions	Eigenvalue	Percentage of variance	Cumulative variance
Dimension 1	2.6722	22.2687	22.2687
Dimension 2	1.5194	12.6619	34.9306
Dimension 3	1.2942	10.7847	45.7153
Dimension 4	1.1349	9.4577	55.173
Dimension 5	1.0539	8.7821	63.9551
Dimension 6	0.9585	7.9874	71.9425

Table 5. Factor analysis for mixed data: factors

	Coordinates					
Factors	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
duration	0.022	0.3723	0.0027	0.1524	0.0091	0.0541
intensity	$5.0e{-}04$	0.0159	0.0896	0.2643	0.3295	0.0569
pitch	0.634	0.0527	0.0038	0.0267	0.0011	$7.0e{-}04$
F1	0.2052	0.1508	0.0973	0.1524	0.0391	0.0457
F2	0.3496	0.001	0.2209	0.0822	0.0091	0.05
F3	0.477	0.0026	0.1911	0.0121	0.0031	0.0067
sex	0.7507	0.0067	0.007	0.0026	0.022	0.0055
stress	0.1005	0.3114	0.1983	0.0033	0.0016	0.018
syllable	0.0827	0.5822	0.1743	0.1956	0.5636	0.6776
style	0.0499	0.0239	0.3091	0.2433	0.0756	0.0433
	Contributions					
duration	0.8221	24.5012	0.2064	13.4281	0.8653	5.644
intensity	0.0197	1.0454	6.9252	23.2887	31.2654	5.9345
pitch	23.7242	3.4675	0.2904	2.3565	0.108	0.0729
F1	7.6795	9.9277	7.5216	13.4307	3.7119	4.7688
F2	13.084	0.0635	17.0712	7.2386	0.8635	5.2204
F3	17.8518	0.1728	14.7661	1.0634	0.2957	0.6949
sex	28.0923	0.4396	0.5433	0.2254	2.0917	0.5686
stress	3.7612	20.4917	15.3238	0.2923	0.153	1.8756
syllable	3.0964	38.32	13.4673	17.2374	53.476	70.6982
style	1.8689	1.5706	23.8847	21.4391	7.1697	4.522
	Cos^2					
duration	$5.0e{-}04$	0.1386	0	0.0232	$1.0e{-}04$	0.0029
intensity	0	$3.0e{-}04$	0.008	0.0699	0.1086	0.0032
pitch	0.4019	0.0028	0	$7.0e{-}04$	0	0
F1	0.0421	0.0228	0.0095	0.0232	0.0015	0.0021
F2	0.1222	0	0.0488	0.0067	$1.0e{-}04$	0.0025
F3	0.2276	0	0.0365	$1.0e{-}04$	0	0
sex	0.5635	0	0	0	$5.0e{-}04$	0
stress	0.0101	0.0969	0.0393	0	0	$3.0e{-}04$
syllable	0.0023	0.113	0.0101	0.0128	0.1059	0.1531
style	0.0025	$6.0e{-}04$	0.0955	0.0592	0.0057	0.0019

Table 6. Principal component analysis

	Coordinates					
Factors	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
duration	0.1482	0.6101	0.0517	0.3904	0.0955	−0.2326
intensity	−0.0229	−0.126	0.2994	0.5141	0.574	0.2385
pitch	0.7962	−0.2295	−0.0613	0.1635	0.0337	−0.0264
F1	0.453	0.3884	−0.312	0.3904	−0.1978	−0.2138
F2	0.5913	−0.0311	0.47	−0.2866	0.0954	0.2237
F3	0.6907	−0.0512	0.4371	−0.1099	−0.0558	0.0816
	Contributions					
duration	0.8221	24.5012	0.2064	13.4281	0.8653	5.644
intensity	0.0197	1.0454	6.9252	23.2887	31.2654	5.9345
pitch	23.7242	3.4675	0.2904	2.3565	0.108	0.0729
F1	7.6795	9.9277	7.5216	13.4307	3.7119	4.7688
F2	13.084	0.0635	17.0712	7.2386	0.8635	5.2204
F3	17.8518	0.1728	14.7661	1.0634	0.2957	0.6949
	Cos^2					
duration	0.022	0.3723	0.0027	0.1524	0.0091	0.0541
intensity	$5.0e{-}04$	0.0159	0.0896	0.2643	0.3295	0.0569
pitch	0.634	0.0527	0.0038	0.0267	0.0011	$7.0e{-}04$
F1	0.2052	0.1508	0.0973	0.1524	0.0391	0.0457
F2	0.3496	0.001	0.2209	0.0822	0.0091	0.05
F3	0.477	0.0026	0.1911	0.0121	0.0031	0.0067

Table 7. Multiple correspondence analysis

			Coordinates			
Factor levels	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
f	1.4449	−0.1028	−0.0973	0.055	−0.1555	−0.0737
m	−1.3883	0.0988	0.0935	−0.0528	0.1494	0.0708
Stressed	1.0799	−1.4333	−1.0557	0.1279	−0.0859	−0.2735
Unstressed	−0.2487	0.3301	0.2431	−0.0294	0.0198	0.063
Closed	−0.4813	−0.2301	1.0161	−0.5364	−0.7029	−1.1198
Closed_Final	−0.6821	1.4095	−0.173	1.4854	−4.5233	4.4434
Open	−0.151	−0.6211	−0.2399	0.3219	0.1771	0.1352
Open_Final	0.8372	1.653	−0.1781	−0.5387	0.526	0.1339
Speech	−0.1834	−0.0956	0.3176	0.2638	−0.1417	−0.1023
WordList	0.7276	0.3793	−1.2597	−1.0466	0.562	0.406
			Contributions			
f	14.3266	0.2242	0.2771	0.115	1.0667	0.29
m	13.7657	0.2154	0.2662	0.1105	1.025	0.2786
Stressed	3.0571	16.6559	12.4554	0.2376	0.1244	1.5245
Unstressed	0.704	3.8358	2.8684	0.0547	0.0286	0.3511
Closed	0.5802	0.41	11.0258	3.9949	7.9568	24.416
Closed_Final	0.1355	1.7899	0.0372	3.563	38.316	44.6991
Open	0.1842	9.6339	1.9806	4.6382	1.6274	1.1465
Open_Final	2.1965	26.4862	0.4237	5.0413	5.5757	0.4366
Speech	0.3763	0.3162	4.8087	4.3164	1.4435	0.9104
WordList	1.4926	1.2544	19.076	17.1227	5.7262	3.6116
			Cos^2			
f	0.9337	0.0047	0.0042	0.0014	0.0108	0.0024
m	0.9337	0.0047	0.0042	0.0014	0.0108	0.0024
Stressed	0.2113	0.3722	0.2019	0.003	0.0013	0.0136
Unstressed	0.2113	0.3722	0.2019	0.003	0.0013	0.0136
Closed	0.0483	0.011	0.2153	0.06	0.103	0.2615
Closed_Final	0.0097	0.0415	6.0e−04	0.0461	0.4274	0.4125
Open	0.027	0.4566	0.0681	0.1226	0.0371	0.0216
Open_Final	0.1557	0.6068	0.007	0.0644	0.0615	0.004
Speech	0.1165	0.0317	0.3492	0.2411	0.0695	0.0363
WordList	0.1165	0.0317	0.3492	0.2411	0.0695	0.0363
			v.test			
f	30.0264	−2.8324	−2.9061	1.7528	−5.1453	−2.5585
m	−30.0264	2.8324	2.9061	−1.7528	5.1453	2.5585
Stressed	10.9868	−19.3375	−15.433	1.996	−1.3917	−4.6466
Unstressed	−10.9868	19.3375	15.433	−1.996	1.3917	4.6466
Closed	−4.7619	−3.0187	14.4466	−8.1433	−11.0745	−18.501
Closed_Final	−2.1076	5.7755	−0.7683	7.0425	−22.2544	22.9234
Open	−3.7364	−20.3754	−8.5262	12.2186	6.9743	5.5828
Open_Final	9.5299	24.9536	−2.9129	−9.4089	9.5351	2.5445
Speech	−7.7446	−5.3535	19.2676	17.0945	−9.526	−7.2149
WordList	7.7446	5.3535	−19.2676	−17.0945	9.526	7.2149

Appendix C   Mixed Effects Models

Table 8. F1 Freq. linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	CI: lower	CI: upper	p
			F1		
(Intercept)	438.00	16.07	406.50	469.50	< 0.001*
vowel [a]	280.51	8.75	263.35	297.67	< 0.001*
vowel [e]	29.18	10.62	8.38	49.99	0.006*
vowel [o]	78.76	10.08	58.99	98.52	< 0.001*
vowel [u]	43.01	12.06	19.37	66.64	0.002*
sex [m]	−62.80	19.92	−101.85	−23.75	0.002*
syllable [Closed_Final]	−10.21	15.31	−40.21	19.79	0.505
syllable [Open]	−5.57	5.76	−16.85	5.72	0.334
syllable [Open_Final]	−5.03	7.55	−19.82	9.77	0.505
stress [Unstressed]	−1.32	6.14	−13.35	10.70	0.829
style [WordList]	10.46	7.57	−4.39	25.30	0.167
vowel [a] * sex [m]	−90.54	13.03	−116.07	−65.00	< 0.001*
vowel [e] * sex [m]	−22.92	14.49	−51.33	5.49	0.114
vowel [o] * sex [m]	−38.12	13.69	−64.94	−11.29	0.005*
vowel [u] * sex [m]	−8.50	15.88	−39.63	22.63	0.593

Random Effects

σ^2	4161.30
$\tau_{00\ word}$	796.26
$\tau_{00\ speaker}$	2377.81
ICC	0.43
$N_{speaker}$	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.630 / 0.790

Table 9. F1 Freq. linear mixed effects regression - Intercept /u/

	F1				
Predictors	Estimates	std. Error	CI: lower	CI: upper	p
(Intercept)	481.01	16.28	449.10	512.92	$< 0.001^*$
vowel [a]	237.51	11.04	215.87	259.14	$< 0.001^*$
vowel [e]	−13.83	12.04	−37.43	9.78	0.251
vowel [i]	−43.01	12.06	−66.64	−19.37	$< 0.001^*$
vowel [o]	35.75	11.96	12.31	59.19	0.003*
sex [m]	−71.30	20.41	−111.31	−31.29	$< 0.001^*$
syllable [Closed_Final]	−10.21	15.31	−40.21	19.79	0.505
syllable [Open]	−5.57	5.76	−16.85	5.72	0.334
syllable [Open_Final]	−5.03	7.55	−19.82	9.77	0.505
stress [Unstressed]	−1.32	6.14	−13.35	10.70	0.829
style [WordList]	10.46	7.57	−4.39	25.30	0.167
vowel [a] * sex [m]	−82.04	14.78	−111.01	−53.07	$< 0.001^*$
vowel [e] * sex [m]	−14.42	15.96	−45.70	16.85	0.366
vowel [i] * sex [m]	8.50	15.88	−22.63	39.63	0.593
vowel [o] * sex [m]	−29.62	15.41	−59.82	0.57	0.055

Random Effects

σ^2	4161.30
$\tau_{00 \text{ word}}$	796.26
$\tau_{00 \text{ speaker}}$	2377.80
ICC	0.43
N_{speaker}	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.630 / 0.790

Table 10. F2 Freq. linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	F2 CI: lower	CI: upper	p
(Intercept)	2360.07	34.74	2291.99	2428.15	$< 0.001^*$
vowel [a]	−596.35	24.65	−644.66	−548.04	$< 0.001^*$
vowel [e]	−150.70	29.80	−209.12	−92.28	$< 0.001^*$
vowel [o]	−912.19	28.43	−967.91	−856.47	$< 0.001^*$
vowel [u]	−1066.07	33.02	−1130.78	−1001.35	$< 0.001^*$
sex [m]	−368.60	42.59	−452.07	−285.13	$< 0.001^*$
syllable [Closed_Final]	−119.21	43.31	−204.11	−34.32	0.006^*
syllable [Open]	−1.39	16.37	−33.47	30.68	0.932
syllable [Open_Final]	17.81	21.32	−23.97	59.59	0.403
stress [Unstressed]	−41.00	17.18	−74.66	−7.33	0.017
style [WordList]	35.35	20.89	−5.59	76.29	0.091
vowel [a] * sex [m]	123.15	36.74	51.14	195.15	0.001^*
vowel [e] * sex [m]	40.90	40.95	−39.36	121.16	0.318
vowel [o] * sex [m]	150.10	38.64	74.37	225.82	$< 0.001^*$
vowel [u] * sex [m]	228.22	44.15	141.68	314.76	$< 0.001^*$

Random Effects

σ^2	32038.01
$\tau_{00 \text{ word}}$	8261.45
$\tau_{00 \text{ speaker}}$	4637.05
ICC	0.29
N_{speaker}	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.765 / 0.832

Table 11. F2 Freq. linear mixed effects regression - Intercept /u/

Predictors	Estimates	std. Error	CI: lower	CI: upper	p
			F2		
(Intercept)	1294.00	34.47	1226.44	1361.57	$< 0.001^*$
vowel [a]	469.71	30.04	410.84	528.59	$< 0.001^*$
vowel [e]	915.37	32.91	850.85	979.88	$< 0.001^*$
vowel [i]	1066.07	33.02	1001.35	1130.78	$< 0.001^*$
vowel [o]	153.88	32.81	89.57	218.18	$< 0.001^*$
sex [m]	−140.38	43.26	−225.17	−55.60	0.001^*
syllable [Closed_Final]	−119.21	43.31	−204.11	−34.32	0.006^*
syllable [Open]	−1.39	16.37	−33.47	30.68	0.932
syllable [Open_Final]	17.81	21.32	−23.97	59.59	0.403
stress [Unstressed]	−41.00	17.18	−74.66	−7.33	0.017^*
style [WordList]	35.35	20.89	−5.59	76.29	0.091
vowel [a] * sex [m]	−105.07	40.73	−184.90	−25.24	0.010^*
vowel [e] * sex [m]	−187.32	44.31	−274.16	−100.48	$< 0.001^*$
vowel [i] * sex [m]	−228.22	44.15	−314.76	−141.68	$< 0.001^*$
vowel [o] * sex [m]	−78.12	42.66	−161.74	5.50	0.067

Random Effects

σ^2	32038.01
$\tau_{00 \ word}$	8261.45
$\tau_{00 \ speaker}$	4637.05
ICC	0.29
$N_{speaker}$	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.765 / 0.832

Table 12. F3 Freq. linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	F3 CI: lower	CI: upper	p
(Intercept)	3072.74	52.86	2969.14	3176.35	< 0.001*
vowel [a]	−242.40	37.90	−316.67	−168.12	< 0.001*
vowel [e]	−120.36	45.73	−209.99	−30.72	0.008*
vowel [o]	−335.84	43.60	−421.28	−250.39	< 0.001*
vowel [u]	−366.71	50.51	−465.72	−267.71	< 0.001*
sex [m]	−404.09	64.51	−530.53	−277.64	< 0.001*
syllable [Closed_Final]	−88.03	66.05	−217.48	41.43	0.183
syllable [Open]	−22.08	24.82	−70.74	26.57	0.374
syllable [Open_Final]	−30.75	32.51	−94.47	32.96	0.344
stress [Unstressed]	16.36	26.41	−35.40	68.12	0.536
style [WordList]	−0.96	30.75	−61.22	59.31	0.975
vowel [a] * sex [m]	0.04	56.15	−110.01	110.08	0.999
vowel [e] * sex [m]	53.53	62.48	−68.92	175.99	0.392
vowel [o] * sex [m]	32.81	59.06	−82.93	148.56	0.578
vowel [u] * sex [m]	6.13	67.19	−125.56	137.83	0.927

Random Effects

σ^2	80133.05
$\tau_{00\ word}$	12056.78
$\tau_{00\ speaker}$	9476.69
ICC	0.20
$N_{speaker}$	27
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.348 / 0.481

Table 13. F3 Freq. linear mixed effects regression - Intercept /u/

Predictors	Estimates	std. Error	F3 CI: lower	CI: upper	p
(Intercept)	2706.03	52.58	2602.97	2809.09	< 0.001*
vowel [a]	124.32	45.98	34.21	214.43	0.007*
vowel [e]	246.36	50.24	147.89	344.82	< 0.001*
vowel [i]	366.71	50.51	267.71	465.72	< 0.001*
vowel [o]	30.88	50.18	−67.47	129.22	0.538
sex [m]	−397.95	65.64	−526.60	−269.30	< 0.001*
syllable [Closed_Final]	−88.03	66.05	−217.48	41.43	0.183
syllable [Open]	−22.08	24.82	−70.74	26.57	0.374
syllable [Open_Final]	−30.75	32.51	−94.47	32.96	0.344
stress [Unstressed]	16.36	26.41	−35.40	68.12	0.536
style [WordList]	−0.96	30.75	−61.22	59.31	0.975
vowel [a] * sex [m]	−6.10	62.19	−127.98	115.78	0.922
vowel [e] * sex [m]	47.40	67.33	−84.56	179.36	0.481
vowel [i] * sex [m]	−6.13	67.19	−137.83	125.56	0.927
vowel [o] * sex [m]	26.68	65.05	−100.81	154.17	0.682

Random Effects

σ^2	80133.05
$\tau_{00\ word}$	12056.78
$\tau_{00\ speaker}$	11021.20
ICC	0.22
$N_{speaker}$	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.342 / 0.489

Table 14. Vowel pitch linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	pitch CI: lower	CI: upper	p
(Intercept)	248.96	9.75	229.86	268.06	$< 0.001^*$
vowel [a]	−3.86	3.50	−10.72	3.01	0.271
vowel [e]	1.82	4.29	−6.59	10.24	0.671
vowel [o]	3.05	4.05	−4.89	10.99	0.452
vowel [u]	7.22	4.95	−2.47	16.92	0.144
sex [m]	−127.91	11.52	−150.50	−105.33	$< 0.001^*$
syllable [Closed_Final]	−16.71	6.22	−28.91	−4.51	0.007^*
syllable [Open]	1.70	2.40	−3.00	6.40	0.478
syllable [Open_Final]	−6.77	3.11	−12.85	−0.68	0.029^*
stress [Unstressed]	−9.10	2.48	−13.97	−4.24	$< 0.001^*$
style [WordList]	19.19	3.40	12.53	25.86	$< 0.001^*$
vowel [a] * sex [m]	−0.78	5.33	−11.23	9.68	0.884
vowel [e] * sex [m]	−2.31	5.95	−13.96	9.35	0.698
vowel [o] * sex [m]	−5.36	5.61	−16.35	5.62	0.338
vowel [u] * sex [m]	−8.14	6.66	−21.19	4.91	0.222

Random Effects

σ^2	598.66
$\tau_{00\ word}$	260.81
$\tau_{00\ speaker}$	1380.39
ICC	0.73
$N_{speaker}$	26
N_{word}	604
Observations	1173
Marginal R^2 / Conditional R^2	0.670 / 0.912

Table 15. Vowel pitch linear mixed effects regression - Intercept /u/

Predictors	Estimates	std. Error	pitch CI: lower	CI: upper	p
(Intercept)	256.18	9.88	236.81	275.56	$< 0.001^*$
vowel [a]	−11.08	4.53	−19.97	−2.19	0.015^*
vowel [e]	−5.40	4.98	−15.16	4.36	0.278
vowel [i]	−7.22	4.95	−16.92	2.47	0.144
vowel [o]	−4.18	4.90	−13.78	5.43	0.394
sex [m]	−136.05	11.78	−159.15	−112.96	$< 0.001^*$
syllable [Closed_Final]	−16.71	6.22	−28.91	−4.51	0.007^*
syllable [Open]	1.70	2.40	−3.00	6.40	0.478
syllable [Open_Final]	−6.77	3.11	−12.85	−0.68	0.029^*
stress [Unstressed]	−9.10	2.48	−13.97	−4.24	$< 0.001^*$
style [WordList]	19.19	3.40	12.53	25.86	$< 0.001^*$
vowel [a] * sex [m]	7.36	6.21	−4.81	19.54	0.236
vowel [e] * sex [m]	5.83	6.74	−7.37	19.03	0.387
vowel [i] * sex [m]	8.14	6.66	−4.91	21.19	0.222
vowel [o] * sex [m]	2.77	6.47	−9.92	15.46	0.668

Random Effects

σ^2	598.66
$\tau_{00 \text{ word}}$	260.81
$\tau_{00 \text{ speaker}}$	1380.39
ICC	0.73
N_{speaker}	26
N_{word}	604
Observations	1173
Marginal R^2 / Conditional R^2	0.670 / 0.912

Table 16. Vowel duration linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	duration CI: lower	CI: upper	p
(Intercept)	88.26	7.43	73.70	102.82	< 0.001*
vowel [a]	12.25	6.31	−0.11	24.61	0.052
vowel [e]	−12.55	7.49	−27.23	2.12	0.094
vowel [o]	−13.05	7.24	−27.25	1.15	0.072
vowel [u]	−16.78	7.76	−32.00	−1.57	0.031*
sex [m]	−10.82	8.33	−27.15	5.50	0.194
syllable [Closed_Final]	1.90	10.93	−19.51	23.31	0.862
syllable [Open]	−1.92	4.10	−9.95	6.11	0.639
syllable [Open_Final]	24.05	5.35	13.56	34.54	< 0.001*
stress [Unstressed]	2.23	4.34	−6.28	10.74	0.608
style [WordList]	−16.14	4.51	−24.98	−7.30	< 0.001*
vowel [a] * sex [m]	4.46	9.26	−13.70	22.61	0.631
vowel [e] * sex [m]	17.74	10.28	−2.40	37.88	0.084
vowel [o] * sex [m]	17.05	9.77	−2.10	36.20	0.081
vowel [u] * sex [m]	8.48	10.59	−12.27	29.23	0.423

Random Effects

σ^2	2353.34
$\tau_{00 \text{ word}}$	183.33
$\tau_{00 \text{ speaker}}$	44.65
ICC	0.09
N_{speaker}	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.097 / 0.177

Table 17. Vowel duration linear mixed effects regression - Intercept /u/

Predictors	duration				
	Estimates	std. Error	CI: lower	CI: upper	p
(Intercept)	71.47	7.03	57.69	85.26	$< 0.001^*$
vowel [a]	29.04	6.96	15.40	42.68	$< 0.001^*$
vowel [e]	4.23	7.62	−10.70	19.17	0.579
vowel [i]	16.78	7.76	1.57	32.00	0.031^*
vowel [o]	3.74	7.74	−11.44	18.92	0.630
sex [m]	−2.35	8.19	−18.40	13.71	0.775
syllable [Closed_Final]	1.90	10.93	−19.51	23.31	0.862
syllable [Open]	−1.92	4.10	−9.95	6.11	0.639
syllable [Open_Final]	24.05	5.35	13.56	34.54	$< 0.001^*$
stress [Unstressed]	2.23	4.34	−6.28	10.74	0.608
style [WordList]	−16.14	4.51	−24.98	−7.30	$< 0.001^*$
vowel [a] * sex [m]	−4.02	9.65	−22.94	14.89	0.677
vowel [e] * sex [m]	9.26	10.52	−11.35	29.87	0.378
vowel [i] * sex [m]	−8.48	10.59	−29.23	12.27	0.423
vowel [o] * sex [m]	8.57	10.18	−11.39	28.52	0.400

Random Effects

σ^2	2353.34
$\tau_{00\ word}$	183.33
$\tau_{00\ speaker}$	44.65
ICC	0.09
$N_{speaker}$	26
N_{word}	616
Observations	1202
Marginal R^2 / Conditional R^2	0.097 / 0.177

Table 18. Vowel intensity linear mixed effects regression - Intercept /i/

Predictors	Estimates	std. Error	CI: lower	CI: upper	p
			intensity		
(Intercept)	71.25	1.44	68.43	74.06	$< 0.001^*$
vowel [a]	0.39	0.50	−0.59	1.37	0.437
vowel [e]	−0.34	0.62	−1.55	0.88	0.586
vowel [o]	−0.36	0.59	−1.51	0.79	0.542
vowel [u]	−0.32	0.72	−1.73	1.09	0.654
sex [m]	0.75	1.80	−2.79	4.28	0.679
syllable [Closed_Final]	−1.22	0.93	−3.05	0.61	0.191
syllable [Open]	0.77	0.37	0.05	1.49	0.035^*
syllable [Open_Final]	−0.39	0.46	−1.29	0.50	0.390
stress [Unstressed]	−0.59	0.36	−1.29	0.11	0.099
style [WordList]	−0.49	0.59	−1.63	0.66	0.406
vowel [a] * sex [m]	−0.06	0.79	−1.60	1.49	0.940
vowel [e] * sex [m]	0.70	0.88	−1.02	2.42	0.426
vowel [o] * sex [m]	0.33	0.82	−1.28	1.94	0.687
vowel [u] * sex [m]	−0.07	1.00	−2.03	1.89	0.943

Random Effects

σ^2	10.46
$\tau_{00 \text{ word}}$	14.70
$\tau_{00 \text{ speaker}}$	27.14
ICC	0.80
N_{speaker}	26
N_{word}	616
Observations	1201
Marginal R^2 / Conditional R^2	0.013 / 0.803

Table 19. Vowel intensity linear mixed effects regression - Intercept /u/

Predictors	Estimates	std. Error	intensity CI: lower	CI: upper	p
(Intercept)	70.92	1.45	68.09	73.76	$< 0.001^*$
vowel [a]	0.71	0.66	−0.58	2.01	0.281
vowel [e]	−0.02	0.73	−1.46	1.42	0.983
vowel [i]	0.32	0.72	−1.09	1.73	0.654
vowel [o]	−0.03	0.72	−1.44	1.38	0.962
sex [m]	0.67	1.83	−2.91	4.26	0.712
syllable [Closed_Final]	−1.22	0.93	−3.05	0.61	0.191
syllable [Open]	0.77	0.37	0.05	1.49	0.035^*
syllable [Open_Final]	−0.39	0.46	−1.29	0.50	0.390
stress [Unstressed]	−0.59	0.36	−1.29	0.11	0.099
style [WordList]	−0.49	0.59	−1.63	0.66	0.406
vowel [a] * sex [m]	0.01	0.91	−1.77	1.80	0.989
vowel [e] * sex [m]	0.77	1.00	−1.20	2.74	0.443
vowel [i] * sex [m]	0.07	1.00	−1.89	2.03	0.943
vowel [o] * sex [m]	0.40	0.95	−1.47	2.27	0.673

Random Effects

σ^2	10.46
$\tau_{00 \ word}$	14.70
$\tau_{00 \ speaker}$	27.14
ICC	0.80
$N_{speaker}$	26
N_{word}	616
Observations	1201
Marginal R^2 / Conditional R^2	0.013 / 0.803