Crosslinguistic Comparison of Acoustic Variation in the Vowel Sequences /ia/ and /io/ in Four Romance Languages

Johanna Cronenberg[1], *Ioana Chitoran*[1], *Lori Lamel*[2], *Ioana Vasilescu*[2]

[1]Université Paris Cité, France
[2]CNRS, Université Paris-Saclay, Laboratoire Interdisciplinaire des Sciences du Numérique, France
johanna.cronenberg@u-paris.fr, ioana.chitoran@u-paris.fr, lamel@lisn.fr, ioana@limsi.fr

Abstract

According to theoretical accounts, Romance languages differ with respect to the pronunciation of vowel sequences such as /ia/ and /io/. Italian produces these sequences as diphthongs, i.e. /ja/ and /jo/, while Portuguese prefers hiatuses. Spanish and Romanian are claimed to use a mix of diphthongs and hiatuses. These accounts are based on phonological criteria or on small samples of carefully read isolated words. This study proposes to investigate the realisation of /ia/ and /io/ in large corpora of fluent speech, focusing on their acoustic properties along the whole formant trajectory. The results of the functional and statistical analyses show extensive acoustic variation with respect to the duration of the sequences as well as their formant dynamics. The languages clearly differ from one another, but the analysis shows that there remains more to learn about the distinction between diphthongs and hiatuses. We discuss the inclusion of further factors in future investigations.
Index Terms: acoustic phonetics, diphthong, hiatus, Romance languages, functional data analysis

1. Introduction

1.1. Diphthongs vs. Hiatuses

This study is concerned with acoustic variations related to the distinction between diphthongs and hiatuses of rising sonority. While diphthongs consist of a glide or semi-vowel followed by a full vowel (e.g. /ja, jo/), hiatuses consist of two steady-state vowels (e.g. /ia, io/). Henceforth we will use the term vowel sequences as a hypernym for diphthongs and hiatuses. The phonological importance of this distinction lies in its effect on the syllabification of words: hiatuses are typically considered to be heterosyllabic while diphthongs are tautosyllabic [1]. When the first element of the vowel sequence is stressed, as e.g. in Portuguese *academias* /ɐ.kɐ.də.ˈmi.ɐʃ/, it is a syllable nucleus and the sequence is thus heterosyllabic. Significant variations in the production and perception of these sequences only occur when their first element is unstressed [2, 3]. In other words, when the second vocalic element is stressed or the entire vowel sequence is unstressed, either a diphthong or a hiatus can occur.

Independently of the studied language, diphthongs are typically described in terms of duration and formant measurements [4, 5, 6, 7, 8, 9, 10, 11, 12]. The transition between the glide and vowel is sometimes quantified by means of the formants' rate of change [13, 14, 15]. There are, however, only very few studies about the contrast between diphthongs and hiatuses, and with the exception of [2], they are all on Spanish. For Spanish, it has been shown that hiatuses are approximately 50ms longer than diphthongs [16], even though there is a substantial overlap between the duration distributions of the two [1]. It is in particular the initial section of hiatuses that contributes to this longer duration [3]. Hiatuses not only have a greater curvature of the second formant frequency than diphthongs [3], but differences can be observed along the whole formant trajectory. Using the same functional data analysis technique as the study at hand, Gubian and colleagues found that Spanish diphthongs display a gradual movement of F1 and F2 without clear steady states, whereas hiatuses are characterised by stable formants at the start and end with steeper formant slopes in between [16].

1.2. Vowel Sequences in Romance Languages

Vowel sequences in Spanish present an interesting case, precisely because both hiatuses and diphthongs occur. In this study, we will broaden the view from Spanish to three further Romance languages, namely Italian, Romanian, and Portuguese. There are two main sources from which vowel sequences have evolved in Romance languages [17]. The first is the "breaking" of the Late Latin stressed mid vowels /ɛ/ and /ɔ/, e.g. Latin *petra* (engl. stone) > Italian *pietra* /pjetra/, Spanish *piedra* /pjedra/, and Romanian *piatră* /pjatrə/. Portuguese did not undergo this process, thus *pedra* /pɛdrɐ/. The second source is the palatalisation of Latin /l/, e.g. Latin *clave* (engl. key) > Italian *chiave* /kjave/ and Romanian *cheie* /kjeje/. This process developed further into palatal laterals and fricatives for Spanish (*llave* /ʎave/) and Portuguese (*chave* /ʃave/).

Vowel sequences derived from these two historical sources are produced as diphthongs in almost all cases, but the actual realisation of vowel sequences from other sources can be subject to adaptations and variation. In Spanish and Romanian, word-initial sequences tend to be produced as hiatuses, e.g. Spanish *biologo* /biˈologo/ vs. *radiologo* /raˈdjologo/[2]. Cabré and Prieto point out that diphthongs become more likely in Spanish with increased distance from the stressed syllable, e.g. *diálogo* /diˈalogo/ (engl. dialogue) and *dialogo* /diaˈlogo/ (engl. I converse), but *dialogó* /djaloˈgo/ (engl. she/he conversed) [18]. In Italian and Spanish, hiatuses may occur if the high vowel is stressed in a morphologically related word, e.g. Italian *spianti* /ˈspjanti/ (engl. you uproot) vs. *spiante* /spiˈante/ (engl. spying) which is related to *spia* /ˈspia/ (engl. spy) [18, 1]. Phonetic context seems to play a limited role in affecting the contrast, but /ia, io/ may be produced as hiatus after trills in Spanish [1]. In general, Spanish and Italian hiatuses remain relatively rare [19, 18]. Portuguese is different from the other Romance languages because its vowel sequences do not derive from the breaking of Latin vowels or the palatalisation of Latin /l/. Thus, all vowel sequences in Portuguese are phonologically classified as hiatuses. A diphthongal realisation can, however, occur in informal speech or in post-tonic position [2]. According to [1], the actual realisation of vowel sequences in Romance languages remains somewhat unpredictable even when controlling for all the historical, morphological, phonological, and phonetic factors mentioned above.

1.3. Aim and Hypotheses

The aim of this study is to take the first step towards a comprehensive account of acoustic variations related to the distinction between diphthongs and hiatuses in the big Romance languages as spoken in Europe. To this end, we analyse large corpora of fluent speech instead of small and controlled production data, which will allow us to observe this variation in more naturalistic settings and bypass the Observer's Paradox [20]. At the same time, we will limit the available material and scope for this study to certain conditions in order to reduce the complexity of the analysis. More specifically, we will focus on vowel sequences in tonic position in which the second element is stressed, and we will not investigate the effects of morphological factors or phonetic context on the configuration of the vowel sequences.

In line with the literature review, we expect Portuguese to show the clearest tendency towards hiatuses which should present as long vowel sequences with steady-state portions at the start and end as well as steep formant transitions. Italian and Spanish, at the other end of the spectrum, should prefer diphthongs, i.e. shorter, less curvy and flatter formants. Romanian has the most robust contrast between diphthongs and hiatuses and may thus show the largest variation with respect to the acoustic measurements. In Spanish and Romanian, initial vowel sequences may be longer than medial ones, and based on [2] there should be no initiality effect for Portuguese.

2. Method

2.1. Data Selection

The corpora used for this study consist of radio and television show recordings in Italian (168 hours), Portuguese (114 hours), Romanian (300 hours), and Spanish (223 hours) which aired between 1992 and 2012 [21]. The corpora were compiled as training material for automatic speech recognition systems, i.e. they were not intended for linguistic investigations. The recordings are of broadcast news, i.e. read or semi-read speech, as well as interviews and debates, i.e. spontaneous speech. We selected the European variants of the four Romance languages wherever possible. Given that the shows were intended for a broad audience, we do not expect there to be a lot of dialectal, non-Standard variation.

The recordings were orthographically transcribed and phonemically segmented via forced alignment with word-context independent phone models [22, 23]. Given the large amount of data, the alignment was not checked manually. For each corpus, there was a pronunciation dictionary from which we selected all words that contained sequences of /i/ or /j/ followed by /a/ or /o/ (as well as some language-specific pronunciation variants, such as /ɐ/ and /ɔ/). Note that the symbols /i/ and /j/ were not necessarily indications of the phonological status of the vowel sequence as a diphthong or hiatus, but rather of the transcription conventions that differed between the languages. Henceforth, we will refer to these sequences as /ia/ and /io/, respectively. The vowel sequences had to be preceded and followed by at least one phone, so that they were neither in absolute initial nor final position within the word. Lexical stress and proximity of the vowel sequence to the stressed syllable were annotated semi-automatically by devising and applying stress rules and using available resources [24, 25]. For the present study, only sequences in which the second element was stressed were analysed. Finally, we extracted all occurrences of the words of interest – i.e. those that contained /ia/ or /io/ in non-word-initial or non-word-final tonic position – plus/minus one word from the audio and segmentation files. Table 1 shows the count of analysed vowel sequences per language. The number of Portuguese vowel sequences is relatively low compared to the other three languages because Portuguese did not undergo the breaking of Latin mid vowels and because words with the sequence /io/ in the other three languages are often produced with /iu/ in Portuguese, e.g. Spanish *diarios* /diarios/ vs. Portuguese *diários* /diariuʃ/ (engl. diaries).

Table 1: *Count of /ia/ and /io/ sequences analysed in the four languages.*

Language	**/ia/**	**/io/**	
Italian	20,186	26,033	**46,219**
Portuguese	4,116	415	**4,531**
Romanian	26,773	2,783	**29,556**
Spanish	16,474	64,046	**80,520**
	67,549	**93,277**	

2.2. Data Processing

The first two formants were extracted between the start and end of the vowel sequence using the `forest` function from the R package `wrassp` [26] with default settings, i.e. a 30 ms window with a 5 ms shift to extract a maximum of four formants. Subsequently, the formants were z-scaled by language.

The duration of the vowel sequences was extracted from the automatic segmentation. In order to normalise the duration for speech rate differences, we computed articulation rate as the number of phones per second (all silences, pauses, breaths, and hesitations excluded) for every three-word-sequence. The duration of a vowel sequence was then divided by the articulation of the three-word-sequence it appeared in. The signals' normalised duration was included in the following functional analysis by means of time warping [27]. That is, the duration was transformed into a horizontal signal $r(t)$ with the intercept at $-log(\frac{dur}{rate})$, i.e. the negative logarithm of the normalised duration.

Separately for /ia/ and /io/, the normalised F1 and F2 as well as the normalised and time-warped duration $r(t)$ of the vowel sequences were then submitted to Functional Principal Components Analysis (FPCA; [16, 28]). Given time-varying signals as input, FPCA returns (i) the mean signals, (ii) K Principal Components (PCs) that capture the main dimensions of variation in the signals' shapes, and (iii) a score or weight s for every input signal i and every PCk that expresses how the variation captured by the PC is manifested in the input signal. The linear decomposition of the input signals is achieved using the following formula:

$$x_i(t) \approx \mu_x(t) + \sum_{k=1}^{K} s_{k,i} \cdot PCk_x(t) \tag{1}$$

where $x(t)$ is either normalised F1, normalised F2, or the normalised duration signal $r(t)$, $\mu_x(t)$ is the mean of the signal, $PCk_x(t)$ are K Principal Components ($k = 1, ..., K$), and $s_{k,i}$ are the PC scores that modulate each PCk differently for each input signal triplet $(F1_i(t),\ F2_i(t),\ r_i(t))$. FPCA has three main advantages. First, it analyses variation across the full time-varying signal, not just e.g. the start, end, or mid point, while still being computationally efficient. Second, FPCA, unlike Generalised Additive Models (GAMs) can cope with multi-dimensional signals. Note that the PC scores $s_{k,i}$ are independent of $x(t)$, i.e. the scores modulate the formants and duration signals together. And third, the resulting parametrisation captures the variation that is actually contained in the input signals [29], and not some default properties such as the signals' slope and curvature which are returned by a discrete cosine transform.

FPCA was applied using the R package `fda` [30]. After careful inspection, the first $K = 4$ PCs (which together captured 75.5% of all variance in the signals' shapes) were analysed statistically for /ia/, and PC1, PC2, and PC6 (which captured 54.2% of all variance) were analysed for /io/.

2.3. Statistical Analysis

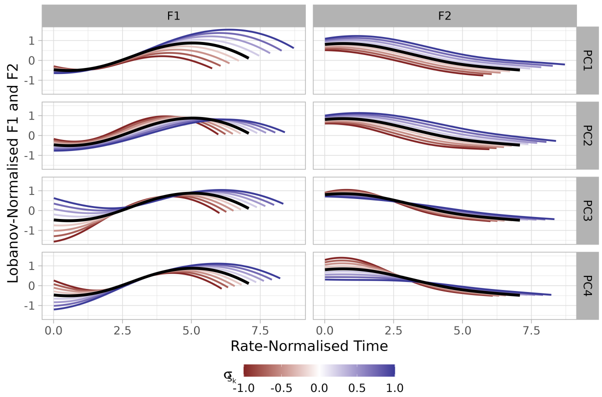
Since initiality was an important factor in determining the production of vowel sequences in [2], we encoded the sequence here as initial when no more than two phones preceded the vowel sequence and as medial in all other cases.

For each PC score that was considered relevant to the distinction between diphthongs and hiatuses, we ran a linear mixed effect regression model with the PC score as the dependent variable, language (4 levels) and initiality (2 levels) as well as their interaction as fixed factors, and a random intercept for word (3849 levels for /ia/, 3765 levels for /io/). We were unable to include a random effect structure for speaker since the corpora did not identify individual speakers. We believe that the found statistical effects would emerge in an even clearer way if the models had been able to take care of speaker-related random variability. The LMERs were computed with `lmerTest` [31]. Post-hoc comparisons were calculated with estimated marginal means, using the R package `emmeans` [32]. Since almost all comparisons were statistically highly significant, we refrain from mentioning them all in the results.

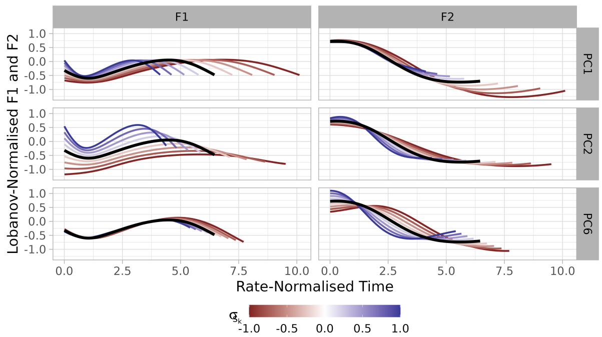
3. Results

Eq. (1) can be used to inspect the variation captured by the PCs by setting s_k to a range of values between $\pm\sigma_{s_k}$ (i.e. the standard deviation of the PCk score) and all other scores to zero. So in Figure 1a the thick black lines represent the mean formants μ_{F1} and μ_{F2} for /ia/ across rate-normalised time, which are the same for all PCs. In the top row, the red lines are the result of $\mu_{F1} - \sigma_{s_1} \cdot PC1_{F1}(t)$ (left) and $\mu_{F2} - \sigma_{s_1} \cdot PC1_{F2}(t)$ (right), while the blue lines are the result of adding instead of subtracting the s_1 standard deviation multiplied by the PC1 curve.

For /ia/ (Figure 1a), PC1 captured mostly the variation of duration along with the height of the F1 peak and the overall height of F2. PC2 captured variation in the rate of change of F1 as well as the overall height of F2. PC3 captured differences in the height of F1 at start, and consequently differences in F1 rate of change, with minimal covariation in F2. Finally, PC4 captured variation in the rate of change in F2 and in F1 height at start. Figure 1b was constructed in the same way as Figure 1a, but for /io/. PC1 mainly captured differences in duration. PC2 captured variation in the curviness of F1 and, less prominently, variation in the rate of change of F2. PC6 captured variation in the length of steady states at the start and end of the F2 trajectory with no covariation in F1. For both /ia/ and /io/, the variations in the formant dynamics were inextricably linked to



(a)



(b)

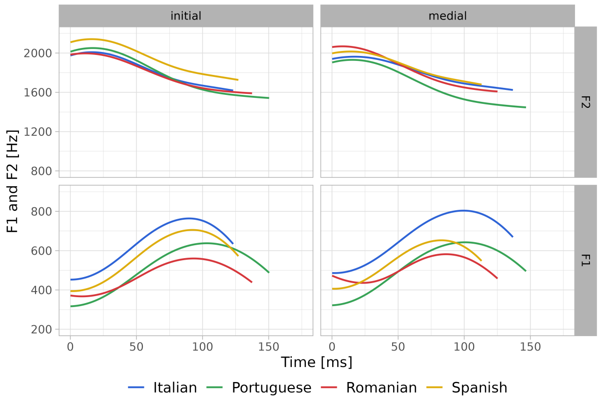
Figure 1: *Variation in duration and formant dynamics captured by the relevant Principal Components for /ia/ (1a) and /io/ (1b).*

variations in the duration of the trajectories, showing that separate analyses of duration and formant measurements may not be appropriate to study these vowel sequences.

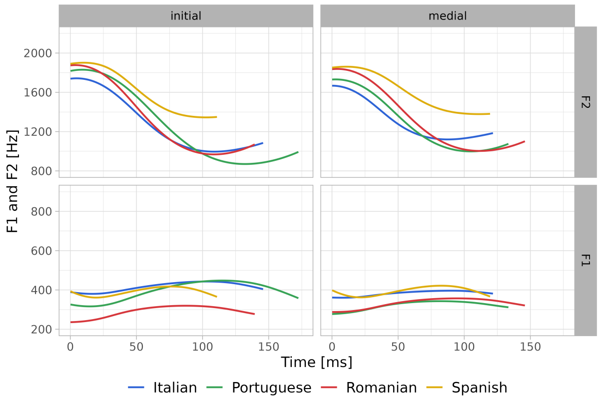
Eq. (1) was used once again to reconstruct F1 and F2, this time separately for the languages and position within the word. In order to do so, we inserted the estimated marginal means for all combinations of the factors language and initiality in the place of s_k (where $k = 1, 2, 3, 4$ for /ia/ and $k = 1, 2, 6$ for /io/). Figure 2 can thus be interpreted as the estimated or predicted F1 and F2 trajectories of the vowel sequences word-initially and word-medially for each language.

From Figure 2a it is obvious that Portuguese /ia/ is longer than that of the other three languages both initially and medially. Portuguese /ia/ also has the steepest F2 transition. Word-initially, Portuguese /ia/ shows a slightly longer F2 plateau at the start of the vowel sequence compared to the other languages. Romanian /ia/ has a long steady state at the start, especially in initial position. Word-medially, Spanish has the flattest and shortest F2 trajectory. While there is no duration difference between initial and medial /ia/ for Portuguese, medial /ia/ is shorter than initial /ia/ for Spanish and Romanian; the reverse is true for Italian. The statistical analysis showed no significant difference between Spanish and Italian /ia/ word-initially.

Figure 2b shows that word-initial /io/ is predicted to be the longest and have the curviest formants for Portuguese. Spanish, on the other hand, has the shortest word-initial /io/ with the least steep F2 trajectory. Romanian /io/ has the steepest F2 transitions both initially and medially compared to the other



(a)



(b)

Figure 2: *Reconstructed F1 and F2 for /ia/ (2a) and /io/ (2b), separately for each language (colour-coded) and for initial (left) vs. medial vowel sequences (right). Both plots use the same axes for easier comparison.*

languages. In word-medial position, /io/ is shorter and has less steep formants in Spanish and Italian than in Portuguese and Romanian. The F2 trajectories for Italian /io/ in both positions show no steady state at the start. In Italian and Portuguese, /io/ is shorter in word-medial than in word-initial position. There was no statistical difference between initial and medial /io/ for both Spanish and Romanian. We did not find considerable differences in the variation of the PC scores predicted by the LMERs nor in the 95% confidence interval predicted by estimated marginal means for Romanian compared to the other languages.

4. Discussion

The analysis showed that the acoustic configuration of the vowel sequences in Portuguese is hiatus-like, i.e. long and curvy formants with steep transitions between two steady states. There were no differences between initial and medial /ia/ in Portuguese which coincides with the findings by [2], but /io/ was longer initially than medially. Thus, it may be the case that word-medial /io/ is more likely to be produced as a diphthong than word-initial /io/ in Portuguese. Vowel length is of course influenced by several prosodic and phonetic factors, so in this limited study, the initiality effect cannot be linked unequivocally to the diphthong/hiatus distinction. In line with the expectations, Spanish and Italian show a preference for diphthongal vowel sequences, i.e. /ia/ and /io/ are produced with barely any steady state at the start, short durations, and relatively flat formant trajectories. For /ia/ in Spanish we also found the expected initiality effect, i.e. word-medial sequences may be more resistant to being produced as a hiatus than word-initial ones. However, there was no difference between initial and medial /io/ in Spanish. For Italian, the analysis showed that /io/ was longer in initial than in medial position and /ia/ was longer in medial than in initial position. This duration difference did not coincide with a longer steady state at the start of the sequence or with steeper formant transitions, so the differences in duration may not necessarily be attributable to the distinction between diphthongs and hiatuses. Romanian, which has the most robust contrast between diphthongs and hiatuses according to the literature, showed mostly hiatus-like patterns with relatively long steady states at the start of the vowel sequences, especially in initial position. We found the expected initiality effect for /ia/, but not for /io/. However, we were not able to identify any measure of variability that was larger for Romanian than for the other languages which would have indicated clearly defined, separate configurations for hiatuses and diphthongs.

To our knowledge, this study is the largest cross-linguistic comparison of vowel sequences to date. Using large speech corpora which were never intended for linguistic use allowed us to investigate the acoustic characteristics of /ia/ and /io/ in a naturalistic setting. This kind of data makes it possible "to test whether effects that arise in experimental or intuition-based studies are widespread and meaningful" [33, 8]. Our findings confirmed some of the general trends that had been found in smaller experimental studies, but also revealed some heretofore unknown details and effects. This was largely due to the fact that we were able to observe differences between the languages along the whole formant trajectory – and not only at a few chosen points in time – and describe how the formants' dynamics covaried with the duration of the sequences. The functional data analysis thus offers entirely new insights into the acoustics of diphthongs and hiatuses in Romance languages.

Nevertheless, using these corpora and analysis techniques has limitations. First, there are still a lot of factors that might influence the production of /ia, io/ which were not explored here. We expect there to be differences between read and spontaneous speech (i.e. between broadcast news and interviews or debates), but this information is only available for parts of the corpora. Similarly, we do not have information on morphological boundaries or paradigmatic relations for all 7614 words which were included in the analysis, yet this might be important to find clearer hiatus- and diphthong-like patterns. Including these morphological factors may also help to disentangle the impact of the sequences' initiality on their realisation from typical prosodic patterns which lead to longer vowels in initial than in medial position. Because this was the first analysis of these corpora, we limited the data to those cases in which the second element of the vowel sequence was stressed. In future investigations, we plan to include both lexical stress and proximity to the stressed syllable as factors since there are clear predictions about their impact on /ia/ and /io/ [18, 2]. Furthermore, we would like to extend our analyses to the vowel sequence /ie/ which was not considered here for reasons of space. Finally, to complete the comprehensive picture of diphthongs and hiatuses in Romance languages we aim for, future analyses will control for phonetic context.

5. Acknowledgements

This study was funded as part of the project DIPVAR (ANR-21-CE38-0019), PI Ioana Vasilescu. We thank the reviewers for their insightful comments which helped to improve this paper.

6. References

[1] J. I. Hualde and M. Prieto, “On the diphthong/hiatus contrast in Spanish: Some experimental results,” *Linguistics*, vol. 40, no. 2, pp. 217–234, 2002.

[2] I. Chitoran and J. I. Hualde, “From hiatus to diphthong: The evolution of vowel sequences in Romance,” *Phonology*, vol. 24, no. 1, pp. 37–75, 2007.

[3] L. Aguilar, “Hiatus and diphthong: Acoustic cues and speech situation differences,” *Speech Communication*, vol. 28, pp. 57–74, 1999.

[4] E. L. Asu, P. Lippus, E. Niit, and H. Türk, “The Acoustic Characteristics of Monophthongs and Diphthongs in the Kihnu Variety of Estonian,” *Linguistica Uralica*, vol. 48, no. 3, pp. 161–170, 2012.

[5] M. J. Ball, “The centring diphthongs in Southern English – a sound change in progress,” *Journal of the International Phonetic Association*, vol. 14, no. 1, pp. 38–44, 1984.

[6] ——, “A spectrographic investigation of three Welsh diphthongs,” *Journal of the International Phonetic Association*, vol. 13, no. 2, pp. 82–89, 1983.

[7] F. Clermont, “Spectro-temporal description of diphthongs in F1–F2–F3 space,” *Speech Communication*, vol. 13, pp. 377–390, 1993.

[8] R. Collier, F. Bell-Berti, and L. J. Raphael, “Some Acoustic and Physiological Observations on Diphthongs,” *Language and Speech*, vol. 25, no. 4, pp. 305–323, 1982.

[9] M. Gottfried, J. D. Miller, and D. J. Meyer, “Three approaches to the classification of American English diphthongs,” *Journal of Phonetics*, vol. 21, no. 3, pp. 205–229, 1993.

[10] M. Lindau, K. Norlin, and J.-O. Svantesson, “Some cross-linguistic differences in diphthongs,” *Journal of the International Phonetic Association*, vol. 20, no. 1, pp. 10–14, 1990.

[11] A. Pycha and D. Dahan, “Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /aɪ/,” *Journal of Phonetics*, vol. 56, pp. 15–37, 2016.

[12] M. Weirich and A. P. Simpson, “Individual differences in acoustic and articulatory undershoot in a German diphthong – Variation between male and female speakers,” *Journal of Phonetics*, vol. 71, pp. 35–50, 2018.

[13] A. M. Borzone de Manrique, “Acoustic Study of /i,u/ in the Spanish Diphthong,” *Language and Speech*, vol. 19, no. 2, pp. 121–128, 1976.

[14] I. Chitoran, “A perception-production study of Romanian diphthongs and glide-vowel sequences,” *Journal of the International Phonetic Association*, vol. 32, no. 2, pp. 203–222, 2002.

[15] S. K. Jha, “Acoustic analysis of the Maithili diphthongs,” *Journal of Phonetics*, vol. 13, pp. 107–115, 1985.

[16] M. Gubian, F. Torreira, and L. Boves, “Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts,” *Journal of Phonetics*, vol. 49, pp. 16–40, 2015.

[17] T. Alkire and C. Rosen, “The evolution of stressed vowels,” in *Romance Languages: A Historical Introduction*. Cambridge: Cambridge University Press, 2010, pp. 5–25.

[18] T. Cabré and P. Prieto, “Exceptional hiatuses in Spanish,” in *Optimality-Theoretic Advances in Spanish Phonology*, S. Colina and F. Martínez-Gil, Eds. Amsterdam: John Benjamins Publishing Company, 2007, pp. 205–238.

[19] A. Herrero de Haro and A. Alcoholado Feltstrom, “Anti-hiatus tendencies in Spanish: Rate of occurrence and phonetic identification,” *Linguistics*, pp. 1–26, 2023.

[20] W. Labov, “The Social Stratification of /r/ in New York City Department Stores,” in *Sociolinguistic Patterns*, W. Labov, Ed. Philadelphia: University of Pennsylvania Press, 1972, pp. 43–69.

[21] I. Vasilescu, Y. Wu, A. Jatteau, M. Adda-Decker, and L. Lamel, “Alternances de voisement et processus de lénition et de fortition: Une étude automatisée de grands corpus en cinq langues romanes,” *Traitement Automatique des Langues*, vol. 61, no. 1, pp. 11–36, 2020.

[22] L. Lamel and J.-L. Gauvain, “Continuous Speech Recognition at LIMSI,” in *Proceedings of the Final Review of the DARPA ANNT Speech Program*, Stanford, 1992, pp. 1–7.

[23] M. Adda-Decker and L. Lamel, “Pronunciation variants across system configuration, language and speaking style,” *Speech Communication*, vol. 29, pp. 83–98, 1999.

[24] G. D. Garcia, “Weight gradience and stress in Portuguese,” *Phonology*, vol. 34, no. 1, pp. 41–79, 2017.

[25] A.-M. Barbu, “Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries,” in *Proceedings of the 6th Language Resources and Evaluation Conference*, Marrakesh, 2008, pp. 1937–1941.

[26] R. Winkelmann, L. Bombien, M. Scheffers, and M. Jochim, *wrassp: Interface to the 'ASSP' Library*, 2024, r package version 1.0.5. [Online]. Available: https://CRAN.R-project.org/package=wrassp

[27] M. Gubian, L. Boves, and F. Cangemi, “Joint analysis of f0 and speech rate with Functional Data Analysis,” in *Proceedings of the 36th International Conference on Acoustics, Speech, and Signal Processing*, Prague, 2011, pp. 4972–4975.

[28] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. New York: Springer New York, 2010.

[29] J. Wilson Black, J. Brand, J. Hay, and L. Clark, “Using principal component analysis to explore co-variation of vowels,” *Language and Linguistics Compass*, vol. 17, no. 1, p. e12479, 2022.

[30] J. Ramsay, *fda: Functional Data Analysis*, 2024, r package version 6.1.7. [Online]. Available: https://CRAN.R-project.org/package=fda

[31] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest package: Tests in linear mixed effects models,” *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[32] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2024, r package version 1.10.0. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[33] A. C. Cohn and M. E. L. Renwick, “Doing phonology in the age of big data,” in *Cornell Working Papers in Phonetics and Phonology 2019*, Ithaca, 2019, pp. 1–36.