PROMINENCE AND CENTRALIZATION: A TWO-WAY CONTRAST BETWEEN LEXICAL AND EPENTHETIC NON-LABIAL HIGH VOWELS IN ALGHERESE CATALAN

PROMINENCIA Y CENTRALIZACIÓN: UN DOBLE CONTRASTE ENTRE VOCALES ALTAS NO LABIALES LÉXICAS Y EPENTÉTICAS EN EL CATALÁN DE L'ALGUER

JESÚS JIMÉNEZ
Universitat de València
jesus.jimenez@uv.es

MARIA-ROSA LLORET
Universitat de Barcelona
mrosa.lloret@ub.edu

Artículo recibido el día: 09/01/2015
Artículo aceptado definitivamente el día: 12/06/2015
Estudios de Fonética Experimental, ISSN 1575-5533, XXIV, 2015, pp. 171-204

ABSTRACT

Algherese Catalan has the peculiarity of presenting an inserted [i] vowel across words in order to avoid certain consonant codas. In this study, we compare five acoustic features of this epenthetic segment (i.e., duration, intensity, and the three first formants) with those of lexical /i/ vowels, both stressed and unstressed. The results indicate that the three vowels present differences only with respect to duration, F2 and, to a lesser extent, F3. The second formant values decline progressively from the lexical stressed vowels to the inserted unstressed segments, with lexical unstressed segments at an intermediate point. This gradation mirrors the relative prominence of each vowel. The differences in F2 between lexical stressed and unstressed vowels can be attributed to the shorter duration of the later segments. Lexical unstressed and inserted vowels, however, have an equivalent duration, challenging the idea that inserted vowels are more centralized due to their shorter duration. All in all, the data point to a double contrast: first, between lexical stressed and unstressed segments and, second, between unstressed segments.

Keywords: *acoustic phonetics*, *centralization*, *duration*, *epenthesis*, *formant values*, *vowels*, *Algherese Catalan*.

RESUMEN

El catalán de L'Alguer presenta la peculiaridad de insertar una vocal [i] entre palabras con el objetivo de evitar ciertas codas consonánticas. En este estudio, comparamos cinco rasgos acústicos de las vocales epentéticas (duración, intensidad y los tres primeros formantes) con los de las vocales /i/ léxicas, tónicas y átonas. Los resultados indican que las tres vocales solo difieren con respecto a duración, F2 y, en menor grado, F3. Los valores del segundo formante descienden desde las vocales léxicas tónicas hasta las vocales insertadas, con las vocales léxicas átonas en un punto intermedio, una escala que reproduce el grado de prominencia relativa de cada vocal. Las diferencias de F2 entre las vocales léxicas tónicas y átonas se pueden atribuir a la menor duración de estas últimas. En cambio, las vocales átonas léxicas y epentéticas tienen una duración equivalente, lo que cuestiona la idea de que la mayor centralización de las vocales insertadas se deba a su menor duración. En conjunto, los datos apuntan a un doble contraste: por un lado, entre las vocales léxicas átonas y las vocales léxicas tónicas y, por otro, entre las vocales átonas léxicas y las vocales insertadas.

Palabras clave: *fonética acústica*, *centralización*, *duración*, *epéntesis*, *valores formánticos*, *vocales*, *catalán de L'Alguer*.

1. INTRODUCTION

The variety of Catalan spoken in the Sardinian town of Alghero has the peculiarity of displaying an inserted [i] vowel (underlined in the examples below) across words to avoid certain consonant codas that would otherwise arise phrasally (1a) (Kuen, 1932; Loporcaro, 1997; Lloret & Jiménez, 2005, 2006, 2008, 2010)[1]. This internal epenthesis is not notated orthographically and always gives rise to open syllables. As shown in (1b), the final inserted [i] vowel does not appear to satisfy minimal word requirements but repairs ill-formed syllabic structures. Hence, words like *típic* or *dolç* are realized with a final consonant –without a flanking final vowel– in isolation or phrase-finally and before a vowel initial word (1b), but are followed by an inserted [i] before a consonant initial word (1a)[2]. Inserted [i] vowels coexist with lexical [i] vowels (i.e., underlying /i/'s), which can occur in stressed or unstressed positions, as in *típic* (1c)[3].

(1) a. típic dolç de Pasqua ['ti.pi.k̲i̲.'dol.t͡s̲i̲.de.'pas.kwa]

[1] For the sake of completeness, one of the reviewers mentions that in the pronunciation of non-schooled people one can hear inserted [i]'s within words in recent loans and learned words that display the same problematic consonant contacts, as in *còc*[i]*tel* 'cocktail', *ap*[i]*nea* 'apnea', and even in the acronym *INPS* ['impis] (*Istituto Nazionale della Previdenza Sociale* 'Social Security National Institute'), where the inserted vowel avoids the regular consonant simplification that applies elsewhere (cf. *camp* ['kamp] 'field', *camps* ['kanˈs] 'fields'; Cabrera, 2013:69). Exceptionally, one might also hear a short [i] added phrase-finally, especially in emphatic contexts (*típic!* ['ti.pikˈ]). As the reviewer points out, these marginal pronunciations do not alter the results of the investigation carried out in the present study.

[2] Lloret & Jiménez (2008:68-69) point out that, unlike Algherese Catalan, Sardinian does not usually show vowel insertion across words to repair illicit consonant contacts (final /t/, in third person singular verb forms, for instance, is elided before all consonants, with reinforcement or gemination of the second consonant: *màndhicat su casu* ['mandika ssu 'kazu] '(s/he) eats the cheese'; cf. Jones, 1988:322, 326). However, the vowel [i] is inserted word-initially for syllabic reasons ([i]*scola* 'school'; cf. [a]*scola* and also [a]*spaguets* 'spaghetti' in Algherese Catalan) and a copy of the preceding vowel is inserted in absolute final position (*tempus* ['tempuzu] 'time', *cantat* ['kantata] '(s/he) sings'; cf. Jones, 1988:326). Recent data show that this copy-vowel epenthesis may be maintained across words after *-s* (*rosas sardas* [ˌrɔza 'zardaza]~[ˌrɔzaza 'zardaza] 'Sardinian roses'; cf. Torres-Tamarit *et al*, forthcoming).

[3] For the purposes of the paper we do not distinguish between primary and secondary stress.

	‘typical Easter sweet’		
b.	típic	[ˈti.pik]	‘typical’
	és típic	[es.ˈti.pik]	‘it is typical’
	típic alguerès	[ˈti.pi.kal.ga.ˈres]	‘typical Alguerese’
	dolç	[ˈdol͡ts]	‘sweet’
	és dolç	[ez.ˈdol͡ts]	‘it is sweet’
	dolç alguerès	[ˈdol.t͡sal.ga.ˈres]	‘Alguerese sweet’
c.	típic	[ˈti.pik]	

According to work by Kuen (1932, 1934), in the 1930s vowel insertion was applied optionally as an alternative to consonant simplification (e.g. *los plats de plata* [lus.ˈpra.t͡si̲.de.ˈpra.ta]~[lus.ˈpraz.de.ˈpra.ta] ‘the silver dishes’; Kuen, 1932: 173)[4]. Loporcaro (1997) suggests that, at that time, its status was more that of an *excrescent* vowel than that of an inserted *epenthetic* vowel (cf. Levin, 1987; Hall, 2006), because it is reported as having a variable phonetic nature[5]: it is described as lower and/or more centralized than [i] and shorter than lexical vowels (see also Recasens, 1991:67; Bosch, 2002:123; Ballone, 2008, 2010). Kuen (1932:156-157) asserts, though, that the syllable containing this vowel already counted as a full unit for metrical parsing in traditional poetry and folk songs. He illustrates his claim by highlighting the rhythm of the traditional song in (2), with nine syllables counting until the last stressed syllable for the meter: there are five instances of [i] insertion (i.e., [tot͡si̲ ma]) alternating with one case of consonant simplification in the same context (i.e. [toz ma daˈzid͡ʒan] in the penultimate line). The example in (3), from the poet Rafael Catardi (Catardi, 1971:32), born in 1893, proves further that an extra vowel (noted as [i̲] in the example) was pronounced and metrically parsed to obtain a decasyllable in the second line of the verse.

[4] We adapt Kuen’s (1932, 1934) transcriptions to the IPA notation. All the English translations, and the Catalan transliteration in (2), are ours. In (2), following the suggestion of one of the reviewers, we have changed the original transcription [ˈdɔnan] to [ˈdonan], which we consider a lapse in the source.

[5] Among the characteristics that Hall (2006:391) mentions for the excrescent, intrusive, vowels are the following: the fact that they are phonologically invisible; their quality is either a schwa or a copy of a nearby vowel; they generally occur in heterorganic clusters; they are likely to be optional, have a highly variable duration, or disappear at fast speech rates, and they do not seem to have the function of repairing illicit contacts.

(2) […]
tots me miren, tots m'adoren,
tots me fanen los compliments,
tots me desitgen, tots me volen,
tots me donen l'apuntament.

Phonetic transcription
[to͡tsi̲ ma ˈmiɾan | to͡tsi̲ m aˈdɔɾan |
to͡tsi̲ ma ˈfanan lus kumpɾiˈmen͡ts |
oz ma daˈzi͡ʤan | to͡tsi̲ ma ˈvɔɾan |
to͡tsi̲ ma ˈdonan l apuntaˈment]

'[…]
all look at me, all adore me,
all give compliments to me,
all desire me, all want me,
all make a date with me.'

(3) O joia del meu cor, prenda (e)stimada,
la nit[i̲] que t'he dat aquesta rosa.

'Oh joy of my heart, my dear darling,
the night I have given you this rose.'

Since the 1990s, [i]-insertion has been considered to be categorical. Therefore it is not analyzed as motivated by purely low-level phonetic grounds, but is treated as a true phonological epenthesis (Loporcaro, 1997; Lloret & Jiménez, 2005, 2006, 2008, 2010)[6]; however, there are only two short tentative experimental studies by Ballone (2008, 2010), based on the same corpus, which compare the quality and the duration of the inserted vowel with those of lexical vowels[7].

[6] Bosch (2002, 2011), Ballone (2008, 2010), and Lloret & Jiménez (2010), who base their studies on spontaneous speeches and hence report more variability than in controlled elicitations, describe sporadic cases of consonant maintenance without vowel insertion; e.g. *Si me vol*[t͡s f] *fer aquesta caritat* 'If you want to do me this favor', but also *quant*[i̲] *la criatura era sola* 'when the child was alone', elicited by the same informant in the same text (Bosch, 2002:200); *jo no me recor*[t k]*uant* 'I do not remember when', but also *no li he dit*[i̲] *que* ...'I didn't tell him that ...', elicited by the same informant in the same text as well (Ballone, 2008:78-79).

[7] Since Ballone (2008) is an improved version of Ballone (2010), from now on we will refer exclusively to the 2008 work. More recently, in his dissertation (Ballone, 2013) he carried out a complete acoustic analysis of the vowels of Algherese Catalan, without paying attention to the specific characteristics of the inserted [i].

The present study has two main aims. On the one hand, we expand Ballone's (2008) study and analyze the acoustic traits that characterize the three different non-labial high vowels found in Algherese Catalan; namely, the lexical (underlying) stressed [i], the lexical (underlying) unstressed [i], and the inserted (epenthetic) unstressed [i], paying special attention to the comparison of the features of the two unstressed vowels. On the other hand, we investigate if there is a correlation between the acoustic characteristics of these vowels and the prominence they have according to their nature and the position in which they occur, in order to show that, for some properties, there is a gradual path going from the most salient element (i.e., lexical stressed [i]) to the most marginal one (i.e., inserted unstressed [i]).

2. METHOD

In this section, we first describe the corpus on which the analysis is based (section 2.1) and then present the criteria used for the segmentation of the phonetic units and the selection of the vowels targeted in the study (section 2.2).

2.1. Corpus

The data analyzed were extracted from an interview conducted in the city of Alghero in 1997, which was published in Viaplana & Perea (2003) as part of the *Corpus Oral Dialectal* (COD) (2003-2014) of the Universitat de Barcelona (available at http://www.ub.edu/cccub/corpusoraldialectal-cod.html; the text is also available online at http://diposit.ub.edu/dspace/handle/2445/11637). The COD material in-cludes the sound file as well as the phonetic transcription, which we have revised with the help of two native speakers. The informant is a housewife in her forties, born and raised in Alghero. The interview was recorded with a Digital Audio Tape at the informant's house and deals with common topics of her daily life, such as local feasts, food, and family; it is thus an example of semi-spontaneous speech. The sound file is 09'11'' long.

2.2. Selection and segmentation of the target vowels

As noted above, inserted vowels across words in Algherese Catalan only appear in open syllables, unlike lexical vowels, which can also occur in closed syllables, as in *fills* [ˈfiʎ͡ʦ] 'sons', or in the first syllable of *cinquanta* [siŋˈkwanta] 'fifty'. Hence, to homogenize the corpus, for the analysis we selected only non-labial high vowels

appearing in open syllables. To simplify matters, we omitted lexical vowels in hiatus, such as *pregadoria* [pɾegaɾuˈɾia] 'prayer' or *diem* [diˈem] '(we) say'. Although the text mostly contains declarative sentences, with no special emphasis, some segments were discarded because they were clearly pronounced with an expressive or emphatic lengthening –in particular, the words *sí* 'yes' and *i* 'and'. Furthermore, among the initial set of vowels, a statistical test was run to detect multivariate outliers, taking into account the five variables considered in the study: length, intensity, and the three first formants. As a result of this test, the stressed vowel in the word *(al)bergínia* 'aubergine' was identified as an outlier (length: 215 ms; Mahalanobis D^2=24.28, $p<0.001$) and was removed from the analysis. Finally, other vowels were excluded due to interference of several kinds; e.g., vowels emitted when the subject was laughing or while the interviewers were also talking. We eventually obtained a corpus of 174 items appearing in an equivalent syllabic context: 56 lexical stressed [i], 69 lexical unstressed [i], and 49 inserted unstressed [i][8].

The acoustic analysis was carried out with the help of the Praat software (Boersma & Weenink, 2014). Target vowels were segmented and labeled manually based on spectrogram and waveform. Then, a Praat script was used to automatically extract the following parameters: duration of the whole segment (in ms), and intensity (in dB) and the three first formants (F1, F2 and F3, in Hz) as measured at the center of the vowel. Due to the design of the study, we could not control the environment in which every single item was produced; hence, in order to reduce the influence of the neighboring consonants in the analysis, the formants of the vowels were measured at the midpoint of each segment, which is considered to be the closest to the vowel's target. Since the data are taken from a single speaker, we did not need to normalize the formant values to mitigate inter-speaker variation. As for the intensity, the differences between taking the value referring to the midpoint of the vowel or to the whole segment were so small that we decided to measure this parameter at the midpoint of the segment as well. (The values of these variables for each input vowel are summarized in table 1; the specific values corresponding to the targeted items are reported in Appendix 1).

[8] The main difference with respect to Ballone's (2008) work, which is also based on an interview with a single female informant (his subject was older than ours, but the interview was conducted in similar circumstances), is that his study analyzes a smaller set of elements: 7 lexical stressed [i] (2 in closed syllables, 5 in open syllables), 4 lexical unstressed [i] (1 in an open syllable and 3 in closed syllables), and 4 inserted unstressed [i] (in open syllables). However, he includes 11 tokens of lexical stressed [e] (6 in closed syllables, 5 in open syllables) and 5 lexical unstressed [e] (2 in closed syllables, 3 in open syllables) in the measurements for the sake of comparison.

	Input vowel		
Variables	lexical stressed	lexical unstressed	inserted unstressed
Length (ms)	105.86 (36.04)	57.77 (11.71)	48.94 (11.90)
F1 (Hz)	355.84 (45.78)	352.71 (41.42)	372.88 (33.05)
F2 (Hz)	2200.27 (100.95)	2087.55 (165.08)	1973.45 (156.37)
F3 (Hz)	2745.82 (158.03)	2675.41 (209.74)	2610.41 (132.70)
Intensity (dB)	69.54 (4.12)	70.67 (3.15)	69.14 (3.25)

Table 1. *Characterization of the three input vowels (the standard deviation of each variable is shown in parentheses beneath the variable means).*

3. RESULTS

In the following sections, two different approaches were used to investigate the similarities and the differences between the three input vowels. First, a cluster analysis of the whole dataset was performed in order to classify the input segments in groups (section 3.1). Second, a series of Kruskal-Wallis tests were run on the vowels surrounded by coronal consonants to isolate the variables for which the input segments differed (section 3.2).

3.1. First approach to the data: cluster analysis of the whole dataset

To explore the data, a two-step cluster analysis was conducted using SPSS, version 22.0 (IBM Corp, 2013), on all 174 vowels of the corpus. The aim of this first approach was twofold: on the one hand, to classify the vowels into homogeneous categories taking into account the five continuous variables previously obtained from the acoustic analysis, i.e., length, intensity, F1, F2, and F3; on the other, to

gain insight into the variables that are important in determining group membership. In order to avoid bias in the analysis due to the intrinsic organization of the data (with three different subsets of vowels in the input: lexical stressed, lexical unstressed, and inserted unstressed vowels), the cases were introduced in the dataset in the same order in which they appear in the interview (see Appendix 1), that is, randomly with respect to the input segment involved. The cluster analysis yields three relatively well defined groups, with 39 tokens in cluster 1 (22.41% of the tokens), 76 in cluster 2 (43.68%), and 59 in cluster 3 (33.91%), as shown in figure 1.

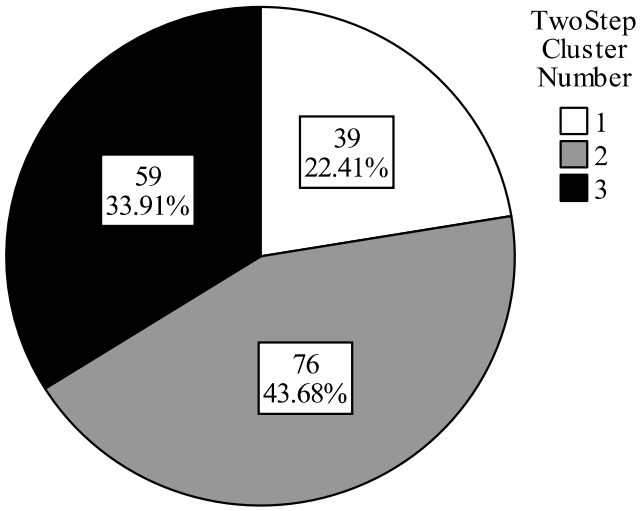

Figure 1. *Distribution and size of the clusters*.

In figure 2 we present the relative contribution of each variable to the clustering. The variable Length contributes the most to differentiating the three clusters, closely followed by F2. F1 is the least relevant variable, with Intensity in a slightly higher position. The variable F3 appears in an intermediate position in the ranking. However, variables F2 and F3 are strongly correlated ($r=0.581$, $p<0.001$); hence, the effects of F3 are likely to overlap with those of F2, so that the potential discriminatory power of F3 might be hidden by the variable F2.

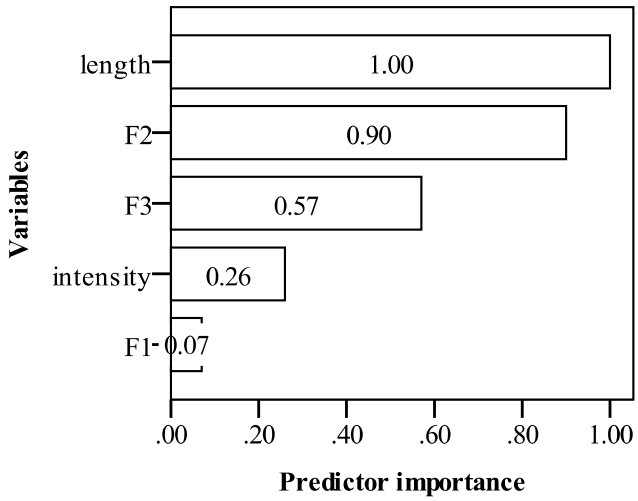

Figure 2. *Relative importance of each variable in estimating the model.*

The details of the three clusters are given in table 2. The first cluster, which fills 22.41% of the tokens, mostly contains long vowels with the highest values for F2 and F3. Cluster 2 consists of vowels whose F2 and F3 values are slightly lower than those of the vowels in cluster 1 and are considerably shorter than these segments. Finally, cluster 3 comprises the shortest segments, with the lowest values for F2 and F3 as well. As previously noted, Length is the best defining factor, though not by much; this small distance is mirrored in the fact that the duration is only the most important factor to differentiate clusters 1 and 3, whereas F2 is the most important cue to define cluster 2.

	Cluster		
Variables	1	2	3
Length (ms)	*118.74* *(34.88)*	60.43 (15.27)	*52.34* *(13.05)*
F2 (Hz)	2222.41 (100.74)	*2160.67* *(90.67)*	1916.44 (137.18)

F3 (Hz)	2795.97 (130.70)	2745.83 (134.94)	2517.85 (148.07)
Intensity (dB)	67.46 (3.87)	71.61 (2.54)	69.24 (3.38)
F1 (Hz)	343.56 (42.79)	358 (40.04)	371.66 (39.05)

Table 2. *Characterization of the three clusters (the standard deviation of each variable is shown in parentheses beneath the variable means; the factor that best defines each cluster is marked in italics).*

We conducted a chi-square test in order to assess the link between the three input vowels and the group membership defined by the model, that is, in order to demonstrate that the three input vowels are not distributed in the clusters at random. The results reveal a significant association between the variables Input vowel and Cluster ($\chi^2(4)=65.94$, $p<0.001$, Cramer's $V=0.435$). Indeed, as shown by the data in table 3 and figure 3, unstressed vowels (whether lexical or inserted), which are usually shorter than stressed vowels, tend to concentrate in clusters 2 and 3, with only a few items in cluster 1. In contrast, more than 50% of the stressed vowels (57.1%) belong to cluster 1, which groups the longest segments and the segments with highest values for F2 and F3.

		Input vowel			Total
		lexical stressed	lexical unstressed	inserted unstressed	
TwoStep Cluster Number	1	32 (57.1%)	5 (7.2%)	2 (4.1%)	39 (22.4%)
	2	18 (32.1%)	39 (56.5%)	19 (38.8%)	76 (43.7%)
	3	6 (10.7%)	25 (36.2%	28 (57.1%)	59 (33.9%)
Total		56 (100%)	69 (100%)	49 (100%)	174 (100%)

Table 3. *TwoStep Cluster Number / Input vowel Crosstabulation.*

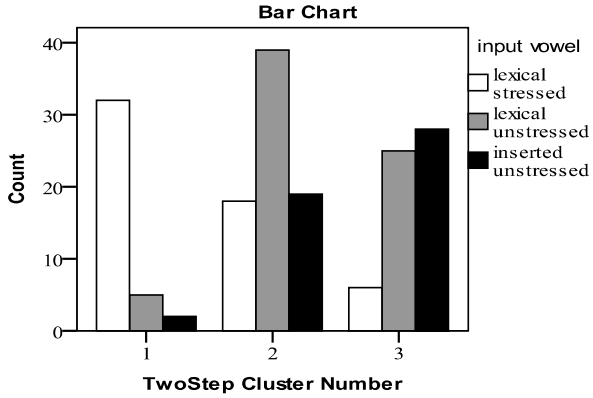

Figure 3. *Representation of the distribution of the three input vowels in the three clusters.*

We ran a second chi-square test to compare the distribution of the unstressed input vowels (lexical and inserted) in clusters 2 and 3. Again, there was a significant association between the variables Input vowel and Cluster, although the effect was clearly weaker ($\chi^2(1)=4.57$, p=0.033, Cramer's V=0.203). As figure 4 shows, there was a slightly greater concentration of lexical unstressed vowels in cluster 2 and of inserted unstressed vowels in cluster 3.

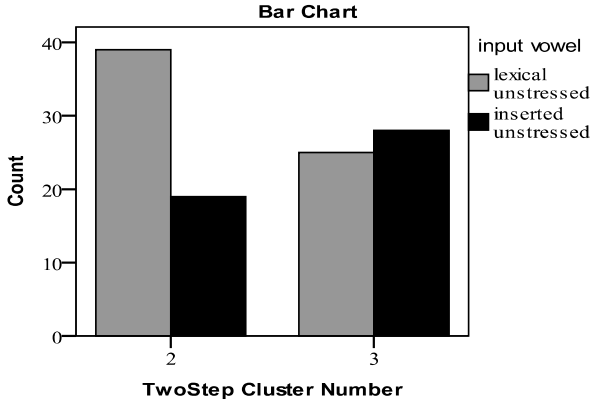

Figure 4. *Representation of the distribution of the unstressed input vowels among clusters 2 and 3.*

All in all, in this first approach to the data, the two-step cluster analysis identified three distinct clusters based mainly on the variables Length and F2. These clusters partially coincide with the three input vowels. So it seems that each segment presents specific acoustic features. Now that we have shown that the three input vowels tend to be treated as different segments by the cluster analysis, we will focus on the identification of the features that best define each unit.

3.2. Second approach to the data: Kruskal-Wallis tests of vowels surrounded by coronal consonants

We selected a subset of the original database to determine the features that characterize each segment and the features that all vowels share. This subset only contains the vowels that are surrounded by coronal consonants, with the aim of maximally unifying their segmental context (recall from section 2.2 that all vowels in the database appear in the same syllabic environment; i.e., an open syllable). The selected segments (60 in total, marked in italics in Appendix 1) include 26 lexical stressed, 14 lexical unstressed, and 20 inserted unstressed vowels.

We carried out a series of Kruskal-Wallis tests on this corpus, taking Input vowel (with three factors: lexical stressed, lexical unstressed, and inserted unstressed vowels) as independent variable and with the same measures studied in section 3.1 as dependent variables (i.e., Length, F1, F2, F3, and Intensity)[9]. (The mean values and the standard deviation of the variables Length, F1, F2, F3, and Intensity for each group of vowels are summarized in table 4). When the overall test yielded significant results (p=0.05), Mann-Whitney tests including all pairwise comparisons were run to follow up this finding. As usual in this kind of analysis, a Bonferroni correction was applied so that all effects are reported at a 0.0167 level of significance (one third of the standard level, p=0.05, since three different Mann-Whitney tests were performed with respect to each significant variable: one per each pair of vowels).

[9] A non-parametric approach was chosen in order to address the lack of normality in the distributions of the five parameters.

	Input vowel		
Variables	lexical stressed	lexical unstressed	inserted unstressed
Length (ms)	106.46 (34.08)	54.50 (10.35)	52.05 (13.61)
F1 (Hz)	348.69 (42.29)	347.79 (33.99)	368 (36.11)
F2 (Hz)	2215.46 (102.68)	2052.14 (102.15)	1903.90 (147.51)
F3 (Hz)	2814 (139.09)	2697 (134.67)	2643.45 (104.65)
Intensity (dB)	69.42 (3.99)	70.86 (3.82)	68.90 (3.46)

Table 4. *Mean of the variables Length, F1, F2, F3 and Intensity in the subset of vowels surrounded by coronal consonants (the standard deviation of each variable is shown in parentheses beneath the variable means).*

The Kruskal-Wallis test yielded non-significant results for the variables F1 ($H_{(2)}$=3.99, p=0.136; Overall mean=354.92 Hz, SD=38.98 Hz; see figure 5) and Intensity ($H_{(2)}$=5.13, p=0.077; Overall mean=69.58 dB, SD=3.79 dB; see figure 6). Although in the last case the value of *p* was close to the level of significance (p=0.05), we can ignore the effects associated with this variable because, as shown in table 4 and figure 6, the absolute differences in intensity between the input vowels were rather small and, therefore, not really relevant.

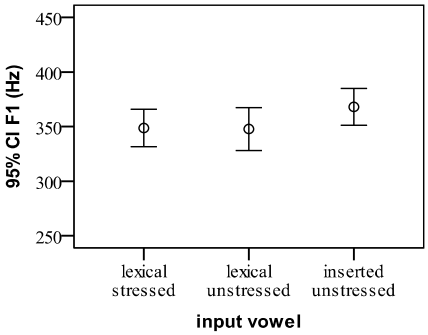

Figure 5. *F1, in Hz, of the input vowels surrounded by coronal consonants*.

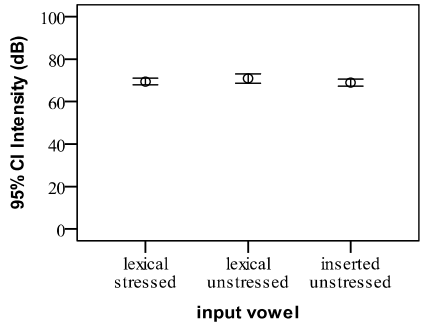

Figure 6. *Intensity, in dB, of the input vowels surrounded by coronal consonants.*

On the other hand, the analysis of the other three variables did produce significant results. To start with, there was a significant association of the variables Input vowel and Length ($H_{(2)}$=38.26, p<0.001). In the pairwise comparison, the Mann-Whitney tests indicated that, as expected, lexical stressed vowels (Mean=106.46 ms, SD=34.08 ms) appear to be longer than both lexical unstressed vowels (Mean=54.50 ms, SD=10.35 ms; U=11, p<0.001, r=–0.77) and inserted unstressed vowels (Mean=52.05 ms, SD=13.61 ms; U=19, p<0.001, r=–0.79). By contrast,

there were no significant differences between the two kinds of unstressed vowels in duration (U=104.5, p=0.213, r=–0.21), as figure 7 shows.

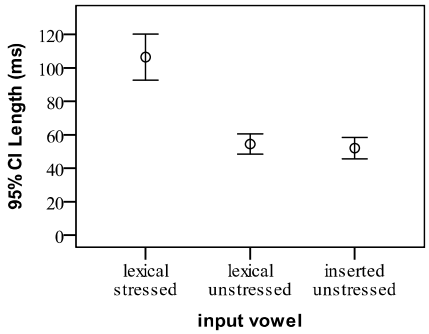

Figure 7. *Length, in ms, of the input vowels surrounded by coronal consonants.*

The test also yielded a significant effect of the variable Input vowel on the second formant of the segments ($H_{(2)}$=34.44, p<0.001). The follow-up Mann-Whitney tests revealed that the three comparisons are significant: the F2 value of the lexical stressed vowels (Mean=2215.46 Hz, SD=102.68 Hz) was higher than both the value of the lexical unstressed vowels (Mean=2052.14 Hz, SD=102.15 Hz; U=46, p<0.001, r=–0.61) and the value of the inserted unstressed vowels (Mean=1903.90 Hz, SD=147.51 Hz; U=27, p<0.001, r=–0.76); in turn, the value of the lexical unstressed vowels was higher than that of the inserted unstressed vowels (U=54, p=0.003; r=–0.52). Therefore, as figure 8 illustrates, there is a progressive decline in the F2 values of the input vowels, going from the lexical stressed segments on the left to the inserted unstressed vowels on the right.

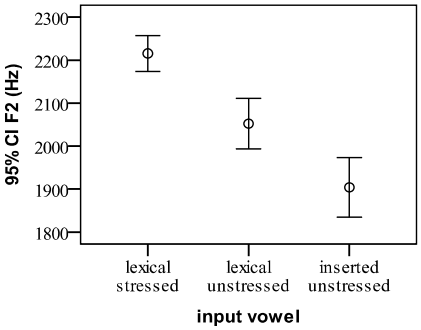

Figura 8. *F2, in Hz, of the input vowels surrounded by coronal consonants.*

There was also an effect of the variable Input vowel on the third formant of the segments ($H_{(2)}$=17.17, p<0.001). As with duration, the Mann-Whitney tests were significant when comparing the stressed vowels to the unstressed ones, with the lexical stressed vowels (Mean=2814 Hz, SD=139.09 Hz) displaying higher values than the lexical unstressed vowels (Mean=2697 Hz, SD=134.67 Hz; U=92, p=0.011, r=–0.40) and the inserted unstressed vowels (Mean=2643.45 Hz, SD=104.65 Hz; U=79, p<0.001, r=–0.59). The values of the lexical unstressed vowels and the inserted unstressed vowels, though, did not differ statistically (U=116, p=0.401, r=–0.14). These results are illustrated in figure 9, where we can see a descending pattern from the stressed vowels to the two unstressed vowels again, although these segments now present a similar F3 value.

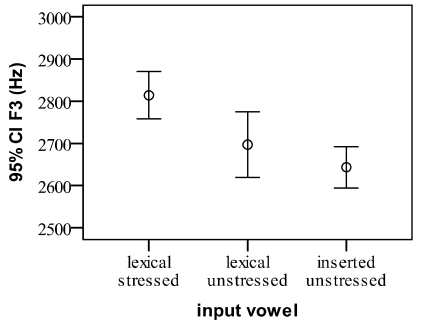

Figure 9. *F3, in Hz, of the input vowels surrounded by coronal consonants.*

To sum up, differences in the input vowels were found with respect to the two most important predictors in the cluster analysis, i.e., duration and F2. Additionally, we found differences with respect to the third formant, with moderate size effects, which confirm that the impact of this variable may be nuanced by the influence of F2 in the cluster analysis. In contrast, the role of the variables F1 and Intensity in defining the input segments was virtually irrelevant, as it was in the cluster analysis.

4. DISCUSSION

The present study compares the acoustic features of three non-labial high vowels in Algherese Catalan: lexical stressed, lexical unstressed, and inserted unstressed vowels. The two-step cluster analysis run on all the data and the Kruskal-Wallis tests performed on the subset of vowels surrounded by coronal consonants indicate that the variety maintains a triple contrast between these segments, mostly based on differences in length and F2 and, secondarily, on differences in F3. The group membership defined by the cluster analysis partially coincides with the distribution of input vowels, as can be seen by comparing the similarities between the two following scatter plots, which were obtained by crossing the variables F2 and Length referred to the three input vowels –in figure 10– and to the three clusters –in figure 11.

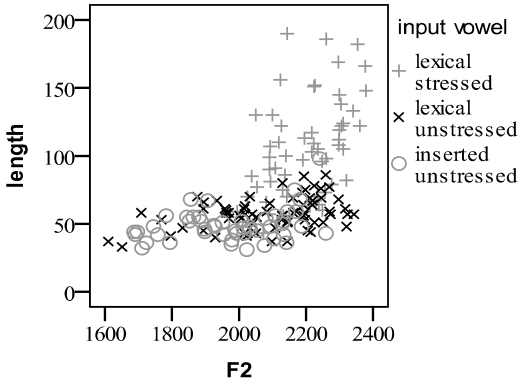

Figure 10. *Scatter plot of the three input vowels, crossing the variables F2, in Hz, and Length, in ms.*

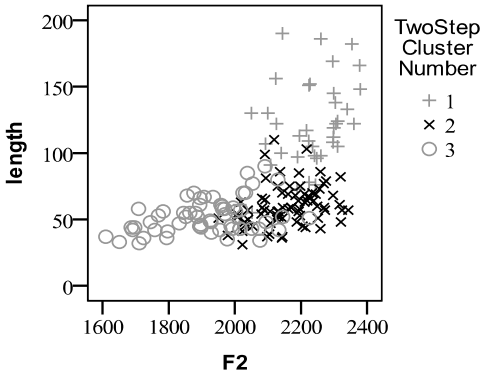

Figure 11. *Scatter plot of the three clusters, crossing the variables F2, in Hz, and Length, in ms.*

To shed light on the properties of the input vowels, we focused specifically on the subset of vowels surrounded by coronal consonants. Firstly, the results revealed that neither the openness of the vowels –their F1 value– nor their intensity were relevant for distinguishing the three input segments. Therefore, our data support the claim, already proposed by Ballone (2008), that the epenthetic segment appearing between certain consonants is a high vowel, exactly like lexical –stressed or unstressed– /i/'s studied here.

As for the contrasting features, our data also corroborated the progressive decline in the second formant values of the input segments, similar to that found in previous work (cf. Ballone, 2008). Thus, we found a pattern of gradual centralization (represented in figure 12) that goes from the lowest degree of centralization –i.e., the highest F2 values– in the lexical stressed segments, on the right, to the highest degree –i.e., the lowest F2 values– in the inserted unstressed vowels, on the left. The F3 values displayed a similar gradation, but in this case the two classes of unstressed segments had an equivalent F3 value, different from the higher value typical of the lexical stressed vowels.

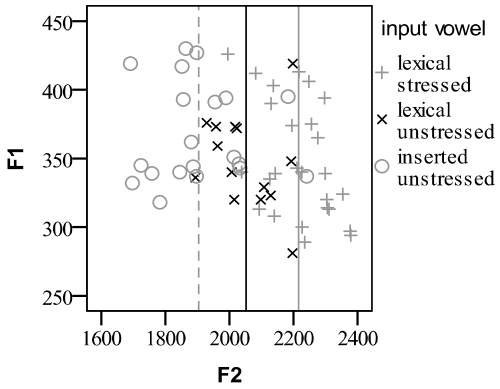

Figure 12. *Scatter plot of the input vowels surrounded by coronal consonants, crossing the variables F2 and F1, in Hz. The vertical lines indicate the mean F2 value for each input segment: lexical stressed, continuous grey line; lexical unstressed, continuous black line, and inserted unstressed, dashed grey line.*

Our findings on the duration of lexical (stressed and unstressed) vowels are also in accordance with earlier studies, which reported the stressed segments to be longer than the unstressed ones (see, e.g., Ballone, 2008). If we focused on lexical segments, the results would then support the standard hypothesis that longer segments are more likely to achieve their ideal value target, whereas shorter segments are more likely to deviate from that value (in our case, to be more centralized; cf. Lindblom, 1963).

However, contrary to previous studies, in which inserted unstressed vowels were described as shorter than lexical unstressed ones (cf. Ballone, 2008), our data did not show a significant difference between the duration of the two unstressed vowels. This finding, further supported by the fact that inserted unstressed vowels count for metrical parsing (see section 1), implies that the differences in F2 between the two unstressed segments cannot be exclusively attributed to differences in length between these vowels, since, as figure 13 shows, both kinds of unstressed segments have a similar duration[10]. That is, in the case of unstressed

[10] Interestingly enough, Hall (2013) also finds that F2 is the most reliable cue to distinguish between lexical and epenthetic vowels in Lebanese Arabic, although in a previous study by Gouskova & Hall (2009) duration contributed to reinforce the contrast as well.

vowels there does not seem to be any correlation between length and centralization[11].

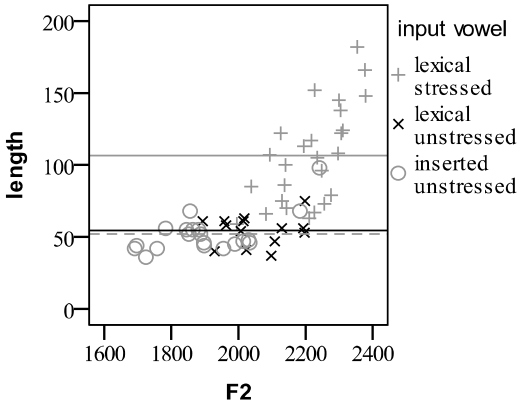

Figure 13. *Scatter plot of the input vowels surrounded by coronal consonants, crossing the variables F2, in Hz, and length, in ms. The horizontal lines indicate the mean duration of each input segment: lexical stressed, continuous grey line; lexical unstressed, continuous black line, and inserted unstressed, dashed grey line.*

In conclusion, the difference in F2 between the two unstressed segments suggests that it is the nature of the vowels, and not their duration, that determines the degree of centralization that they present. Therefore, according to our data, Algherese Catalan has an underlying high front vowel /i/, with two distinct realizations depending on the stress –namely, a lexical stressed /i/ and a lexical unstressed /i/– and, additionally, a second non-labial high vowel, which appears in epenthetic

[11] In fact, if we took the regression line correlating length and F2 in the lexical (stressed and unstressed) vowels to predict the F2 values of the inserted segments from their length, their predicted F2 values would be significantly higher than the actual ones. To address this issue, a paired t-test was carried out to compare the actual F2 value of the inserted vowels with the F2 value predicted by the correlation between length and F2 in the lexical vowels. The results show that, on average, the predicted F2 value (Mean=2076.66, SD=24.06) is clearly higher than the actual F2 value (Mean=1973.45, SD=156.37; $t(48)=-4.795$, $p<0.001$). (The predicted F2 values of the inserted vowels and the differences between their actual and their predicted F2 values are reported in Appendix 2.)

contexts and which does not seem to have a defined Front specification. As for the correlation between prominence and acoustic features, the gradation in the F2 values points to a double contrast as well: firstly, between lexical stressed and unstressed segments, with higher values in the more prominent vowels (i.e., the stressed ones), and, secondly, between unstressed segments, with higher values in the lexical vowels, which are relatively more prominent due to their underlying (and more stable) nature.

ACKNOWLEDGEMENTS: This research was supported by the Spanish Ministerio de Economía y Competitividad (project number FFI2013-46987-C3-1-P: www.ub.edu/GEVAD) and by the Catalan Government (research group 2014SGR918: www.ub.edu/GEV); the authors are also members of the research group of the Universitat de València GIUV2013-137. We thank Francesco Ballone and Luca Scala for their help with the interpretation of some of the data, as well as the audience at the VI Congreso Internacional de Fonética Experimental (5-7 November 2014, Universitat de València) for providing useful suggestions. We are also very grateful to Ricard Herrero for his help with the Praat scripts, to José-Ignacio Hualde for valuable comments on a first draft of the work, and to two anonymous reviewers for their insightful remarks.

5. REFERENCES

BALLONE, F. (2008): «Català de l'Alguer: anàlisi instrumental d'un text oral», *Insula: Quaderno di cultura sarda*, 4, pp. 71-89.

BALLONE, F. (2010): «Estudi instrumental sobre la qüestió de la vocal epentètica del català de l'Alguer», *Interlingüística*, XX. http://filcat.uab.cat/clt/XXIVAJL/Interlinguistica/Encuentro%20XXIV/Ballone_REVF.pdf [21/12/2014].

BALLONE, F. (2013): *An Acoustic Study of Sardinian and Algherese Catalan Vowels*, doctoral dissertation, Bellaterra, Universitat Autònoma de Barcelona.

BOERSMA, P. & D. WEENINK (2014): *Praat: doing phonetics by computer [Computer program]*, version 5.4.02. http://www.fon.hum.uva.nl/praat/ [26/11/2014].

BOSCH I RODOREDA, A. (2002): *El català de l'Alguer*, Barcelona, Publicacions de l'Abadia de Montserrat.

BOSCH I RODOREDA, A. (2011): «Algunes consideracions a l'entorn de la distribució i la sonoritat de les consonants africades en alguerès», in M.-R. Lloret & C. Pons (eds.): *Noves aproximacions a la fonologia i la morfologia del català*, Alacant, Institut Interuniversitari de Filologia Valenciana, pp. 63-76.

CABRERA, M. (2013): «*Tot va bé si acaba bé*: Aplecs consonàntics finals en la flexió nominal patrimonial algueresa», in E. Clua & M.-R. Lloret (eds.): *Qüestions de morfologia flexiva i lèxica del català*, Alacant, Institut Interuniversitari de Filologia Valenciana, pp. 63-87.

CATARDI, R. (1971): *Rimes alguereses*, Barcelona, Barcino.

COD (2003-2014): *Corpus Oral Dialectal*, Barcelona, Departament de Filologia Catalana, Universitat de Barcelona.
http://www.ub.edu/ccub/corpusoraldialectal-cod.html [21/12/2014]
File from Algherese Catalan available at:
http://diposit.ub.edu/dspace/handle/2445/11637 [21/12/2014]

GOUSKOVA, M. & N. HALL (2009): «Acoustics of epenthetic vowels in Lebanese Arabic», in S. Parker (ed.): *Phonological Argumentation. Essays on evidence and motivation*, London, Equinox, pp. 203-225.

HALL, N. (2006): «Cross-linguistic patterns of vowel intrusion», *Phonology*, 23, pp. 387-429.

HALL, N. (2013): «Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic», *Journal of Phonetics*, 41, pp. 133-143.

IBM Corp. (2013): *IBM Statistics for Windows*, version 22.0, Amonk, New York, IBM Corp.

JONES, M. (1988): «Sardinian», in M. Harris & N. Vincent (eds.): *The Romance languages*, London, Routledge, pp. 314-350.

KUEN, H. (1932): «El dialecto de Alguer y su posición en la historia de la lengua catalana [1]», *Anuari de l'Oficina Romànica de Lingüística i Literatura*, V, pp. 121-177.

KUEN, H. (1934): «El dialecto de Alguer y su posición en la historia de la lengua catalana [2]», *Anuari de l'Oficina Romànica de Lingüística i Literatura*, VII, pp. 41-112.

LEVIN, J. (1987): «Between epenthetic and excrescent vowels (or what happens after redundancy rules)», in M. Crowhurst (ed.): *West Coast Conference on Formal Linguistics*, 6, Stanford, Stanford Linguistic Association, pp. 187-201.

LINDBLOM, B. (1963): «Spectrographic study of vowel reduction», *Journal of the Acoustic Society of America*, 35, pp. 1773-1781.

LLORET, M.-R. & J. JIMÉNEZ (2005): «Two kinds of vowel epenthesis in Alguerese Catalan», Paper presented at *2nd Phonetics and Phonology in Iberia (Bellaterra, 20-21 juny 2005)*, Bellaterra, Universitat Autònoma de Barcelona.

LLORET, M.-R. & J. JIMÉNEZ (2006): «Prominence-driven epenthesis: evidence from Catalan», Ms., Universitat de Barcelona and Universitat de València. http://roa.rutgers.edu, n. 806 [21/12/2014].

LLORET, M.-R. & J. JIMÉNEZ (2008): «Marcatge posicional i prominència en el vocalisme àton», *Caplletra*, 45, pp. 55-91.

LLORET, M.-R. & J. JIMÉNEZ (2010): «Sobre les vocals afegides de l'alguerès», *Insula: Quaderno di cultura sarda*, 7, pp. 111-139.

LOPORCARO, M. (1997): «On vowel epenthesis in Alguer Catalan», in P. Bertinetto *et al.* (ed.): *Certamen Phonologicum III. Papers from the Third Cortona Phonology Meeting, April 1996*, Torino, Rosenberg & Sellier, pp. 205-227.

RECASENS, D. (1991): *Fonètica descriptiva del català. (Assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX)*, Barcelona, Institut d'Estudis Catalans.

TORRES-TAMARIT, F.; K. LINKE & M. M. VANRELL (forthcoming): «Opacity in Campidanian Sardinian: Metaphony and copy-vowel epenthesis», *Natural Language & Linguistic Theory*.

VIAPLANA, J. & M. P. PEREA (2003): *Textos orals dialectals del català sincronitzats. Una selecció*, Barcelona, PPU.

APPENDIX I. DATASET, CONTAINING ALL TOKENS

Vowels of the corpus, ordered (*Case* column) as they appear in the interview. The second column (*Word / Context*) displays the corresponding orthographic form; targeted lexical vowels are indicated in capital letters, and inserted vowels between square brackets. The phonetic context surrounding the vowel is indicated in the third column (*IPA*). The number in the fourth column (*IV=Input vowel*) refers to the kind of vowel in the input: lexical stressed vowel, 1; lexical unstressed vowel, 2, and inserted unstressed vowel, 3. The number in the fifth column (*Cl.=Cluster*) indicates the cluster to which the vowel is assigned in the analysis. For each segment, the following variables are reported: length of the whole segment (in ms), as well as the intensity (in dB) and the three first formants (F1, F2, and F3, in Hz) as measured at the center of the vowel. The items marked in italics correspond to the vowels surrounded by coronal consonants that yield the subset of segments analyzed in section 3.2.

Case	Word / Context	IPA	IV	Cl.	ms	F1	F2	F3	dB
1	bellÍsim	[ʎis]	1	2	81	306	2094	2790	69
2	PòllIna	[ʎin]	2	2	50	329	2040	2922	75
3	PòllIna	[ʎin]	2	3	58	426	1969	2775	67
4	*tens [i] de*	*[t͡sid]*	*3*	*3*	*36*	*345*	*1724*	*2634*	*68*
5	*localItat*	*[lit]*	*2*	*3*	*58*	*359*	*1963*	*2483*	*70*
6	localitat [i] que és	[tik]	3	3	34	404	2076	2485	68
7	PòllIna	[ʎin]	2	2	59	353	2083	3035	71
8	paisagístIc i	[tik]	2	2	85	311	2194	2672	73
9	[*Pause*] I tu	[it]	2	2	86	324	2259	2795	75
10	*tens [i] d'anar*	*[t͡sid]*	*3*	*3*	*44*	*332*	*1697*	*2639*	*71*
11	*vuit [i] de*	*[tid]*	*3*	*2*	*48*	*346*	*2030*	*2607*	*75*
12	mIssa	[mis]	1	2	103	360	2217	2778	72
13	històrIc i	[ɾik]	2	2	57	441	2039	2684	73
14	històric [i] de	[kid]	3	3	48	413	1928	2535	71
15	*costalers [i] se*	*[t͡sis]*	*3*	*3*	*42*	*419*	*1691*	*2720*	*69*
16	vIgília	[vid͡ʒ]	2	2	58	283	2162	2465	72
17	vigÍlia	[d͡ʒiɾ]	1	1	169	301	2296	2567	72
18	allÍ [*Pause*]	[ʎi]	1	1	190	372	2144	2821	73

19	prImera	[pɾim]	2	3	33	372	1651	2199	69
20	tengut [i] campanyes	[tik]	3	2	45	382	2054	2579	75
21	*presIdent*	*[zid]*	*2*	*3*	*61*	*336*	*1894*	*2573*	*76*
22	[*Pause*] I doncs	[id]	2	2	59	315	2270	2746	70
23	*record I tu*	*[tit]*	*2*	*2*	*53*	*281*	*2197*	*3020*	*72*
24	tengarIves	[ɾiv]	1	2	70	329	2234	2657	73
25	vint-I-quatre	[tik]	2	3	45	343	1894	2378	70
26	*sÍ, sant*	*[sis]*	*1*	*1*	*113*	*374*	*2195*	*2803*	*66*
27	mIg agost	[mit͡ʃ]	1	2	56	316	2132	2464	73
28	agost [i] que és	[tik]	3	3	42	334	2133	2573	68
29	*agost I lego*	*[til]*	*2*	*3*	*61*	*373*	*1958*	*2579*	*63*
30	*vuit [ɪ] són*	*[tis]*	*3*	*3*	*55*	*340*	*1845*	*2507*	*67*
31	*de sòlIta* (It.)	*[lit]*	*2*	*2*	*63*	*373*	*2018*	*2616*	*72*
32	*focs (de) [ɪ] Sant*	*[t͡sis]*	*3*	*3*	*68*	*393*	*1856*	*2655*	*70*
33	*comarats [ɪ] nous*	*[t͡sin]*	*3*	*3*	*46*	*337*	*1897*	*2423*	*68*
34	tot [i] fan	[tif]	3	3	36	410	1794	2479	64
35	*la nIt [ɪ] del*	*[nit]*	*1*	*2*	*75*	*390*	*2130*	*2718*	*82*
36	*la nit [ɪ] del*	*[tid]*	*3*	*3*	*44*	*427*	*1898*	*2702*	*76*
37	gIten	[d͡ʒit]	1	2	64	306	2238	2594	78
38	l'augurI que	[ɾik]	2	3	70	391	2032	2437	77
39	donguI providència	[gip]	2	3	60	400	1962	2202	74
40	se gItaven	[d͡ʒit]	2	3	56	328	2002	2516	74
41	emparant *(It.)* [i] que	[tik]	3	2	54	368	2080	2641	75
42	*anant [ɪ] tornar*	*[tit]*	*3*	*3*	*56*	*318*	*1783*	*2516*	*70*
43	sÍ [*Pause*]	[si]	1	2	110	342	2119	2730	73
44	sÍ [*Pause*]	[si]	1	1	91	372	2108	2738	67
45	temps [i] bo	[t͡sib]	3	2	52	330	2008	2708	71
46	*no se pot [ɪ] tanta*	*[tit]*	*3*	*3*	*55*	*430*	*1864*	*2449*	*67*
47	pIló	[piɾ]	2	3	57	324	2018	2322	66
48	*casadInes*	*[ɾin]*	*1*	*1*	*122*	*335*	*2126*	*2765*	*72*
49	*casadInes*	*[ɾin]*	*1*	*1*	*96*	*406*	*2248*	*2801*	*68*

50	casadines I casquetes	[zik]	2	2	53	376	2128	2682	70
51	tÍpic [i]	[tip]	1	1	105	321	2311	2823	70
52	típIc [i]	[pik]	2	3	43	323	2051	2474	66
53	típic [i] dolç	[kid]	3	2	43	339	2259	2747	67
54	*dolç [i] de*	*[t͡sid]*	*3*	*1*	*47*	*351*	*2014*	*2704*	*61*
55	*casadInes*	*[ɾin]*	*1*	*1*	*145*	*339*	*2299*	*2781*	*71*
56	[*Pause*] I les	[il]	2	1	76	351	2244	2931	68
57	*Tots los Sants [i] los*	*[t͡sil]*	*3*	*3*	*45*	*394*	*1989*	*2721*	*68*
58	*pabassInos*	*[sin]*	*1*	*1*	*122*	*314*	*2306*	*2611*	*68*
59	*pabassInos*	*[sin]*	*1*	*1*	*138*	*320*	*2304*	*2766*	*70*
60	*galetInes*	*[tin]*	*1*	*1*	*105*	*289*	*2235*	*3054*	*62*
61	blanc I les	[kil]	2	3	53	313	1769	2639	67
62	*facilÍssim*	*[ris]*	*1*	*3*	*85*	*340*	*2038*	*2568*	*67*
63	*casadInes*	*[ɾin]*	*1*	*1*	*108*	*394*	*2297*	*2850*	*66*
64	t'explIc [i]	[pik]	1	3	77	454	2054	2511	64
65	t'explic [i] lego	[kil]	3	2	52	389	2132	2678	70
66	*casadInes*	*[ɾin]*	*1*	*2*	*66*	*412*	*2082*	*2819*	*77*
67	[*Pause*] I sigui	[is]	2	2	61	270	2317	2836	71
68	siguI de	[gid]	2	2	69	390	2211	2696	72
69	*meitat [i] de*	*[tid]*	*3*	*2*	*68*	*395*	*2183*	*2770*	*72*
70	*meitat [i] de*	*[tid]*	*3*	*3*	*42*	*391*	*1955*	*2681*	*68*
71	condIment	[dim]	2	3	70	364	1876	2663	71
72	condiment hI va	[tiv]	2	2	58	428	2196	2816	69
73	va tot [i] ben	[tib]	3	2	59	365	2170	2774	69
74	macInato *(It.)*	[t͡ʃin]	2	2	44	350	2214	3111	70
75	[*Pause*] I tot	[it]	2	2	68	363	2297	2978	70
76	tot [i] condit	[tik]	3	3	48	344	2094	2495	70
77	*condIt així*	*[dit]*	*1*	*2*	*67*	*340*	*2226*	*2826*	*71*
78	part [i] prepar	[tip]	3	2	36	411	2142	2609	71
79	*dIre*	*[diɾ]*	*1*	*2*	*79*	*365*	*2276*	*3117*	*70*
80	quIlo	[kiɾ]	1	2	60	360	2249	2663	71

81	*farIna*	*[ɾin]*	*1*	*1*	*148*	*294*	*2379*	*3056*	*66*
82	[*Pause*] hI pos	[ip]	2	2	64	323	2192	2793	72
83	*olI de*	*[ɾid]*	*2*	*2*	*54*	*340*	*2007*	*2693*	*73*
84	[*Pause*] I lego	[il]	2	2	57	324	2273	2834	70
85	[*Pause*] I lego	[il]	2	2	48	379	2321	2756	71
86	tenc [i] la	[kil]	3	2	31	404	2023	2725	70
87	maquIneta	[kin]	2	2	57	343	2343	2973	69
88	formes I fas	[zif]	2	3	47	348	1832	2689	65
89	*casadInes*	*[ɾin]*	*1*	*1*	*124*	*313*	*2311*	*2992*	*68*
90	tenc [i] lo	[kil]	3	3	35	355	1977	2497	70
91	tenc [i] lo	[kil]	3	3	43	352	2016	2368	70
92	*marIt [i] prepara*	*[ɾit]*	*1*	*1*	*100*	*308*	*2140*	*2764*	*67*
93	marit [i] prepara	[tip]	3	2	56	339	2107	2645	68
94	[*Pause*] I lego	[il]	2	2	57	425	2324	2786	74
95	[*Pause*] I lego	[il]	2	3	80	474	2130	2512	71
96	blanc [i] que	[kik]	3	2	75	378	2166	2744	72
97	*casadInes*	*[ɾin]*	*1*	*1*	*107*	*313*	*2093*	*2678*	*64*
98	blanc [i] pròprio *(It.)*	[kip]	3	2	48	351	2187	2771	70
99	mIdó	[miɾ]	2	2	77	319	2269	2611	74
100	*en Italià*	*[nit]*	*2*	*2*	*56*	*348*	*2193*	*2777*	*75*
101	llImó	[ʎim]	2	2	69	350	2161	3162	71
102	coagulat [i] ve	[tiv]	3	2	50	369	2128	2751	71
103	[*Pause*] I posada	[ip]	2	2	45	362	2205	2703	71
104	[*Pause*] I són	[is]	2	1	73	284	2189	2908	61
105	CarraixalI fem	[ɾif]	2	2	65	383	2092	2696	73
106	brInyols	[bɾiɲ]	2	3	66	378	1895	2609	70
107	brinyols [i] mosaltros	[t͡sim]	3	3	67	378	1907	2650	72
108	*farIna*	*[ɾin]*	*1*	*1*	*182*	*324*	*2354*	*3030*	*68*
109	rIbell	[rib]	2	3	41	384	1797	2492	71
110	*emparat* (It.) *[i] sempre*	*[tis]*	*3*	*3*	*52*	*417*	*1852*	*2608*	*70*
111	[*Pause*] sÍ [*Pause*]	[si]	1	1	109	370	2223	2803	66

112	*farIna*	*[rin]*	*1*	*2*	*73*	*375*	*2256*	*2799*	*71*
113	rIbell	[rib]	2	3	37	342	1610	2318	73
114	lievito *(It.)*	[vit]	2	3	52	313	2145	2421	71
115	bIrra	[bir]	1	1	151	298	2224	2483	71
116	l'anIce *(It.)*	[nit͡ʃ]	2	1	78	382	2224	2789	67
117	*escurrIda*	*[rir]*	*1*	*1*	*166*	*297*	*2377*	*2859*	*69*
118	*escurrIda*	*[rir]*	*1*	*1*	*152*	*300*	*2227*	*2768*	*71*
119	*posant-hI-li*	*[tir]*	*2*	*2*	*75*	*419*	*2198*	*2787*	*71*
120	dIventar	[div]	2	2	64	361	2219	2832	69
121	te puc [i] dire	[kid]	3	3	49	347	1925	2345	70
122	*dIre*	*[dir]*	*1*	*2*	*70*	*339*	*2143*	*2756*	*69*
123	quant [i] passa	[tip]	3	2	39	412	2104	2694	69
124	bollIr [*Pause*]	[ʎi]	1	1	186	396	2260	2779	73
125	[*Pause*] I tu fas	[it]	2	2	67	364	2220	2821	71
126	brInyols	[briɲ]	2	3	67	307	1934	2533	69
127	brInyols	[briɲ]	2	3	58	407	1709	2182	70
128	mIra	[mir]	1	2	57	495	2089	2785	73
129	a mI [*Pause*]	[mi]	1	1	156	459	2124	2824	70
130	[*Pause*] sÍ [*Pause*]	[si]	1	1	130	328	2100	2884	75
131	famÍlia	[mir]	1	1	119	395	2299	2868	72
132	mInyones	[miɲ]	2	2	37	370	2144	2789	74
133	*fadrInes*	*[rin]*	*1*	*1*	*117*	*413*	*2217*	*2811*	*72*
134	[*Pause*] I lo	[il]	2	2	69	316	2245	2573	69
135	mInyó	[miɲ]	2	3	51	319	2225	2403	71
136	fIlosofia	[fil]	2	3	43	341	2016	2459	73
137	[*Pause*] I l'altra	[il]	2	2	51	324	2144	2877	68
138	bravÍssims	[vis]	1	2	99	339	2091	2588	73
139	fills [i] bravos	[t͡sib]	3	3	32	380	1711	2449	63
140	prImer	[prim]	2	3	54	404	1970	2463	77
141	tots [i] mos	[t͡sim]	3	2	51	431	1951	2735	73
142	mIg [i] de	[mit͡ʃ]	1	3	70	390	2025	2298	73

143	mig [i] de	[t͡ʃid]	3	2	38	351	1979	2657	73
144	*verd [i] del*	*[tid]*	*3*	*1*	*98*	*337*	*2241*	*2747*	*66*
145	*sIlenci*	*[sil]*	*2*	*2*	*37*	*320*	*2098*	*2763*	*69*
146	*silencI de*	*[t͡sid]*	*2*	*2*	*56*	*323*	*2129*	*2614*	*71*
147	*tranquil·lItat*	*[lit]*	*2*	*2*	*47*	*329*	*2108*	*2707*	*74*
148	[*Pause*] I perquè	[ip]	2	2	59	426	2171	2687	72
149	salut [i] viure	[tiv]	3	2	59	343	2145	2673	66
150	aquÍ [*Pause*]	[ki]	1	1	133	362	2340	2933	67
151	*marIt és*	*[rit]*	*1*	*2*	*86*	*403*	*2137*	*2593*	*73*
152	famÍlia	[miɾ]	1	1	97	354	2190	2581	62
153	*tens [i] també*	*[t͡sit]*	*3*	*3*	*42*	*339*	*1758*	*2827*	*63*
154	dIfendre	[dif]	2	1	55	290	2052	2763	65
155	mIra	[miɾ]	1	1	98	413	2260	2812	67
156	*dIre avui*	*[diɾ]*	*1*	*2*	*63*	*343*	*2211*	*2785*	*70*
157	só eixIda	[ʃiɾ]	1	3	43	360	2073	2656	68
158	*francs I só tornada*	*[t͡sis]*	*2*	*1*	*61*	*320*	*2015*	*2830*	*63*
159	te puc [i] dire	[kid]	3	3	44	398	1687	2260	65
160	*dIre amb*	*[diɾ]*	*1*	*3*	*59*	*426*	*1995*	*2794*	*67*
161	un poc [i] de	[kid]	3	3	43	394	1993	2417	65
162	un poc [i] de	[kid]	3	3	48	440	1745	2357	64
163	dIfícil	[dif]	2	2	56	344	2131	2577	71
164	difÍcil	[fis]	1	3	90	376	2091	2532	69
165	*difícIl especialment*	*[sil]*	*2*	*3*	*40*	*376*	*1929*	*2618*	*72*
166	*quant [i] los joves*	*[til]*	*3*	*3*	*52*	*344*	*1887*	*2623*	*71*
167	dIpenent	[dip]	2	2	51	314	2251	2860	71
168	*doncs [i] tens*	*[t͡sit]*	*3*	*2*	*46*	*343*	*2033*	*2696*	*70*
169	*tens [i] de*	*[t͡sid]*	*3*	*3*	*55*	*362*	*1881*	*2640*	*68*
170	*tIrar*	*[tiɾ]*	*2*	*2*	*41*	*372*	*2023*	*2698*	*71*
171	equilIbre	[lib]	1	1	112	325	2298	2787	67
172	fIlla	[fiʎ]	1	2	82	377	2320	2574	68
173	llIbres	[ʎib]	1	1	122	381	2360	2772	59

174	aixÍ [*Pause*]	[ʃi]	1	1	130	304	2050	2507	61

Glosses. *1, 'very beautiful'; 2, 3, 7, '(a place name)'; 4, 169, '(you) have to'; 5, 'place'; 6, 'place that is'; 8, 'with beautiful scenery and'; 9, 'and you'; 10, '(you) have to go'; 11, 'eighth of'; 12, 'Mass'; 13, 'historic'; 14, 'historic of'; 15, 'inhabitants them.REFL'; 16, 17, 'eve'; 18, 'there'; 19, 'first.FEM.SG'; 20, 'had seasons'; 21, 'president'; 22, 'and then'; 23, '(I) remember and you'; 24, '(you) should'; 25, '24'; 26, 'yes, Saint'; 27, 'mid August'; 28, 'August, which is'; 29, 'August and then'; 30, 'eighth are'; 31, 'often'; 32, 'Saint [John]'s fireworks'; 33, 'new sisterhood members'; 34, 'all, (they) do'; 35, 36, 'the night of'; 37, '(they) throw'; 38, 'the wish that'; 39, '(it) lets'; 40, '(they) were thrown'; 41, 'learning that'; 42, 'going back'; 43, 44, 111, 130, 'yes'; 45, 'good weather'; 46, '(it) cannot so much'; 47, '(weight measure)'; 48, 49, 50, 55, 63, 66, 89, 97, '(typical sweets)'; 51, 52, 53, 'typical'; 53, 'typical sweet'; 54, 'sweet of'; 56, 'and the.FEM.PL'; 57, 'All Saints Day the.MASC.PL'; 58, 59, '(typical sweet)'; 60, 'little cookies'; 61, 'white and the.FEM.PL'; 62, 'very easy'; 64, '(I) will explain'; 65, '(I) will explain later'; 67, 'and either'; 68, 'either of'; 69, 70, 'half of'; 71, 'seasoning'; 72, 'seasoning goes'; 73, '(it) all goes well'; 74, 'grounded'; 75, 'and everything'; 76, 'everything seasoned'; 77, 'seasoned like this'; 78, '[for my] part (I) prepare'; 79, 122, 'to say'; 80, 'kg'; 81, 108, 112, 'flour'; 82, '(I) add'; 83, 'oil of'; 84, 85, 94, 95, 'and then'; 86, '(I) have the.FEM.SG'; 87, '(kitchen appliance)'; 88, 'forms and (you) do'; 90, 91, '(I) have the.MASC.SG'; 92, 93, 'husband prepares'; 96, 'white that'; 98, 'white just'; 99, 'starch'; 100, 'in Italian'; 101, 'lemon'; 102, 'thickened, (it) comes'; 103, 'and put.FEM.SG'; 104, 'and (they) are'; 105, 'Carnival (we) do'; 106, 126, 127, '(kind of donut)'; 107, '(kind of donut) we'; 109, 113, '(kind of bowl)'; 110, 'learned always'; 114, 'yeast'; 115, 'beer'; 116, 'the aniseed'; 117, 118, 'drained'; 119, 'putting it there'; 120, 'to become'; 121, 159, '(I) can say'; 123, 'when (it) passes'; 124, 'to boil'; 125, 'and you do'; 128, 155, 'look'; 129, 'to me'; 131, 152, 'family'; 132, 'girls'; 133, 'single'; 134, 'and the.MASC.SG'; 135, 'boy'; 136, 'philosophy'; 137, 'and the other.FEM.SG'; 138, 'very good'; 139, 'good sons'; 140, 'first.MASC.SG'; 141, 'all us.REFL'; 142, 143, '[in the] middle of'; 144, 'green, of'; 145, 'silence'; 146, 'silence of'; 147, 'calm'; 148, 'and because'; 149, 'health living'; 150, 'here'; 151, 'husband is'; 153, '(you) have as well'; 154, 'to fend for'; 156, 'to say today'; 157, '(I) went out'; 158, 'francs and (I) came back'; 160, 'to say, with'; 161, 162, 'a little of'; 163, 164, 'difficult'; 165, 'difficult especially'; 166, 'when the youth'; 167, 'depending on'; 168, 'then (you) have'; 170, 'to manage'; 171, 'balance'; 172, 'daughter'; 173, 'books'; 174, 'that way'.*

APPENDIX II. INSERTED VOWELS, REAL AND PREDICTED F2 VALUE

Inserted vowels, ordered (*Case* column) as they appear in the interview. The second column (*Word / Context*) displays the corresponding orthographic form; targeted inserted vowels appear between square brackets. The numbers in the third column (*Length*) and in the fourth column (*Actual F2*) indicate the real values of duration, in ms, and F2, in Hz, of the vowel respectively. The number in the next column (*Predicted F2*) refers to the F2 value, in Hz, predicted by the linear equation correlating F2 and Length in all lexical vowels (linear equation: Intercept=1977.71; B=2.02). The last column reports the difference between the actual F2 and the predicted F2. In the last two lines, we have included the mean values and the standard deviation of the variables Length, F2, Predicted F2, and Difference (F2 – Predicted F2) in the subset of inserted vowels.

Case	Word / Context	Length	Actual F2	Predicted F2	Difference (Actual F2–Predicted F2)
4	tens [i] de	36	1724	2050.50	–326.50
6	localitat [i] que és	34	2076	2046.45	29.55
10	tens [i] d'anar	44	1697	2066.67	–369.67
11	vuit [i] de	48	2030	2074.76	–44.76
14	històric [i] de	48	1928	2074.76	–146.76
15	costalers [i] se	42	1691	2062.63	–371.63
20	tengut [i] campanyes	45	2054	2068.70	–14.70
28	agost [i] que és	42	2133	2062.63	70.37
30	vuit [i] són	55	1845	2088.92	–243.92
32	focs (de) [i] Sant	68	1856	2115.20	–259.20
33	comarats [i] nous	46	1897	2070.72	–173.72
34	tot [i] fan	36	1794	2050.50	–256.50
36	la nit [i] del	44	1898	2066.67	–168.67
41	emparant *(It.)* [i] que	54	2080	2086.89	–6.89
42	anant [i] tornar	56	1783	2090.94	–307.94
45	temps [i] bo	52	2008	2082.85	–74.85
46	no se pot [i] tanta	55	1864	2088.92	–224.92
53	típic [i] dolç	43	2259	2064.65	194.35
54	dolç [i] de	47	2014	2072.74	–58.74
57	Tots los Sants [i] los	45	1989	2068.70	–79.70

65	t'explic [i] lego	52	2132	2082.85	49.15
69	meitat [i] de	68	2183	2115.20	67.80
70	meitat [i] de	42	1955	2062.63	–107.63
73	va tot [i] ben	59	2170	2097.00	73.00
76	tot [i] condit	48	2094	2074.76	19.24
78	part [i] prepar	36	2142	2050.50	91.50
86	tenc [i] la	31	2023	2040.39	–17.39
90	tenc [i] lo	35	1977	2048.48	–71.48
91	tenc [i] lo	43	2016	2064.65	–48.65
93	marit [i] prepara	56	2107	2090.94	16.06
96	blanc [i] que	75	2166	2129.36	36.64
98	blanc [i] pròprio (*It.*)	48	2187	2074.76	112.24
102	coagulat [i] ve	50	2128	2078.81	49.19
107	brinyols [i] mosaltros	67	1907	2113.18	–206.18
110	emparat (*it.*) [i] sempre	52	1852	2082.85	–230.85
121	te puc [i] dire	49	1925	2076.78	–151.78
123	quant [i] passa	39	2104	2056.56	47.44
139	fills [i] bravos	32	1711	2042.41	–331.41
141	tots [i] mos	51	1951	2080.83	–129.83
143	mig [i] de	38	1979	2054.54	–75.54
144	verd [i] del	98	2241	2175.86	65.14
149	salut [i] viure	59	2145	2097.00	48.00
153	tens [i] també	42	1758	2062.63	–304.63
159	te puc [i] dire	44	1687	2066.67	–379.67
161	un poc [i] de	43	1993	2064.65	–71.65
162	un poc [i] de	48	1745	2074.76	–329.76
166	quant [i] los joves	52	1887	2082.85	–195.85
168	doncs [i] tens	46	2033	2070.72	–37.72
169	tens [i] de	55	1881	2088.92	–207.92
	Mean	48.94	1973.45	2076.66	–103.21
	Standard deviation	11.90	156.37	24.06	150.67