LEXICAL VS. EPENTHETIC VOWELS IN ALGHERO CATALAN:
A PHONETIC STUDY OF NON-LABIAL HIGH VOWELS

13èmes Rencontres du Réseau Français de Phonologie, Bordeaux 2015, June 29 – July 1

Jesús Jiménez
Universitat de València
jesus.jimenez@uv.es

Maria-Rosa Lloret
Universitat de Barcelona
mrosa.lloret@ub.edu

SUBJECT: Algherese Catalan displays, in addition to **lexical stressed** and **lexical unstressed** high vowels (as in *típic* [típik] 'typical'), an **epenthetic** (**inserted unstressed**) high vowel which is added across words whenever a word ends in a stop or an affricate and the following word starts with a consonant:

típic dolç de Pasqua [típiki dóltsi de páskwa] 'typical Easter sweet' (cf. *típic* [típik] 'typical'; *dolç* [dóltʃ] 'sweet')

GOALS: **(a)** To analyze the acoustic features that characterize the three different non-labial high vowels found in Algherese Catalan.

(b) To investigate if there is a correlation between the acoustic characteristics of these vowels and their relative prominence.

1. METHODOLOGY

- **Materials**: an interview conducted in the city of Alghero in 1997, published in Viaplana & Perea (2003) as part of the *Corpus Oral Dialectal* (COD) of the Universitat de Barcelona.
- **Informant & topics**: a housewife in her forties, born and raised in Alghero, dealing with common topics of her daily life → semi-spontaneous speech.
- **Selection of the segments**:
  - only non-labial high vowels appearing in open syllables
  - lexical vowels in hiatus, as well as segments with an expressive or emphatic lengthening, were discarded
- **Corpus**: **56 lexical stressed [i]**, **69 lexical unstressed [i]**, **49 inserted unstressed [i]**
- The acoustic analysis was carried out with PRAAT (Boersma 2001). **Segmentation and labeling** of the target vowels were done manually, based on spectrogram and waveform.
- A Praat script was used to automatically extract the following **parameters**:
  - **duration** of the whole segment (in ms)
  - **intensity** (in dB) and the three first formants (**F1**, **F2** and **F3**, in Hz) as measured at the center of the vowel
- All **statistical tests** were carried out using SPSS, version 22.0 (IBM Corp. 2013).

2. CLUSTER ANALYSIS OF THE WHOLE DATASET

➢**Statistical test**: **Two-step cluster** analysis, **on all 174 vowels** of the corpus (**56 lexical stressed [i]**, **69 lexical unstressed [i]**, and **49 inserted unstressed [i]**), considering the five continuous variables obtained from the acoustic analysis, i.e., length, intensity, F1, F2, and F3.

➢The cluster analysis yielded **three relatively well defined groups**, with 39 tokens in cluster 1, 76 in cluster 2, and 59 in cluster 3.

➢**Predictors' importance**: The variable **Length** contributes the most to differentiating the three clusters (PI=1), closely followed by **F2** (PI=.9). F1 is the least relevant variable (PI=.07), with Intensity in a slightly higher position (PI=.26). The variable F3 appears in an intermediate position in the ranking (PI=.57).

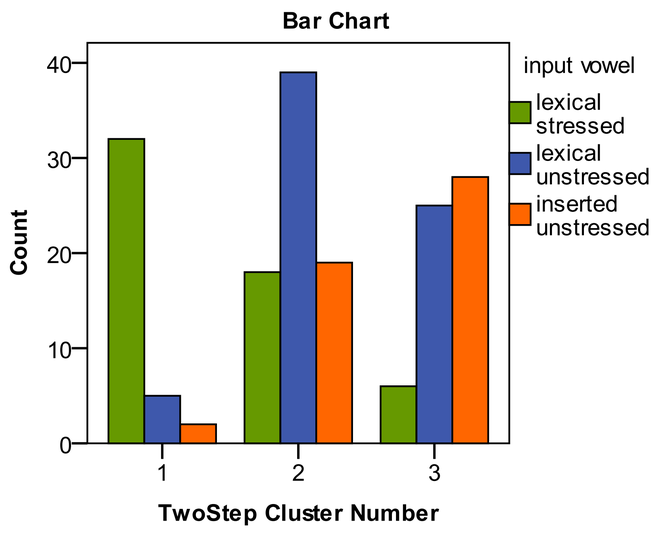
➢**Cluster composition**: The first cluster (22.41% of the tokens) mostly contains long vowels with the highest values for F2 and F3. Cluster 2 (43.68%) consists of vowels whose F2 and F3 values are slightly lower than those of the vowels in cluster 1 and are considerably shorter than these segments. Finally, cluster 3 (33.91%) comprises the shortest segments, with the lowest values for F2 and F3 as well.

➢**Correlation between the clusters and the input vowels.** 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.

✓The results revealed a **significant association** between the variables Input vowel and Cluster ($\chi^2(4)=65.94$, $p < .001$, Cramer's V=.435).

➢We ran a second 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 (unstressed) Input vowel and Cluster, although the effect was clearly weaker ($\chi^2(1)=4.57$, p=.033, Cramer's V=.203).



3. A STUDY OF VOWELS SURROUNDED BY CORONAL CONSONANTS

➢**Corpus**: a subset of the original database, containing only vowels surrounded by coronal consonants: **26 lexical stressed**, **14 lexical unstressed**, and **20 inserted unstressed** vowels.

➢**Main statistical test**: Kruskal-Wallis test. **Follow-up test**: Mann-Whitney pairwise tests.

➢**Variables**:

✓**Independent variable**: Input vowel (lexical stressed, lexical unstressed, & inserted unstressed)
✓**Dependent variables**: Length, F1, F2, F3, and Intensity (the table summarizes the mean values and the standard deviation of these variables).

	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)

➢**Statistical results**:

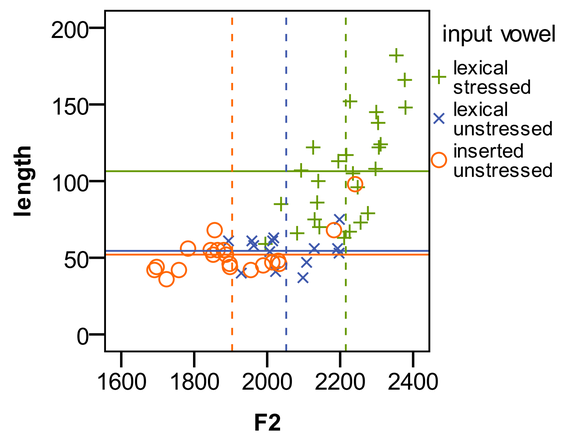
✓**Non-significant** results for **F1** ($H_{(2)}=3.99$, p=.136) and **Intensity** ($H_{(2)}=5.13$, p=.077).
✓**Significant** association of the variable Input vowel and the variables **Length** ($H_{(2)}=38.26$, $p < .001$), **F2** ($H_{(2)}=34.44$, $p < .001$) and **F3** ($H_{(2)}=17.17$, $p < .001$).

	Pairwise comparisons (Mann-Whitney test)		
Variables	**lex. stressed** vs. **lex. unstr.**	**lex. stressed** vs. **ins. unstr.**	**lex. unstr.** vs. **ins. unstr.**
Length (ms)	U=11, p < .001, r=–.77	U=19, p < .001, r=–.79)	U=104.5, p=.213 (n.s.), r=–.21
F2 (Hz)	U=46, p < .001, r=–.61	U=27, p < .001, r=–.76	U=54, p=.003, r=–.52
F3 (Hz)	U=92, p=.011, r=–.40	U=79, p < .001, r=–.59	U=116, p=.401 (n.s.), r=–.14

✓**Lexical stressed vowels are longer** than both lexical unstressed vowels and inserted unstressed vowels. But there are no significant differences between the two kinds of unstressed vowels.

✓The **same contrast** (lexical stressed vs. all unstressed vowels) is found with respect to **F3**.

✓Instead, there is a **triple contrast in F2**, with a **progressive decline in the values** of the input vowels, going from the lexical stressed segments to the inserted unstressed vowels, with lexical unstressed vowels at an intermediate point.



4. CONCLUDING REMARKS

- **Clear differences** between the three input vowels were found, but **not for every measure**. Neither the openness of the vowels — their F1 value — nor their intensity were relevant for distinguishing the three input vowels (which is in accordance with Ballone 2008).
- As generally claimed, **stressed segments came out longer** than the unstressed ones. However, although earlier studies found that inserted unstressed vowels were shorter than lexical unstressed ones (cf. Ballone 2008), **our data did not show a significant difference in duration between lexical and inserted unstressed vowels**.
- **Regarding F2, we found a pattern of gradual centralization** that goes from the lowest degree of centralization — i.e., the highest F2 values — in the lexical stressed segments to the highest degree — i.e., the lowest F2 values — in the inserted unstressed vowels.
- The contrast in F2 between lexical stressed and unstressed segments follows from 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.
- The gradation in the F2 values from lexical to inserted unstressed vowels, though, suggests that **it is the nature of these vowels**, and not their duration, that determines the degree of centralization that they present, which goes in line with Hall's (2013) findings on Lebanese Arabic. Further research is needed to determine if the leveling of duration among these segments while maintaining their F2 differences is a **stage preceding their complete merging**, both in Algherese and cross-linguistically.
- To sum up, the data point to a **double contrast**: firstly, between lexical stressed and unstressed vowels, with longer duration and higher F2 values in the more prominent vowels (i.e., the stressed ones), and, secondly, between unstressed segments, with higher F2 values in the lexical vowels, which are relatively more prominent due to their underlying (and more stable) nature.

REFERENCES

● BALLONE, F. (2008): «Català de l'Alguer: anàlisi instrumental d'un text oral», *Insula,* 4, 71-89. ● BALLONE, F. (2013): *An Acoustic Study of Sardinian and Algherese Catalan Vowels*, doctoral dissertation, Bellaterra, Universitat Autònoma de Barcelona. ● BOERSMA, P. (2001): «Praat, a system for doing phonetics by computer». *Glot int.* 5, 341–345. ● COD=*Corpus Oral Dialectal*, Barcelona, Departament de Filologia Catalana, Universitat de Barcelona. File from Algherese Catalan available at http://diposit.ub.edu/dspace/handle/2445/11637. ● 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, 203-225. ● HALL, N. (2006): «Cross-linguistic patterns of vowel intrusion», *Phonology*, 23, 387-429. ● HALL, N. (2013): «Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic», *Journal of Phonetics*, 41, 133-143. ● IBM Corp. (2013): *IBM Statistics for Windows*, version 22.0, Amonk, NY, IBM Corp. ● JIMÉNEZ, J. & M.-R. LLORET (forthcoming): «Prominence and centralization: A two-way contrast between lexical and epenthetic non-labial high vowels in Algherese Catalan», *Estudios de Fonética Experimental*, 34. ● 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, 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, 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, 187-201. ● LINDBLOM, B. (1963): «Spectrographic study of vowel reduction», *Journal of the Acoustic Society of America*, 35, 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. ● LLORET, M.-R. & J. JIMÉNEZ (2010): «Sobre les vocals afegides de l'alguerès», *Insula: Quaderno di cultura sarda*, 7, 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, 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. ● VIAPLANA, J. & M. P. PEREA (2003): *Textos orals dialectals del català sincronitzats. Una selecció*, Barcelona, PPU.

Research funded by Spanish MINECO (project FFI2013-46987-C3-1-P) and the Catalan Government (2014SGR918)

Poster available at http:/www.ub.edu/GEVAD