

Phonological categorization, sound change and speech corpora

José Ignacio Hualde

Univ. of Illinois at Urbana-Champaign

WORKSHOP “LINGUISTIQUE ET BIG DATA”

30/Nov/2017

Maison des Sciences de l’Homme

Université Paris-Saclay

Research question

Phonological recategorization in sound change (e.g. /t/ > /d/) presupposes allophonic variation, but not all allophony results in sound change.

How and when does recategorization take place? What can we learn about sound change from the analysis of variation in speech corpora?

Two topics: Intervocalic lenition of stops and final devoicing.

Phonological recategorization in sound change: Western Romance Intervocalic Lenition

/p t k/ > /b d g/ / V____V

	<u>Latin</u>	<u>Italian</u>	<u>Spanish</u>
p	sa p ĕre	<i>sapere</i>	<i>sab</i> er
t	vĭ t a	<i>vita</i>	<i>vida</i>
k	amĭ c a	<i>amica</i>	<i>amiga</i>

Stage I (Latin)

/p/ **p**orta ‘door’ , cor**p**us ‘body’ , lu**p**um ‘wolf’

/b/ **b**ucca ‘mouth’, her**b**a ‘grass’

Stage II (Spanish)

/p/ **p**uerta, cuer**p**o

/b/ **b**oca, hier**b**a, lobo

Phonetics vs phonology in sound change

1. /p/ → [b] / V__V voice assimilation
2. /p/ > /b/ How?

Corpus study: Rome Italian

There has been no phonological recategorization of Latin /p t k/ but intervocalic phonological /p t k/ (*lupo, lato, amico*) often realized with *some degree* of voicing (Loporcaro, 2005; Marotta, 2005)

> Why hasn't voicing lenition resulted in sound change? Could it?

Rome Italian

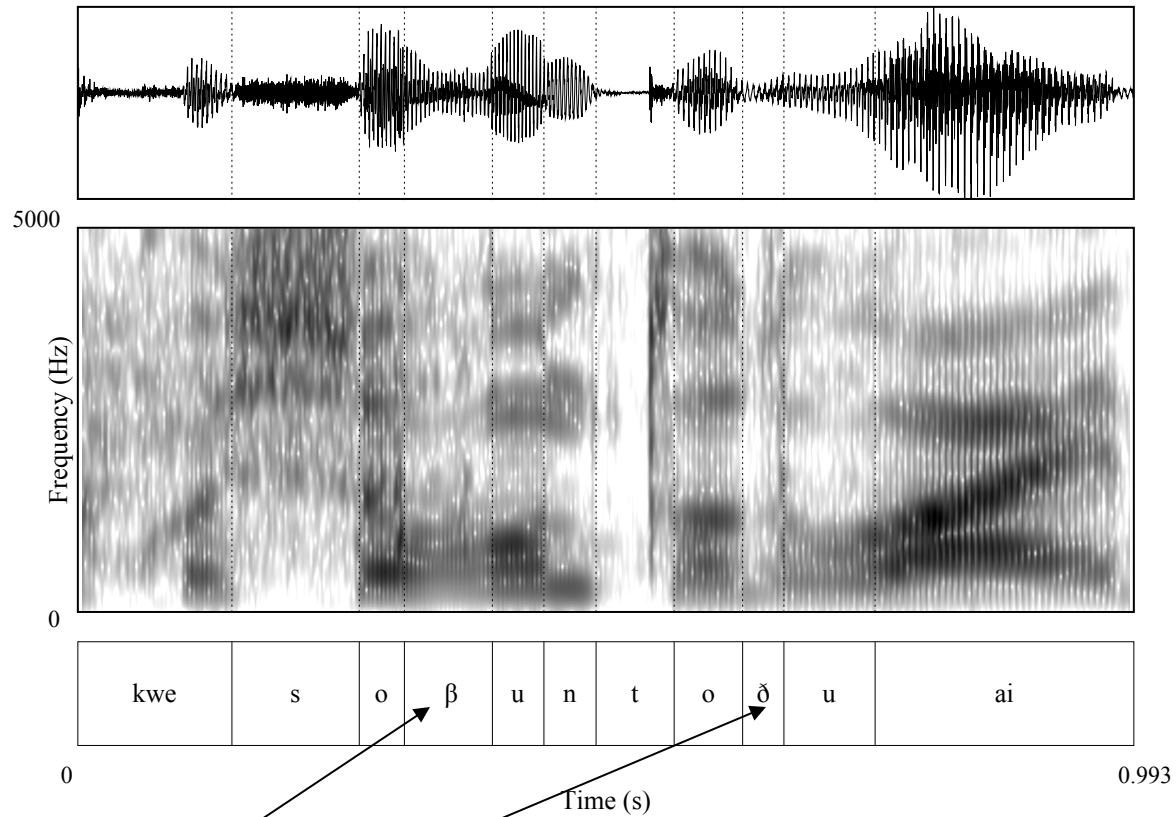
(Hualde & Nadeu, 2011. *Phonetica* 68)

Corpora e Lessici dell'Italiano Parlato e Scritto (CLIPS), directed by Federico Albano Leoni

http://www.clips.unina.it/it/documenti/2_tecniche_di_elicitazione_dialogica.pdf

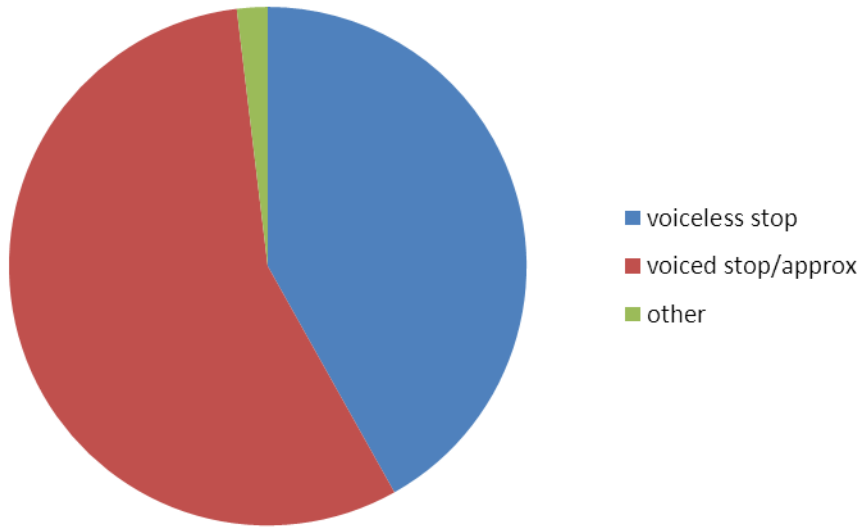
- Rome data from three map tasks and three “find the differences” games

Rome Italian

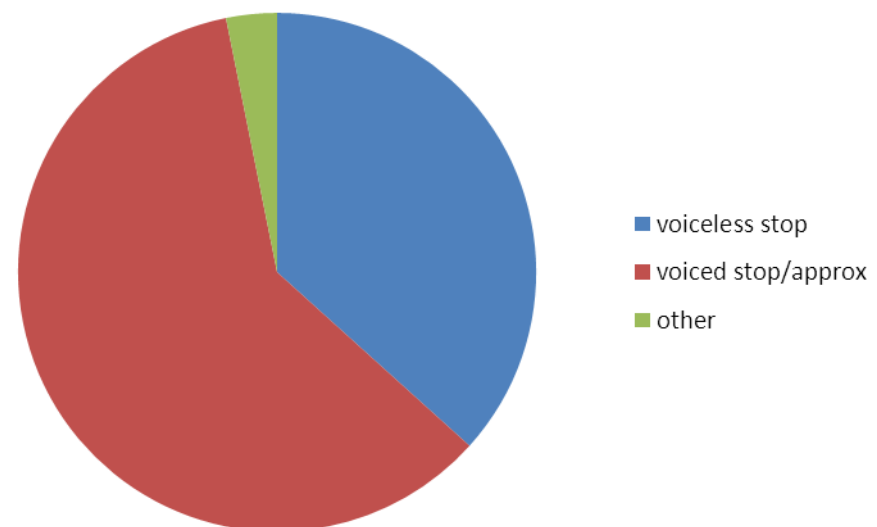


a questo punto tu hai

Rome Italian: realization of intervocalic /p t k/



**Word-initial /p t k/
(postvocalic)**



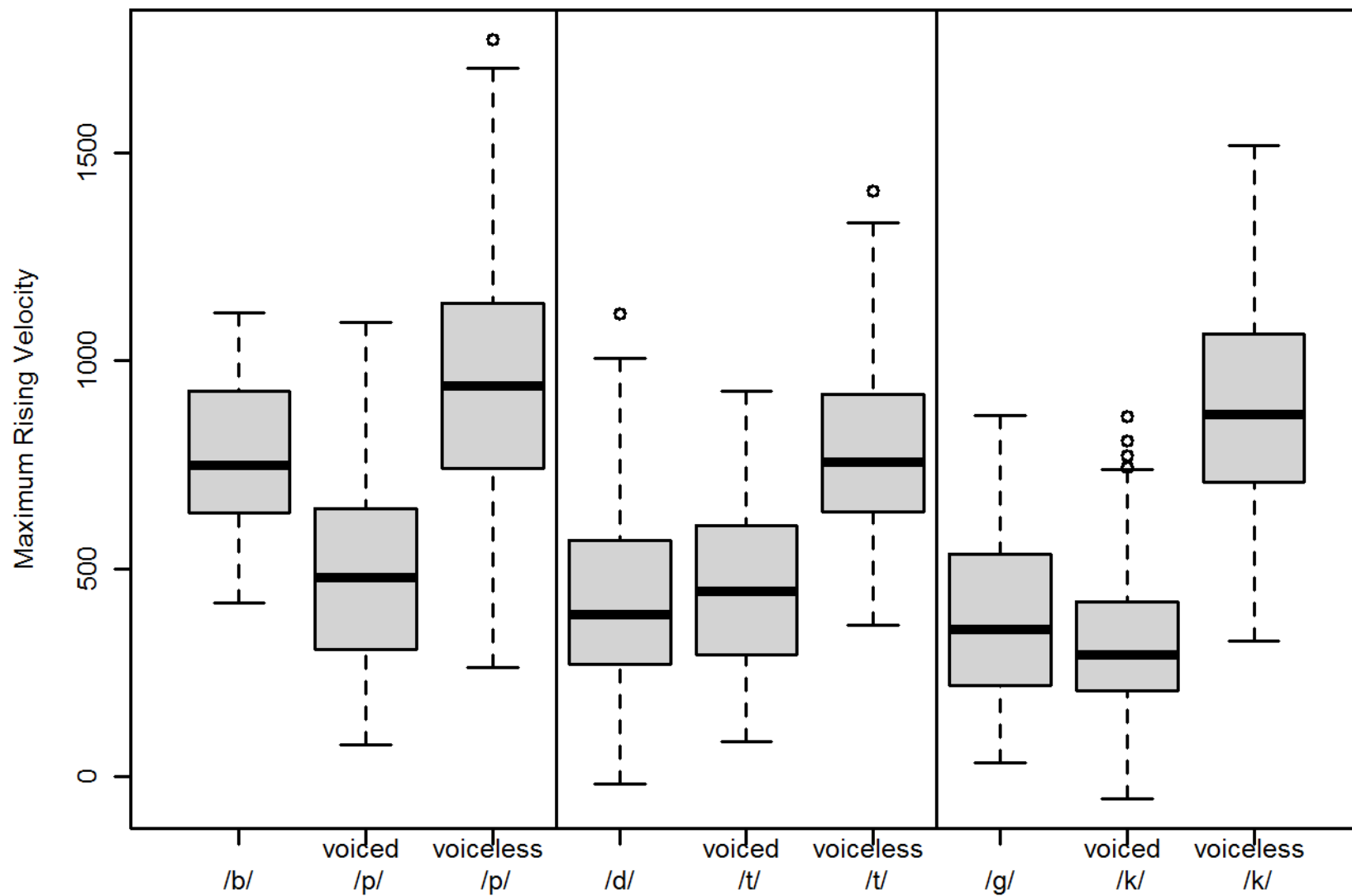
**Word-internal /p t k/
(postvocalic)**

Intervocalic /ptk/

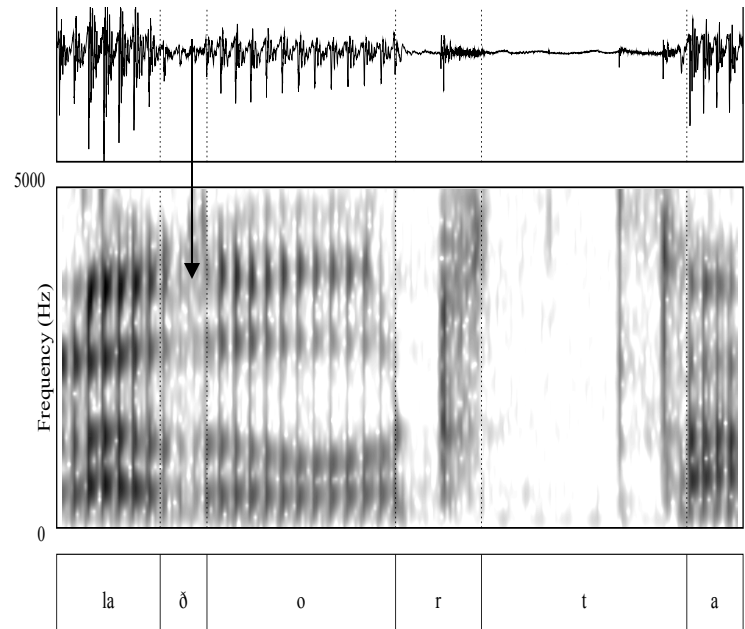
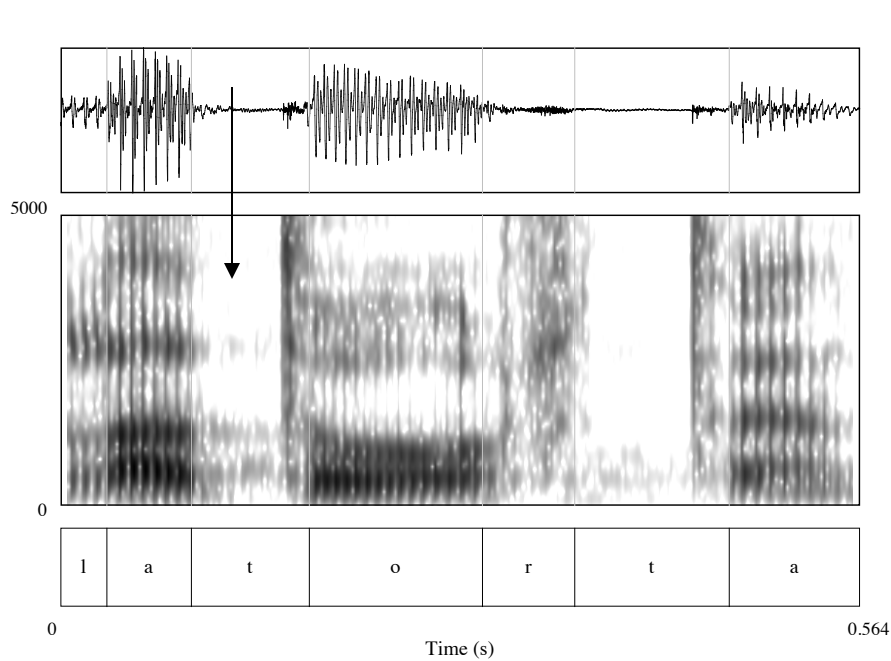
<u>Latin</u>	<u>Spanish</u>	<u>Latin</u>	<u>Spanish</u>
sa p ēre	sa b er	illa p orta	la p uerta
vī t a	vi d a	illa t urre	la t orre
amī c a	ami g a	illa c asa	la c asa

Voicing word-internally only

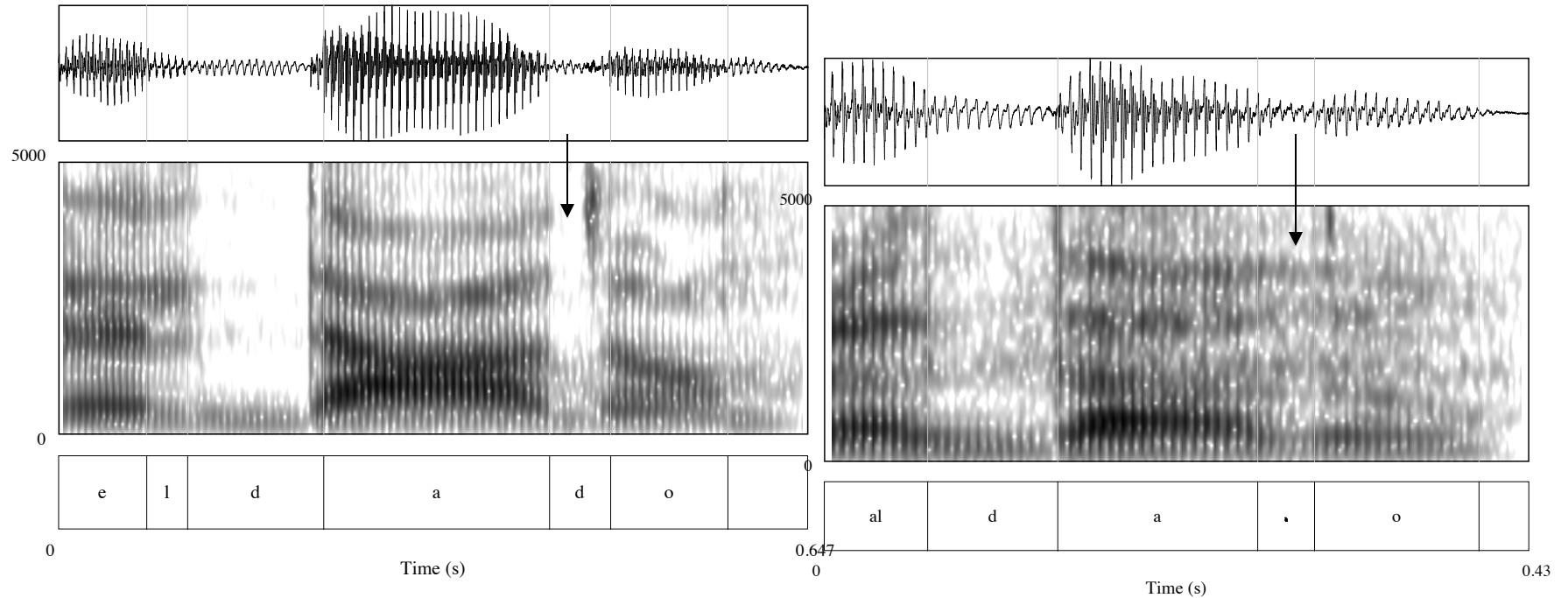
Rome Italian: Acoustic estimation of degree of constriction



Phonologically voiceless plosives



Phonologically voiced plosives



Large amount of phonetic overlap between
/t/-/d/, /k/-/g/

But /p t k/ also have voiceless realizations

Rome Italian vs Western Romance

Lat. *latu(m)* >

- **Rome Italian:**

lato [lato] ~ [lado] ~ [laðo]

- **Portuguese:**

lado [lado] ~ [laðo] **[lato]

- **Spanish:**

lado [laðo] ~ [lao] **[lato], ??[lado]

(cf Cravens. T. 2002. *Comparative historical dialectology: Italo-Romance clues to Ibero-Romance sound change*. Amsterdam: Benjamins)

From Latin to Spanish and French

From Latin to Portuguese:

mutare > /mudar/ [mudar] (~ [-ð-])

From Latin to Spanish:

mutare > /mudar/ [d] > /mudar/ [muðar]

latu > /lado/ > /lado/ [laðo] > /lado/ [laðo] -
[lao]

From Latin to French:

mutare > muder > muðer > muer [mɥe]

Main lesson from this study: Allophonic lenition vs phonological recategorization

1. At a first stage lenition (e.g. voicing) is optional and morphological boundaries are ignored (phonemes are affected in phonologically defined contexts).
2. Subsequent recategorization (with exclusively voiced realizations) may affect individual lexical items. Word-initial phonemes are less likely to be recategorized than word-medial ones.

Two distinct targets for voiceless stops
(voiceless and voiced) or a single wide
window of allophonic voicing &
constriction?

Cf allophony of /bdg/ in Madrid vs Costa
Rica Spanish

Spirantization as a synchronic process in Spanish

<i>boca</i>	[bóka]	<i>vista</i>	[bísta]
<i>la boca</i>	[laβóka]	<i>la vista</i>	[laβísta]

<i>día</i>	[día]
<i>de día</i>	[deðía]
<i>gato</i>	[gáto]
<i>mi gato</i>	[miɣáto]

Two distinct allophones or continuum of allophonic realizations?

Allophony of /b d g/ in Spanish

Navarro Tomás, Quilis, Harris, etc.

Stops [b d g]

-after pause [bóka]

pafter a homorganic nasal or lateral

[mb], [nd], [ŋg], [ld]: [embóka]

Approximants [β ð γ] elsewhere:

[sáβe], [la βóka]

[árβol] [ezβélto]

- > More recent work has argued that there is a continuum of realizations conditioned by preceding context, stress and other factors.

Central American Spanish

Preference for voiced stops everywhere except in the intervocalic context

Canfield, D.L. *La pronunciación del español en América* (1962:78)

*las **barbas**, el **buey** **volvió***

*es **verdad**, la deuda*

*hay **galgos**, rasgos*

Montes Giraldo (1975); Canfield (1981:5-6); Amastae (1989, 1995); Lipski, (1994:210, 223, 258, 290); Quesada Pacheco (1996: 103, 1998); Quesada & Vargas (Costa Rica), Utgård (Guatemala), Rosales Solís (2010), Hernández Torres (Honduras), Azcúnaga López (El Salvador), Cardona Ramírez (2010).

Fernandez, J. (1982) : The allophones of /b,d,g/ in Costa Rican Spanish. *Orbis* 31: 121-146 .

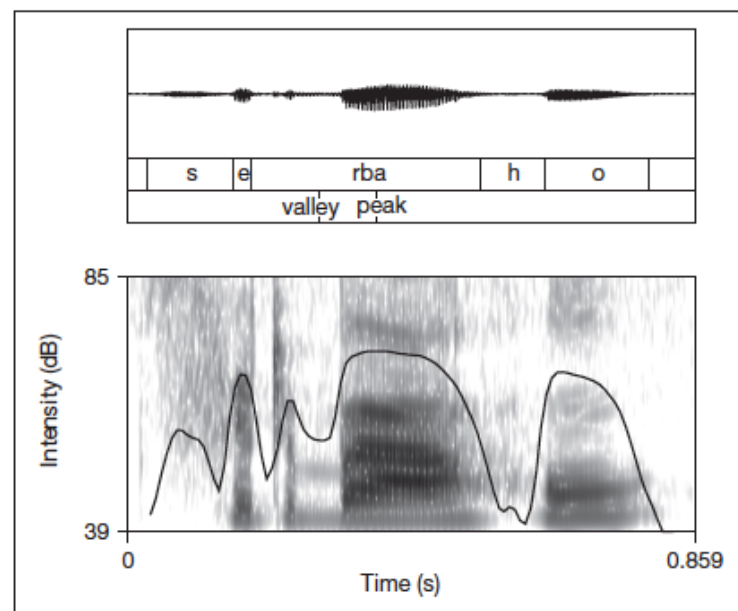
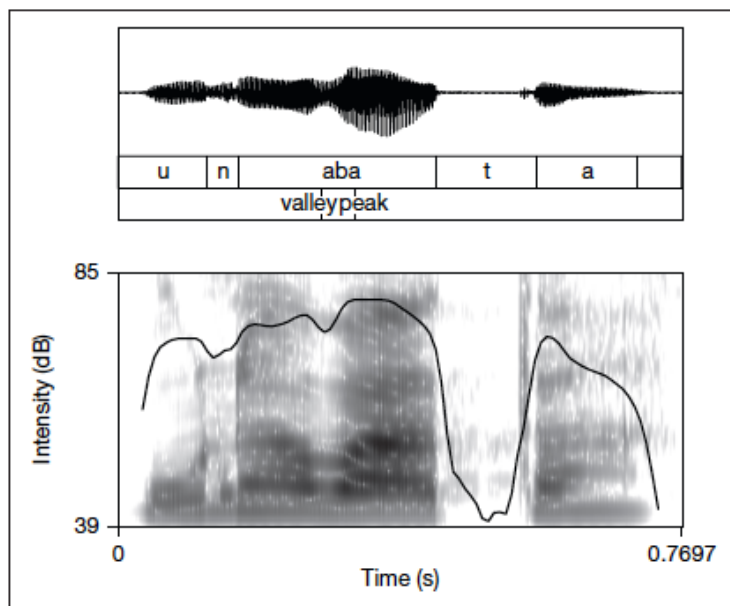
12 Costa Rica speakers

Realizations of /b d g/ classified as stop or fricative (approx.) based on auditory impression and some visual inspection of spectrograms

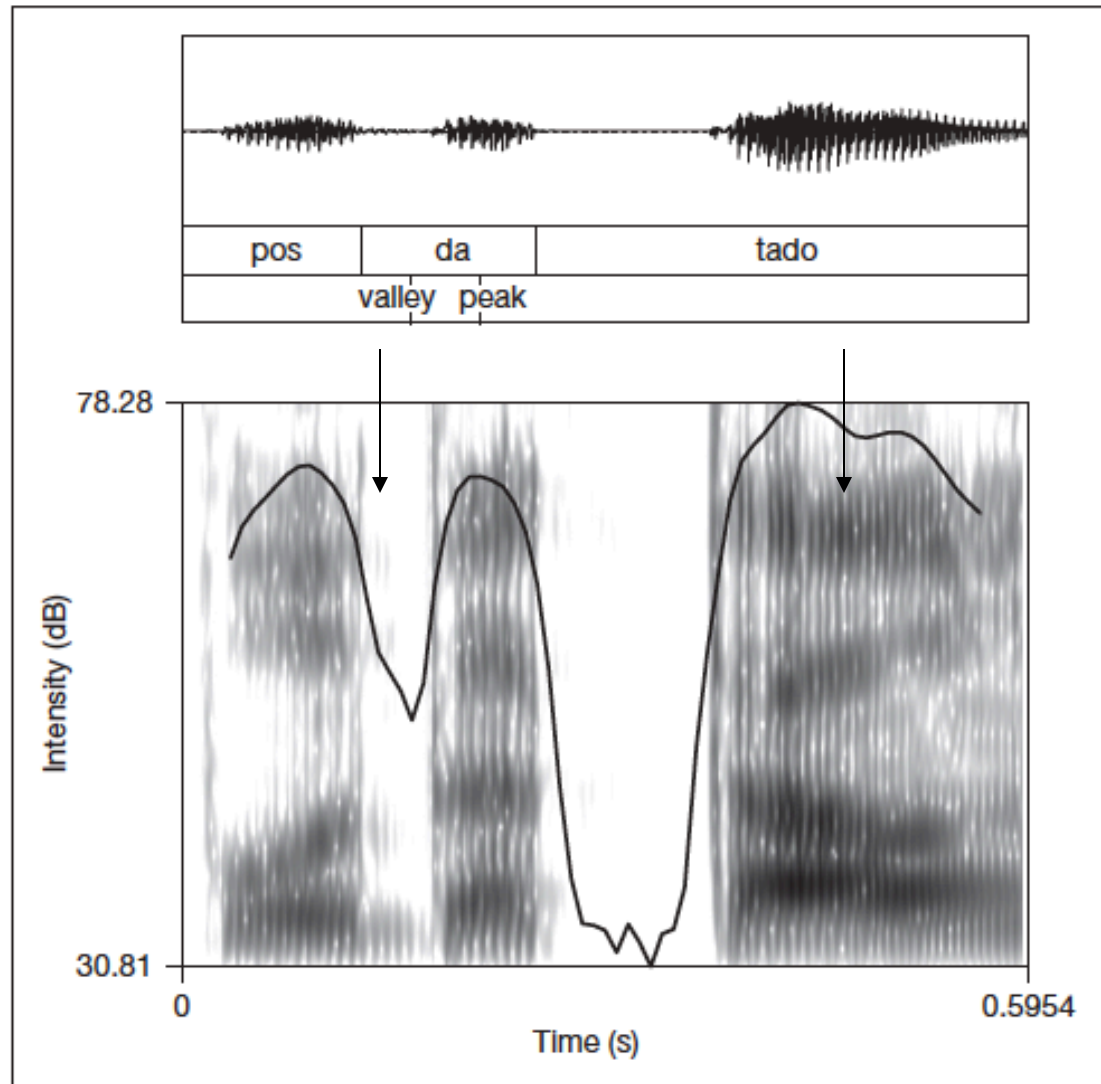
Result: of 280 tokens de /b d g/ after a consonant or glide, 221 stops (79%)

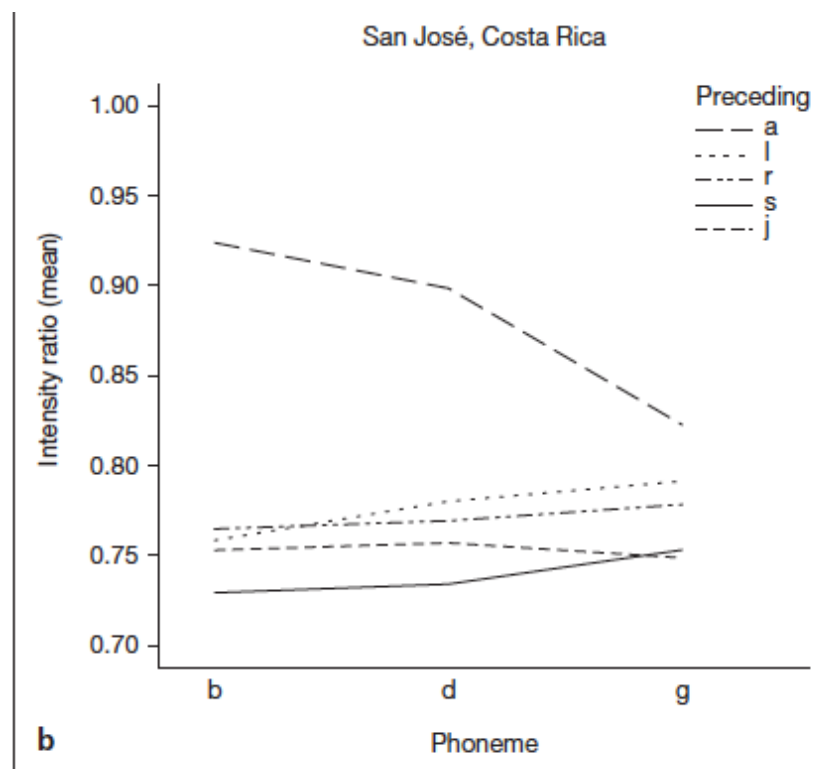
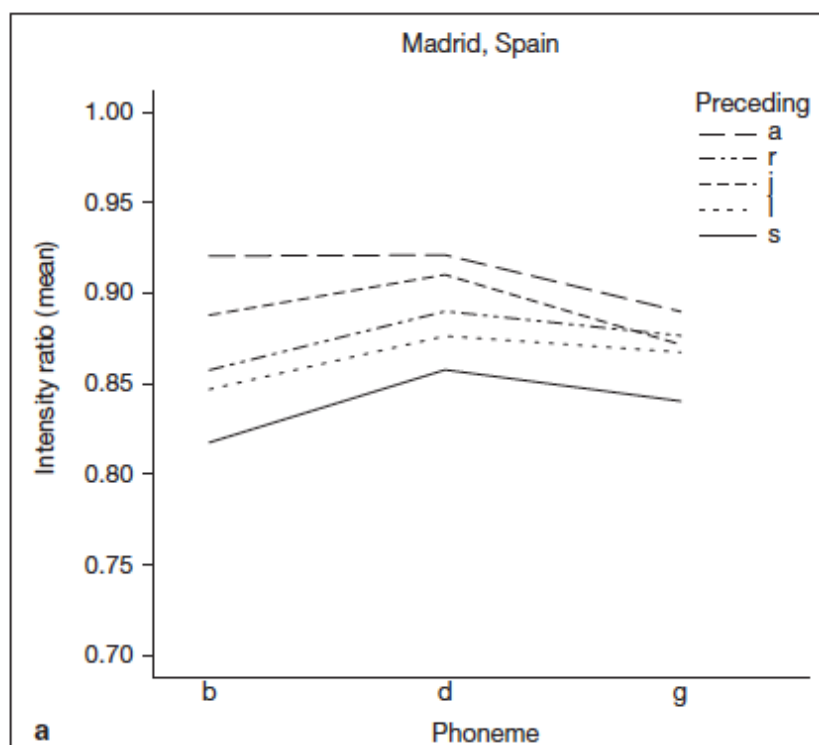
Conclusion: variation, with a strong tendency to choose the stop allophone after any cons. or glide.

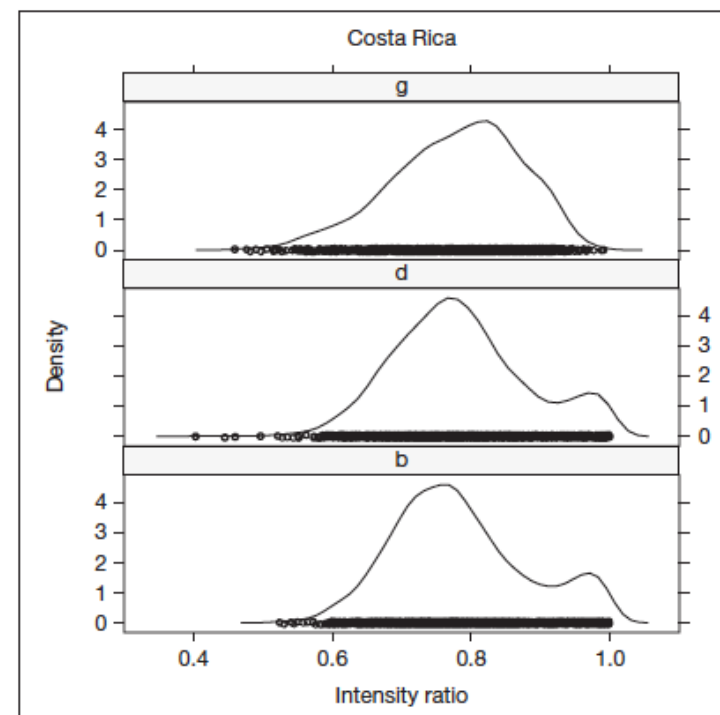
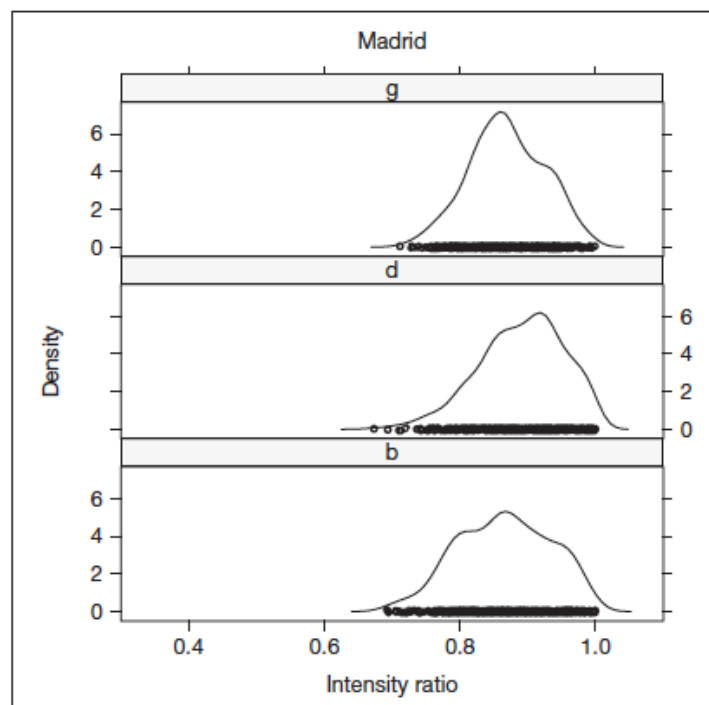
Costa Rica vs Madrid (Carrasco, Hualde & Simonet 2012 – Phonetica 69: 149-179)



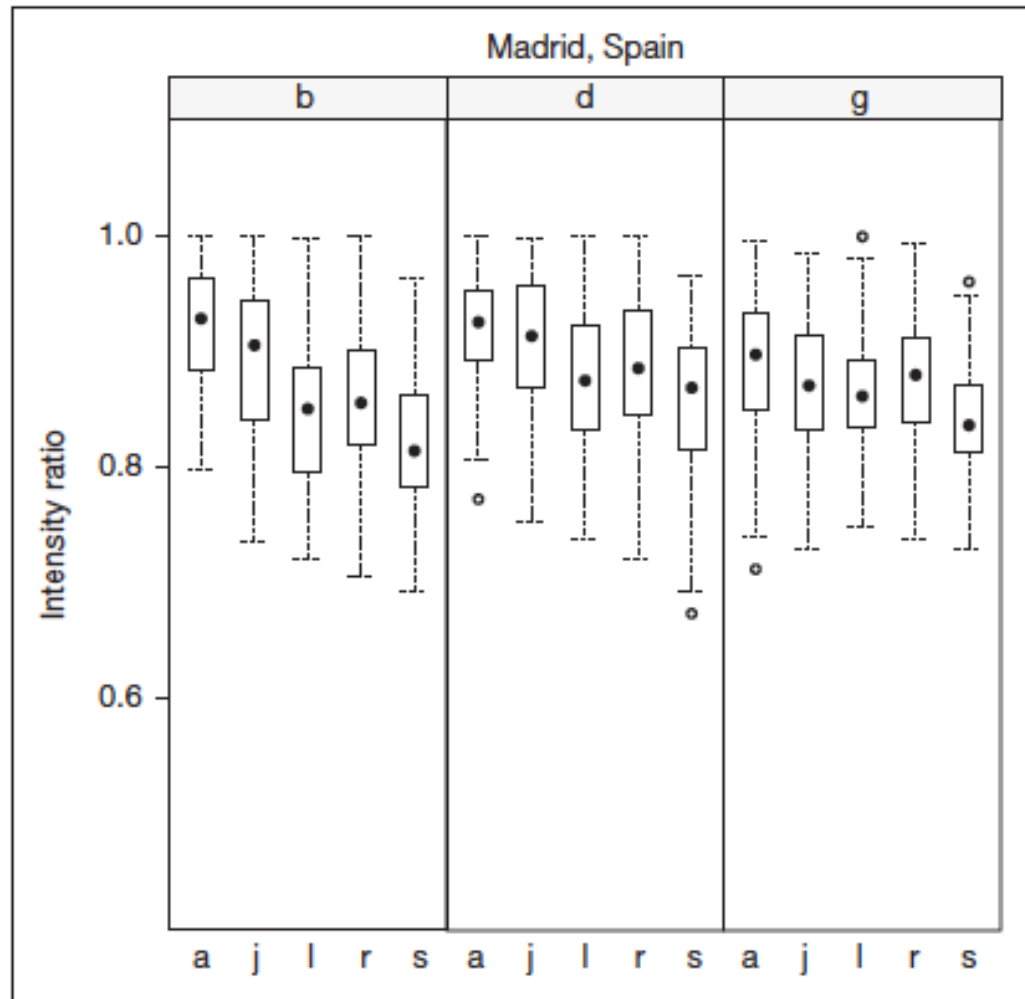
Costa Rica: token with two /d/'s, postconsonantal and postvocalic (Carrasco, Hualde & Simonet 2012 – *Phonetica* 69:149-179)



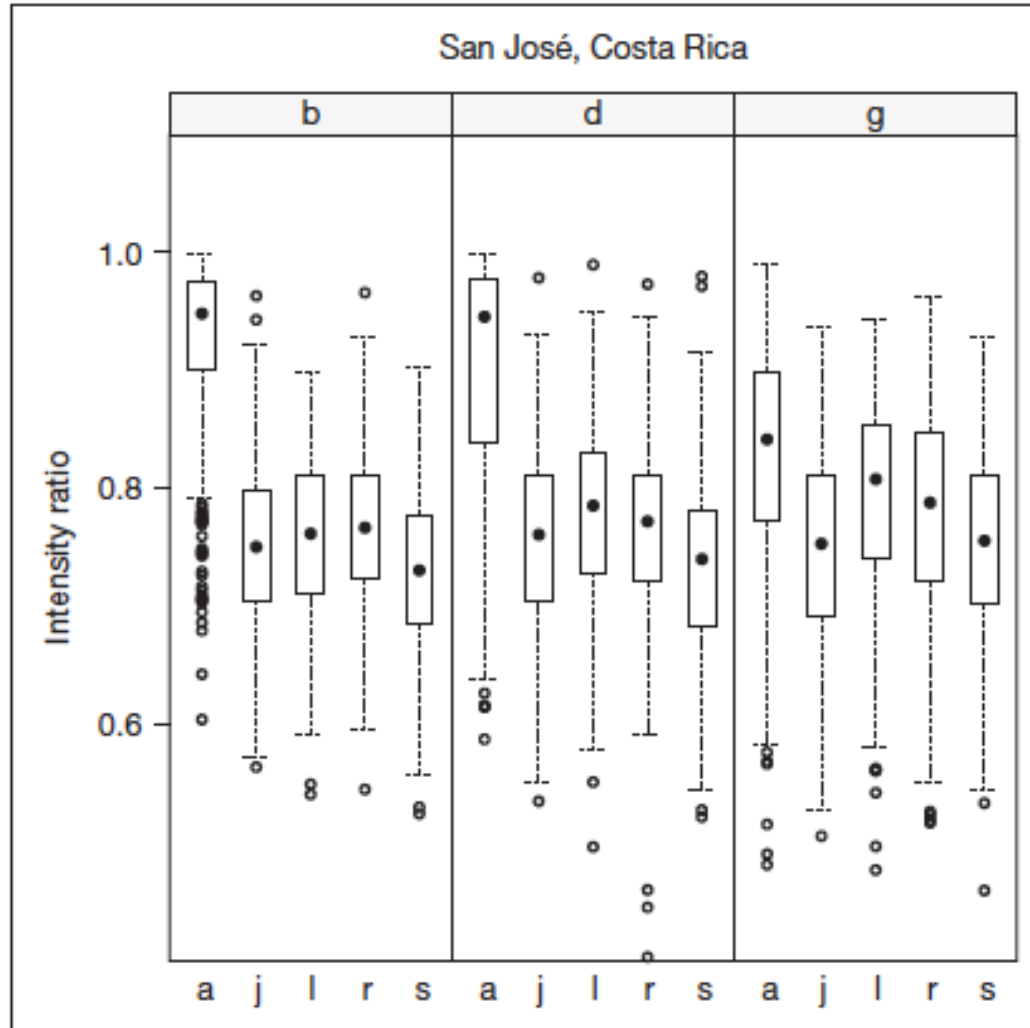




Degree of constriction by preceding context: Madrid



Degree of constriction by preceding context: Costa Rica



Study II: Word-final devoicing in Castilian Spanish (Hualde & Eager 2016)

Syllable-initial: /p-b/, /t-d/, /k-g/

Word-final: /p/, /t/, /k/

➤ Common phenomenon:

e.g. Catalan:

amic/amiga 'friend m/f' /-k/ ~ /-g-/

ric/rica 'rich m/f' /-k/ ~ /-k-/

Also: German, Russian, Turkish, etc

Historical origin?

Unlike other frequent sound changes (assimilation, lenition) it does not have a straightforward phonetic origin.

Assimilation < phonetic coarticulation

Lenition < target undershoot

Word-final devoicing <?

Historical origin of word-final devoicing?

- 1. Phonologization of phonetics: phrase-final, pre-pausal 'phonetic' devoicing.
- 2. Analogical extension.

(Hock 1991, etc)

Hypothesized diachronic stages

Phrase-final	Phrase-medial
Stage I: mad #	mad en
Stage II: mat ^t #	mad en
Stage III: mat ^t #	mat ^t en

- Problem: surprisingly little evidence for Stage II languages.

North-Western Castilian Spanish

Word-final /d/ variably realized as [θ] (not as [t])
red [reθ] ‘net’, *Madrid* [maðriθ]

(other stops are rare word-finally: *zigzag*, *bulldog*, *blog*)

Why [θ]?

Postvocalic /d/ → [ð]

[-ad] > [-að] > [-aθ]

But /ada/ [aða] vs /aθa/

Cf. Cat. *canta*[t] ‘sung, m.’, *canta*[ð]a ‘sung, f.’

Research question

➤ Is devoicing more common phrase-finally?

La ciuda[θ] ‘the city’ vs

La ciuda[ð̥] *es* ‘the city is??’

Cf. *ciudades* ‘cities’

-if so, this could be taken as evidence for the hypothesis of analogical extension from the prepausal environment

Additional process: Deletion

$/-d/ \rightarrow 0$

verda(d), Madri(d)

How does it interact with devoicing?

Our corpus

Glissando (<http://eca-simm.uva.es/glissando/node/10>)

Garrido et al. 2013: <https://link.springer.com/article/10.1007/s10579-012-9213-0>)

20+ speakers, 3 styles

High-quality recordings

> conversational speech subcorpus

Methods

All tokens of word-final /d/ (N=574) were segmented and visually classified as

stops,

voiced approximants,

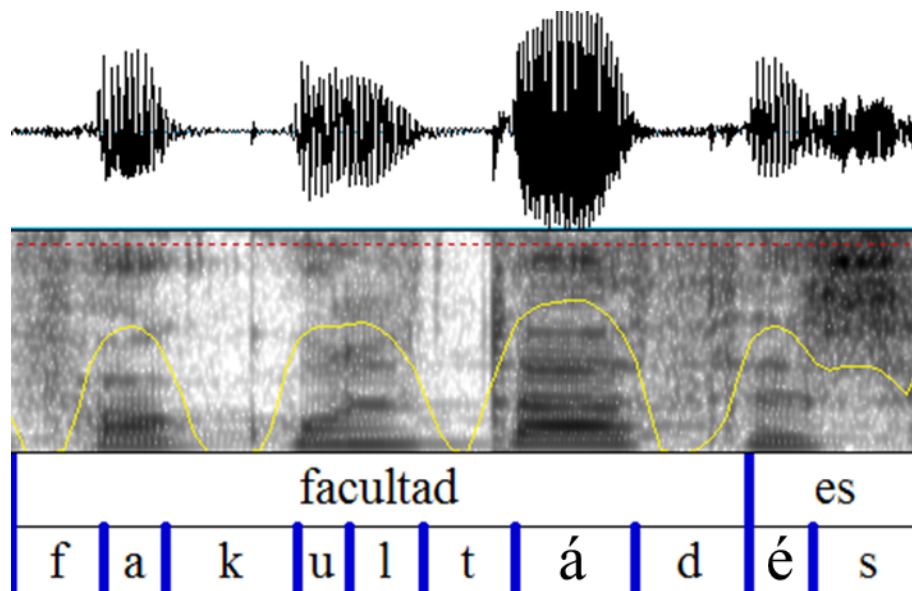
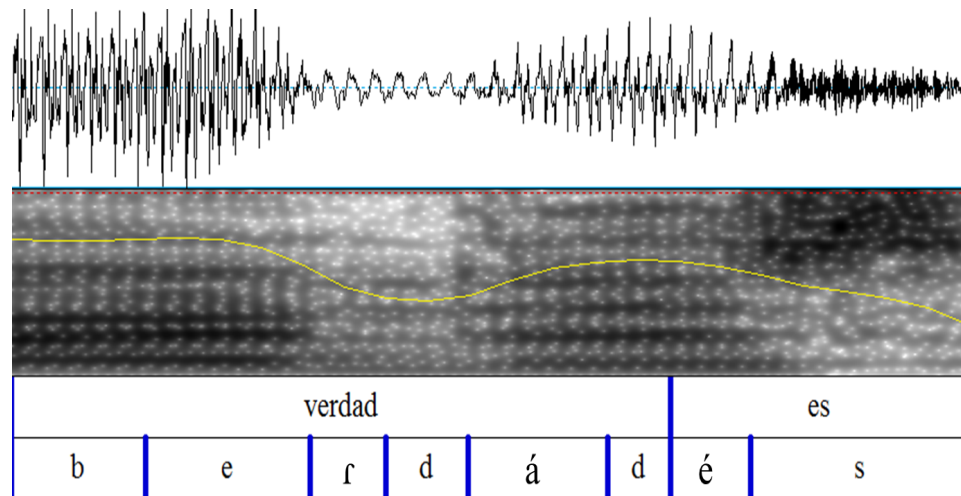
voiceless fricatives or

deleted.

(independent coding by two trained labelers, 91% agreement)

For comparison (i.e. possible phonological neutralization), all tokens of word-final /θ/ were also extracted (N= 79).

For non-deleted segments duration and percent of voicing were measured with a script.

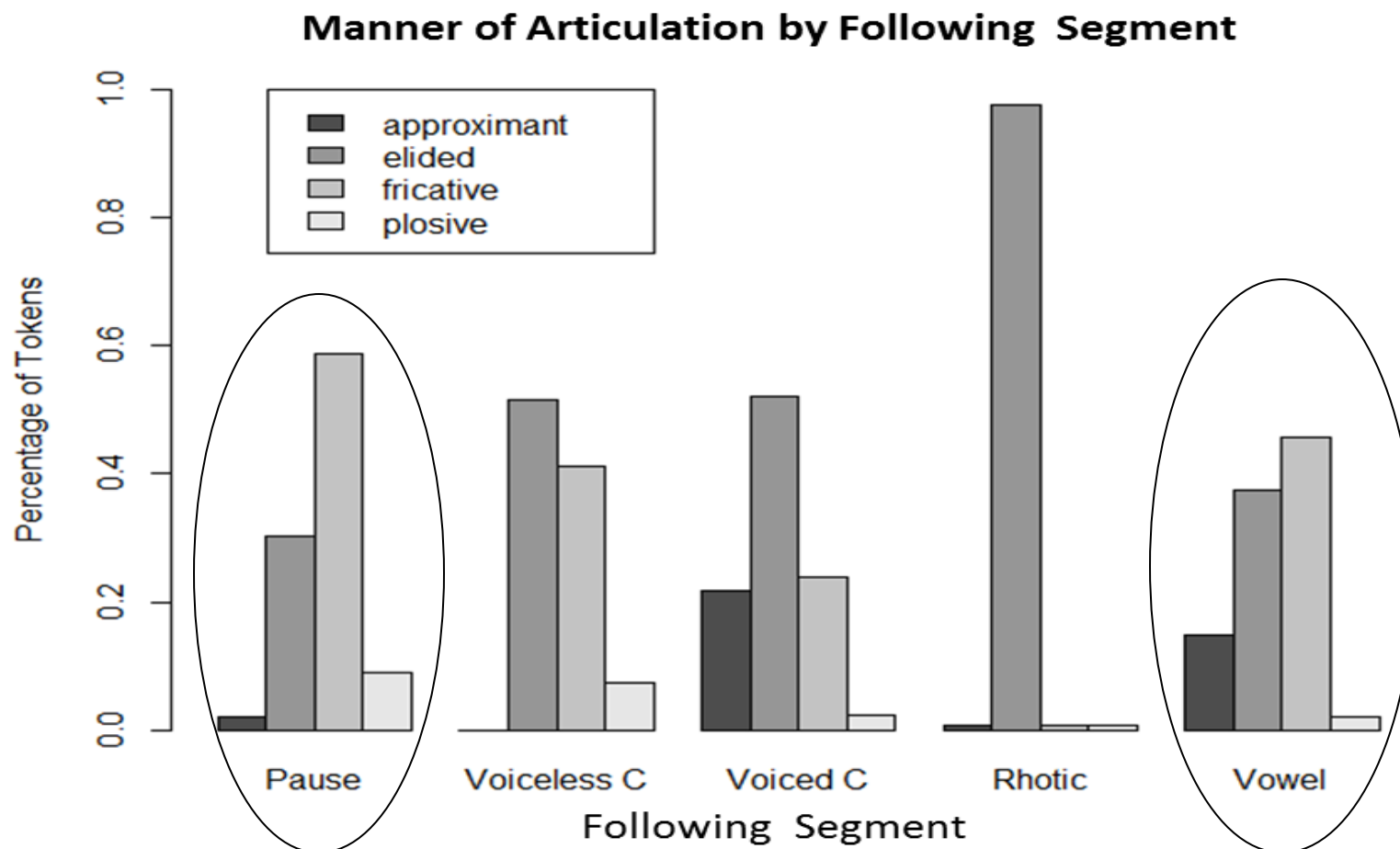


Effects of following context

Contexts:

- Before pause
- Before a voiceless consonant
- Before a voiced consonant
- Before /r/
- Before a vowel

Distribution of final /d/ realizations



Fricatives [θ] & deletions are the most common realizations.
Approximant [ð] and stop [d] realizations are surprisingly infrequent.

Is devoicing more common before
pause than before a vowel?

ANSWER: YES

although fricative realizations are frequent
before a vowel, they are significantly less
frequent in this context than before pause

(Mixed effects logistic regression)

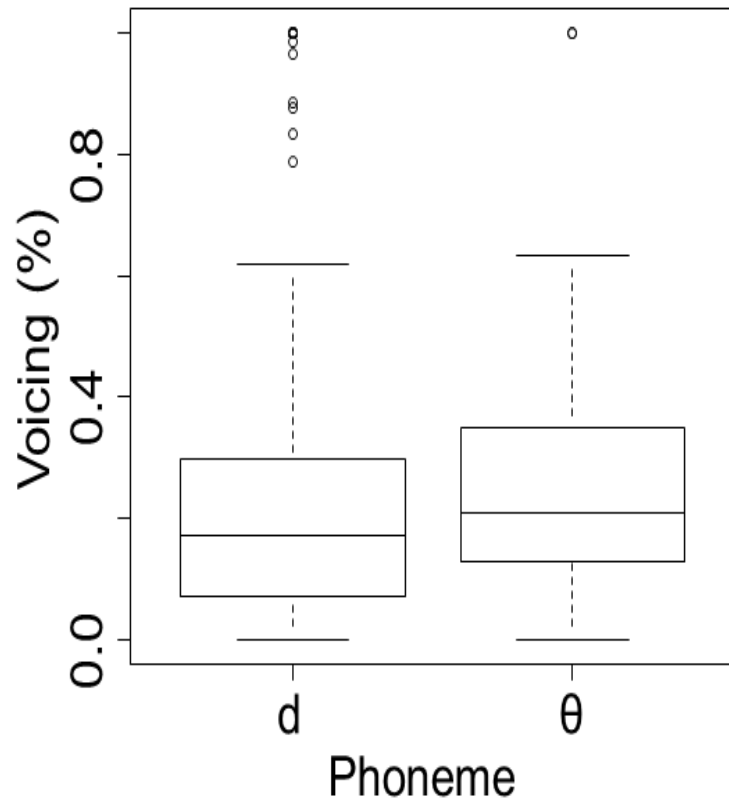
Is the devoicing of /-d/ neutralizing?

YES, when /-d/ is devoiced it is identical to phonemic /θ/

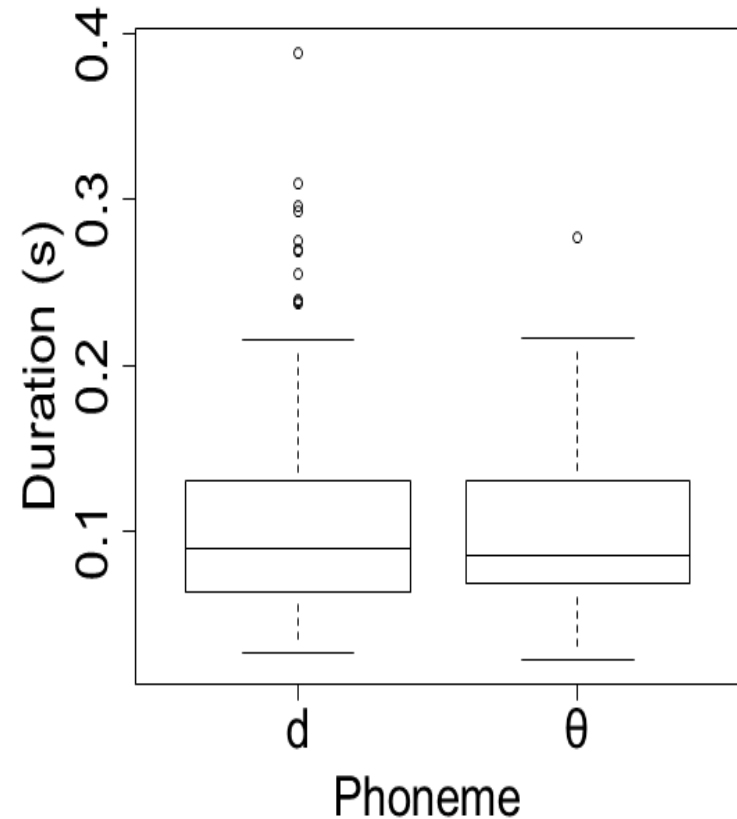
but /-d/ has other realizations not shared with /-θ/, in particular, zero.

Fricative /d/ vs Fricative /θ/

Voicing of Fricative /-d/ and /-θ/



Duration of Fricative /-d/ and /-θ/



Impossible evolution

$[-d] > [\text{ð}] > [\theta] > \emptyset$

Because phonemic /θ/ is not subject to deletion in this dialect.

verdad /berdad/ [berðáθ] ~ [berðá] ‘truth’

VS

capaz /kapaθ/ [kapáθ] ‘able’ **[kapá]

Two different and competing processes targeting word-final /-d/:

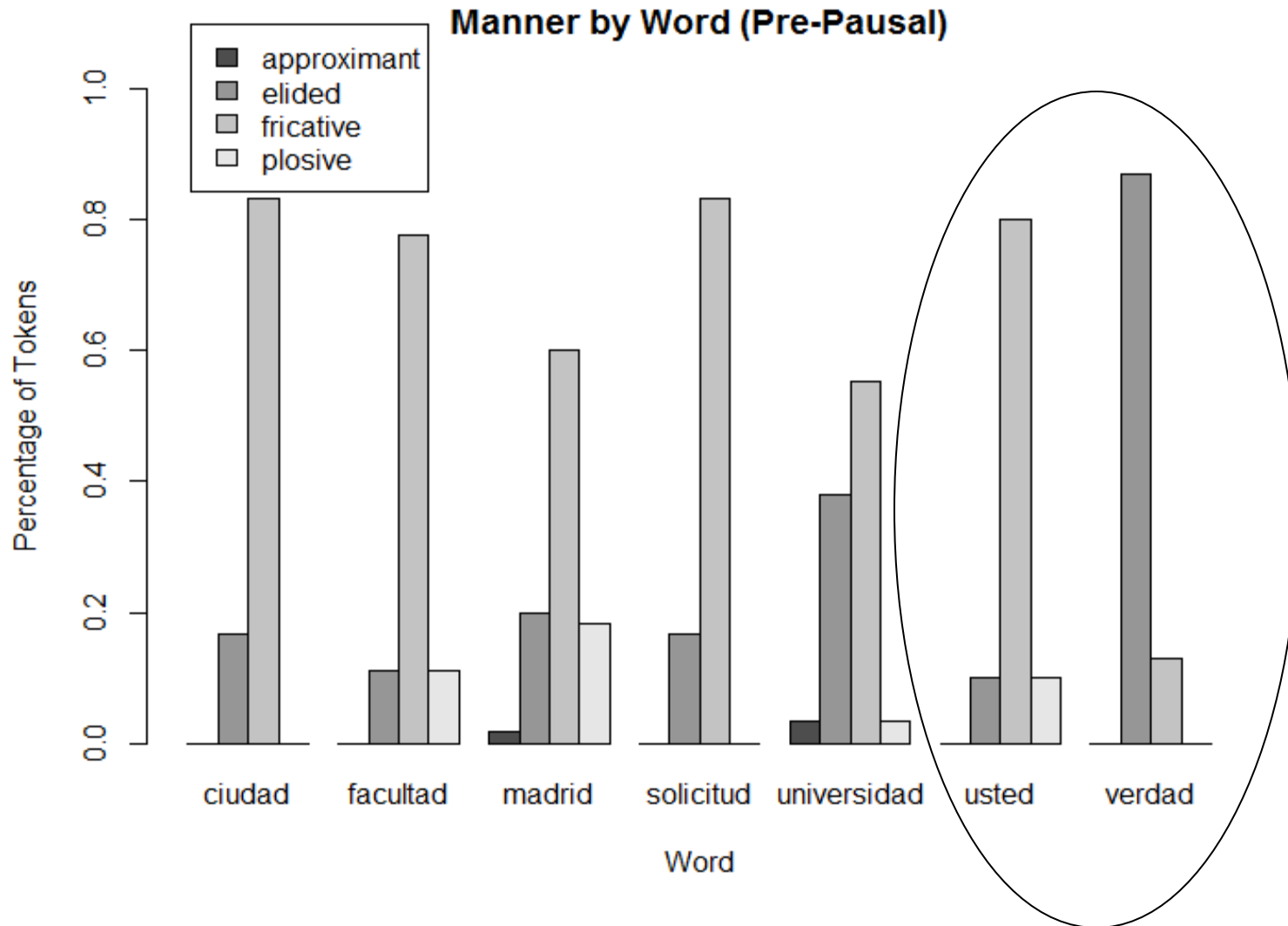
- Deletion

- Devoicing

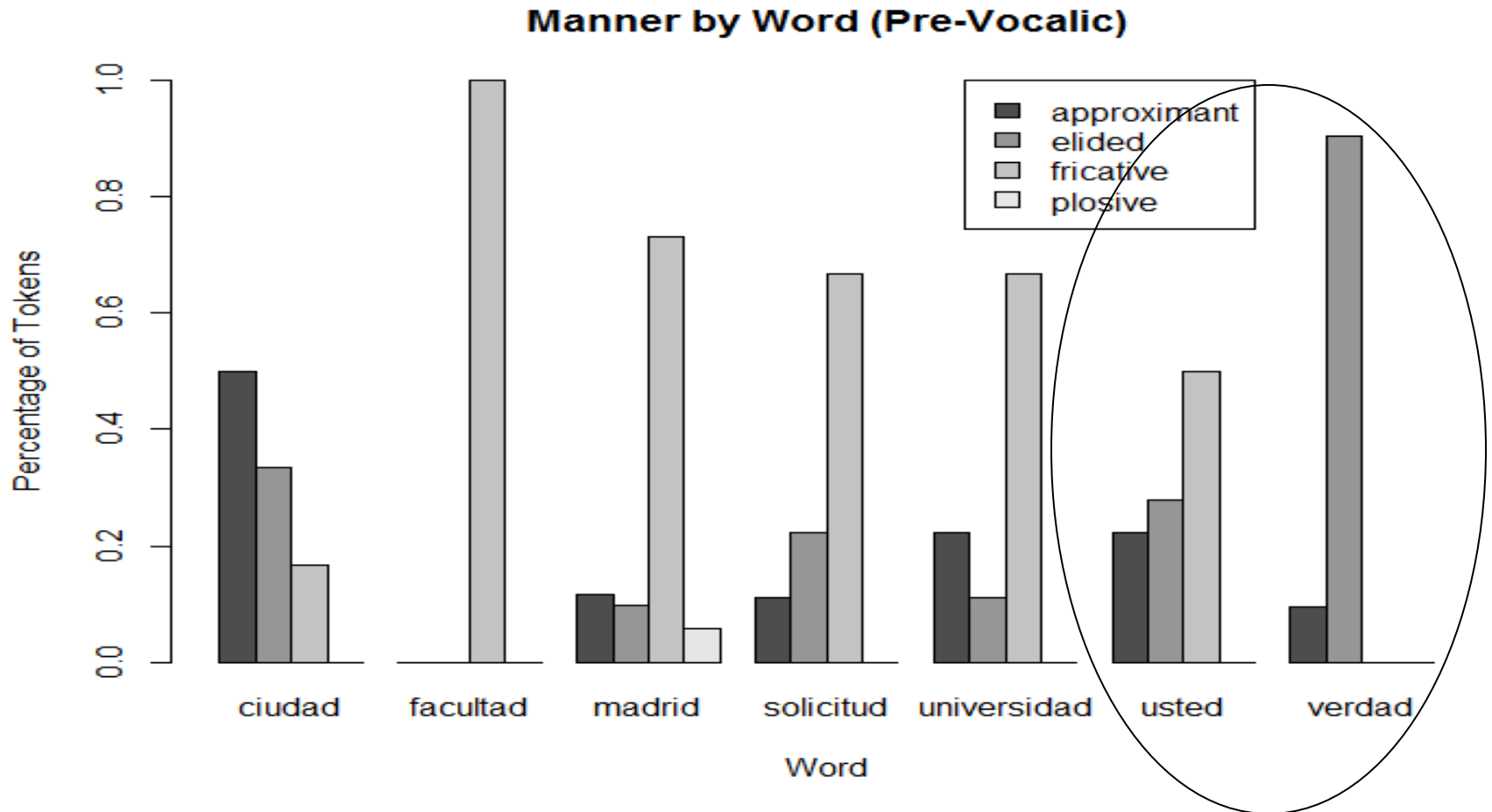
> Devoicing cannot be an intermediate stage towards deletion because /θ/ does not delete and [θ] from /d/ is identical to [θ] from /θ/

Word-specific phonetics?

Words with at least 10 tokens: before pause



before a vowel



Conclusion Final Devoicing in Castilian Spanish

Path I

/d/ → [ð] → 0

Path II

/d/ → [ð] → [θ] / phrase-final → word-final

With word-specific preferences

Final Summary

- Both intervocalic voicing and phrase final devoicing on consonants may be phonologized at the word level.
- The phonologization of intervocalic voicing involves the disappearance of voiceless realizations as an option in specific words. More likely word internally.
- The phonologization of phrase-final devoicing involves the lexicalization of word variants with final devoicing and their use in word-medial contexts.

Model of sound change

- 1) **“Online” effects**: gesture reduction and overlap
/apa/ [apa] ~ [aba] ~ [aβa]
~ [aφa]
- 2) **Conventionalization**: /apa/ [aba]
(Neogrammarian, phonemes are affected,
across word boundaries)
- 3) **Phonemic recategorization**:
[aba] /-p-/ > [aba] /-b-/ (lexically gradual, words
change) → word-boundary effects at this stage.
>> spread of change throughout the lexicon may
be aided by sociolinguistic linkage.

Thank you!