An Acoustic Description of Central Catalan Vowels Based on Real and Nonsense Word Data
Dylan Herrick

AN ACOUSTIC DESCRIPTION OF CENTRAL CATALAN VOWELS BASED ON REAL AND NONSENSE WORD DATA

DYLAN HERRICK

ABSTRACT

This paper examines the extent to which vowel height data taken from real words differs from data taken from nonsense words, and it finds no significant differences. As a result, it provides quantitative acoustic data for the seven stressed and three unstressed vowels of Standard Catalan (as uttered by female speakers). The data are drawn from three distinct phonetic contexts, i.e., /bVp/, /bVt/, and /bVk/, and the /bVp/ context consists entirely of nonsense words (the other contexts were all real words). A comparison and statistical analysis of the data for each vowel phoneme show that there are neither considerable nor statistically significant differences in the vowel height (F1 values) among the data from the three different phonetic contexts. In terms of vowel height, nonsense words provide as accurate a picture of the Catalan data as real words do.

INTRODUCTION

Phonetics and laboratory phonology investigations typically make use of rigorously defined data sets which attempt to control for factors such as word size, syllable count, stress patterns, speech rate, register, focus, intonation patterns, consonantal context, and more. Often, the requirements of these controls prove so restrictive that it becomes impossible to create an appropriate word-list comprised entirely of real, meaningful, and commonly known words. When this happens, investigators must choose either to make use of rare or nonsense words, to relax the degree to which they are controlling the data, or both, and all three of these options can be observed in the literature (Ladefoged, *Phonetic* 3-12).

However, since rare and nonsense words should have a zero (or at the least, an extremely low) rate of occurrence, and since there are measurable frequency effects (e.g., Benki; Bybee "Usage-Based," "Word"; Pierrehumbert), it begs the methodological question: Is it fair to draw conclusions about phonetic or phonological properties based on nonsense words?

Of course, the scope of this question is too large to be addressed in a single paper, so to narrow it down into something more manageable, this study focuses on a single aspect —vowel height in Catalan.

The research presented here is designed to accomplish two tasks; to provide quantitative acoustic data for the Catalan vowel system based on both real and nonsense word data, and in doing so, to examine the methodological question: to what extent is the vowel height data collected from nonsense words different from data collected from real words? In terms of vowel height, the paper finds that there are no statistically significant differences between the three data sets (two from real words, one from nonsense words), and this suggests quite strongly that, at least in terms of acoustics, the production of nonsense is real.

In the past few years, there have been a number of investigations concerned with phonetic properties of the Catalan vowel system. Aguilar et al. (2769-72) discuss acoustic data for /a/ and schwa in both Catalan and Russian vowel reduction. Crosswhite makes crucial use of Catalan in her dissertation on a general theory of vowel reduction ("Vowel Reduction in Optimality Theory" 100-02, 111-14, 141-67; see also Crosswhite's book-article "Vowel Reduction" for a more succinct version of her theory). Herrick's dissertation provides acoustic data for the vowels from six Catalan dialects ("An Acoustic" 113-53), describes the way the vowels are dispersed in a perceptually based F_1 x F_2 vowel space (185-214), and discusses the relevance of this data for our understanding of phonological vowel reduction (214-23). Herrick ("Mid") also discusses the status of the mid-vowels and schwa in rural dialects of Catalan. In addition, Recasens and Espinosa provide data for four dialects of Catalan, discuss the dispersion of vowels in the F_1 x F_2 vowel space, and examine the relevance of that data to our understanding of schwa (645-66). And this is only a partial list which has not touched on the considerable literature from cognitive science (on bilingualism and language acquisition, see, for example, Bosch, Pallier, and Sebastián-Gallés), papers on stress and intonation such as Prieto or Astruc and Prieto, work on coarticulation like Recasens ("Patterns"), or descriptive work on Catalan phonetics such as Recasens (*Fonètica*) or Planas' dissertation. This paper fits into that literature by providing quantitative acoustic data (F_1 and F_2) for the stressed and unstressed vowel phonemes for three female speakers of Central Catalan.

The data in (1) provide an illustration of the vowels of Central Catalan (also considered to be Standard Catalan). Central Catalan allows seven vowels (/i, e, ɛ, a, ɔ, o, u/) in stressed position but only three vowels (/i, ə, u/) in unstressed position (though /e, o/ may appear under conditions involving vowel hiatus). For more information on this and other varieties of Catalan, see Herrick ("An Acoustic"); Mascaró (*Catalan*, "El sistema"); Moll; Recasens (*Estudis*, *Fonètica*); Wheeler (*Phonology*, *The Phonology*); or Wheeler, Yates, and Dols. (Dim. = diminutive.)

(1) Illustration of the stressed and corresponding unstressed vowels of Central Catalan[1]

	STRESSED	UNSTRESSED	CHANGE	SPELLING	GLOSS
a.	'bigə	bi'getə	i → i	biga	'beam' / (dim.)
b.	'bek	bə'kɛt	e → ə	bec	'beak' / (dim.)
c.	'bɛkə	bə'kɛtə	ɛ → ə	beca	'grant' / (dim.)
d.	'bakə	bə'kɛtə	a → ə	vaca	'cow' / (dim.)
e.	'pɔkə	pu'kɛtə	ɔ → u	poca	'little, small' / (dim.)
f.	'bokə	bu'kɛtə	o → u	boca	'mouth' / (dim.)
g.	'buk	bu'kɛt	u → u	buc	'hull, beehive' / (dim.)

In the leftmost column of (1), the initial syllable bears stress, but in the second column of words, the diminutive suffix /-'ɛtə/ (/-'ɛt/ for masculine) has been added, and the stress shifts away from the initial syllable revealing the unstressed realization of each vowel phoneme. By comparing the initial vowels of the words in the first two columns, we can observe that in unstressed syllables, the high front vowel /i/ is realized as [i], the non-high, unrounded vowels /e, ɛ, a/ are realized as [ə], and the rounded vowels /ɔ, o, u/ are realized as [u]. For example, in (1d), /'bakə/ 'cow,' the first vowel /a/ is stressed, but after adding the diminutive affix /-'ɛtə/, the stress shifts away from the initial syllable and the /a/ surfaces as [ə] ([bə'kɛtə]); stressed /a/ corresponds to unstressed [ə]. The neutralization pattern is illustrated in (2).

(2) Central Catalan (Standard Catalan) vowel neutralization pattern

	STRESSED	UNSTRESSED
a.	i ————————→	i
b.	e	
c.	ɛ	
d.	a ————————→	ə
e.	ɔ	
f.	o	
g.	u ————————→	u

---
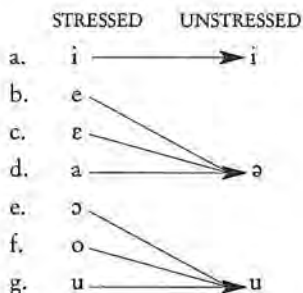
[1] Since this data set is designed to focus on the changes that vowels undergo, spirantization of voiced stops has been omitted. For example, in a more narrow transcription, the /g/ in (1a) would spirantize to [ɣ] since it immediately follows a vowel. Likewise, the word-initial /b/'s found in all the examples except (1e) would spirantize to [β] if uttered in a phrase and immediately preceded by a vowel. For a more details on spirantization in Catalan, the reader should consult general grammars such as Wheeler (*The Phonology*), and references therein.

This neutralization pattern is a form of vowel reduction, a phenomenon which has received a considerable amount of attention in the literature recently (see, for example, the dissertations of Barnes, Crosswhite, or Herrick; book articles by Flemming or Crosswhite; or recent papers by Padgett or Padgett and Tabain). A phonetically based phonological explanation of vowel reduction typical of this literature can be summarized roughly as follows: a) unstressed vowels are shorter in duration than stressed vowels; b) a decrease in duration makes it more difficult to achieve the jaw lowering necessary for low vowels (this results in considerable F_1 formant undershoot for low vowels); c) F_1 formant undershoot effectively raises the floor of the vowel space causing the overall vowel space to shrink (phonetic vowel reduction); d) in this shrunken vowel space, the perceptual distance between vowels also shrinks making neighboring vowels more easily confused; and e) ultimately, this confusion leads to neutralization (phonological vowel reduction).

With this explanation, the boundary between phonetics and phonology has blurred and phonologists have found themselves needing access to detailed acoustic phonetic data in order to test their theories.

For example, at the core of the preceding analysis is the idea that phonological vowel reduction depends in large part on a correlation between (physical constraints on) jaw lowering and F_1 values, and this gives a prediction that, since F_2 does not correlate with jaw lowering, phonological vowel reduction will exhibit a primary reduction of F_1 and not necessarily F_2.

This prediction is testable with quantitative acoustic data (giving F_1 and F_2 values for all the vowels), and while Herrick ("An Acoustic," "Eastern") shows that this is substantiated empirically, the data used to test the prediction come entirely from vowels embedded in nonsense words.

Given that we know real words behave differently from one another based on neighborhood density and frequency effects (Benki; Bybee, "Usage-Based," "Word"; Pierrehumbert), is it fair to assume that the vowel height data taken from nonsense words accurately reflects Catalan? The remainder of this paper suggests that, at least in the case of Catalan vowel height, nonsense is real.

METHODOLOGY

For this study, three native speakers were recorded uttering a combination of real and nonsense words containing the seven vowel phonemes of Central Catalan in both stressed and unstressed position.

The real-word task provided 420 stressed vowel tokens (7 vowel phonemes x 2 word lists x 10 repetitions x 3 speakers = 420 tokens) and 180 unstressed vowels tokens (3 vowels x 2 word lists x 10 repetitions x 3 speakers = 180 tokens). The nonsense-word task provided 210 stressed vowel tokens (7 vowels x 1 word list x 10 repetitions x 3 speakers = 210 tokens) and 90 unstressed vowel tokens (3 vowels x 1 word list x 10 repetitions x 3 speakers = 90 tokens) giving a grand total of 900 tokens for this study.

All target vowels were taken from the initial syllable of the word and preceded by a bilabial consonant; of the 900 tokens, 600 were from real words (300 were followed immediately by a velar consonant, 300 by an alveolar) and 300 from nonsense words (all followed immediately by a bilabial consonant).

All native speakers were female college students (between the ages of 21 and 22) attending the Universitat Autònoma de Barcelona. In an effort to minimize the influence of Spanish, the selected speakers came from the relatively rural regions of Bages (speakers 1 and 2) and Berguedà (speaker 3) instead of urban centers such as Barcelona, where Spanish has a stronger presence. In addition, each speaker stated that Catalan was not only her native language, but her primary language, that both of her parents were native speakers of Catalan, and that Catalan was the language of daily use both at home and with friends. Speakers were paid for their participation.

The speakers participated in three tasks; they provided a short sample of free speech, they read through a list of Catalan real words, and they read through a list of nonsense words. The real and nonsense words were embedded in carrier phrases and the tasks were performed in this order to maximize the quality of recording while minimizing the amount of instructions necessary for each task. The free-speech task was intended to put the speaker in a Catalan-speaking mode and to give her time to get accustomed to speaking into a microphone. In addition, it allowed for adjustments to the recording levels to optimize the quality of the recordings.

The recordings were digitized at 44.1KHz on a SONY TCD-D8 Digital Audio Tape recorder using an AudioTechnica ATM-75 headworn cartoid microphone and analyzed with Boersma and Weenink's *Praat* phonetics software package (version 4.4.30). Formant measurements were made for the 50ms surrounding the midpoint of the vowel, the data were normalized, and the statistical analysis was carried out using SPSS (version 10.0).

Individual differences in physiology (in the shape of the vocal tract) have an effect on the formant values for a given speaker, and as a consequence of this, a reviewer suggests following the lead of Recasens and Espinosa in using Nearey's log-mean based CLIH (Constant Log

Interval Hypothesis) algorithm to normalize the data before it is combined and analyzed.[2]

Disner finds that "a log-mean normalization appears to be the best technique in general" (260), and Adank (3106) suggest that Nearey's CLIH performs well at preserving sociolinguistic variation while reducing individual speaker variation (which, as they point out, agrees with Labov's conclusion about the CLIH given in *Principles of Linguistic Change*). A drawback, however, is that the numerical results of the CLIH fall between negative and positive 1 (in the case of this data set, the range falls roughly between –.5 and .5). In order to make these results more familiar (and easier to interpret and analyze), Disner scales the CLIH data into an idealized F1 x F2 vowel plot (260). This paper follows Disner in that respect and idealizes the F1 range as falling between 300Hz and 1000Hz and the F2 range as falling between 600Hz and 3000Hz (roughly the attested maximum and minimum values from the raw data of the three speakers).[3]

Real-word task

For the real-word task, speakers were presented with a list of real words (written on index cards and randomized) and asked to utter them in a carrier phrase. The list of words contained two near minimal sets, each set containing all seven of the Catalan vowel phonemes in stressed position. In both sets of real words, the phonetic shape is either CVC or CVC, and the first consonant is always a bilabial (usually /b/). For the first set, shown in (3), consonant following the target vowel is velar (usually /k/), and for the second set, shown in (4), the consonant following the target vowel is alveolar (usually /t/).

In an effort to make the word sets as similar as possible, each word begins with a bilabial consonant and contains the target vowel in the initial syllable. The bilabial consonant was chosen since it will interfere

---

[2] The CLIH formula works by subtracting the average natural log-transformed frequency of F1 and F2 for all vowels for speaker k from the natural logarithm of a given formant i for a given vowel j of a given speaker k:

$$F_{ijknorm} = G_{ijk} - \underline{G}$$

where, G stands for the natural logarithm (of formant i, for vowel j, and speaker k) and $\underline{G}$ stands for the average of the natural logarithm of F1 and F2 for all the vowels of speaker k.

[3] For those concerned that the scaling of the data would interfere with the statistical analysis, the statistics were run not only on these scaled data, but on data scaled exactly as Disner did (F1 = 250-750; F2 = 800-2400), on the Nearey transformed data, and on the raw data. The results for F1 do not change —as discussed previously, there are no statistically significant differences between the different contexts for each vowel phoneme.

minimally with the dorsal gesture required to articulate the target vowel. In words of two syllables, the initial unstressed vowel of the second syllable is always schwa.

In addition, all of these words can take the diminutive affix /-'ɛtə/ (or /-'ɛt/ for masculine words), and since this affix bears stress, the diminutive form of each word reveals the unstressed realization of each of the seven underlying vowel phonemes. As such, these two sets of words, combined with the diminutive forms, provide two complete sets of stressed and corresponding unstressed realizations of all seven Catalan vowel phonemes.

(3) Real-word list 1 (velar context): bVk

	VOWEL	WORD	DIMINUTIVE	GLOSS
a.	/i/	'bigə	bi'gɛtə	'beam'
b.	/e/	'bek	bə'kɛt	'beak'
c.	/ɛ/	'bekə	bə'kɛtə	'grant'
d.	/a/	'bakə	bə'kɛtə	'cow'
e.	/ɔ/	'pɔkə	pu'kɛtə	'little, small'
f.	/o/	'bokə	bu'kɛtə	'mouth'
g.	/u/	'buk	bu'kɛt	'hull, beehive'

(4) Real-word list 2 (alveolar context): bVt

	VOWEL	WORD	DIMINUTIVE	GLOSS
a.	/i/	'bit	bi'tɛt	'bit (computer)'
b.	/e/	'pesə	pə'sɛtə	'piece'
c.	/ɛ/	'bɛta	bə'tɛtə	'ribbon'
d.	/a/	'bat	bə'tɛt	'Watt'
e.	/ɔ/	'pɔtə	pu'tɛtə	'foot, paw'
f.	/o/	'botə	bu'kɛtə	'shoe, boot'
g.	/u/	'budə	bu'dɛt	'Buddha'[4]

The list of fourteen words were written (in standard orthography) on index cards, shuffled, and presented to the speakers. After checking the word list for any unfamiliar terms (there were questions about two words, (3g) /buk/ and (4a) /bit/, but these questions disappeared immediately after each speaker saw the words used in context), the speakers were asked to read each word in the carrier phrase shown in (5).

---

[4] On a few occasions, speakers uttered [bu'dɛtə] (followed by a brief pause) instead of the expected [bu'dɛt] implying that the speakers may have been treating *buda* as a feminine noun instead of a masculine one.

(5)   Carrier phrase for real-word task
       Ella diu la paraula _____ a la nena.
       [eʎe diw lə pəraulə _____ ə lə nɛnə]
       'She says the word _____ to the girl.'

In order to minimize vowel-to-vowel coarticulation effects the target word is both immediately preceded and followed by schwa (a neutral vowel) in the carrier phrase. This task was repeated ten times resulting in 140 stressed tokens, and then repeated another ten times with the diminutive form of each word giving another 140 unstressed tokens for each speaker. Since Herrick ("An Acoustic," "Neutralization") has shown that the neutralization of unstressed vowels is complete in Catalan, only unstressed realizations of /i/, /u/, and /a/ (realized as [i, u, ə]) were analyzed (giving 60 unstressed and 140 stressed tokens per speaker).[5]

When working with real words, there are often difficulties in perfectly controlling the word lists. In this study, there were both one and two syllable words for the stressed context and two and three syllable words for the unstressed context (where maintaining a constant syllable length would have been preferred). In addition, the bilabials were drawn from the set of /p, b/, coronals were drawn from the set of /t, d, s/, and velars from the set of /k, g/ (where drawing consonants exclusively from /p, t, k/ would have been ideal).

Thus, the set of words used for the real-word task introduces variables with respect to word length, consonantal context, and lexical context (word choice and frequency). However, when we examine the actual pronunciation of these words within the carrier phrase, some of these concerns become less critical. For example, the word list varies between one and two syllable words (/CVC/ and /CVCə/) at the underlying level, but at the surface level, these word types were both pronounced as [CV.Cə]. This is because the target word was always followed by schwa (from _____ "*a la nena*" [ə lə nɛnə]), and in the CVC words, resyllabification joined the word-final C with the following schwa to give a pronunciation of [CV.Cə] ([CV.Cə lə nɛnə]). On the other hand, with the CVCə shaped words, the word-final schwa and the schwa from the carrier phrase merged and were pronounced as a

---

[5] In the pilot study for this experiment, which only made use of nonsense words, several speakers asked for additional instruction on inserting nonsense words into a carrier phrase. However, after including an initial real-word task with a carrier phrase, these questions disappeared. Speakers were willing to repeat and modify real words without questioning the motivation, and when it came time for the nonsense words, they seemed comfortable modifying the nonsense words in a similar fashion. They had nearly no need for instructions, and they did not question the degree of repetition involved in the task.

single schwa, which also results in a [CV.Cə] pronunciation. Thus, though not a perfect minimal set when comparing words in isolation and at the underlying level, the surface level realizations of the target words are largely identical in syllable structure and length.[6]

Nonsense-word task

The nonsense-word task was designed to elicit vowel reduction data from a set of nonsense verbs differing only in the quality of the target vowel. Using nonsense words, the consonantal context, word length, and syllable structure could be controlled perfectly; in this case, the vowels were surrounded by bilabial sounds so that the consonantal gesture would not interfere with the dorsal gesture of the target vowel. In addition, by making the nonsense words verbs, the stressed and unstressed context could be manipulated through a conjugation task without adding or subtracting a syllable (like the diminutive suffix) to the word length.[7]

For this task, the difference between the infinitival form and the third person singular present tense form of Catalan verbs (ending in *-ar*) was exploited, allowing changes in word stress without changes in word length.[8] (6) gives the verbal paradigm for present tense verbs, and (7) highlights how stress changes for the verbs *cantar, robar,* and *tapar.*

(6)   Regular verb conjugation: *-ar* verbs

Infinitive: cantar [kən'ta]

		SINGULAR		PLURAL	
a.	1st p.	canto	['kantu]	cantem	[kən'tɛm]
b.	2nd p.	cantes	['kantəs]	canteu	[kən'tɛw]
c.	3rd p.	canta	['kantə]	canten	['kantən]

---

[6] This is true of most of the utterances recorded. Some utterances, however, were slower, more deliberate, and contained pauses between the target word and the carrier phrase. In these few instances, there were differences in the syllabic structure and or word-length of the target words.

[7] Eulàlia Bonet and Joan Mascaró (p.c.) point out that this would be possible for nouns and adjectives by using place names and modifying them to mean 'person from (place name)'; e.g., *Roma* ['romə] 'Rome' : *romà* [ru'ma] 'Roman', or with nonsense words *Bipa* ['bipə] : *bipà* [bi'pa]. Again, this allows us to shift the word stress without changing the length of the word. The consonantal context remains constant, so the only change is one of stress.

[8] Controlling the word length is important since vowel reduction has often been linked to duration (as in Moon's and Lindblom's studies), and since word-length affects the duration of syllables (see, for example, Ladefoged, *Phonetic*).

(7) Stress shift between infinitival and 3rd person singular present tense forms

3RD P. SG. PRESENT TENSE FORM		INFINITIVAL FORM		GLOSS
a.	canta [ˈkantə]	cantar	[kənˈta]	'sing'
b.	roba [ˈrobə]	robar	[ruˈba]	'rob'
c.	tapa [ˈtapə]	tapar	[təˈpa]	'cover'

The speakers were given a randomized set of nonsense verbs which had been conjugated in the second person singular present tense. This list is provided in (8) (and only the italicized portion was shown to the speakers). The speakers were told that the verbs in (8) were regular verbs from the *-ar* paradigm, and they were asked to conjugate these verbs appropriately so that they fit in the target sentences shown in (9).

(8) The second person singular form of the nonsense verbs

a.	*bipes* [ˈbipəs]	e.	*bòpes* [ˈbɔpəs]
b.	*bépes* [ˈbepəs]	f.	*bópes* [ˈbopəs]
c.	*bèpes* [ˈbɛpəs]	g.	*bupes* [ˈbupəs]
d.	*bapes* [ˈbapəs]		

(9) Carrier phrase "Always ___ at night" for the stressed form of nonsense words

a.	Tu sempre _____ a la nit.	[tu semprə _____ ə lə nit]
b.	Ell sempre _____ a la nit.	[eʎ semprə _____ ə lə nit]
c.	Ella sempre _____ a la nit.	[eʎə semprə _____ ə lə nit]
d.	Elles sempre _____ a la nit.	[eʎəs semprə _____ ə lə nit]

When presented with a form such as *bipes*, a speaker would respond "Tu sempre [ˈbipəs] a la nit; Ell sempre [ˈbipə] a la nit; Ella sempre [ˈbipə] a la nit; Elles sempre [ˈbipən] a la nit." Thus, the speaker produces four tokens: [ˈbipəs] from (9a), [ˈbipə] from both (9b) and (9c), and [ˈbipən] from (9d). Of these tokens, the first and last were discarded leaving the two identical tokens ([ˈbipə]) from (9b) and (9c).

The speaker was asked to read through the four sentences in (9) for each of the words in (8). The word list in (8) contains all seven of the Catalan stressed vowels, so reading through (9) for each word results in two tokens per vowel. This process was then repeated four more times, giving ten tokens per vowel (7 vowels x 2 third-person singular tokens x 5 total repetitions = 70 tokens). On a few occasions, a speaker skipped a word during one of the repetitions. In this case, the missing token would be replaced by the token from (9a).

Once the conjugation task was completed, the speakers were asked

to read each nonsense verb in its infinitival form in the carrier phrases shown in (10).

(10)   Carrier phrase for the infinitive form of the nonsense words
  a. Sempre _____ a la nit.   [semprə _____ ə lə nit]
  b. La frase és sempre _____ a la nit. [lə frazə es semprə _____ ə lə nit]

When presented with the target word *bipes*, the subject would respond "Sempre [bi'pa] a la nit; La frase és sempre [bi'pa] a la nit." The word-list was repeated ten times, but only the token embedded in (10b) was used for analysis.

Although the token embedded in (10a) was typically discarded, it did serve several purposes. First, it allowed the speaker a chance to practice each new word. The hope was that by interjecting a practice phrase (10a) between each targeted token, any possible long distance influence of the previous target word on the following target word could be minimized. In addition, speakers often took breaths when switching to a new word (and almost never between (10a) and (10b)), and the initial utterance after inhaling was produced with greater intensity (loudness) than following utterances. In practice, the greatest increase in intensity following inhalation was borne by the words uttered in (10a), and thus, keeping (10a) as a practice phrase helped control (minimize) variations in intensity for the targeted tokens embedded within (10b). Finally, if a speaker failed to produce enough tokens (e.g., by skipping a word or by omitting the final repetition) the overall number of tokens for a vowel would be made up by using tokens embedded in the (10a) phrase. This task gave 10 tokens for each unstressed vowel (70 tokens overall: 7 vowels x 10 repetitions). However, since the neutralization of unstressed vowels is complete (see Herrick, "An Acoustic," "Neutralization") only 30 tokens are distinct (those for [i, u, ə]), so only 30 tokens were included in the data for this study. (Those included were the unstressed forms of /i/, /u/, and /a/; /a/ surfaces as [ə] when unstressed.)

Higher formants

Consonants are known to have a coarticulatory effect on adjacent vowels (for a general discussion of English, see Ladefoged, *A Course*; for Catalan, see, for example, Recasens, "An Acoustic"). In this study, the consonant immediately preceding the target vowel is always a bilabial; however, the immediately following consonant varies between a bilabial (300 tokens), an alveolar (300 tokens), and a velar (300 tokens). According to Ladefoged (*A Course* 197), the acoustic

signature of each place of articulation on the C-to-V transitions is as follows: a bilabial consonant exhibits a lowering of the second and third formants, alveolars show an F_2 locus near 1800Hz, and velars show a common origin for both F_2 and F_3.[9] Thus, while we should expect to see similar F_1 values for all three vowel contexts, the F_2 values for the /bVp/ nonsense words (surrounded by bilabials) could be slightly lower than for the other vowels. In addition, the F_2 values for the /bVt/ context (the alveolar context) should be raised among the back vowels (since F_2 will be drawn higher towards 1800Hz in the back vowels).

The principal aim of this paper, however, is to study vowel height, so we will not be concerned with differences in F_2 (frontness, backness). Furthermore, by measuring a 50ms average of the midpoint of the vowel, instead of a single point (or a series of points), the variation in F_2 may not appear to be as prominent in this data as it does in coarticulation studies (which tend to track series of points).

RESULTS

The 900 vowel tokens from the three speakers were measured, recorded in a spread sheet, normalized, graphed as a series of F_1 x F_2 vowel plots, and analyzed statistically (with SPSS version 10.0). The results show that the differences in vowel height that exist between real and nonsense word data are not statistically significant.

Data

Figure 1 shows an F_1 x F_2 vowel plot of all 630 stressed vowels; a diamond (◆) indicates a vowel taken from the /bVk/ context (a real word), a square (■) indicates the /bVt/ context (real word), and the triangle (▲) indicates the /bVp/ context (nonsense word). This figure is interesting for two reasons. First, it illustrates that for each phoneme there is a high degree of overlap between the diamonds, triangles, and squares, and this is exactly what we would expect if the vowels from the three different contexts are similar enough to be representative of one another. In short, it suggests that there are no great differences between real and nonsense words.

---

[9] Ladefoged (*A Course* 196) is careful to point out that the exact value of 1800Hz is merely a guideline; the actual value will vary speaker to speaker.

FIGURE 1
F1 x F2 vowel plot of all 630 stressed vowels
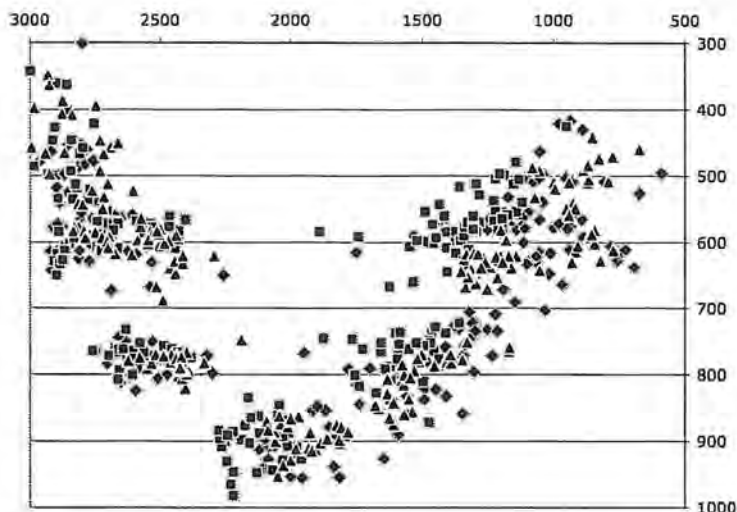
A second observation is that there is no clear boundary between the groupings for the high vowels /i~e/ or /u~o/. This could potentially be a problem since, a priori, one might expect seven distinct groupings —one for each vowel phoneme. Some factors which might lead to a greater degree of overlap among vowel sets are as follows. First, while the Nearey normalization algorithm helps eliminate individual speaker differences, Disner (259-60) also reports that it does well at preserving some sociolinguistic variation. As such, when data are combined for multiple speakers, individual differences in variation may still result in a blurring of the boundaries between vowels. Second, the boundary between /u/ and /o/, in particular, is less distinct than for other vowel pairs for all three speakers on an individual level, so it makes sense that this pattern will reemerge in the combined data. Third, speaker 2 made minimal use of F1 (but great use of F2) to distinguish between /i/ and /e/ (and /o/ and /u/), and this adds to the blurring of the /i~e/ and /o~u/ groupings.

While the vowel plot of Figure 1 gives an idea of the range of variation found in the data set, Figure 2 plots the average values of each vowel phoneme for the three different contexts being compared, and the position of the seven vowel phonemes in the vowel space becomes much more clear. Examining each grouping (by phoneme, for example

the three points ◆ /bVk/, ■ /bVt/, ▲ /bVp/ for /i/) reveals that almost all the groupings agree in vowel height.

FIGURE 2
F1 x F2 plot of stressed vowel averages
(◆ = /bVk/, ■ = /bVt/, ▲ = /bVp/)
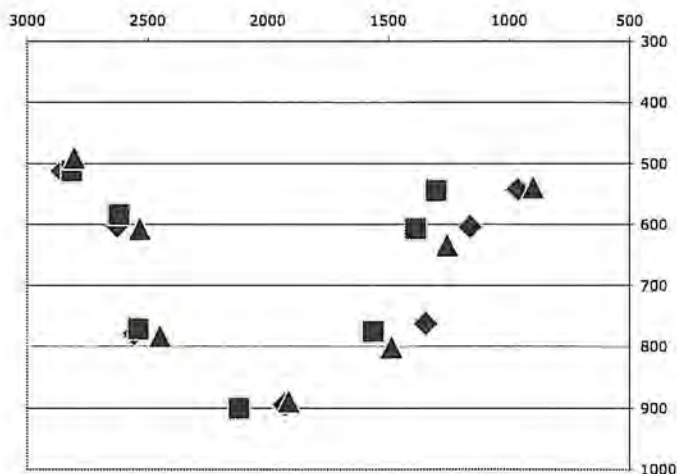
TABLE 1
DATA BY CONTEXT FOR STRESSED VOWEL PHONEMES

Vowel	Context	F1	F2
i	◆ big	513 (84)	2855 (63)
	■ bit	512 (88)	2817 (270)
	▲ bip	491 (75)	2805 (97)
e	◆ bek	603 (43)	2629 (170)
	■ pes	584 (18)	2621 (122)
	▲ bep	607 (27)	2533 (99)
ɛ	◆ bɛk	779 (20)	2559 (117)
	■ bɛt	771 (16)	2541 (161)
	▲ bɛp	782 (15)	2450 (79)
a	◆ bak	895 (43)	1932 (165)
	■ bat	900 (37)	2121 (128)
	▲ bap	889 (27)	1916 (139)
ɔ	◆ pɔk	762 (50)	1345 (110)
	■ pɔt	775 (37)	1563 (106)
	▲ bɔp	801 (36)	1486 (129)

(Table 1 Continuation)

Vowel	Context	F_1	F_2
ɔ	◆ bok	604 (51)	1163 (173)
	▣ bot	608 (38)	1388 (80)
	▲ bop	634 (20)	1258 (58)
u	◆ buk	543 (65)	964 (309)
	▣ bud	545 (40)	1304 (193)
	▲ bup	539 (57)	900 (94)

The numerical values for the vowel points shown in Figure 2 are given in Table 1, which provides the average value of F_1 and F_2 for each context (bilabial, alveolar, and velar) for each vowel phoneme. The standard deviation is indicated within parentheses immediately following the vowel average. For example, looking at the top of the table, the phoneme /i/ was measured in three contexts: /bVk/, /bVt/, and /bVp/. For the velar context, the F_1 value was 513Hz with a standard deviation of 84, and the F_2 value was 2855Hz with a standard deviation of 63. For the alveolar context, the F_1 was 512Hz with a standard deviation of 88 while the F_2 was 2817Hz with a standard deviation of 270. The bilabial context had an F_1 of 491Hz (with a standard deviation of 75) and an F_2 of 2805Hz (with a standard deviation of 97). The remaining vowel data can be read in the same way. Remember also that these are the normalized data, which have been scaled into a vowel space with F_1 ranging from 300-1000Hz and F_2 from 600-3000Hz. For those interested in the raw data, the unnormalized Hz averages are provided at the end of this section.

Interestingly, the most divergent cases appear among the mid back vowels /ɔ/ and /o/. Table 1 indicates that in the nonsense context /ɔ/ is 39Hz higher than the velar context (compare ▲ /bɔp/ = 801Hz to ◆ /bɔp/ = 762Hz; the alveolar context ▣ /pɔt/ falls between the two at 775Hz). This could be attributed to the nonsense vs. real difference, but it seems more likely that it is a combination of factors. Based on standard deviations, the low and low mid back vowels often exhibit more variation than other vowels in Catalan (see, for example, the data in Herrick, "An Acoustic" or in Recasens and Espinosa; although, in the data presented here, /i/ exhibits a greater degree of variation than /ɔ/ based on standard deviations).

An additional point is that for the mid-low back vowel /ɔ/, the nonsense context /bɔp/ begins with a voiced consonant while the realword contexts /pɔt/ and /pɔk/ both begin and end with voiceless consonants,[10] and this voiced/voiceless difference may induce a

---

[10] To follow the exact context for real words in this study, we would need to use

slightly lower Hz value before voiceless consonants. This is something which is seen, for example, in Canadian Raising, where vowels raise — lower their F1 values— in a similar context (see, for example, Chambers; Moreton).

Figure 3 presents an F1 x F2 vowel plot for all the unstressed vowel tokens. Again, a diamond (◆) indicates a vowel taken from the /bVk/ context (real words), a square (■) indicates the /bVt/ context (real words), and a triangle (▲) indicates the /bVp/ context (nonsense words). Figure 4 and Table 2 present the averaged values (and standard deviations) for F1 and F2 for each of the three contexts.

FIGURE 3
F1 x F2 plot of unstressed vowel tokens
(◆ = /bVk/, ■ = /bVt/, ▲ = /bVp/)
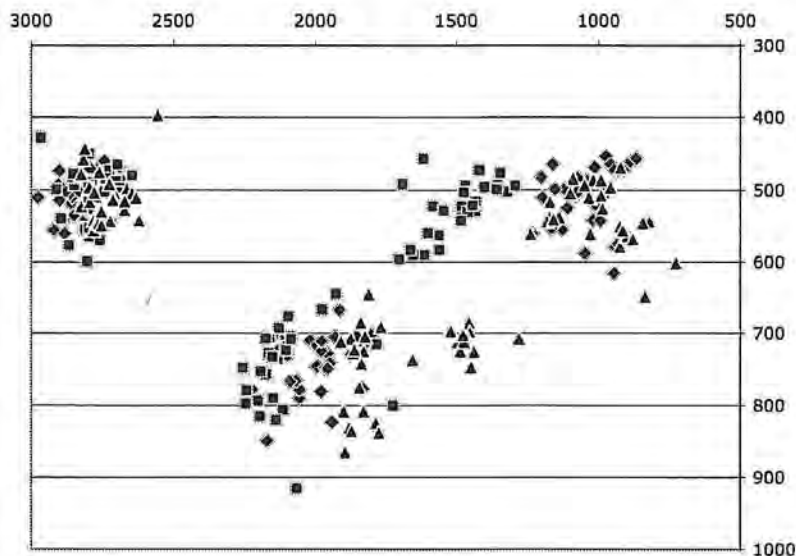
/bɔt/, /bɔtə/, /bɔk/, or /bɔkə/. Unfortunately, these forms are not real or commonly used words. If we change the initial /b/ to a voiceless /p/, we get the real words /pɔt/ and /pɔk/, and these forms were the ones used in this study.

FIGURE 4
F1 x F2 plot of unstressed vowel averages
(◆ = /bVk/, ■ = /bVt/, ▲ = /bVp/)
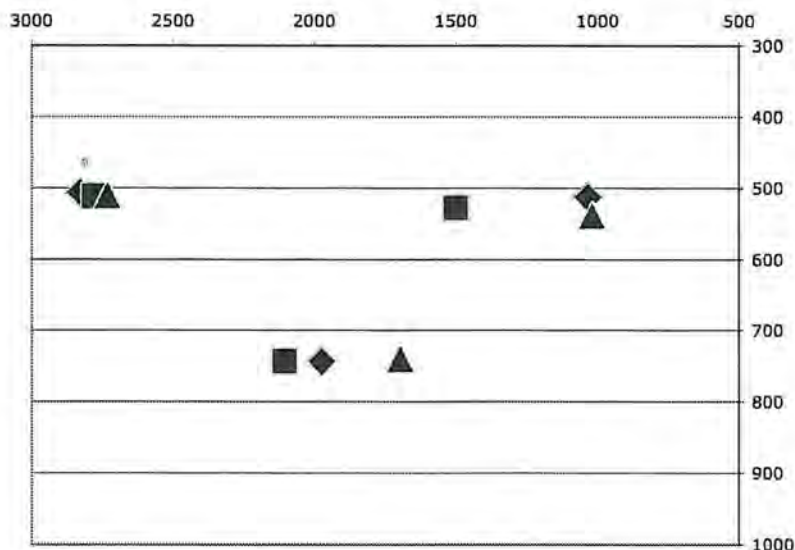
TABLE 2
DATA BY CONTEXT FOR UNSTRESSED VOWEL PHONEMES

Vowel	Context	F1	F2
i	◆ big	506 (28)	2830 (73)
	■ bit	511 (41)	2783 (75)
	▲ bip	509 (34)	2735 (67)
ɔ	◆ bək	743 (43)	1977 (21)
	■ bət	743 (56)	2108 (119)
	▲ bəp	739 (56)	1697 (193)
u	◆ buk	512 (46)	1034 (104)
	■ bud	527 (39)	1502 (116)
	▲ bup	538 (49)	1020 (149)

The situation with the unstressed vowel data is similar to the stressed data. F1 is roughly equal for the three different contexts, and the differences in F2, which are notable in the graphs and tables,

conform to what we expect from coarticulation. What is surprising in the unstressed data is the extent to which the different contexts agree in vowel height for the three vowels; for the vowel /i/, the maximum difference between contexts is 5Hz, 4Hz for schwa, and 26Hz for /u/.

The interim conclusion then is that while there are some differences in vowel height (F1) for the three different phonetic contexts, these differences are, for the most part, very slight. The differences in F1 for data from the real vs. nonsense words appear to be negligible. (For the sake of completeness, the normalized data for all three contexts and speakers has been averaged together and presented as both tables and vowel plots in the appendix.)

Statistical analysis

Statistical analysis provides another way to examine the data. If the data collected from real words and nonsense words are the same (at least with respect to vowel height), then there should be no statistically significant differences between the data collected from the three contexts. Conversely, if we begin with the expectation that there will be differences between the real and nonsense word data, then the lack of statistically significant differences would be surprising. This section shows that there are no statistically significant differences between the three contexts for each phoneme.

All tokens (stressed and unstressed; all three contexts) were subjected to an analysis of variance (ANOVA) with a post-hoc pair-wise comparison (Tamhane's post-hoc comparison assuming non-homogenous variance among the groups, $p < 0.01$) of all groups (/bVp/, /bVt/, and /bVk/ for each vowel phoneme).[11] Table 3 shows the results of the post-hoc pair-wise comparison when we examine the two real-word contexts (/bVt/ and /bVk/). The conclusion is that for all vowel phonemes, there were no statistically significant differences between the two real-word contexts when examining vowel height (F1); i.e., the p-value was less than 0.01 for each of the compared phoneme pairs.

---

[11] The focus of this paper is vowel height (F1), so an examination of F2 (and F3) does not seem relevant to the discussion. However, a statistical analysis of the data show that there are significant differences between the /bVk/ and /bVt/ sets for the back vowels /a/, /ɔ/, and /o/, between the /bVt/ and /bVp/ sets for the vowels /ɛ/, /a/, /o/, and /u/, and between the /bVk/ and /bVp/ sets for the vowel /e/. This is largely consistent with what we would expect; bilabials and velars are the most similar while alveolars exhibit a strong effect on F2 (especially for back vowels).

TABLE 3
STATISTICALLY SIGNIFICANT DIFFERENCES
BETWEEN REAL WORDS ($p < 0.01$)

	/bVt/ compared to /bVk/ (real to real)
F_1 /i/	Not statistically significant
F_1 /e/	Not statistically significant
F_1 /ɛ/	Not statistically significant
F_1 /a/	Not statistically significant
F_1 /ɔ/	Not statistically significant
F_1 /o/	Not statistically significant
F_1 /u/	Not statistically significant

The interesting question at this point is whether the data collected from nonsense words (/bVp/) follow the same pattern as above or not. Table 4 shows that for vowel height, there are no statistically significant differences between the nonsense-word data and either of the real-word data sets —the data collected from nonsense words is the same as that for the two real-word contexts; i.e., again, the p-value was less than 0.01 for all the phoneme pairs that were compared.

TABLE 4
STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN
REAL AND NONSENSE WORDS ($p < 0.01$) FOR F_1

	/bVp/ compared to /bVt/ (nonsense to real)	/bVp/ compared to /bVk/ (nonsense to real)
F_1 /i/	Not statistically significant	Not statistically significant
F_1 /e/	Not statistically significant	Not statistically significant
F_1 /ɛ/	Not statistically significant	Not statistically significant
F_1 /a/	Not statistically significant	Not statistically significant
F_1 /ɔ/	Not statistically significant	Not statistically significant
F_1 /o/	Not statistically significant	Not statistically significant
F_1 /u/	Not statistically significant	Not statistically significant

In short, the statistical analysis supports the same conclusions presented in the previous sub-section. For a given vowel phoneme, nonsense-word data does not show statistically significant differences compared to real-word data in terms of vowel height.

Alternatives

A potential concern might be that the results of Nearey's CLIH are not specifically designed to be scaled so they fit into a more traditional view of the F_1 x F_2 vowel space. Another concern could be that while individual speaker differences add a variable to the data, so too does any normalization algorithm (though, in the case of normalization, the idea is to lessen individual speaker variation).

Examining these alternatives in detail would exceed the scope of this paper; however, the statistical analysis described above was also carried out on: a) the un-normalized raw data; b) the raw data converted to the bark scale (itself a type of normalization); c) the unscaled results of the CLIH; and d) scaled results of the CLIH (following Disner exactly and using a smaller F_1, F_2 range more typical of a male vocal tract). In every case the results were the same; the Tamhane's post-hoc pair-wise comparison showed that there were no statistically significant differences in vowel height between the nonsense-word context (/bVp/) and the two real-word contexts (/bVk/ and /bVt/).

For the sake of comparison and completeness, the data tables for the raw, un-normalized data are given in Tables 5 and 6.

TABLE 5
RAW DATA BY CONTEXT FOR STRESSED
VOWEL PHONEMES — NOT NORMALIZED

Vowel	Context		F_1	F_2
i	◆	big	415 (66)	2685 (107)
	■	bit	415 (72)	2648 (313)
	▲	bip	396 (58)	2605 (163)
e	◆	bek	493 (47)	2335 (229)
	■	pes	473 (17)	2318 (177)
	▲	bep	495 (28)	2188 (139)
ɛ	◆	bɛk	700 (28)	2227 (164)
	■	bɛt	688 (22)	2206 (204)
	▲	bɛp	704 (20)	2074 (103)
a	◆	bak	884 (75)	1496 (154)
	■	bat	893 (68)	1685 (132)
	▲	bap	872 (48)	1478 (126)
ɔ	◆	pɔk	679 (69)	1025 (74)
	■	pɔt	696 (54)	1178 (80)
	▲	bɔp	732 (55)	1123 (89)

(Taula 8 Continuation)

Vowel	Context	F_1	F_2
o	◆ bok	495 (56)	916 (115)
	■ bot	497 (40)	1052 (56)
	▲ bop	523 (21)	968 (36)
u	◆ buk	439 (56)	820 (216)
	■ bud	438 (34)	1004 (24)
	▲ bup	435 (51)	771 (46)

TABLE 6

RAW DATA BY CONTEXT FOR UNSTRESSED
VOWEL PHONEMES — NOT NORMALIZED

Vowel	Context	F_1	F_2
i	◆ big	405 (23)	2644 (122)
	■ bit	410 (34)	2565 (125)
	▲ bip	407 (27)	2488 (106)
ə	◆ bək	653 (58)	1535 (112)
	■ bət	655 (77)	1670 (118)
	▲ bəp	649 (77)	1290 (154)
u	◆ bud	411 (39)	840 (57)
	■ bud	423 (34)	1134 (84)
	▲ bup	433 (46)	834 (83)

CONCLUSION

This paper has provided quantitative acoustic data for the seven stressed and three unstressed vowels of Catalan (as uttered by female speakers). The data are drawn from three distinct phonetic contexts (/bVp/, /bVt/, and /bVk/), and the /bVp/ context consisted entirely of nonsense words (the other contexts contained all real words). A comparison and statistical analysis of the data for each vowel phoneme show that there are neither considerable nor statistically significant differences in the vowel height (F_1 values) among the data from the three different phonetic contexts. Looking at the graphs and data tables, the variation between vowel sets appears primarily along the F_2 dimension — not F_1.

As the boundary between phonetics and phonology continues to blur, the research presented here provides some justification for the use of nonsense words in studies concerning vowel height where complete sets of real words are either unavailable, inappropriate, or impractical.

DYLAN HERRICK
UNIVERSITY OF OKLAHOMA

WORKS CITED

ADANK, Patti. *Vowel Normalization. A Perceptual-Acoustic Study of Dutch Vowels.* Wageningen: Ponsen & Looijen, 2003.

AGUILAR, Lourdes, Sergio BALARI, Joan CASTELLVÍ, Rafael MARÍN, and Teresa VALLVERDÚ. "On the Phonological Interpretation of Vowel Reduction: The Case of Russian and Catalan." *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona, 2003. 2769-772. CD-ROM.

ASTRUC, Lluïsa, and Pilar PRIETO. "Stress and Accent: Acoustic Correlates of Metrical Prominence in Catalan." *ExLing-2006* (2006): 73-76.

BARNES, Jonathan. "Positional Neutralization: A Phonologization Approach to Typological Predictions." Diss. U of California, Berkeley, 2002.

BENKI, Jose. "Quantitative Evaluation of Lexical Status, Word Frequency, and Neighborhood Density as Context Effects in Spoken Word Recognition." *Journal of the Acoustic Society of America* 113 (2003): 1689-705.

BOERSMA, Paul, and David WEENINK. "Praat (Version 4.4.30)," 2006 <http://www.praat.org>.

BOSCH, Laura, Christophe PALLIER, and Núria SEBASTIÁN-GALLÉS. "A Limit on Behavioral Plasticity in Speech Perception." *Cognition* 64 (1997): B9-B17.

BYBEE, Joan. "Usage-Based Phonology." *Functionalism and Formalism in Linguistics. Vol. I: General Papers.* Ed. Michael Darnell, Edith A. Moravscik, Michael Noonan, Frederick J. Newmeyer, and Kathleen Weatley. Amsterdam: Benjamins, 1999. 211-42.

—. "Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change." *Language Variation and Change* 14 (2002): 261-90.

CHAMBERS, Jack. "Canadian Raising." *Canadian Journal of Linguistics* 18 (1973): 113-35.

CROSSWHITE, Katherine. "Vowel Reduction in Optimality Theory." Diss. UCLA, 1999.

—. "Vowel Reduction." *Phonetically Based Phonology.* Ed. Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge UP, 2004. 191-231.

DISNER, Sandra Ferrari. "Evaluation of Vowel Normalization Procedures." *Journal of the Acoustic Society of America* 67 (1980): 253-61.

FLEMMING, Edward. "Contrast and Perceptual Distinctiveness." *Phonetically Based Phonology.* Ed. Bruce Hayes, Robert Kirchner, and Donca Steriade. Cambridge: Cambridge UP, 2004. 232-76.

HERRICK, Dylan. "An Acoustic Analysis of Phonological Vowel Reduction in Six Varieties of Catalan." Diss. U of California, Santa Cruz, 2003.

—. "Neutralization is Complete in Catalan Vowel Reduction." *The Proceedings of the 2004 Linguistic Society of Korea International Conference, Yonsei University, Seoul, Republic of Korea, July 28ᵗʰ-30ᵗʰ* (2004): 91-100.

—. "Mid Vowels and Schwa in Eastern Catalan: Five Non-Barcelona Dialects." *New Perspectives on Romance Linguistics.* Ed. Jean-Pierre Montreuil. Amsterdam: Benjamins, 2006. 113-26.

—. "Eastern Catalan Vowel Reduction is Characterized by Raising — Not Centralization." *Studies in Language Sciences 6.* Tokyo: Kurosio, 2007. 227-41.

LABOV, William. *Principles of Linguistic Change. Vol. II: Social Factors.* Oxford: Blackwell, 2001.

LADEFOGED, Peter. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques.* Oxford: Blackwell, 2003.

—. *A Course in Phonetics.* 5ᵗʰ ed. Boston: Thomson Wadsworth, 2006.

LINDBLOM, Björn. "Spectrographic Study of Vowel Reduction." *Journal of the Acoustic Society of America* 35 (1963): 1773-81.

MASCARÓ, Joan. *Catalan Phonology and the Phonological Cycle.* Bloomington: Indiana Linguistics Club, 1978.

—. "El sistema vocàlic: reducció vocàlica." *Gramàtica del català contemporani.* Dir. Joan Solà, Maria-Rosa Lloret, Joan Mascaró, and Manuel Pérez Saldanya. Vol. 1. Barcelona: Empúries, 2002. 89-123.

MOLL, Francesc de Borja. *Gramàtica històrica catalana.* València: U de València, 1991.

MOON, Seung-Jae, and Björn LINDBLOM. "Interaction Between Duration, Context, and Speaking Style in English Stressed Vowels." *Journal of the Acoustic Society of America* 96 (1994): 40-55.

MORETON, Elliot. "Realization of the English Postvocalic [voice] Contrast in F1 and F2." *Journal of Phonetics* 32 (2004): 1-33.

NEAREY, Terrance. *Phonetic Feature Systems for Vowels.* Bloomington: Indiana Linguistics Club, 1978.

PADGETT, Jaye. "Russian Vowel Reduction and Dispersion Theory." *Phonological Studies* 7 (2004): 81-96.

PADGETT, Jaye, and Marija TABAIN. "Adaptive Dispersion Theory and Phonological Vowel Reduction in Russian." *Phonetica* 62 (2005): 14-54.

PIERREHUMBERT, Janet. "Word-Specific Phonetics." *Papers in Laboratory Phonology 7.* Ed. Carlos Gussenhoven and Natasha Warner. Berlin: Mouton de Gruyter, 2002. 101-40.

PLANAS MORALES, Sílvia. "Identificació de les vocals tòniques del català." Diss. U de Barcelona, 2000.

PRIETO, Prieto. "Tune-Text Association Patterns in Catalan: An
    Argument for Hierarchical Structure of Tunes." *Probus* 14 (2002):
    173-204.
RECASENS, Daniel. *Estudis de fonètica experimental del català oriental
    central.* Barcelona: Abadia de Montserrat, 1986.
—. "An Acoustic Analysis of V-to-C and V-to-V Coarticulatory
    Effects in Catalan and Spanish." *Journal of Phonetics* 15 (1987): 299-
    312.
—. *Fonètica descriptiva del català. (Assaig de caracterització de la
    pronúncia del vocalisme i consonantisme del català al segle XX.).* 2nd
    ed. Barcelona: Institut d'Estudis Catalans, 1996.
—. "Patterns of Jaw Coarticulatory Direction and Linguomandibular
    Coordination in VCV Sequences." *ZAS Papers in Linguistics* 28
    (2002): 83-92.
RECASENS, Daniel, and Aina ESPINOSA. "Dispersion and Variability of
    Catalan Vowels." *Speech Communication* 48 (2006): 645-66.
WHEELER, Max W. *Phonology of Catalan.* Oxford: Blackwell, 1979.
—. *The Phonology of Catalan.* Oxford: Oxford UP, 2005.
WHEELER, Max W., Alan YATES, and Nicolau DOLS. *Catalan: A Com-
    prehensive Grammar.* London: Routledge, 1999.

APPENDIX

NORMALIZED DATA FOR ALL THREE CONTEXTS
AND SPEAKERS AVERAGED TOGETHER

For clarity, the stressed and unstressed vowels are kept separate; the stressed vowels are shown in Figure 1 and Table 1, and the unstressed vowels are shown in Figure 2 and Table 2.

FIGURE 1
COMBINED (NORMALIZED) DATA FOR ALL SPEAKERS AND CONTEXTS
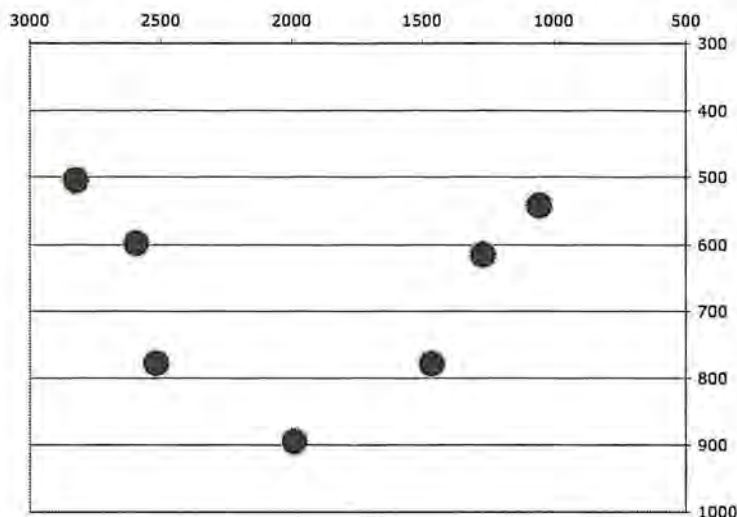—STRESSED VOWELS

FIGURE 2
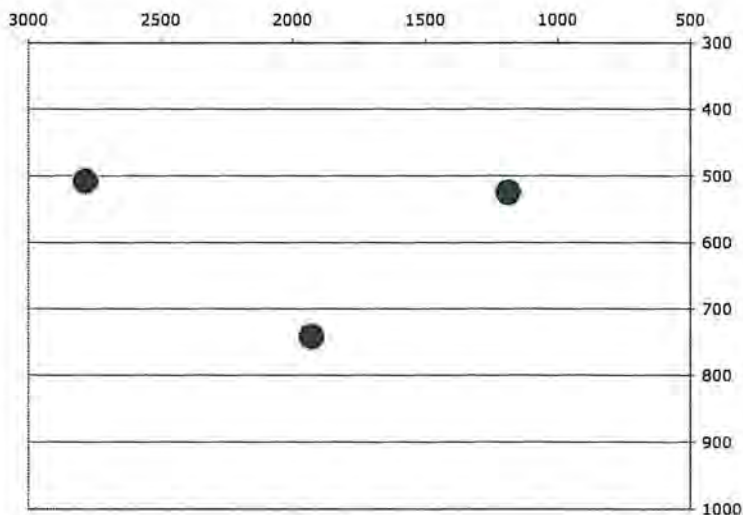COMBINED (NORMALIZED) DATA FOR ALL SPEAKERS AND CONTEXTS
—UNSTRESSED VOWELS

TABLE 1
COMBINED DATA FOR ALL SPEAKERS AND CONTEXTS —STRESSED VOWELS

Vowel	Average F1	Average F2
i	505 (82)	2862 (169)
e	598 (64)	2594 (204)
ɛ	778 (17)	2516 (131)
a	894 (37)	1989 (172)
ɔ	779 (44)	1465 (146)
o	615 (41)	1269 (147)
u	542 (54)	1056 (279)

TABLE 2
COMBINED (NORMALIZED) DATA FOR ALL SPEAKERS —UNSTRESSED VOWELS

Vowel	Average F1	Average F2
i	509 (34)	2783 (81)
ə	742 (51)	1927 (225)
u	526 (46)	1185 (257)