Moments of moments: Acoustic phonetic character and within-category variability of the Basque three-sibilant contrast

Zachary Jaggers and Melissa Baese-Berk

University of Oregon
zjaggers@uoregon.edu, mbaesebe@uoregon.edu

ABSTRACT

The Basque language has a rare three-sibilant contrast: post-alveolar, apico-alveolar, and lamino-alveolar. Previous work has identified the existence of this distinction; however, relatively little attention has been paid to documenting and analyzing its acoustic phonetic character. The current study analyzes the production of this distinction by 28 native Basque speakers, paying special attention to within-category variability.

Statistical analysis confirms the three-way contrast. Multiple acoustic properties that could distinguish the categories are examined, finding that only one seems necessary: spectral centre of gravity (COG). Within-category variability seems to be influenced by this three-way contrast, where distribution skewness is more positive for categories of lower COG means, and distribution of the central category exhibits less variance. To consider how neighboring categories (and the number thereof) relate within the acoustic space of interest, results are compared with the English two-sibilant system.

Keywords: speech production, variability, dispersion, sibilants, under-documented languages

1. INTRODUCTION

The Basque sound system has a six-way distinction of sibilants that is considered typologically rare. There are post-alveolar, apico-alveolar, and lamino-alveolar categories, which have fricative and affricate counterparts that are contrastive with each other.

Previous work has identified the existence of this distinction, as well as deliberated its historical development and regional attestation.[7,8,13] For example, in some varieties, there also appear to be contrastive voiced counterparts.[7,8] And, regarding some varieties, some have suggested that the two anterior categories may have merged while in others they remain distinct.[7,13] Research has also examined how non-native speakers learn the perception of this contrast.[2] However, relatively little attention [cf.1] has been paid to documenting and analyzing the acoustic phonetic character of this distinction in speech production.

At least two other languages are well known to have three-sibilant systems: Polish and Mandarin. However, these systems differ from that of interest here, with the three categories usually characterized as dental, alveopalatal, and retroflex.[5,11,18] Given that this contrast is typologically rare and phonetically under-documented, the current study analyzes the voiceless fricative members of this differently populated sibilant system. Special attention is given to comparing multiple acoustic properties that could distinguish the three categories and to their variability within the acoustic space, as well as comparing this variability to the English two-way sibilant system. While sibilants have previously been considered quite consistent and stable in their realization [4,16], further work has identified that they still exhibit within-category variability, both of static and dynamic acoustic measurements [9,10,17].

For example, in a Dispersion-Theoretic account [5,12], we might expect this three-way system to lead to a more widely dispersed use of the acoustic space in comparison to a two-way system. However, Recasens and Espinosa [15] observe that a more crowded phonemic system leads not necessarily to more phonetic dispersion but to contrast-sensitive limitations on variability: The Majorcan variety of Catalan is identified to have a phonemic /ə/ category while other varieties exhibit [ə] only allophonically. While the peripheral vowel categories neighboring the acoustic space of /ə/ are not further dispersed in this variety than in other varieties, they exhibit less variability encroaching on its acoustic space.

We may, in the case at hand, see a similar influence of this more crowded phonemic system on category variability. It may not be that the three-way system has more widely dispersed category means than the two-way system. Rather, categories may avoid encroaching upon each other's acoustic spaces by means of within-category variability, such as by outer categories exhibiting non-normal distributions.

2. METHODS

Adult native speakers of Basque (N=28) performed a speech production task in a sound-attenuated room, audio-recorded at a 44.1kHz sampling rate. All speakers were also bilingual in Spanish. The task was designed to elicit utterances of the three categories: post-alveolar [ʃ], apico-alveolar [s̺], and lamino-alveolar [s̻] (henceforth referred to by their orthographic representations: <x>, <s>, and <z> respectively). Each was the first segment in a nonce syllable with a consistent [Cɑ] format. The purpose of this task was to elicit endpoints of three sound-continua that would be synthesized for a future perception study: <s>-<x>, <s>-<z>, and <s>-<ts> (<ts> representing the affricate counterpart). These were randomly cycled through 10 times, leading to 10 utterances of <x, z> and 30 of <s>.

For the comparison English data, adult native speakers of American English (N=27) performed a word list reading task in a sound-attenuated room, audio-recorded at a 44.1kHz sampling rate. For each category (alveolar [s] and post-alveolar [ʃ], henceforth referred to by orthographic <s> and <sh>), utterances of 4 separate words of [ˈCɑ…] form were elicited in 3 randomized cycles (including fillers).

Sound files were segmented in Praat by trained assistants, found to be highly consistent (95% CI of .965 < ICC < .99). The start of each fricative of interest was identified as the onset of aperiodic noise at the start of the single-word utterance, and the end was identified as the onset of periodicity attributable to the following vowel. A Praat script was used to extract measurements from each segment: the first four spectral moments centered around the midpoint of the segment—centre of gravity, variance, skewness, and kurtosis—as well as segment duration.

3. ANALYSIS

First, the Basque data will be discussed, for documentation and analysis of within-category variability. Then, the variability of the three-place Basque system will be compared with the variability of the categories in the English two-place system.

3.1. Character of the distinction

The five measurements of each segment utterance, along with its category ID, were submitted to a Linear Discriminant Analysis (LDA) model in the R statistical programming environment using the `lda()` function provided by Displayr's `flipMultivariates` package [6]. This model uses all dependent variables provided to disperse *n* number of categories as widely as possible in a space of *n*-1 dimensions. Each dependent variable can then be assessed for how strongly it contributed to the final model's category discrimination. Table 1 provides the results of this assessment along with the modeled means of each category. What the results suggest is that, in spite of using all measurements provided, only spectral centre of gravity (henceforth COG) significantly contributes to category discrimination. In other words, while a multidimensional approach to distinguishing the three categories is possible, they can be distinguished by COG in a unidimensional approach. In light of these results, the remainder of this study will analyze only COG as the acoustic dimension of this contrast. The successive {x < s < z} increase in COG parallels previous findings [1][i] and will motivate further referring to the categories of <x> and <z> as the "outer" categories and <s> as the "inner" category in this shared acoustic space.

Table 1: Category means and evaluation of significance of contribution to LDA model.

	<x> n=276	<s> n=826	<z> n=277	*p*
COG	7200Hz	7507Hz	7780Hz	<.001
variance	2029	2039	2079	.999
skewness	1.41	1.38	1.25	.141
kurtosis	4.86	5.15	4.27	.276
duration	0.19s	0.18s	0.18s	.1

Linear mixed-effects regression (LMER) modeling corroborates that this is a three-way distinction (Table 2). Using the `lmer()` function provided by the `lme4` package [3] in R, a model of COG treating 'category' as a three-level factor was generated alongside an identical model treating it as a two-level factor. In the two-level model, the anterior <s> and <z> categories were collapsed, given some suggestions that these two categories may be merged in at least some varieties.[8,13] That appears to not be the case in this dataset, given that an ANOVA test between the two models identifies the three-level model to provide a significantly better fit of the data ($\chi^2(1)=7.759$, $p=.0053$).

Table 2: LMER of COG including sound category and speaker sex as fixed effects, with random terms for speaker intercept and speaker-by-category slope; *p*-values (here and in all following tables) come from ANOVA tests between the final model and that excluding the fixed effect of interest and any related random effect.

COG	Est.	Std.Err.	*t*	*p*
Intercept: <x>, F	7701Hz	136Hz	56.54	
category: <s>	312Hz	95Hz	3.28	.0015
category: <z>	575Hz	121Hz	4.77	
sex: M	-931Hz	150Hz	-6.22	<.0001

3.2. Within-category variability

Now, we turn to analyzing the *statistical* moments of the acoustic spectral moment identified as necessary for distinguishing the three categories: COG. For example, henceforth skewness does not refer to the shape of the spectrum of a single utterance and which frequency ranges exhibit greater amplitude. Instead, it refers to the shape of the distribution of variability across the utterances pertaining to a sound category and whether COG values higher or lower than the category's mean COG are more common.

For the following analyses of these statistical moments of variability, the variability of concern is within-speaker, within-category variability. That is, for every speaker, there is one value per category which describes a different aspect of that category's variability across that speaker's realizations of it.

Figure 1: Statistical moments of within-speaker, within-category COG variability.

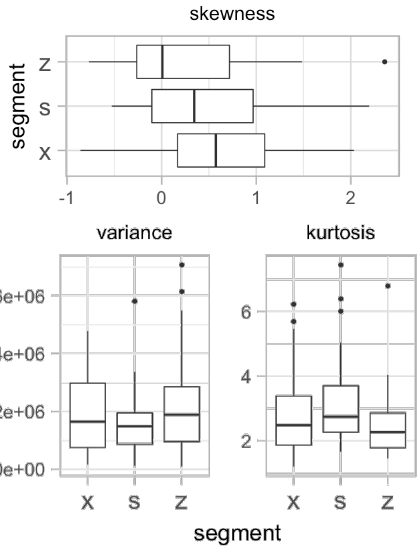

Table 3: LMER of statistical moments of within-speaker, within-category COG variability—category as the binary factor of "inner" vs. "outer" for analyses of variance and kurtosis.

SKEWNESS	Est.	Std.Err.	*t*	*p*
Intercept: <x>	.645	.139	4.62	
category: <s>	-.146	.137	-1.07	.028
category: <z>	-.367	.137	-2.69	
VARIANCE	Est.	Std.Err.	*t*	*p*
Intercept: <s>	1680357	289378	5.81	
category: <x, z>	433726	208726	2.08	.039
KURTOSIS	Est.	Std.Err.	*t*	*p*
Intercept: <s>	3.23	.25	12.87	
category: <x, z>	-.53	.25	-2.13	.035

3.2.1. Category distribution skewness

As discussed above regarding the Majorcan variety of Catalan with a phonemic central /ə/ category, the categories neighboring it exhibit less variance into the central acoustic space than those same categories do in varieties without phonemic /ə/.[15] Given this, it is hypothesized that the variance of the outer <x> and <z> categories in this three-sibilant system will favor COGs further from the mean of the inner category, meaning that the lower-COG <x> category will have a higher skewness and vice versa.[ii]

This hypothesis is corroborated. While the variance of all three categories tends toward a more positive skewness, this skewness is inversely correlated with the mean COG of the category. That is, the <x> category, with a lower mean COG than the inner category, exhibits a distribution of variability more commonly favoring realizations of a lower COG. And, vice versa, the <z> category, with a higher mean COG than the inner category, exhibits not only a higher mean COG but also a distribution of variability more commonly favoring realizations of a higher COG. This is apparent in the results plotted in Figure 1 and identified as significant (Table 3).

3.2.2. Category crowding

The inner <s> category of the Basque three-place system is hypothesized to exhibit a crowding effect: i.e., It does not vary as much as the outer categories since, if it did, it would encroach upon the acoustic space associated with a neighboring category and be potentially confoundable with that category.

This crowding effect is apparent in the results regarding variance. The inner <s> category exhibits significantly less variance (Table 3). Or, at least, it exhibits less variance *of variance* across speakers (see Figure 1): i.e., If a speaker does exhibit more variance of a category, it will be an outer category. It was also hypothesized that <s> would exhibit a lower degree of kurtosis, straying from its mean less commonly; however, the reverse was found, where the outer categories exhibit significantly less kurtosis. Therefore, while the crowding of the inner category appears to confine its range of variability, it does not appear to confine that variability's shape.

3.3. Comparison with a two-place system

The acoustic phonetic realization of the English system mirrors the Basque system, where the more anterior <s> ([s]) is distinguished from <sh> ([ʃ]) by a higher COG, as previously documented [4,10,i.a.] and confirmed in the LMER results provided in Table 4. Two immediate observations run counter to what a Dispersion-Theoretic account [5] would predict. For one, the range between the two categories appears greater than that between the outer members of the Basque three-category system. Also, the two categories' mean COGs are lower than would be expected: not neatly situated between Basque's three. Given this, the most acoustically similar anterior categories across the two systems will be treated as analogous counterparts in the cross-linguistic comparison to come: each language's <s> categories.

Figure 2 plots the statistical moments of within-speaker, within-category variability in English. Table 5 provides the results of LMER comparisons between these attributes of the two English categories. Of peak interest in this study with regards to English, however, is how the within-category variability across categories in the two-place system compares to that of the three-place Basque system. Table 6 provides the results of LMER analysis of within-speaker, across-category differences of these statistical moments, considering how these may differ across the two languages.

Table 4: English: LMER of COG including segment category and speaker sex as fixed effects.

COG	Est.	Std.Err.	*t*	*p*
Intercept: <sh>, F	5558Hz	84Hz	66.09	
category: <s>	2318Hz	153Hz	15.07	<.0001
sex: M	-562Hz	147Hz	-3.83	<.0001

Figure 2: English: Statistical moments of within-speaker, within-category COG variability.

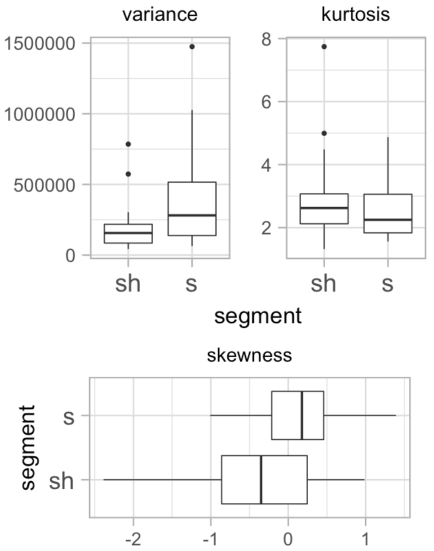

Table 5: English: LMER of statistical moments of within-speaker, within-category COG variability.

Skewness	Est.	Std.Err.	*t*	*p*
Intercept: <sh>	-.393	.134	-2.93	
category: <s>	.553	.184	3.01	.004
Variance	**Est.**	**Std.Err.**	***t***	***p***
Intercept: <sh>	190378	54441	3.50	
category: <s>	20837	76991	2.71	.0076
Kurtosis	**Est.**	**Std.Err.**	***t***	***p***
Intercept: <sh>	2.90	.21	13.69	
category: <s>	-.38	.30	-1.28	.196

Table 6: Across-language, within-speaker, across-category statistical moments and COG range, where Intercept = Basque.

Skew Diff	Est.	Std.Err.	*t*	*p*
Basque: <x-z>	.37	.16	2.25	
English: <sh-s>	-.92	.23	-3.95	.0002
Var Diff	**Est.**	**Std.Err.**	***t***	***p***
Basque: <x-s>	285758	173572	1.65	
English: <sh-s>	-494128	247730	-1.99	.0512
COG Range	**Est.**	**Std.Err.**	***t***	***p***
Basque: <z-x>	579Hz	136Hz	4.25	
English: <s-sh>	1737Hz	194Hz	8.94	<.0001

Regarding skewness, English shows the opposite pattern, where skewness is directly correlated with COG mean rather than inversely as observed for Basque. What this may mean is that, for the Basque outer categories, there is a favoring of realizations further from the inner category mean, and outliers are considered those that encroach upon the inner category's acoustic space. For English, the same constraint does not seem to be at play, and each category is actually more commonly realized with a more central COG; outliers may be tokens especially dispersing or hyper-distinguishing the two categories. This difference in skewness between the two English categories is identified as significantly different from that between the Basque outer categories (Table 6).

Regarding crowding, recall that the Basque inner category <s> exhibited significantly less variance than its outer neighbors. This was compared to the variance of its analogous counterpart <s> category in English, not flanked on both sides by neighboring categories but only by one. Indeed, the variance of the English <s> category seems less constrained than that of its analogous <s> category in Basque, which has a significantly greater variance than its own neighboring <sh> category (Table 5). When comparing each language's <s> category with its respective posterior neighboring category, this across-language difference is identified as near-significant (Table 6). Finally, while not an aspect of within-category variability, it was also identified as significant (Table 6) that the range of the acoustic space differs across the two systems, where the difference between the two English categories' is greater than that between the outer Basque categories.

4. CONCLUSIONS

The Basque three-sibilant system seems to be affected by its more crowded acoustic space. The inner category exhibits more constrained within-category variability, both when compared to the outer categories in the same system and to a similar category in a two-sibilant system not flanked on both sides by contrastive neighbors. Additionally, the skewness of each outer category's distribution of variability suggests that realizations further from the inner category's acoustic space are more commonly favored. Meanwhile, in a two-sibilant system, within-category skewness is the opposite, where each category appears to more commonly favor more central realizations in the shared acoustic space, in spite of such realizations being closer to the acoustic space of a neighboring category. However, the boundaries of the shared acoustic space itself may not be as sensitive to the number of contrasts within it: The range across which the three contrastive categories span is tighter than that of a two-sibilant system, similar to previous findings [15] suggesting that phonemic crowding may affect within-category variability more than across-category dispersion.

5. REFERENCES

[1] Álvarez, J. L. 1982. Some acoustic data about the three Basque sibilants. *Proc. First Int. Basque Conf. in North America*, 18–34.

[2] Baese-Berk, M., Samuel, A. 2016. Listeners beware: Speech production may be bad for learning speech sounds. *J. Memory and Language* 89, 23–36.

[3] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear-mixed-effects models using lme4. *J. Statistical Software* 67, 1–48.

[4] Behrens, S., Blumstein, S. 1988. Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *J. Phonetics* 16, 295–298.

[5] Boersma, P., Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25, 217–270.

[6] Displayr. 2018. R package: flipMultivariates. https://github.com/Displayr/flipMultivariates

[7] Egurtzegi, A. 2013. Phonetics and Phonology. In: Martínez-Areta, M. (ed.), *Basque and Proto-Basque: Language-internal and typological approaches to linguistic reconstruction*. Bern: Peter Lang, 119–172.

[8] Hualde, J.I., 2004 [1991]. *Basque phonology*. Abingdon: Routledge.

[9] Iskarous, K., Shadle, C., Proctor, M. 2011. Articulatory-acoustic kinematics: The production of American English /s/. *J. Acoust. Soc. Am.* 129, 944–954.

[10] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252–1263.

[11] Li, F., Jan, E., Beckman, M. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. *Proc. 16th ICPhS* Saarbrücken, 917–920.

[12] Liljencrants, J., Lindblom, B. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839–862.

[13] Michelena, L. 1985 [1961]. *Fonética histórica vasca* (3rd ed.). San Sebastián: Diputación de Guipúzcoa.

[14] Jurado Noriega, M. 2011. Caracterización de sibilantes fricativas vascas y su percepción en el sistema fonético español. *Int. J. Basque Linguistics and Philology* XLV, 81-137.

[15] Recasens, D., Espinosa, A. 2006. Dispersion and variability of Catalan vowels. *Speech Comm.* 48, 645–666.

[16] Recasens, D., Espinosa, A. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *J. Acoust. Soc. Am.* 125, 2288–2298.

[17] Reidy, P. 2016. Spectral dynamics of sibilant fricatives are contrastive and language specific. *J. Acoust. Soc. Am.* 140, 2518–2529.

[18] Zygis, M., Hamann, S. 2003. Perceptual and acoustic cues of Polish coronal fricatives. *Proc. 15th ICPhS* Barcelona, 395–398.

---

[i] These successive {x < s < z} COG results also resemble how some have described the Basque <s> category as slightly posterior, resembling the Castilian Spanish /s/ [1], and the <z> category as dental [2,14]. If considering COG to be directly correlated with anteriority of articulation and length of the chamber of turbulence [9], these acoustic results further corroborate such descriptions and motivate future articulatory analyses.

[ii] Of course, there are different ways that statistical skewness can be interpreted. For example, it could be predicted that the lower-COG category should have a lower skewness, since a lower skewness would mean that more of the 'tails' or 'outliers' of the distribution fall below the category mean rather than above it, therefore not encroaching as much on the acoustic space of the inner of the three categories. However, we make our prediction with respect to where the majority of the non-outlier data fall within its range of variability, rather than the 'tails' of the distribution. In this respect, the lower-COG category is predicted to have a higher skewness because, within the range of possible realizations, most of the time realizations will be favored that are further away from the inner category: hence the median being further from the inner category than the mean.