A NASALANCE-BASED STUDY OF THE /h/ VS. /h̃/ OPPOSITION IN ZUBEROAN BASQUE

Ander Egurtzegi,[1,2] Andrea García-Covelo[2,3,4] & Iñigo Urrestarazu-Porta[1,2,4,5]

[1]CNRS, [2]IKER-UMR5478, [3]IPS-LMU, [4]UPPA & [5]UPV/EHU
ander.egurtzegi@iker.cnrs.fr, andrea.garcia@phonetik.uni-muenchen.de & inigo.urrestarazu-porta@iker.cnrs.fr

ABSTRACT

Eastern varieties of Basque have been described with an extremely rare opposition between an oral /h/ and a nasalized aspirate /h̃/ (1), which some researchers considered impossible (see (2)). This paper presents the first nasalance-based study of the /h/ vs. /h̃/ contrast in Basque, with data from the endangered Zuberoan variety.

We report the production of a reading task by 5 participants from the village of Larraine (4 male, 1 female; age range 60-70) including items with nasalized and oral aspirates (e.g. *ih̃ue* 'no one' vs. *aihai* 'dinner'), with oral aspirates comprising both words with only oral segments and words with a nasal stop that might trigger nasal assimilation (3).

Our results suggest that the /h/ vs. /h̃/ opposition is still present in Larraine Basque, although some lexical items sporadically lost nasality, and some speakers have completely merged both segments by consistently producing /h/.

Keywords: Nasal aspiration; nasalance; rare sound patterns; endangered languages; Zuberoan Basque

1. INTRODUCTION

Two endangered varieties of Basque (basq1248), Zuberoan (1) and Mixean (4; 5), show evidence of a scarcely attested phonological opposition between an oral aspirate /h/ and a nasalized /h̃/ (1; 3; 6). This opposition has been proposed as a potential analysis in only a handful of languages, and some authors deemed it theoretically impossible (2). In the Basque varieties that show it, some aspirates are produced with audible nasalization that spans a whole [VCV] sequence, resulting in sequences phonetically realized as [Ṽh̃Ṽ] (7; 8), where nasality is phonologically analyzed as originating in /h̃/ and then spreading to the surrounding vowels (1; 3). However, due to the difficulty of obtaining recordings of these endangered varieties, there are very few studies that present phonetic evidence of the /h/ vs. /h̃/ opposition (9). This study is the first to present nasalance data of this contrast, with recordings of speakers from Larraine (Larrau in French), a small village in the South of Zuberoa.

1.1. Zuberoan Basque

The historical province of Zuberoa (Soule in French) is located in the South-West of France, in the Pyrénées-Atlantiques department, and the north-eastern corner of the Basque Country. Its population is in clear recession, from 15350 in 2006 to 12716 as of 2019 (10). Zuberoa is formed by 43 villages, the main town being Maule-Lextarre (Mauléon-Licharre, 2947 inhabitants in 2019). The village of Larraine (195 inhabitants in 2019), where we recorded, is in the southern end of the province, bordering with Spain.

Zuberoan is usually considered the most deviant variety of Basque. It has historically been in contact with Bearnese Gascon to the North and Spanish to the South, but the importance of French has increased exponentially in the last century, severely threatening this variety, which is currently highly endangered. As a result, the number of speakers of Zuberoan Basque is much lower than the population of the area; the passive knowledge of the language being estimated at 22.8%, and the effective usage at 5.8%. In addition, its speakers are among the oldest population, making the study of this variety very pressing.

1.2. Typological background

The oral glottal approximant /h/ can be found in most of the world's languages, being part of 88% (279/317) of the inventories in the UPSID data-base (11) and 56% (1703/3020) of the inventories in the PHOIBLE data-base (12). On the other hand, the nasalized glottal approximant /h̃/, its nasal counterpart, while not unheard of, is by no means equally common, being attested in just a handful of languages including Krim (Niger-Congo, bomk1234), Lisu (Tibeto-Burman, lisu1250) and Pirahã (isolated, Amazon, pira1253). This rareness

is reflected in the fact that its cell is shaded in gray on the IPA chart, a feature used to signal impossible segments.

The impossibility attributed to /h̃/ is rooted in the aerodynamic definition of nasality; whether or not enough air-stream can go through the nasal cavity once it has produced glottal friction and after it has been divided between the nasal and the oral tract. However, nasalized aspirates are not as problematic from an articulatory perspective, where any sound produced with a lowered velum can be considered nasal (2).

From the fact that /h̃/ is an uncommon segment follows that the phonological opposition between /h/ and /h̃/ is exceedingly rare, the set of the languages that include /h/ and /h̃/ being a small subset of the languages that include /h̃/. In fact, a phonological opposition between /h/ and /h̃/ has only been proposed for 3-4 languages: kwangali (kwan1273) and thimbukushu (a.k.a. mbukushu, mbuk1240) both Bantu languages from northern Namibia (13), Seimat (seim1238) an Oceanic language from the Admiralty Islands (14), and Basque. Thanks to recent research, Basque is now arguably the clearest example of this opposition (15).

1.3. Previous research

Early work on Zuberoan Basque by Larrasquet (7) described a number of vowel-aspirate-vowel sequences that were nasalized throughout (now transcribed as [Ṽh̃Ṽ], see (16)), in contrast to other comparable [VhV] sequences which were not nasalized. The initial phonological analysis for this sequence of sounds assumed that both vowels were underlyingly nasal, while /h/ was similar to other oral /h/s in the language (17). This analysis was likely conditioned by the typological rarity of the /h/ vs. /h̃/ opposition. Later on, Hualde proposed a different analysis (1), namely that /h̃/ was phonologically nasal, and the surrounding vowels were only phonetically nasalized, given that in Basque nasality spreads to all vowels in contact with a nasal consonant (18). Recent research has found strong phonological evidence in support of this proposal (3). In parallel, historical research has established intervocalic *n as the segment that gave rise to /h̃/ (17; 19). However, phonetic research on this opposition is still scarce. There is only one acoustic analysis of the aspirate opposition in Mixean Basque (9), and only a handful of illustrative spectrograms from Zuberoan speakers can be found in the published literature thus far (16; 6). This paper aims to start filling this gap by means of a quantitative study of nasalance data recorded from speakers from the Zuberoan village of Larraine.

2. EXPERIMENT

We set up a field-lab at a house that was rented in Larraine. The local participants performed an elicitation task mostly including words containing an aspirate and we recorded their productions using a nasalance device or nasometer.

2.1. Participants

We recorded the utterances of 5 volunteer Larraine-Basque speakers (4 male, 1 female; mean age 64.6, range 60-70). Participant profiles were self-reported: they were L1 Larraine-Basque speakers; they were born and grew up in Larraine; they started to learn French (L2) at school at age 5; some left the Basque Country for work, but returned to town later on. To our knowledge, there were no direct family bounds between them.

2.2. Stimuli

Stimuli consisted of isolated words taken from Larrasquet's lexicon (8). We selected 12 words with initial /h/ (e.g. *hiru* 'three'), and around 90 with medial aspirates: 17 with etymologically oral /h/ (e.g. *soho* 'cropland' from Lat. *solu*(*m*)), 43 with etymologically nasalized /h̃/ (e.g. *ah̃ate* 'duck' from Lat. *anate*(*m*)), 7 with an aspirate adjacent to a nasal stop (e.g. *lanho* 'fog'), and 23 including an aspirate and a nasal stop in the same word, which may cause long distance nasalization (e.g. *nahi* 'will' and *ihun* 'nowhere'). Crucially, aspirates were all prompted written with <h>, irrespective of their etymology and alleged phonological status. Thus, any degree of nasalization in the uttered aspirates should be attributed to the participants' intended pronunciation. Finally, we selected another 56 words with no aspirates as fillers. Difficulties came from the fact that some participants were not used to reading. In addition, there were slight divergences in some words between our source (8) and the current local pronunciation, so we included a French translation in the prompts. The translation helped participants understand the target words.

2.3. Procedure

The recordings were conducted using a Glottal Enterprises nasalance with a separator handle (NAS-1 SEP)[1], which consists of two microphones separated by a wooden plate. This plate allows the separation of the acoustic signal coming from

the mouth and the nose. The nasometer stood in a tripod, which was adjusted to a suitable height for each participant, so that they could sit comfortably while the nasometer was pressed against their upper lip. *SpeechRecorder* software was used in order to present the stimuli in a 32 inch TV screen and record the productions. The order of the stimuli was randomized for each participant. The recording started automatically 500 ms after each stimulus was presented and was manually stopped once the participant had read out loud the target word or stated it was unknown to them.

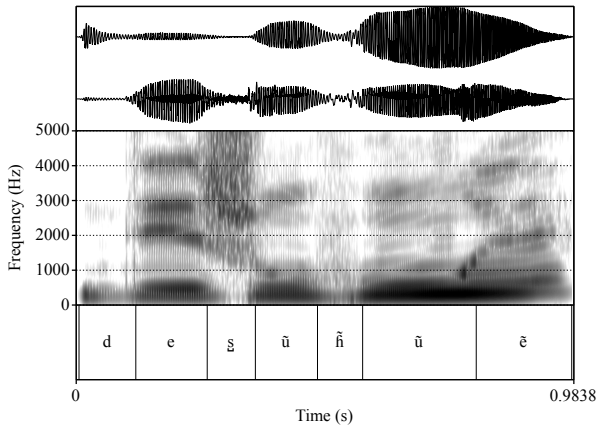

Figure 1: Waveforms and spectrogram of *desuhue* 'dishonor' as produced by speaker 2

2.4. Analysis

The stereo nasalance data was analyzed using Praat. Fig. 1 shows an example of the production of /h̃/. The top waveform corresponds to the nasal channel, while the bottom waveform corresponds to the oral one. For the acoustic analysis, both the nasal and oral channels were band-pass filtered (80 Hz-10000 Hz) and two measurements describing the degree of nasalization were computed: 1. the mean-corrected nasal amplitude (nasal amplitude divided by the mean intensity of the token); and 2. the nasalance (ratio of the nasal amplitude to the sum of the oral and nasal amplitudes, i.e. $A_n/(A_o + A_n) \times 100$) (20; 21). As the description of nasalized aspirates using nasometry is an unexplored topic, the suitability of both metrics for the available data was tested. All measurements of both metrics were then z-transformed (grouped by speaker), and median values of the z-scores of all observations were computed for each aspirate. All recorded words were coded by the etymological category of the aspirate included in them, given that Basque only allows for one aspirated segment in each word (17). A first coding was made following Larrasquet's transcriptions (8), but that needed to be revised, as the variety described by Larrasquet is from the North of Zuberoa and it was spoken 85 years ago.

Two different Bayesian Gaussian models were fitted in *R* using the *brms* interface to *Stan*. Model 1 included nasalance as a response, while Model 2 included mean-corrected nasal amplitude instead. In both cases, etymological category and trial (which was previously min-max normalized) were used as predictors. Speaker and word were added as random effects (including correlated varying intercept and slope) to both models. Both models were fitted in eight chains of 10000 iterations each (including 5000 warm-up iterations). We used *Student's t* distributed priors (df = 5, μ = 0, σ = 2) for all predictors and the intercept, with default priors from *brms* for the random effects and correlation parameter. In both cases, we had to adapt the delta to 0.999 and the maximum tree depth to 12 to ensure correct convergence (no issues, all $\hat{R}$ = 1.00).

2.5. Results

Fig. 2 shows the posterior predictions for Model 1, with nasalance as the response. The posterior distribution plot shows that the median values and overall distributions of the etymologically nasalized /h̃/s and the /h/s nasalized after assimilation are very similar in our data, while the distribution of the oral /h/s falls to lower nasalance values. The main results of the model are in the table in Fig. 2.

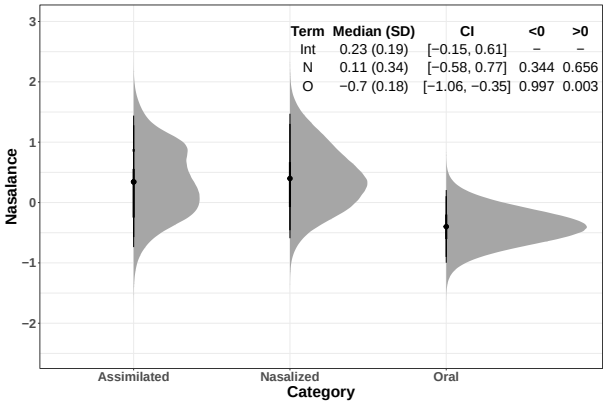

Figure 2: Posterior probability distribution of nasalance by aspirate category

Taking the factor level with the category "assimilated" as the intercept, the "etymologically nasalized" category had a lower 95% credible interval (CI) of -0.58 and an upper 95% CI of

0.77, with a comparable span over both sides of zero, which can be interpreted as the categories "etymologically nasalized" and "nasalized due to assimilation" being quite similar in our data. On the other hand, the oral category had a lower 95% CI of -1.06 and an upper 95% CI of -0.35, showing little overlap with the two nasalized categories and not crossing over 0, which suggests that the "oral" category is less nasalized than the two nasalized categories in our data. The results from model 2 (μ-corrected A_n) were very similar.

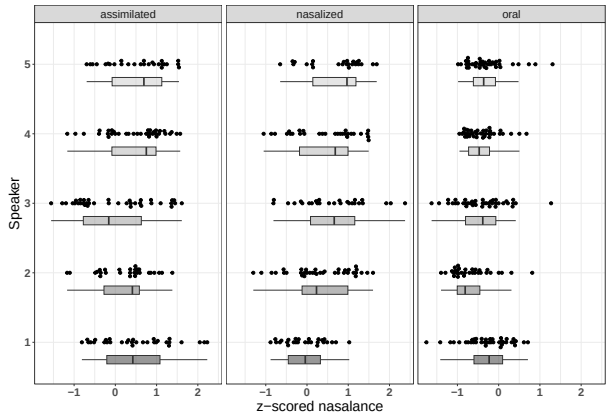

Figure 3: Nasalance by aspirate category

3. DISCUSSION

Our models shows a difference between etymologically nasalized and etymologically oral aspirates in the Zuberoan variety of Larraine. Nonetheless, there are multiple factors that need to be taken into account in the interpretation of these results. First, not all speakers show the same degree of nasalization: while some speakers consistently nasalize the aspirates in the relevant lexical items, others do not produce /h̃/ consistently, if at all. This can be seen in Fig. 3, which presents the nasalance values of the aspirate segments by aspirate category. In parallel to the model results in Fig. 2, the values for assimilated and etymologically nasalized tokens are quite similar, whereas there is a substantial difference between etymologically nasalized and etymologically oral aspirate tokens. Fig. 3 points to big inter-speaker differences: while speaker 1 produced aspirates in all categories with a low nasalance level, the two aspirate groups (nasalized and assimilated vs. oral) are clearly differentiated in speakers 2 or 5.

Second, not all etymologically nasalized aspirates in Larrasquet's lexicon (8) are in fact nasalized in modern-day Larraine. Fig. 4 shows the mean

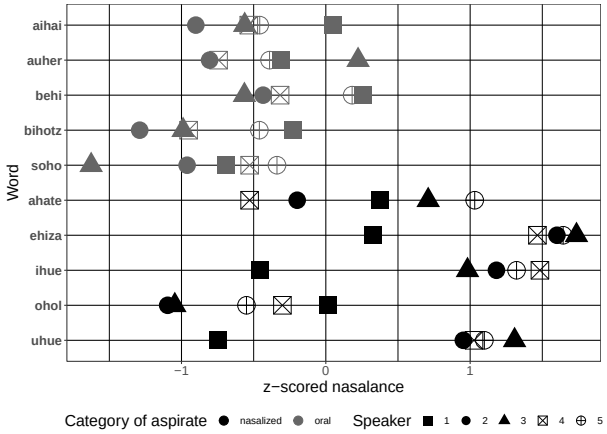

Figure 4: Mean speaker nasalance in example words including an aspirate of oral vs. nasal origin

nasalance value of the aspirate included in ten different lexical roots (i.e. words with the same root but potentially with different declension). The first five roots include etymologically oral /h/s, while the bottom five include aspirates that Larrasquet transcribed as nasal. Each symbol represents the production of a given speaker. When a speaker produced more than one token of the same lexical root, the nasalance values were averaged for this plot. As can be observed in Fig. 4, some nasalized aspirates are indeed produced with a high nasalance value by most speakers (with the recurring exception of speaker 1), such as *eh̃iza* 'hunt', *ih̃ue* 'no one' or *uh̃ue* 'honor', but other tokens are less often nasalized, such as the classic example of /h̃/ *ah̃ate* 'duck', or not at all, as is the case of *ohol* 'board'. In contrast, all aspirates characterized as oral have consistently low nasalance values.

4. CONCLUSION

This research presents evidence that the typologically rare opposition between nasalized and oral aspirates is still present in the Zuberoan Basque variety of Larraine. However, it seems that this contrast is being gradually neutralized as lexical items expected to be nasalized have lost their nasality and some speakers show loss of aspirate nasalization, even in the older generations.

The Zuberoan variety is severely endangered, so further data collection is more difficult by the day. This research reminds us that it is paramount to continue collecting data from endangered languages, given that they might be our only opportunity to witness the rarest phenomena of human language.

REFERENCES

[1] J. I. Hualde, "Topics in Souletin phonology," in *Generative Studies in Basque Linguistics*, J. I. Hualde and J. Ortiz de Urbina, Eds. Amsterdam: John Benjamins, 1993, pp. 289–327.

[2] R. Walker and G. K. Pullum, "Possible and impossible segments," *Language*, vol. 75, pp. 764–780, 1999.

[3] A. Egurtzegi, "On the phonemic status of nasalized /h̃/ in Modern Zuberoan Basque," *Linguistics*, vol. 56, pp. 1353–1367, 2018.

[4] I. Camino, *Amiküze eskualdeko heskuara [The Basque of the region of Amiküze (Mixe)] (Mendaur 11)*. Bilbao: Euskaltzaindia, 2016.

[5] A. Egurtzegi and C. Carignan, "An acoustic description of Mixean Basque," *The Journal of the Acoustical Society of America*, vol. 147, pp. 2791–2802, 2020.

[6] A. Egurtzegi, "/h̃/ hasperen sudurkarituaren inguruan [On the nasalized aspiration /h̃/]," *International Journal of Basque Linguistics and Philology (ASJU)*, vol. 57, 2023.

[7] J. Larrasquet, "Phonétique du basque de Larrajá (Quartier de barcus)," *Revista Internacional de Estudios Vascos*, vol. 23, pp. 153–191, 1932.

[8] ——, *Le basque de la Basse-Soule orientale*. Paris: C. Klincksieck, 1939.

[9] A. Egurtzegi and C. Carignan, "A typological rarity: the /h/ vs. /h̃/ contrast of Mixean Basque," 2020, paper presented at LabPhon 17.

[10] "Populations légales en vigueur á compter du 1er janvier 2022," 2021, Available at: https://www.insee.fr/fr/statistiques/6011060, Accessed: 2022-12-26.

[11] I. Maddieson, *Patterns of sounds*. Cambridge: Cambridge University Press, 1984.

[12] S. Moran and D. McCloy, "Phoible 2.0," http://phoible.org, 2019.

[13] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Oxford: Blackwell, 1996.

[14] R. Blust, *The Austronesian languages (rev. ed.)*. Canberra: Asia-Pacific Linguistics Open Access, 2013.

[15] J. Blevins and A. Egurtzegi, "Refining explanation in evolutionary phonology: Macro-typologies and targeted typologies in action," *Linguistic Typology*, vol. 27, 2023.

[16] A. Egurtzegi, "Diferentes tipos de aspiración en vasco (con análisis espectrales del dialecto suletino actual)," in *Iberia e Sardegna. Legami linguistici, archeologici e genetici dal Mesolitico all'Età del Bronzo*, E. Blasco, P. Francalacci, A. Nocentini, and G. Tanda, Eds. Firenze: Le Monnier, 2013, pp. 151–169.

[17] L. Michelena, "Fonética histórica vasca," in *Obras completas 6*, J. A. Lakarra and I. Ruiz Arzalluz, Eds. Donostia-San Sebastián & Vitoria-Gasteiz: Diputación Foral de Guipúzcoa & University of the Basque Country, 2011.

[18] J. I. Hualde, "Segmental phonology," in *A grammar of Basque*, J. I. Hualde and J. Ortiz de Urbina, Eds. Berlin: Mouton de Gruyter, 2003, pp. 16–65.

[19] I. Igartua, "Diachronic effects of rhinoglottophilia, symmetries in sound change, and the curious case of Basque," *Studies in Language*, vol. 39, pp. 635–663, 2015.

[20] C. Carignan, "Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels," *The Journal of the Acoustical Society of America*, vol. 143, pp. 2588–2601, 2018.

[21] ——, "A practical method of estimating the time-varying degree of vowel nasalization from acoustic features," *The Journal of the Acoustical Society of America*, vol. 149, pp. 911–922, 2021.

[1] We thank the town of Larraine and the first *Basque Summer Tutorial in Language Documentation* for providing us with the setting for this research, Marianne Pouplier for lending us her nasalance device and for her invaluable insights on Praat scripting, Roy Alderton for his help with *SpeechRecorder*, and Shelece Easterday, Aitor Egurtzegi and Francesco Rodriquez for their readiness to answer our questions. All errors are ours. This research has been partially funded by the ANR project MADPAB (ANR-20-CE27-0007).