Evidence for phonological /w/ and /j/ in Chilean Spanish: the case of "hi" and "hu" plus vowel

Lizbeth Gómez, Mauricio A. Figueroa & Gastón F. Salamanca

Universidad de Concepción, Chile
lizgomez@udec.cl, maufigueroa@udec.cl ✉, gsalaman@udec.cl

ABSTRACT[1]

It is generally accepted that the vowel system of Spanish consists of five phonological units, two of them /i/ and /u/, and that [j] and [w] are their gliding variants, present in diphthongs. This account faces challenges, however, when the articulation of the orthographic sequences "hi" plus vowel –as in *hielo* (ice)– and "hu" plus vowel –as in *huevo* (egg)– are considered. The first phonetic element of these sequences displays a wide range of realizations, which, along with a limited contextual distribution, makes it difficult to determine whether they are best interpreted as phonetic variants from phonological vowels or consonants. This study uses quantitative methods to compare the realizations from the aforementioned orthographic sequences to similar vocalic and consonantal structures. Although alternative analyses are also plausible, we conclude that, for Chilean Spanish, these units are better interpreted as variants from two new independent phonological units /w/ and /j/.

Keywords: "hi" and "hu" plus vowel, glides, approximants, Chilean Spanish

1. INTRODUCTION

Spanish glides such as [j] and [w] occur in rising and falling diphthongs in words such as *piano* [ˈpja.no] (piano), *coima* [ˈkoj.ma] (bribe), *puerta* [ˈpwer.t̪a] (door) and *flauta* [ˈflaw.t̪a] (flute) [6]. Although there are some discrepancies regarding whether it is relevant to represent these units differently in rising and falling diphthongs (e.g., [12]), the fact that they are the sole non-syllabic allophones from /i/ and /u/ is generally an uncontested assumption. One notable exception, however, can be found in words beginning with the orthographic sequence "hi" plus vowel, as in the words *hielo* (ice) and *hierba* (grass), and in "hu" plus vowel sequences, as in *huevo* (egg) and *hueso* (bone). These gliding compounds are interesting for several reasons. Firstly, it is not clear whether the pre-nuclear elements in these sequences are consistently articulated as vocoids, contoids, or both, or if variables such as phonetic context are sufficient to explain the observed variability [2]. Secondly, given the aforementioned, it is not clear whether they ought to be analysed as members of the phonological vowels /i/ and /u/ or as allophones of /g/ and /j̞/, present in words like *gato* [ˈga.t̪o] (cat) and *lluvia* [ˈj̞u.β̞ja] (rain) [5]. Finally, it is also of interest that these structures are being flagged orthographically with "h", a grapheme that normally has no phonetic value in Spanish –in words such as *hoja* [ˈo.xa] (leaf) or *herrero* [e.ˈre.ɾo] (blacksmith). Moreover, there are cases of similar sequences spelled without "h" –although all of them are low frequency words– such as *ion* (ion), *iodo* (iodine) and *iota* (iota), and their derivates.

Recent studies of non-peninsular Spanish have shown that, in the Costa Rican variety, both "hi/hu" plus vowel sequences are articulated predominantly using gliding vocoids (although some contoids were found for "hu"); these findings were interpreted as indicating that the first elements of these sequences were allophones of the high vowels /i/ and /u/ [5]. In the case of Chilean Spanish, the variant examined here, [j] and [w] have been identified as allophones of /i/ and /u/, respectively [4]. More recently, however, some studies have included /w/ as an independent phonemic unit, separate from /u/, but no empirical evidence was provided to sustain this claim [19]. The first and only study that has specifically addressed this topic in Chilean Spanish looked at "hu" plus vowel and determined that the opening element was [g], [ɣ], [ɣ̞] or [w], with the approximant [ɣ̞] being more frequent [1].

This study sets out to provide the first quantitative evidence aimed at determining the nature of the sounds articulated in the orthographic sequences "hi" and "hu" plus vowel in Chilean Spanish. In order to do so, a large number of instances were identified and annotated, while analysing a number of their acoustic properties and subsequently comparing them via statistical analyses to those of raising diphthongs and to CV consonant plus vowel syllables, in order to determine whether the opening sounds articulated in "hi/hu" are more closely related to vocoids or to contoids. The results will be discussed in the light of their relevance to our understanding of (a) the functional and phonetic nature of vowels, diphthongs and consonants, (b) the place of glides in syllabic structures, and (c) the status of CV syllables as an unmarked and typologically universal unit.

2. METHODS

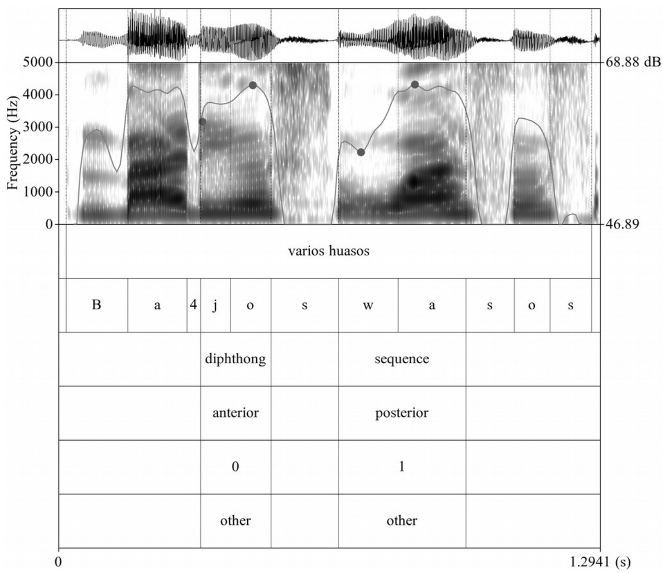
2.1. Participants, elicitation tasks and recordings

The corpus consists of recordings of 64 participants –half female, half male– recruited in 8 Chilean cities, representing all major geographical areas of the country. Participants were all between 18 and 35 years of age, monolingual speakers of Chilean Spanish, and none had completed their secondary education. Participants had to read 28 sentences aloud, of which 24 contained one or more words with "hi/hu" plus vowel, in a variety of phonetic contexts. The sentences also contained a large number of rising diphthongs and CV syllables starting with /ʝ̞/ and /g/ (the first one is articulated most often as and approximant, but sometimes as a voiced affricate; the second one can be articulated as a voiced plosive or as an approximant). Speech was recorded in sound-treated rooms, using Sennheiser EW-152-G3 head-mounted microphones, whose signals were sent to Tascam DR-40 digital recorders, set at a sampling rate of 44100 Hz and a 24 bit depth, in WAV mono format.

2.2. Annotation and data extraction

The audio signals were transliterated and time-aligned at the utterance level in TextGrids from *Praat* [3], and the corpus was then pre-processed using *EasyAlign*, to automatically obtain word, syllable and segment boundaries [11]. Following this process, the entire corpus was manually corrected, via auditory and spectrographic inspections of the signals. All instances of rising diphthongs (i.e., [ja], [je], [jo], [ju], [wa], [we], [wi], [wo]), of the orthographic sequences "hi" and "hu" plus vowel (e.g., the underlined sections of *hierba* "grass" or *huevo* "egg"), and of all instances of /ʝ̞/ and /g/ and their following vowels, were identified and coded in a separate tier, as well as information about preceding phonetic context and whether the structures were located in a stressed or unstressed syllable. In the case of the "hi" plus vowel and "hu" plus vowel sequences, manner of articulation of the opening segment was also registered. The data was then extracted using *Praat* scripts, which also measured the minimum and maximum intensity values found in the durational space of each token, using intensity objects created separately for each speaker with default values. These acoustic measurements were used to calculate normalized intensity differences –subtracting the minimum intensity to the maximum–, which have been suggested as good acoustic correlates of degree of constriction, and assumed to be lower in vocoids and higher in contoids [13, 15]. An example of the annotation and of the intensity landmarks can be seen in Figure 1.

Figure 1: Waveform, spectrogram, intensity contour and TextGrid annotation of the utterance varios huasos ("several farmers"), containing two instances of the target structures. The tiers, from top to bottom, encode: (1) utterance transliteration; (2) segments, in SAMPA; (3) constriction class, that is, whether the token is a diphthong, an orthographic "hi/hu" plus vowel sequence, or a CV structure; (4) position, i.e., whether the first element of the token begins with an anterior or posterior segment; (5) stress; and (6) preceding phonetic context. In the intensity contour of both tokens, the minimum and maximum intensity values have been identified with circles.



2.2. Data, variables and levels

The resulting corpus comprised 8962 tokens. Of these, 3790 instances were rising diphthongs (42%) –referred to from here on as "diphthongs"–, 2426 instances were "hi" or "hu" sequences followed by vowel (27%) –henceforth, "sequences"–, and 2746 were instances of /g/ or /ʝ̞/ followed by vowel (31%) –subsequently, "consonants". These three levels –diphthongs, sequences and consonants– were grouped under the variable *constriction class*. For the specific case of sequences, Table 1 summarizes the manner of articulation found on the opening element. As the table shows, approximant realizations and affricate segments with approximant release predominate in both categories. Regarding place of articulation, in total, 4648 instances were categorized as belonging to the "anterior" group (52%), that is, instances that begin with [j], "hi-" or /ʝ/, and 4304 instances to the "posterior" group (48%), which begin with [w], "hu-" or /g/. These two levels –anterior and posterior– were grouped under the variable *position*. For reasons of space, the variables *stress* and *preceding phonetic context* will not be included in subsequent statistical analyses, but their importance will be addressed briefly in the discussion.

Table 1: Percentages and IPA transcriptions of manners of articulation of the first element of "hi" plus vowel and "hu" plus vowel sequences, ordered by approximate degree of constriction.

Manner of articulation	"hi" sequences		"hu" sequences	
	%	IPA	%	IPA
Plosive	0%		17.0%	[g]
Fricative	0.8%	[ʝ]	5.7%	[ɣ]
Affricate 1 (fricative release)	9.6%	[d͡ʒ]	0%	
Affricate 2 (approximant release)	**34.4%**	[ɟ͡ʝ̞]	0%	
Approximant	**43.8%**	[ʝ̞]	**76.7%**	[ɣ̞]
Glide	11.4%	[j]	0.6%	[w]
Total:	100%	- - -	100%	- - -

3. ANALYSES

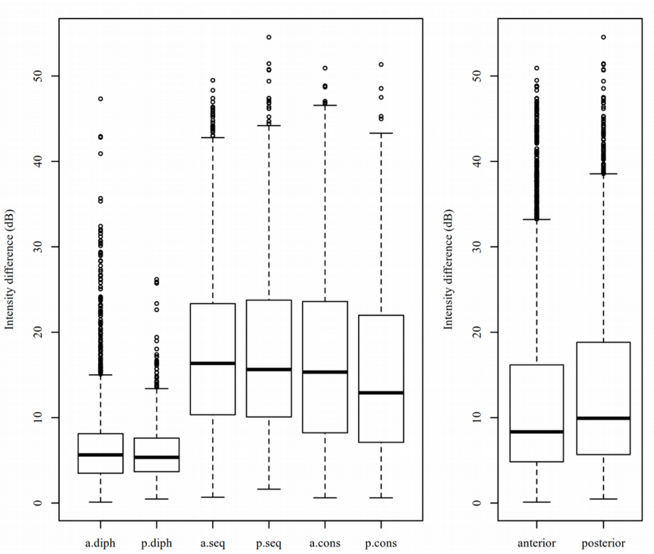
Data was imported into *R* [16], where a linear mixed model was built to evaluate the effects of the variables *constriction class* and *position* (and their interaction) on the dependent variable *intensity differences*. The model was created using the *lmer* function from the *lmerTest* package [14]. The variable *participant* was also included as a random factor. Following [8], a stepwise procedure was used to build the models: first, a null model with only the dependent variable and the random factor was fitted, and then the independent variables and their interaction were included one by one and retained only when they significantly improved the model, as judged by an analysis of variance function (*anova*). Type-II analyses of variance tables for the fixed factors and interactions of each model were produced via the *Anova* function in the *car* package [10] and using the *ranova* function from *lmerTest*.

The best fit mixed-effects model for intensity differences is shown in Table 2. According to this model, there is a main effect for *constriction class* (χ^2 (2) = 3918.085, $p < 0.001$) and *position* (χ^2 (1) = 26.198, $p < 0.001$), and a significant interaction between these two variables (χ^2 (2) = 17.521, $p < 0.001$). The data driving these effects and the interaction can be observed in the left panel of Figure 2. As the model shows, instances of diphthongs ($\bar{x}$ = 6.36, σ = 4.4) and consonants ($\bar{x}$ = 15.83, σ = 10.02) display significantly smaller intensity differences than orthographic sequences ($\bar{x}$ = 17.94, σ = 9.88). However, both the sizes of the coefficients and of the *t* statistics –see Table 2– suggest that the differences between diphthongs and sequences are considerably higher than those between consonants and sequences. In other words, sequences seem to be more closely related acoustically to consonants than to diphthongs. Regarding *position*, although a main effect was detected, suggesting differences between anterior and posterior realizations (see right-hand panel of Figure 2), this effect was greatly diluted when the data of *constriction class* was taken into account (see left panel of Figure 2). Finally, regarding the interaction between *constriction class* and *position*, the difference between the levels anterior ($\bar{x}$ = 6.58, σ = 4.87) and posterior ($\bar{x}$ = 5.95, σ = 3.3) is not significant when diphthongs are compared to sequences, but it is when anterior ($\bar{x}$ = 17, σ = 10.68) and posterior ($\bar{x}$ = 15.09, σ = 9.51) tokens from consonants are compared to orthographic sequences, most likely due to a larger difference between anterior and posterior tokens in consonants than in diphthongs.

Table 2: Best fit mixed-effects model for *intensity differences*, including *constriction class* and *position* as main factors, and *participant* as a random factor.

Fixed Factors	Coefficient	Standard Error	*t*-value	*p*-value
Intercept	18.0009	0.3266	55.119	< 0.001
Constriction class				
Sequence	(ref. level)			
Diphthong	-11.4253	0.2829	-40.380	< 0.001
Consonant	-1.0001	0.3362	-2.975	< 0.01
Position				
Anterior	(ref. level)			
Posterior	-0.1196	0.3203	-0.373	= 0.70890
*Constriction class * Position*				
Sequence / Anterior	(reference level)			
Diphthong / Posterior	-0.5078	0.4176	-1.216	= 0.22402
Consonant / Posterior	-1.7820	0.4444	-4.010	< 0.001
Random Factors	**Log-likelihood**	**Degrees of Freedom**	**Likelihood ratio test statistic**	**p-value**
Participant	-31428	1	320.68	< 0.001

Figure 2: Left-hand panel: box plots of intensity differences by the variables *position* ("diph" = diphthongs; "seq" = sequences; "cons" = consonants) and *constriction class* ("a" = anterior; "p" = posterior). Right-hand panel: box plots of intensity differences by *constriction class*.

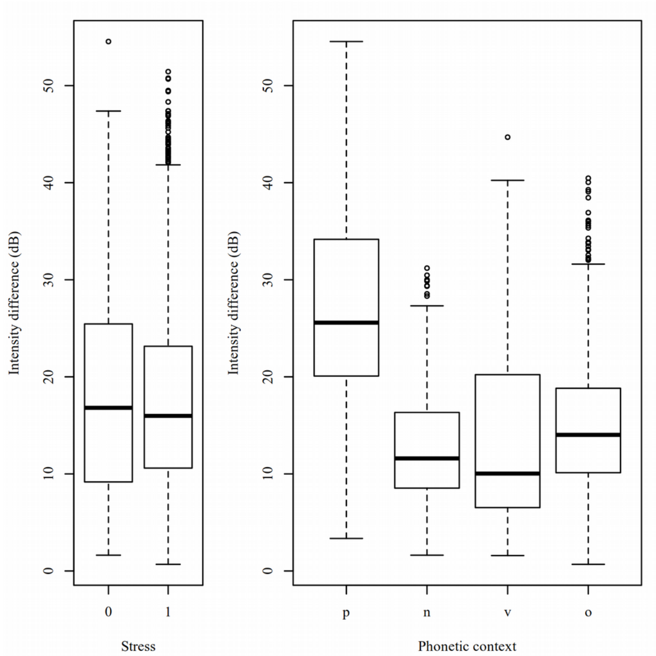


4. DISCUSSION

As mentioned earlier, the acoustic and qualitative evidence (see Table 1) shows that the opening segments of "hi" and "hu" plus vowel sequences are more closely related to contoids such as those found in syllables starting with /ʝ̞/ and /g/, than to glides. We believe that this fact, along with the arguments to follow, allow these segments to be interpreted as belonging to functionally independent units, different from /i/ and /u/, but also from /ɟ͡ʝ/ –notice, not /ʝ̞/– and /g/. This is so, firstly, because it adequately explains the contrast between words such as *llena* ("he/she fills" or "filled"), in which mostly non-aproximant contoids have been observed in the past [7], and *hiena* ("hyena"), in which approximants predominate. Secondly, it explains the spelling that native Chilean Spanish speakers use intuitively to represent swear-words such as *huevón*, written informally as *weón*, but virtually never as **güeón* or **ueón*. Third, from a systemic standpoint, the fact that these segments are articulated more often as approximants is consistent with a general tendency towards lenition in the consonant system of Chilean Spanish, for which there is mounting evidence (e.g., [9], [17], [18]). Fourthly, it allows to restrict a definition of diphthong to tautosyllabic sequences comprising exclusively vocoids. Finally, it makes it possible to assign the realization of these sequences to typologically unmarked CV syllables, instead of VV ones.

Of course, this interpretation is not without its drawbacks. For example, it creates some overlap between the realizations of /j/ and /w/ with those of /ʝ̞/ and /g/ (respectively), which in turn probably requires representing /ʝ̞/ as /ɟ͡ʝ/ (as we did above), and perhaps vocalic glides as [i̯] and [u̯].

Figure 3: Left-hand panel: box plots of intensity differences of "hi" and "hu" plus vowel sequences by *stress* ("0" = unstressed; "1" = stressed). Right-hand panel: box plots of intensity differences of "hi" and "hu" plus vowel sequences by *phonetic context* ("p" = following pauses; "n" = nasals; "v" = vowels; "o" = others).



Regarding some projections, although this study has provided empirical evidence showing that *position* and *constriction class* have an effect in the phonetic realizations of the segments in discussion, several other variables such as *stress*, *phonetic context*, *position in word* and *vowel identity* ought to be taken into consideration in future and more complex incursions on the subject, since it is very likely that they also have a role explaining the realizations of "hi" and "hu" followed by vowel (for preliminary evidence regarding the first two variables, see Figure 3).

5. REFERENCES

[1] Aguilar, E., Salamanca, G. 2014. Fonos que realizan la secuencia grafémica "hu+vocal" en 17 sujetos de Concepción e hipótesis inicial sobre su organización fonológica. *Literatura y Lingüística,* 28, 193-214.

[2] Aguilar, L. 1994. *Los procesos fonológicos y su manifestación fonética en diferentes situaciones comunicativas: La alternancia vocal/semiconsonante/ consonante* (Unpublished doctoral thesis). Universidad Autónoma de Barcelona, Barcelona, España.
[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer software], version 6.0.39. Retrieved from http://www.praat.org/
[4] Borland Delorme, K. 2004. La variación y distribución alofónica en el habla culta de Santiago de Chile. *Onomázein,* 10, 103-115.
[5] Calvo, A. 2008. Las Semiconsonantes y Semivocales en los diptongos del español: propuesta de análisis fonológico. *Filología y Lingüística,* 34, 107-142.
[6] Celdrán, E., Planas, A., Sabaté, J. 2003. Castilian Spanish. *Journal of the International Phonetic Association*, 33(2), 255-259.
[7] Cepeda, G. 1991. *Las consonantes de Valdivia*. Valdivia: Imprenta América.
[8] Chappell, W. 2016. On the social perception of intervocalic /s/ voicing in Costa Rican Spanish. *Language Variation and Change*, 28(3), 357–378.
[9] Figueroa, M. (2016). *Lenition in the production and perception of Chilean Spanish approximant consonants: Implications for lexical access models* (Unpublished doctoral thesis). University College London, London, Uk.
[10] Fox, J., Weisberg, S. 2011. *An {R} Companion to Applied Regression*. Thousand Oaks CA: Sage.
[11] Goldman, J. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. In: *Proceedings of InterSpeech*. Florence, Italy.
[12] Hualde, J. 2014. *Los Sonidos del Español.* New York: Cambridge University Press.
[13] Hualde, J. Simonet, M., Shosted, R., Nadeu, M. 2010. Quantifying Iberian spirantization: Acoustics and articulation. In: *40th Linguistic Symposium on Romance Languages*, Seattle, WA, 26–28.
[14] Kuznetsova, A., Brockhoff, P. Christensen, R. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.
[15] Parrell, B. 2010. Articulation from acoustics: Estimating constriction degree from the acoustic signal. *Journal of the Acoustical Society of America*, 128(4), 2289.
[16] R Core Team. 2018. R: A language and environment for statistical computing [Computer software], version 3.4.4. Retrieved from https://www.R-project.org/
[17] Rogers, B. M. (2016). The influence of linguistic and social variables in the spirantization of intervocalic/b, d, g/in Concepción, Chile. *Studies in Hispanic and Lusophone Linguistics*, *9*(1), 207-237.
[18] Rogers, B. M., & Mirisis, C. A. (2018). Voiceless stop lenition and reduction as linguistic and social phenomena in Concepción, Chile. *Borealis–An International Journal of Hispanic Linguistics*, *7*(2), 187-215.
[19] Sadowsky, S., Salamanca, G. 2011. El inventario fonético del español de Chile: principios orientadores, inventario provisorio de consonantes y sistema de representación (AFI-CL). *Onomázein,* 24, 61-84.

[1] This study was financed by FONDECYT, Chile (Grant 1161466, titled: "Realizaciones fónicas segmentales en el habla subestándar: una contribución a la construcción de un inventario inclusivo y actualizado de fonemas y alófonos del español de Chile"). We would like to thank the anonymous reviewers for their valuable comments and suggestions, which greatly helped us to improve the quality of the manuscript.