An evaluation of several methods for computing lingual coarticulatory resistance using ultrasound

Clara Rodríguez and Daniel Recasens[a)]

Departament de Filologia Catalana, Facultat de Lletres, Universitat Autònoma de Barcelona, Bellaterra 08193, Barcelona, Spain

(Received 2 June 2016; revised 27 September 2016; accepted 13 November 2016; published online 31 July 2017)

The paper evaluates the efficiency of six computation methods in distinguishing lingual coarticulatory resistance among consonants and vowels using ultrasound data. This research goal is tested on a corpus of symmetrical vowel-consonant-vowel sequences composed of 10 consonants and five vowels produced by five Catalan speakers. Results show that, while the coarticulatory resistance hierarchies obtained by all methods conform largely to the predictions of the degree of articulatory constraint model of coarticulation, some (i.e., area of the articulatory zone, mean point-by-point coefficient of variation, and mean nearest neighbour distance) are somewhat more highly predictive than others (i.e., locus equation, mutual information, and highest point of the tongue dorsum). Methods differ mostly regarding the classification of consonants exhibiting intermediate degrees of coarticulatory resistance due to the way the methods have been designed. The implications of these findings for research on dialectal variation are discussed.
[http://dx.doi.org/10.1121/1.4991319]

[SSN] Pages: 378–388

I. INTRODUCTION

The present investigation evaluates different analysis methods of contextual articulatory variability for consonants and vowels using tongue contour data obtained with ultrasound. The ultimate goal of the study is to determine how successful these methods are in predicting the degrees of coarticulatory resistance for consonants and vowels at different tongue regions in speech production. Coarticulatory resistance is inversely related to contextual coarticulatory variability, i.e., the less variable a target phonetic segment is to the coarticulatory effects exerted by the contextual segments, the more resistant may be considered to be to the effects in question.

The analysis of phonetic variability provides crucial information about the degree of articulatory precision with which speakers of a given dialectal community produce vowels or consonants. As a relevant factor bearing on phonetic variability we may mention small but robust differences in place of articulation for consonants. Thus, for example, the fact that the two lingual fricatives [s] (alveolar) and [ʃ] (palatoalveolar) are more anterior and closer to each other in Valencian Catalan than in Eastern Catalan have consequences for coarticulation such as a trend for [s] to become dental before the dental stop [t] in the former dialect vs the latter. Segmental context and word frequency may also have a strong influence on the articulatory implementation of vowels and consonants and even act as sound change triggers. Ultrasound data reported by Lin *et al.* (2014) reveal that dark /l/ vocalization in coda position may be favoured by apical contact loss at constriction location, which is prone to occur before labial and velar consonants but not before alveolars and in high-frequency words such as *help* and *milk* rather than in less frequent ones such as *whelp* and *ilk*. Another factor influencing articulatory precision is language phoneme inventory size. It has been shown in this respect that anticipatory labial activity for rounded vowels may be more extensive and less variable as the number of labial vowels in the language sound inventory increases, as in Swedish vs English (Lubker and Gay, 1982).

Ultrasound is a suitable technique for obtaining data on articulatory variability since it is non-invasive and relatively easy to use, and allows recording lingual contour data for large numbers of subjects (Gick, 2002). It is most useful for studying overall tongue body configuration during vowel and consonant production but less so tongue tip and tongue blade position due to the front of the tongue being often shadowed by the jaw.

The criteria for analyzing tongue contours acquired with ultrasound are still open to debate. The present study evaluates the predictive power of several methods which have been proposed for measuring lingual coarticulatory resistance, namely, mean nearest neighbour distance (NN), highest point of the tongue dorsum (HTD), area of the articulatory zone (AAZ), mean point-by-point coefficient of variation (CVar), locus equation (LE), and mutual information (MI). It will do so by testing the extent to which the patterns of coarticulatory resistance for the front lingual consonants [ð, l, ɾ, t, n, s, r, ʎ, ɲ, ʃ] and the vowels [i, e, a, o, u] of Catalan uttered in symmetrical vowel-consonant-vowel (VCV) sequences by several Catalan native speakers, as computed by the six methods, conform to the predictions of the degree of articulatory constraint (DAC) model of coarticulation (Recasens and Espinosa, 2009). Based on the ways linguistic sounds are produced and on data on coarticulatory resistance available in previous acoustic, electropalatographic (EPG), electromagnetic midsagittal articulography (EMA) and ultrasound studies, the

[a)]Electronic mail: daniel.recasens@uab.es

DAC model predicts the following coarticulation resistance hierarchies for the consonants and vowels under analysis: [ʎ, ɲ, ʃ] (most resistant) > [s, r] > [l, ɾ, t, n] > [ð] (least resistant); [i,e] (most resistant) > [a] > [o, u] (least resistant).

Regarding consonants, the alveolopalatals [ʎ, ɲ, ʃ], which are articulated with the blade and the predorsum, ought to allow the least amount of vowel coarticulation since raising and fronting the tongue dorsum by contracting the genioglossus muscle constrains the tongue body to a larger extent than raising the tongue tip and/or blade for the production of dentoalveolars. Less vowel coarticulation for alveolopalatals than for alveolars has often been reported to occur in the speech production literature, as for [ɲ, ʃ] than for [l, n, s] in Catalan (Recasens and Espinosa, 2009).

Among dentoalveolars, on the other hand, coarticulatory resistance is expected to be highest for the apical trill [r] and the apicolaminal fricative [s] due to the precise aerodynamic and articulatory adjustments required for trilling and frication (Ohala and Solé, 2010), and lowest for the apical approximant [ð], which is the realization of the voiced dental stop in intervocalic position in Catalan, in line with the wide apicodental constriction and little tongue-to-palate contact involved in its production. In agreement with these expectations, coarticulatory resistance has been shown to be greater for [s] than for [n, l] in Catalan and [t, d, l] in German (Recasens and Espinosa, 2009; Hoole *et al.*, 1990), and to proceed in the progression [r] > [l, ɾ] > [ð] in Spanish (Proctor, 2009). Other dentoalveolars should show intermediate degrees of coarticulatory resistance between [ð] and [s, r]: the alveolar nasal [n], the unaspirated laminodentoalveolar [t], the tap [ɾ], and the apicoalveolar [l], which is clear rather than dark in the case of the speakers of the present study.

As for vowels, high front [i, e] ought to be most resistant for similar reasons to those pointed out for alveolopalatal consonants, and [a, o, u] less resistant since they are articulated at the back of the vocal tract thus leaving the tongue front relatively free to adapt to context (see Hoole *et al.*, 1990 for German). Moreover, lingual coarticulation data for the three Catalan vowels in previous studies suggest that [o, u] may be less resistant than [a] in this language (Recasens and Espinosa, 2009).

In line with the coarticulation data reported in Recasens and Rodríguez (2016), we hypothesize that ultrasound data should exhibit analogous patterns of tongue dorsum coarticulation at the palatal zone to those obtained with other techniques. Moreover, ultrasound will allow checking whether the same or highly similar coarticulatory resistance hierarchies hold at more posterior lingual regions. We are also aware that, while the hierarchies in question should proceed in the progression [ʎ, ɲ, ʃ] > [s, r] > [l, ɾ, t, n] > [ð] and [i, e] > [a] > [o, u], there could be differences in coarticulatory variability among consonants or vowels falling within each group depending on the method taken into consideration.

All six methods (NN, HTD, AAZ, CVar, LE, MI) will be used for evaluating coarticulatory resistance for consonants as a function of contextual vowels during the consonant, and only four of them for reasons mentioned below (NN, HTD, AAZ, CVar) also coarticulatory resistance for vowels as a function of contextual consonants during the vowels. Moreover, in order to study coarticulation at phonetically meaningful locations and except for HTD, coarticulatory resistance will be computed separately at the alveolar, palatal, velar and pharyngeal zones on SSANOVA derived splines (see Sec. II for details). A description of the different methods of analysis of coarticulatory resistance follows.

(a) The NN method has been applied by Zharkova and colleagues to the study of vowel-to-consonant coarticulation for front lingual fricatives and stops (Zharkova *et al.*, 2011). It seeks to elicit contextual differences in tongue location for a given segment at all or some x tongue contour points. In order to compute NN we need to calculate the Euclidean distances between each of the x points on a lingual spline and the closest point on another spline for pairs of splines corresponding to sequences sharing the same consonant and differing in the adjacent vowel(s), as exemplified by the thick discontinuous lines connecting given spline points at the pharyngeal zone for the sequence pairs [iɾi]-[eɾe], [iɾi]-[aɾa], [iɾi]-[oɾo], and [iɾi]-[uɾu] in Fig. 1 (top left graph). Then, the resulting distances are averaged thus yielding an index value for each pair of splines being compared. The lower the average index values, the more coarticulation resistant (the less sensitive to the vowel coarticulatory effects) the target consonant may be considered to be. A complication with NN is that it does not yield a single index value for a given phonemic unit but several index values for pairs of splines sharing the same target phonetic segment; thus, for example, there are 10 spline pairs and thus 10 NN values for a given consonant in the context of [i, e, a, o, u].

(b) The HTD method, which has been used for investigating the degree of vowel coarticulation for dental and alveolar consonants, is especially sensitive to contextual changes in tongue dorsum position. HTD values equal the area of the polygons whose vertices correspond to the highest point of the dorsum of the tongue for a given consonant in each contextual vowel condition (Proctor, 2009). The lower the area value of the polygon, the more resistant the consonant is to the vowel-dependent coarticulatory effects. Figure 1 (top right graph) shows the polygon for [ɾ] as delimited by the tongue dorsum Y maxima on the lingual splines for [iɾi, eɾe, aɾa, oɾo, uɾu]. The HTD method should be ideal for analyzing coarticulation for target consonants or vowels whose production involves tongue dorsum activation and thus, vowel-to-consonant coarticulation for velar, palatal and alveolopalatal consonants and consonant-to-vowel coarticulation for [i, e] (palatal), [u] (velar) and [o] (upper pharyngeal). Its application becomes more problematic for studying coarticulatory variability in the case of segmental sequences with [a] since the highest point of the tongue either during this vowel in the context of front lingual consonants or during front lingual consonants in the context of [a] may be located at a much more

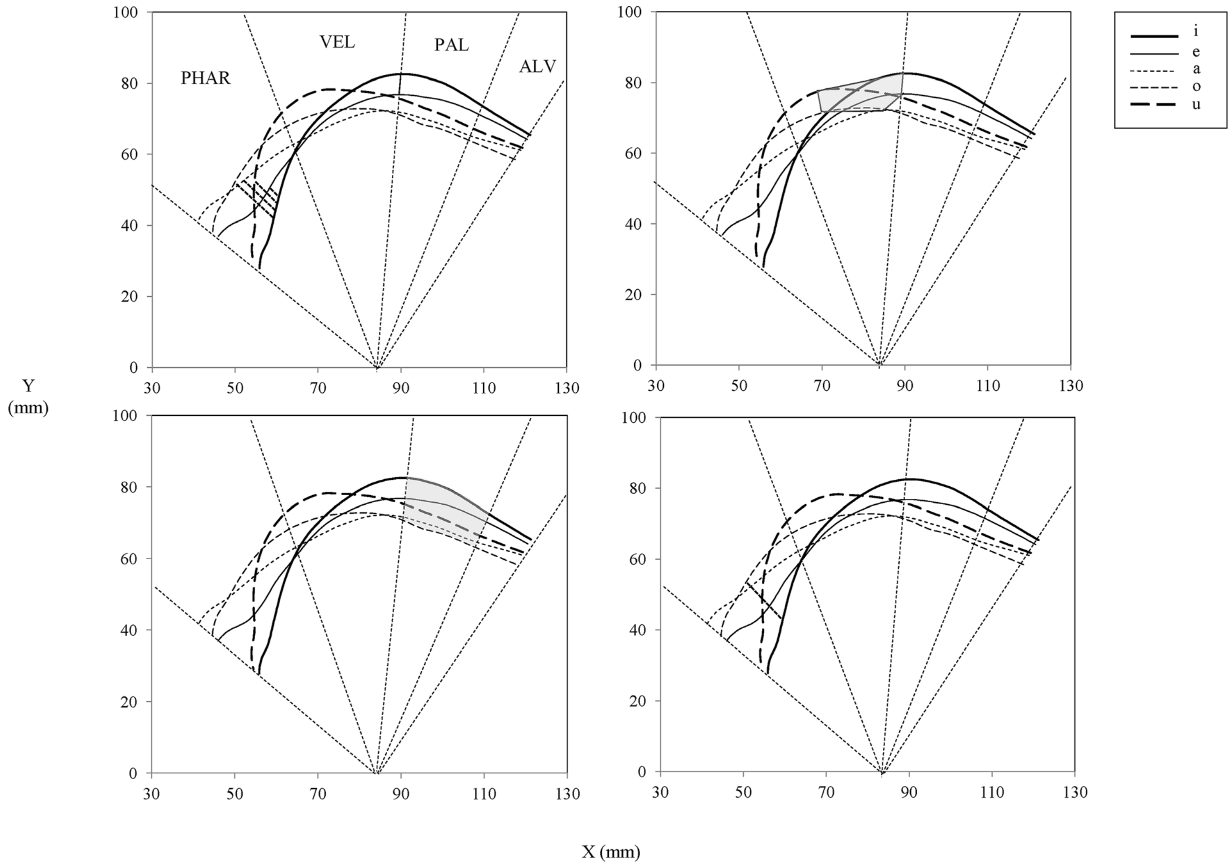

FIG. 1. Exemplification of the four methods NN, HTD, AAZ and CVar for computing coarticulatory resistance. The SSANOVa splines correspond to the [ɾ] midpoint in the sequences [iɾi, eɾe, aɾa, oɾo, uɾu] produced by speaker DR. The front of the vocal tract is on the right of the graph.

anterior position than expected and thus close to the tongue blade or at the tongue blade itself. Another shortcoming of HTD is that it provides no information about coarticulatory variability at the tongue front and at the back of the tongue body.

(c) The AAZ and CVar methods may be applied to the entire tongue contour or a portion of it, and have been used for measuring coarticulatory resistance for consonants as a function of contextual vowels and for vowels as a function of contextual consonants (Recasens and Rodríguez, 2016). The two methods resemble NN in that they take into account all *x* points available. Moreover, AAZ parallels HTD in that it is based on an areal estimate of variability and differs from it in that it calculates coarticulatory variability by estimating the area of the polygon embracing all contextual splines as delimited by the maximal and minimal *y* values at each *x* point. Figure 1 (bottom left graph) shows the polygon for [ɾ] embracing the five splines for [iɾi, eɾe, aɾa, oɾo, uɾu] at the palatal zone. The CVar method, on the other hand, evaluates coarticulatory resistance by computing the mean of the coefficients of variation (CVar = ratio of the standard deviation to the mean) obtained across all contextual splines for a given segment at each *x* point. Figure 1 (bottom right graph) identifies the location of a specific *x* point on the splines for [iɾi, eɾe, aɾa, oɾo, uɾu] with a thick discontinuous line.

(d) In contrast with the other methods, LE and MI evaluate coarticulatory resistance from a dependency measure between articulatory data gathered at two temporal points corresponding to the consonant and the adjacent vowel.

Locus equations are regression lines estimated by predicting a given articulatory position of a target consonant from the corresponding articulatory positions for a subset of contextual vowels. They have been used for studying coarticulatory resistance for consonants in CV sequences using horizontal position data values for the tongue body (TBx) at the consonant release (Iskarous *et al.*, 2010; Noiray *et al.*, 2013) or at the consonant midpoint (Iskarous *et al.*, 2013), and at the vowel midpoint. LE slope values such as those plotted in Fig. 7 of this paper are expected to increase with coarticulation size and thus inversely with coarticulatory resistance, as revealed by differences in TBx slope for [p, b] (about 0.9) > [k, g] (0.7−0.9) > [t, d, s] (0.3−0.7) in English (Iskarous *et al.*, 2013).

MI is a quantification of independence between two variables such that the higher the MI value the more knowledge of the outcome of one variable limits the possible outcomes of the other (see Iskarous *et al.*, 2013 for details). MI has been used for analyzing coarticulatory resistance for consonants in several vowel contexts with tongue position data gathered at the midpoint of the consonantal and vocalic periods in several

VCV and CVC sequences. The lower the MI value the more independent, and therefore, the less predictable and the more coarticulation resistant, the target consonant is considered to be relative to the adjacent vowels. Analogously to the TBx slope values reported above, front dorsum displacement maxima in the X and Y dimensions for German stops embedded in CVC sequences yielded higher MI values for [p] (0.4–0.6) than for [t] (0.1–0.2), those for [k] lying in between (0.2) (Iskarous *et al.,* 2013).

A drawback with the LE and MI methods is that they have not been designed explicitly for measuring vowel coarticulatory resistance (see Sec. IV). Moreover, locus equations have only been applied to the analysis of coarticulation involving phonetic segments placed within the same syllable.

II. METHOD

A. Data recording and spline manipulation

Ultrasound recordings were carried out of all 50 VCV symmetrical combinations of the Catalan consonants [ð, l, ɾ, t, n, s, r, ʎ, ɲ, ʃ] and the Catalan vowels [i, e, a, o, u] with an Echo Blaster unit type EB128CEXT from TELEMED. Ultrasound data for [k] in the sequences [iki, uku] were also recorded so as to determine the boundary between the palatal and velar zones on the lingual splines. The VCV sequences, which always syllabify (V)(CV) in Catalan, were inserted in the Catalan carrier sentence 'Sap___poc' "He/she knows___little" where V1 and V2 occur next to [p] and therefore should not be affected regarding lingual configuration by the preceding and following consonants, respectively. Recordings were made by five native Catalan speakers of 30–60 yr of age who use Catalan in their every day life, i.e., the two men DR (four times) and MO (six times) and the three women ES (six times), JU (seven times) and IM (six times). Subjects were asked to utter the two syllables of the VCV sequences of interest with the same degree of stress at a comfortable rate.

Ultrasound video frames were recorded using a probe with a 90% field of view and a frequency of 2 MHz at a rate of 57 frames per second yielding one ultrasound image every 17.54 ms. Throughout the entire recording session the ultrasound probe was attached to a transducer holder which was positioned under the subject's chin in an Articulate Instruments Stabilization Headset. Image streams were recorded synchronously with the audio signal sampled at 22 050 Hz. Contours of the rear of the alveolar zone and front palate were also recorded by asking speakers to press the tongue against the top of the mouth (Stone, 2005). The tongue contours were tracked automatically at all temporal frames along the entire VCV sequence using the Articulate Assistant Advanced (AAA) software and adjusted manually by the second paper author whenever necessary. Data points for the lingual splines were exported in an ASCII-file as *x-y* coordinates and acoustic files were also exported for taking segmental duration and spectral measures.

Tongue configuration measurements were performed at the midpoint of V1, C, and V2 after identifying V1, C, and V2 onsets and offsets on the spectrographic and waveform displays. V1 onset and V2 offset were taken to occur at the onset and offset of vowel-related formant structure, respectively. The acoustic boundary between the vowels and the intervocalic consonant was identified with the beginning and end of the frication noise for fricatives, of a low intensity formant structure for laterals, nasals, and approximants, and of one short closure for the tap and of two or more short closures for the trill. As for stops, the consonant was equated with the closure period beginning at the offset of the V1 formant transitions and ending at the short stop burst, and V2 onset was located at the first pitch pulse occurring after the burst.

The ultrasound data were converted from Cartesian into polar coordinates by shifting the origin of the ultrasound image to the center of the ultrasound field of view, which was located at X = 86.7 mm and Y = 0 mm. This operation is justified by the fact that the tongue surface typically approximates an arc more closely than a horizontal line (Mielke, 2015). Then SSANOVA computation was applied using the R package gss to find a best fit curve to all spline tokens (Davidson, 2006). The SSANOVA smoothed VCV splines consisted of strings of the same number of points separated by 0.01 radians with the associated standard errors corresponding to the radial coordinate variability at each point.

In order to measure coarticulatory resistance at meaningful tongue regions, the length of the SSANOVA derived splines was divided into four portions corresponding to the alveolar, palatal, velar and pharyngeal articulatory zones applying the following criteria separately to each subject's data. The boundary between the alveolar and palatal zones was set at the approximate constriction location for the alveolar trill as determined visually by a tongue front inflection point occurring on the [VrV] spline curves, and the boundary between the palatal and velar zones at the dorsovelar closure for [k] in the sequence [iki] as determined by the mean Y displacement maximum across tokens. These constriction sites were selected based on EPG data for several subjects showing that in Catalan [r] is generally articulated at the back alveolar zone and that the [k] closure next to [i] occurs consistently at the postpalatal zone just in front of the soft palate. The length of the velar zone, on the other hand, was taken to be 1.26 and 1.51 times that of the palatal zone in the case of the male and female speakers, respectively (Fitch and Giedd, 1999). The only exception was the female subject IM for whom we had to apply the male's ratio since otherwise the velar zone would have been too long relative to the palatal zone and the dorsovelar closure for [uku] much more anterior than expected.

The graphs in Fig. 1 show an example of the output lingual splines for [ɾ] in all five vowel contexts [i, e, a, o, u] after conversion to polar coordinates, SSANOVA smoothing and spline subdivision into articulatory zones.

B. Coarticulation index analysis procedure

NN, HTD, AAZ, and CVar were calculated using the SSANOVA smoothed VCV splines for the 10 consonants at the consonant midpoint and for the five vowels at the V1 and V2 midpoints. On the other hand, LE and MI were evaluated for each of the 10 consonants at the midpoint of the consonant and of the two vowels (MI) or only of V2 (LE) using tongue position measures on unsmoothed splines for the individual VCV tokens in view of the large number of required data points. Moreover, NN, HTD, AAZ, CVar, and LE values were computed for each individual subject and pooling all of the data over all subjects, and MI only for the latter condition.

NN index values for pairs of VCV sequences were obtained by averaging all *x* point distances for each pair of splines at a given articulatory zone, e.g., there were 54, 33, 29, and 17 point distances at the pharyngeal, velar, palatal and alveolar zones, respectively, for speaker DR. This resulted in 10 NN values for each consonant (10 sets of two-spline comparisons) and 45 NN values for each vowel (45 sets). The means across each set of 10 and 45 NN values yielded the unnormalized NN values for the 10 target consonants and five target vowels. A normalization procedure for each consonant, articulatory zone and speaker was carried out by dividing each of the NN values for a given pair of splines by the mean of all NN values across consonants (10 consonants × 10 spline pairs = 100), and averaging the resulting 10 normalized distance values. An analogous normalization procedure was carried out for each vowel: we divided each of the NN values for a given pair of splines by the mean of all NN values across vowels (5 vowels × 45 spline pairs = 225), and averaged the resulting 45 normalized distance values.

As for the HTD method, we measured the area of the polygons delimited by the tongue dorsum Y maxima for the five contextual vowels at the midpoint of each consonant and by the 10 contextual consonants at the midpoint of V1 and V2 for each of the five vowels. Normalization was carried out by dividing the resulting area values by a measure of the maximal available vertical articulatory space for each speaker. This measure was taken to be the difference between the vertical tongue dorsum displacement maxima and minima across all consonants and vowels of the dataset measured at the consonant and vowels midpoints, respectively.

The AAZ values were obtained by computing the area of the polygons delimited by the outermost splines at each articulatory zone. The resulting values were normalized separately for consonants and vowels and speaker at each articulatory zone as follows: we subtracted the mean area value across all consonants or vowels from the area value of each consonant or vowel, divided the difference by the standard deviation of the mean, and rendered positive any resulting negative number. On the other hand, CVar was calculated for each consonant and vowel by averaging all coefficients of variation values across contextual splines obtained at each of the *x* spline points within a given articulatory zone. No normalization procedure was performed in this case since CVar is already a normalized measure of the variation in a data set.

The LE slope and MI values were computed with the *x* tongue position distances between the origin of the ultrasound field of view and the midpoint of each articulatory zone as follows: we first measured on the SSANOVA transformed splines the angle delimited by a line connecting those two points and an intersecting line crossing the origin horizontally; then, we identified the corresponding *x* point on the splines for the individual VCV tokens at that same angle and measured the distance between this *x* point and the origin. The number of distance values across subjects at each articulatory zone was 145 for each consonant (29 tokens × 5 contextual vowels) and 290 for each vowel (29 tokens × 10 contextual consonants). One of the decisions that had to be made for computing MI was the number of bins included in the estimation of the histogram function (see Iskarous *et al.*, 2013 for details). It was found that MI index calculation with five bins discriminated better among the different consonants and vowels than using either three bins or more than five bins.

Tables I, II, and III present the cross-speaker unnormalized and normalized index values for consonants and vowels according to all computation methods. Only the normalized values were submitted to statistical analysis. Separate linear mixed model (LMM) analyses were performed on the NN, AAZ, CVar, and LE data for consonants, and on the NN, AAZ, and CVar data for vowels, using the "mixlm" software package of R version 3.1.2 (R Developmental Core team 2014). The independent variables were "Vowel," "Consonant," and "Zone" for the consonant test, and "Vowel," "Consonant," "Zone," and "Position" for the vowels test. "Subject" was as a random factor. The fixed factor levels were "[ð, l, ɾ, t, n, s, r, ʎ, ɲ, ʃ]" for "Consonant," "[i, e, a, o, u]" for "Vowel," "alveolar," "palatal," "velar" and "pharyngeal" for "Zone," and "V1" and "V2" for "Position." *Post hoc* Tuckey tests were run on those main effects and factor interactions which reached significance. The significance level was set at $p < 0.05$. Given the nature of the study only the relevant statistical results will be reported. No statistical evaluation was carried out of the index values for HTD (there were no data for the different articulatory zones) nor of MI (there were no data for the individual subjects).

III. RESULTS

A. Consonants

The LMM tests performed on the NN, AAZ, and CVar values yielded a significant main "consonant" effect ($F(9, 156) = 88.31$, $p < 0.001$; $F(9, 156) = 79.45$, $p < 0.001$; $F(9, 156) = 38.47$, $p < 0.001$). Figure 2 (top, three left graphs) gives the significant effects for the corresponding pairwise comparisons obtained from the *post hoc* tests. In order to interpret the sign of these differences, Fig. 3 displays for each articulatory zone the proportions between the index values for the individual consonants and the maximal index value of all consonants which happened to be that for [ð] at the palatal, velar and pharyngeal zones and that for [ɾ] at the alveolar zone. In the graphs of the figure the higher the proportion, the more variable and thus the less resistant the consonant is supposed to be.

Inspection of the data plotted in the two figures reveals highly similar consonant-dependent differences in coarticulatory resistance according to the three methods. In all cases coarticulatory resistance was found to decrease significantly across zones in the progression [ʎ, ɲ, ʃ] (least variable, most resistant) > [s, r] > [l, ɾ, t, n] > [ð] (most variable, least resistant), which is largely in agreement with the predictions of the DAC model. Indeed, [ð] differs significantly from all

TABLE I. NN, AAZ, CVar, LE, and MI index values for consonants. unnorm = unnormalized, norm = normalized, PHAR = pharyngeal, VEL = velar, PAL = palatal, ALV = alveolar.

	Consonant	NN unnorm	NN norm	AAZ unnorm	AAZ norm	CVar	LE	MI		Consonant	NN unnorm	NN norm	AAZ unnorm	AAZ norm	CVar	LE	MI
PHAR	ð	5.771	1.427	438.113	2.857	10.215	0.872	0.719	PAL	ð	5.968	1.529	203.739	2.952	7.342	0.905	0.914
	l	4.727	1.198	395.132	2.529	8.708	0.832	0.718		l	4.695	1.180	165.927	2.364	5.515	0.643	0.691
	ɾ	4.493	1.126	365.406	2.165	8.112	0.891	0.580		ɾ	4.409	1.123	164.388	2.351	5.052	0.786	0.588
	t	4.895	1.233	390.337	2.451	9.408	0.820	0.671		t	4.500	1.125	164.321	2.340	5.273	0.727	0.730
	n	4.502	1.115	328.476	1.855	7.989	0.858	0.624		n	4.348	1.089	158.742	2.247	4.997	0.699	0.877
	s	3.948	0.976	381.532	2.334	8.348	0.769	0.576		s	2.986	0.724	111.283	1.573	3.551	0.624	0.635
	r	3.231	0.797	300.972	1.677	6.538	0.704	0.518		r	2.762	0.663	103.841	1.423	3.213	0.565	0.625
	ʎ	2.503	0.613	215.949	0.895	5.938	0.745	0.494		ʎ	1.501	0.374	60.753	0.875	1.605	0.310	0.505
	ɲ	2.188	0.546	150.460	0.310	4.533	0.713	0.563		ɲ	1.322	0.334	54.082	0.761	1.413	0.395	0.577
	ʃ	2.045	0.500	204.225	0.748	4.937	0.629	0.440		ʃ	1.332	0.332	49.829	0.715	1.453	0.343	0.569
VEL	ð	5.908	1.503	231.656	3.258	6.395	0.995	0.830	ALV	ð	3.720	1.168	75.259	2.295	4.503	0.849	0.774
	l	5.314	1.374	198.547	2.692	5.171	0.843	0.763		l	2.668	0.852	55.822	1.877	3.033	0.447	0.544
	ɾ	4.416	1.131	187.828	2.432	4.855	0.876	0.783		ɾ	4.148	1.327	84.254	2.589	4.899	0.655	0.398
	t	4.649	1.183	199.527	2.657	5.360	0.904	0.739		t	2.911	0.882	65.980	1.976	3.324	0.475	0.543
	n	4.643	1.167	192.007	2.534	4.967	0.902	0.825		n	2.269	0.707	48.316	1.703	2.580	0.425	0.546
	s	3.465	0.863	163.032	1.919	4.288	0.845	0.734		s	1.534	0.451	35.567	1.307	1.746	0.546	0.602
	r	3.118	0.763	137.138	1.470	3.371	0.750	0.727		r	2.285	0.708	50.405	1.735	2.561	0.576	0.597
	ʎ	2.090	0.524	110.487	0.923	2.857	0.766	0.637		ʎ	0.905	0.285	21.417	1.058	1.019	0.039	0.285
	ɲ	2.010	0.509	97.240	0.780	2.603	0.862	0.570		ɲ	0.948	0.313	20.891	1.103	1.043	0.245	0.364
	ʃ	1.756	0.434	89.424	0.538	2.373	0.720	0.645		ʃ	0.817	0.262	16.970	0.974	0.913	0.077	0.468

other consonants, while [ʎ, ɲ, ʃ] differ significantly from all dentoalveolars (except for [ʎ] vs [r] in the case of the CVar index) but not among themselves. Regarding dentoalveolars, [s, r] do not differ from each other and are less variable and thus more resistant than [l, ɾ, t, n] with a few exceptions ([s] may not differ significantly from [l, n]).

There was also a significant consonant × zone interaction in the case of the NN and AAZ indices ($F(27, 156) = 2.08$, $p < 0.001$, $F(27, 156) = 3.05$, $p < 0.01$) which, as shown by the graphs of Fig. 3, is associated essentially with more contextual variability and thus less coarticulatory resistance for [s, r] at the velar and pharyngeal zones than at the palatal zone, and with more variability for the apical [ɾ] than for the other consonants at the alveolar zone than at the other zones.

The proportions plotted in the four graphs of Fig. 3 also show that the three methods agree mostly at the palatal zone, and disagree most especially in the case of the palatal consonants [ʎ, ɲ, ʃ] at the alveolar, velar and pharyngeal zones. Moreover, AAZ contributes better than NN and CVar to differentiate between palatal and non-palatal consonants at the pharyngeal and velar regions probably since the former method is less sensitive than the two latter ones to whether the *x* points on the contextual splines lie close to or far apart from each other.

TABLE II. HTD index values for consonants (left) and vowels measured at V1 and V2 (right).

Consonant	unnormalized	normalized	V1	unnormalized	normalized
ð	118.282	7.960	i	3.903	0.277
l	69.897	5.114	e	3.778	0.259
ɾ	60.131	4.113	a	20.987	1.298
t	88.340	6.246	o	9.700	0.641
n	57.326	3.802	u	10.356	0.660
s	57.806	3.816			
r	31.116	2.088	V2	unnormalized	normalized
ʎ	4.590	0.338	i	10.220	0.877
ɲ	4.182	0.298	e	14.210	1.185
ʃ	3.630	0.245	a	43.407	3.543
			o	41.377	3.237
			u	24.998	2.149

The statistical test run on the HTD index values yielded a significant main effect of consonant ($F(9, 36) = 13.28$, $p < 0.001$). As shown by the proportions over the maximal index value displayed in Fig. 4 (top), differences in coarticulatory resistance among consonants are generally consistent with the predictions of the DAC model. Given that the highest point of the tongue dorsum occurs at the palatal or velar zone, the HTD index values of interest should be compared to the NN, AAZ, and CVar values obtained at those two zones. According to results from the *post hoc* tests plotted in Fig. 2 (top, fourth graph), HTD does worse than NN, AAZ, and CVar regarding the classification of [s, r] since these consonants turned out to be generally non-significantly different from [l, ɾ, t, n] and thus showed higher index values than predicted by the DAC model.

As regards to the LE and MI ratios displayed in Fig. 5, differences in coarticulatory resistance among consonants also conform to the initial prediction that palatals should be less variable and thus more resistant than dentoalveolars and [ð] the least resistant of all consonants (statistical results for the LE index values yielded a significant main consonant effect; $F(9, 156) = 14.95$, $p < 0.001$).

The number of significant differences among consonants was less for LE than for NN, AAZ and CVar (see Fig. 2, top, rightmost graph). Analogously to HTD, [s, r] lie closer to the

TABLE III. NN, AAZ, and CVar index values for vowels at V1 (left) and at V2 (right). unnorm = unnormalized, norm = normalized, PHAR = pharyngeal, VEL = velar, PAL = palatal, ALV = alveolar.

	V1	NN unnorm	NN norm	AAZ unnorm	AAZ norm	CVar		V2	NN unnorm	NN norm	AAZ unnorm	AAZ norm	CVar
PHAR	i	1.164	0.863	124.972	1.752	2.433	PHAR	i	1.916	0.789	207.833	1.105	4.263
	e	0.989	0.723	100.382	1.276	1.711		e	1.941	0.803	203.094	1.048	3.831
	a	1.322	0.943	116.271	1.691	1.561		a	2.375	1.005	292.801	2.070	4.349
	o	1.655	1.185	123.443	1.850	1.629		o	2.854	1.175	344.619	2.642	5.466
	u	1.806	1.286	195.345	3.430	2.847		u	2.973	1.228	388.840	3.135	6.721
VEL	i	0.922	0.627	70.574	1.606	1.403	VEL	i	1.325	0.542	113.650	1.329	2.260
	e	0.845	0.560	60.641	1.237	1.251		e	1.478	0.589	115.030	1.525	2.377
	a	1.522	0.992	81.477	1.817	1.853		a	3.034	1.208	124.738	1.664	2.915
	o	2.090	1.371	93.791	2.434	2.141		o	3.567	1.392	143.439	2.344	3.232
	u	2.245	1.449	105.062	2.906	2.040		u	3.269	1.268	157.741	3.139	3.176
PAL	i	0.866	0.509	45.193	0.998	0.988	PAL	i	1.251	0.462	63.875	0.904	1.440
	e	0.806	0.461	52.768	1.191	1.169		e	1.503	0.531	74.561	1.028	1.792
	a	1.685	0.976	79.830	2.149	2.377		a	3.238	1.163	152.886	2.386	4.230
	o	2.331	1.328	96.328	2.454	3.129		o	3.957	1.412	173.572	2.717	5.093
	u	3.002	1.726	126.325	3.208	3.788		u	4.015	1.432	187.692	2.965	5.004
ALV	i	0.931	0.646	23.225	1.237	0.996	ALV	i	1.418	0.641	38.706	1.401	1.720
	e	0.934	0.609	27.244	1.213	1.143		e	1.741	0.774	45.519	1.334	2.074
	a	1.394	0.954	37.866	2.176	2.104		a	2.159	0.968	55.813	1.697	2.703
	o	1.688	1.052	41.614	2.314	2.473		o	2.737	1.231	65.488	2.341	3.469
	u	2.715	1.739	66.233	3.061	3.246		u	3.019	1.386	81.133	3.227	3.821

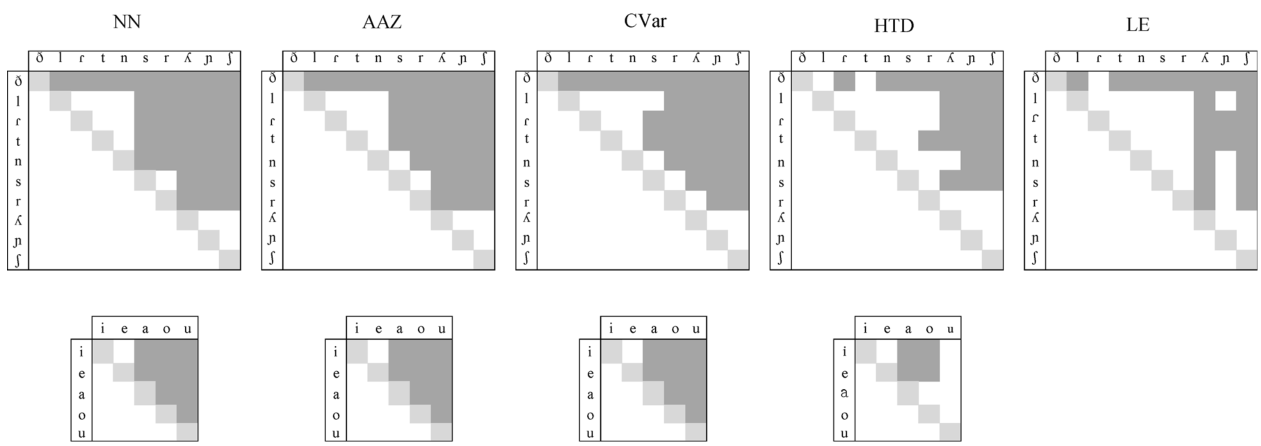

FIG. 2. Results from Tuckey *post hoc* tests run on the consonant (top) and vowel (bottom) main effects. Data correspond to the NN, AAZ, CVar, HTD, and LE indices. The dark shadings indicate significant differences.

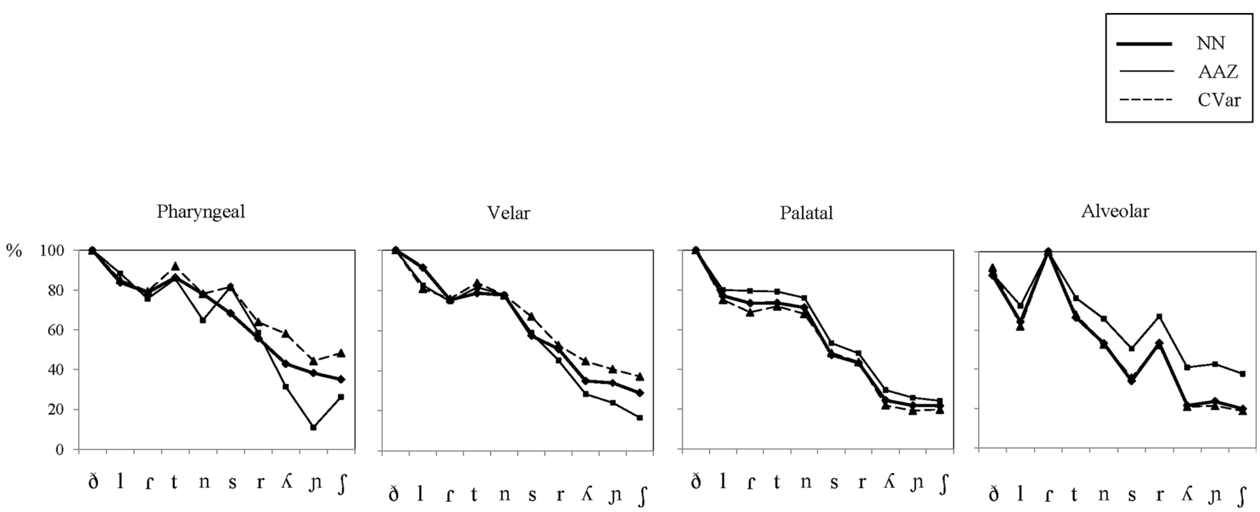

FIG. 3. Cross-speaker NN, AAZ, and CVar ratios over the maximum index value for consonants at the four articulatory zones.

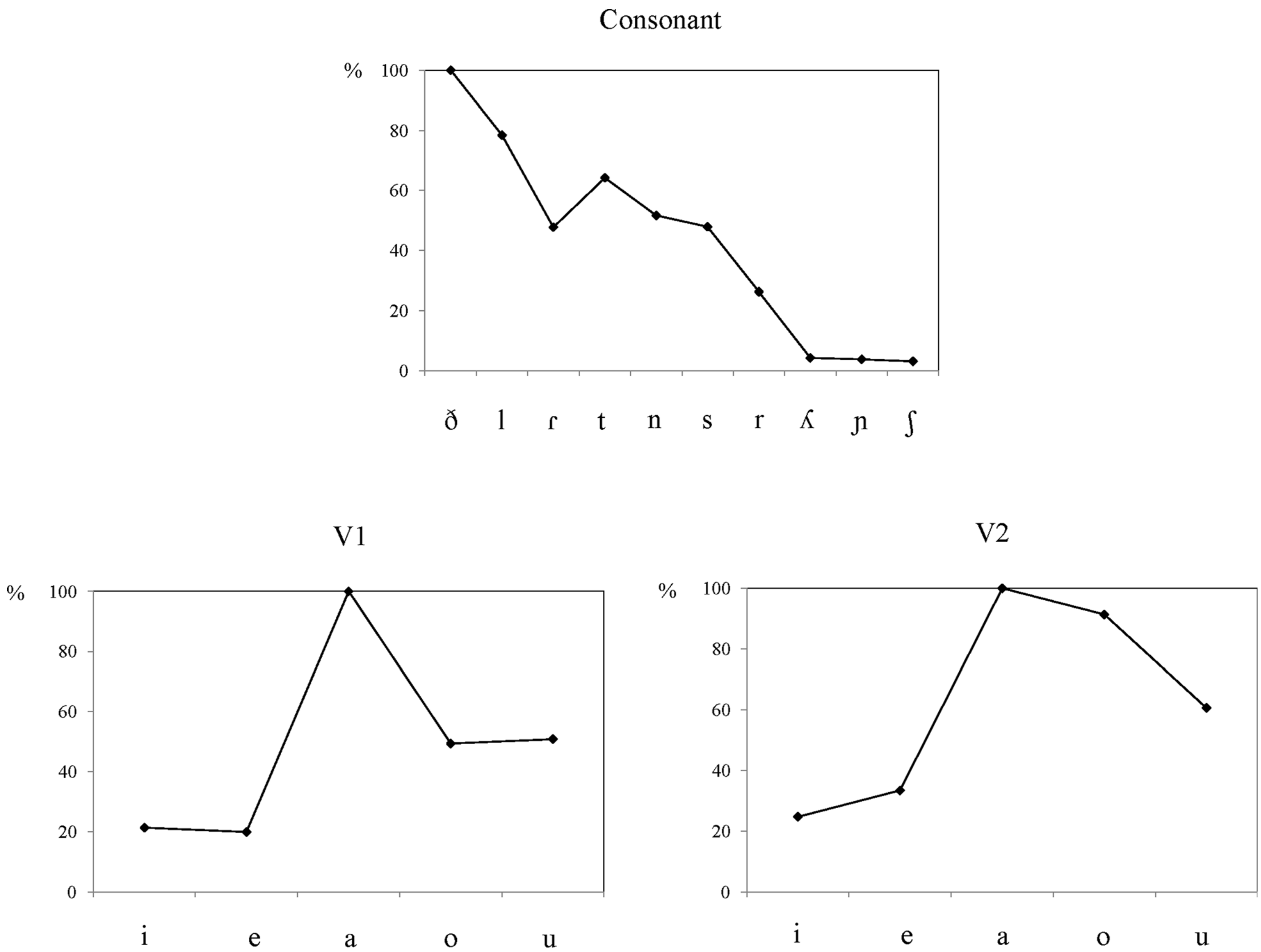

FIG. 4. Cross-speaker HTD ratios over the maximum index value for consonants (top) and vowels at V1 and at V2 (bottom).

other dentoalveolars than to palatals (i.e., they differ significantly from [ʎ, ʃ] though not from [ɲ] nor from [l, ɾ, t, n]) and therefore do not occupy an intermediate position between palatal and non-palatal consonants. On the other hand, [ɲ] is somewhat more variable and thus less resistant than [ʎ, ʃ] which is not consistent with data for the other indices and the predictions from the DAC model.

A similar scenario arises at the different articulatory zones (the LE data yielded a significant consonant × zone interaction; $F(27, 156) = 1.98$, $p < 0.01$). Thus, as shown in Fig. 5, consonant-dependent differences at the velar and pharyngeal zones were relatively small and turned out to be non-significant in the case of the LE method and to a large extent MI as well. Differences at the palatal zone were also

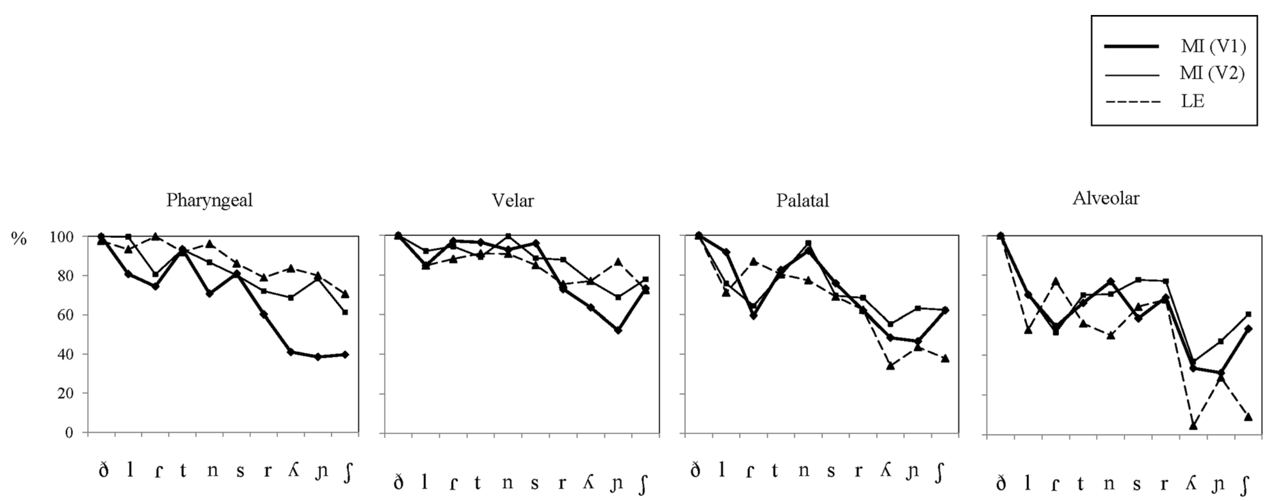

FIG. 5. Cross-speaker MI and LE ratios over the maximum index value for consonants at the four articulatory zones. Separate MI values are given for V1 and V2.

smaller than expected and, in the case of LE, only achieved significance for [ð] vs [ʎ, ɲ, ʃ] and for [ɾ] vs [ʎ, ʃ]; moreover, contrary to the initial expectations, the MI value for [ɾ] at this articulatory zone was as low as that for [r] and lower than that for [s]. Also at the alveolar zone [s, r] were not clearly distinguished from the other dentoalveolar consonants (*post hoc* tests for the LE data at this articulatory region yielded significant differences between [ð, ɾ, s, r] and [ʎ, ɲ, ʃ]).

B. Vowels

LMM tests run on the vowel data yielded a significant main vowel effect in the case of all three methods NN, AAZ and CVar ($F(4, 156) = 144.02$, $p < 0.001$; $F(4, 156) = 80.59$, $p < 0.001$; $F(4, 156) = 53.04$, $p < 0.001$). As shown by results from the *post hoc* tests displayed in Fig. 2 (bottom, three left graphs) and the proportions plotted in Fig. 6, coarticulatory resistance across zones decreases in the progression [i, e] (least variable, most resistant) > [a] > [o] > [u] (most variable, least resistant). This scale is largely in accordance with the initial prediction that palatal vowels ought to allow least coarticulation and back labials the greatest amount. Moreover, a joint inspection of the graphs displayed in Fig. 6 reveals that differences in coarticulatory resistance among vowels obtained with the three different methods are in closer agreement at the palatal zone than at the other articulatory zones. The presence of higher index values for NN than for AAZ at the pharyngeal and velar zones appears to be due to an overestimation of large point-to-point distances for specific pairs of splines.

Statistical tests run on the data for the NN and CVar methods yielded a significant vowel × zone interaction ($F(12, 156) = 4.48$, $p < 0.001$; $F(12, 156) = 5.94$, $p < 0.001$), which was associated with more coarticulatory variability and thus less coarticulatory resistance for palatal vowels, and a smaller contrast between [i, e] and [a], at the pharyngeal zone than at the other zones. There also was a main effect of position for CVar ($F(4, 156) = 204.3$, $p < 0.001$) and a significant vowel × position interaction for NN ($F(4, 156) = 3.32$, $p < 0.01$) and CVar ($F(4, 156) = 2.98$, $p < 0.02$) which, as shown by the proportions for V1 and V2 in Fig. 6, were related to more contextual variability and thus less coarticulatory resistance for [a, o, u] at V2 than at V1. This positional effect was related to greater differences in tongue body configuration as a function of palatal vs non-palatal consonants at V2 than at V1.

Results obtained with HTD differ from those obtained with NN, AAZ and CVar in that the former method yielded greater variability and thus less coarticulatory resistance for [a] than for [o, u] (see Fig. 4, bottom). As referred to in the Introduction, a very large polygon area for the low vowel in this case follows from the tongue dorsum *y* maxima in the context of dentoalveolar consonants being located near the alveolar zone and thus too much dependent on the tongue front raising gesture for these contextual consonants. In agreement with this scenario, results from the *post hoc* tests run on the HTD index values for vowels reveal no significant differences between [a] and [o, u], between the two back rounded vowels nor between [i, e] and [u] (see Fig. 2, bottom rightmost graph).

IV. DISCUSSION

The paper has tested the efficiency of several methods for computing the degree of lingual coarticulatory resistance at several articulatory zones using lingual spline data for 10

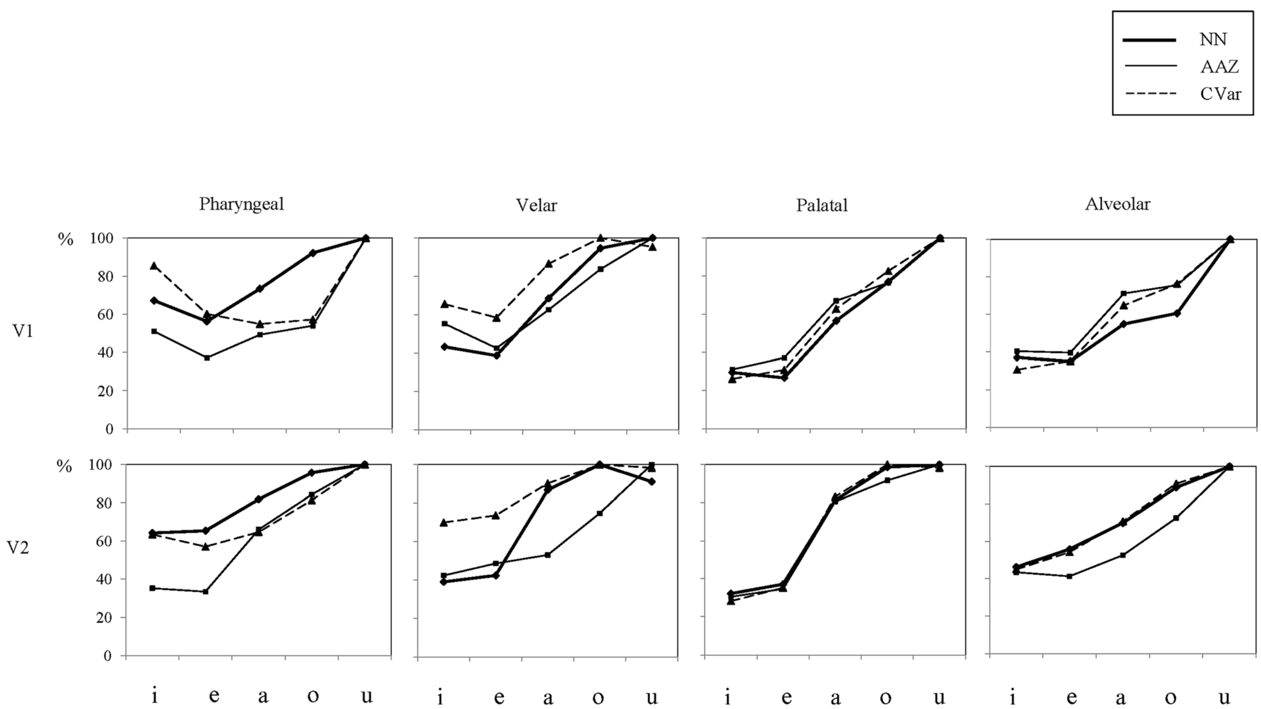

FIG. 6. Cross-speaker NN, AAZ, and CVar ratios over the maximum index value for vowels at the four articulatory zones. Separate index values have been plotted for V1 (top) and V2 (bottom).

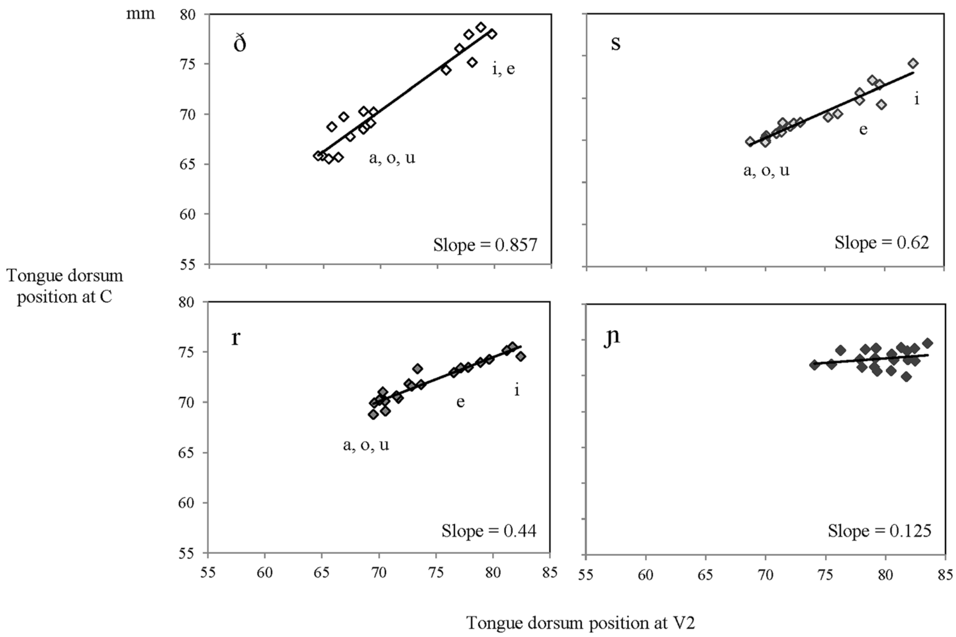

FIG. 7. LE regression slopes for the consonants [ð], [s], [r], and [ɲ] at the palatal zone according to speaker DR. Tongue dorsum position distances from the origin are given for different contextual vowels both at the midpoint of the consonant (Y axis) and at the midpoint of V2 (X axis).

front lingual consonants and five vowels of Catalan recorded in symmetrical VCV sequences.

NN, AAZ, and CVar turned out to distinguish consonants and vowels somewhat more accurately than HTD, LE, and MI. Results for the three former methods were generally consistent with the DAC model of coarticulation in that coarticulatory resistance diminishes in the progression [ʎ, ɲ, ʃ] (least variable, most resistant) > [s, r] > [l, ɾ, t, n] > [ð] (most variable, least resistant) and [i, e] (most resistant) > [a] > [o, u] (least resistant). Moreover, AAZ turned out to do somewhat better than NN and CVar in setting in contrast extreme coarticulatory resistance values for specific consonants and vowels at the rear of the vocal tract.

It was also found that LE and MI are better predictors of coarticulatory resistance for consonants which are most and least resistant than for those which show intermediate degrees of coarticulatory resistance (i.e., [s, r]). The presence of similar degrees of coarticulatory resistance for [s, r] and the other dentoalveolars did not conform to the initial expectations. Moreover, in comparison to NN, AAZ and CVar, LE and MI yielded a smaller distance between the coarticulatory resistance values for dentoalveolars and palatals.

LE slopes for consonants exhibiting high, low and intermediate coarticulatory resistance values will be presented next in order to explore why LE and MI behave the way they do. Highest LE slope values occur for consonants allowing much V-to-C coarticulation and exerting little C-to-V coarticulation, as for [ð]; indeed, the regression slope for the tongue dorsum data at the palatal zone for this consonant plotted in Fig. 7 (top left graph) approaches 1 since the tongue dorsum position both at the consonant and vowel periods is much higher and more anterior in the context of palatal vowels than in the context of low and back rounded vowels. On the other hand, slope values for palatal consonants are close to zero since, as shown by the tongue dorsum data for [ɲ] in the bottom right graph of the same figure, these consonants block vowel coarticulation most and exert the strongest effect on vowels; in this particular case, there are essentially no vowel-dependent differences in tongue dorsum position at the consonant and the position values at the vowel midpoint are much closer to each other than those for [ð] and conform to no obvious vowel pattern.

Regarding consonants with intermediate degrees of coarticulatory resistance, the more common scenario is exemplified by the regression slopes for [s] and [r] in Fig. 7 (top right and bottom left graphs). The slopes in question are flatter than for [ð] and steeper than for [ɲ] since V-to-C coarticulation is less than for the dental and greater than for the palatal while C-to-V coarticulation is greater than for [ð] and less than for [ɲ]. The reason why the LE (and presumably MI) index values for consonants like [s] turn out to be higher than expected appears to be due to the fact that, given that the C-to-V effects exerted by this consonant are not too large and in spite of the corresponding V-to-C effects being relatively small, the vowel-dependent differences in tongue dorsum position occurring during the consonant and the vowel are essentially the same ([i] > [e] > [a, o, u]) and thus still relatively highly correlated. The reason for this outcome is that LE and MI evaluate the relationship between coarticulatory resistance and coarticulatory aggression, i.e., the extent to which consonants affect vowels during the vowel period, and therefore do not provide direct measures of coarticulatory resistance. Therefore, regression values between the two dimensions are prone to reflect coarticulatory resistance for consonants which are maximally and minimally resistant but not for consonants showing intermediate resistance degrees.

The initial goal of this investigation was to come up with a set of index methods for computing coarticulatory resistance for consonants and vowels in dialectal studies using ultrasound data. In view of the results obtained, the most accurate methods appear to be NN, AAZ, and CVar. In spite of using different algorithms, all three methods yield similar results, which are in accordance with the initial predictions of the DAC model, since are based on a point-by-point data evaluation of coarticulatory variability. Moreover, AAZ accounts better than NN and CVar for the consonant- and vowel-dependent differences in coarticulatory resistance. Future work needs to be carried out in order to find out how successful these methods are in dealing with token-to-token and speaker-dependent variability, which should be of greatest relevance to dialectal studies as well.

ACKNOWLEDGMENTS

This research was supported by project FFI2013-40579-P of the Spanish Ministry of Economy and Competitiveness, by the ICREA Academia program, and by project 2014SGR61 of the Catalan Government. We thank Khalil Iskarous for providing us with the MatLab script for computing the MI index.

Davidson, L. (**2006**). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," J. Acoust. Soc. Am. **120**, 407–415.

Fitch, W., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Gick, B. (**2002**). "The use of ultrasound for linguistic phonetic fieldwork," J. Int. Phonet. Assoc. **32**, 113–121.

Hoole, P., Gfroerer, S., and Tillmann., H. G. (**1990**). "Electromagnetic articulography as a tool in the study of lingual coarticulation," Forschungsber. Inst. Phonet. Sprachliche Kommunik. Universität München **28**, 107–122.

Iskarous, K., Fowler, C., and Whalen, D. H. (**2010**). "Locus equations are an acoustic expression of articulator synergy," J. Acoust. Soc. Am. **128**, 2021–2032.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C., Saltzman, E., and Whalen, D. (**2013**). "The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance," J. Acoust. Soc. Am. **134**, 1271–1282.

Lin, S., Beddor, P. S., and Coetzee, A. W. (**2014**). "Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/," Lab. Phonol. **5**, 9–36.

Lubker, J., and Gay, T. (**1982**). "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables," J. Acoust. Soc. Am. **71**, 437–448.

Mielke, J. (**2015**). "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons," J. Acoust. Soc. Am. **137**, 2858–2869.

Noiray, A., Menard, L., and Iskarous, K. (**2013**). "The development of motor synergies in children: Ultrasound and acoustic measurements," J. Acoust. Soc. Am. **133**, 444–452.

Ohala, J. J., and Solé, M. J. (**2010**). "Turbulence and phonology," in *Turbulent Sounds. An Interdisciplinary Guide*, edited by S. Fuchs, M. Toda, and M. Zygis (Mouton de Gruyter, Berlin), pp. 37–97.

Proctor, M. I. (**2009**). "Gestural characterization of a phonological class: The liquids," Ph.D. dissertation, Yale University.

Recasens, D., and Espinosa, A. (**2009**). "An EMA investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan VCV sequences," J. Acoust. Soc. Am. **125**, 2288–2298.

Recasens, D., and Rodríguez, C. (**2016**). "An investigation of lingual coarticulation resistance using ultrasound," J. Phonetics **59**, 58–75.

Stone, M. (**2005**). "A guide to analysing tongue motion from ultrasound images," Clin. Linguist. Phonetics **19**, 455–501.

Zharkova, N., Hewlett, N., and Hardcastle, W. J. (**2011**). "Coarticulation as an indicator of speech motor control development in children: An ultrasound study," Motor Control **15**, 118–140.