THE EFFECT OF VOICING ON TONGUE CONFIGURATION FOR UNASPIRATED STOP SEQUENCES

Daniel Recasens & Clara Rodríguez

Departament de Filologia Catalana, Universitat Autònoma de Barcelona
daniel.recasens@uab.es

ABSTRACT

Data for English aspirated stops reported in the literature reveal a greater expansion of the vocal tract cavities for voiced stops than for voiceless stops, which takes place mostly at the pharynx. Lingual configuration data for the sequences [kt], [gd], [tk] and [dg] in Catalan, where C1 agrees with C2 in voicing and oral stops are unaspirated, also exhibit more tongue postdorsum fronting throughout the entire cluster if voiced than voiceless. In contrast with English, however, the front dorsum is not lower for [gd] than for [kt] and for [dg] than for [tk], which could be related to the articulatory and aerodynamic differences involved in the production of aspirated vs unaspirated stops.

Keywords: stop voicing, ultrasound, pharyngeal cavity expansion, unaspirated stops.

1. INTRODUCTION

Several studies have reported that, in comparison to voiceless stops, voiced stops in English exhibit an active expansion of the supraglottal cavity system during the closure period. Moreover, this volume increase is achieved mainly by fronting the tongue body and thus enlarging the pharyngeal cavity [1, 3, 5], and, albeit less consistently, by lowering the tongue front and thus enlarging the oral cavity [3, 5]. This lingual action facilitates voicing by lowering the intraoral pressure level above the glottis [2]. The present study investigates whether differences in postdorsum fronting at the back of the vocal tract, and perhaps in tongue blade and predorsum height in the oral cavity, also occur in heterosyllabic stop clusters of Catalan where, differently from English, voiceless stops are unaspirated.

Catalan has a regressive voicing assimilation rule by which a syllable-final obstruent assimilates in voicing to the following consonant; by virtue of this assimilatory process, in heterosyllabic stop-stop sequences the syllable-coda stop C1 is supposed to be voiced or voiceless phonetically depending on whether the syllable-onset stop C2 is underlyingly voiced or voiceless, respectively [4]. Consequently, the two consonants should agree in voicing and thus be realized as [kt], [gd] and so on. Under these circumstances it it worth finding out whether a presumable increase in pharyngeal and oral cavity size occurs not only during the stop voicing trigger but during the preceding stop as well.

The effect of voicing on tongue configuration will be analyzed for a selected set of consonant sequences split by a word boundary with ultrasound, which allows tracing the tongue body configuration with the exception of its edges. Even though ultrasound does not allow measuring the actual differences in cavity size that may occur between voiced and voiceless stops, it is believed that such differences may be inferred to a large extent from changes in tongue position. Thus, tongue body fronting ought to reflect an increase in pharyngeal cavity size, while tongue predorsum and blade lowering should correspond to an increase in the size of the oral cavity.

2. METHODOLOGY

Lingual configuration data were collected for four heterosyllabic two-stop sequences preceded and followed by a (mid) low vowel in meaningul Catalan sentences: [kt] (*és un sa**c** **t**ou* 'this is a soft sack'), [gd] (*un conya**c** **d**'anys* 'an old cognac'), [tk] (*un solda**t** **c**alb* 'a bold soldier'), [dg] (*ha compra**t** **g**all* '(s)he has bought rooster'). These sentences were recorded six times by five native Catalan speakers, i.e., two men (DR, RO) and three women (ES, JU, IM), of 40-60 years of age who speak Catalan regularly in their every day life. Ultrasound recordings were performed with an Echo Blaster unit type EB128CEXT from TELEMED and a microconvex Echo Blaster 128 CEXT transducer with a 2 to 4 MHz frequency range and a central curvature of 20 mm. The ultrasound images were acquired using a probe with a 100% of 104° field of view and a frequency of 2 MHz, which was attached to a transducer holder positioned under the subject's chin in an Articulate Instruments Stabilization Headset. The recording sampling rate was 54 frames per second yielding one image every 18.5 ms. Image streams were recorded synchronously with the audio signal sampled at 22,050 Hz with an AKG-D70 microphone. Contours of the back of the alveolar zone and hard palate were also recorded by asking speakers to press the tongue against their hard palate.

Tongue contours were tracked automatically at all temporal frames along all C#C sequence tokens using the Articulate Assistant Advanced (AAA) software and adjusted manually by the first paper author. Data points for all tongue contours were exported in ASCII-files as *x*-*y* coordinates with their origin located at the bottom-left corner of the ultrasound image towards the rear of the vocal tract. Acoustic files were also exported in .wav format for taking segmental duration measures.

Segmentation was carried out on waveform and spectrographic displays. Lingual spline data were processed at the following six temporal points: C1 closure onset, which was located at the offset of formant structure for the preceding vowel; C1 closure midpoint; C1 closure offset, just before the short C1 burst; C2 closure onset, just after the C1 burst; C2 midpoint; C2 offset, just before the C2 burst.

Tongue spline data points were converted from Cartesian to polar coordinates by shifting the origin of the ultrasound image to approximately the center of the ultrasound probe which was located at X= 86.7 mm and Y=0 mm. SSANOVA smoothed splines consisting of strings of points separated by 0.01 radians and the associated standard errors were computed across the splines for all tokens of each consonant sequence using the R package gss to find a best fit curve. The smoothed splines of all consonant sequences had the same number of x points since their rightmost and leftmost edges were determined by entering into the SSANOVA computation procedure the mean angle radian values across tokens of all clusters under analysis.

In order to determine the tongue configuration at different vocal tract regions, the length of the SSANOVA splines displayed in Cartesian coordinates was divided into four portions which correspond roughly to the alveolar (ALV), palatal (PAL), velar (VEL) and pharyngeal (PHAR) articulatory zones. This subdivision procedure was carried out separately for each subject by applying the same criterion as in a previous paper dealing with other consonant sequences [5] since the data for the two studies were acquired in the same recording session. Distances between each of the four lingual regions and the origin of the ultrasound field of view were measured at the six temporal points referred to above. The distance values at the velar and palatal zones were obtained by averaging the distances between the five central points at each zone and the origin. Given that the splines for the consonant sequences subject to analysis could differ in length, the distance values for the two extreme zones, alveolar and pharyngeal, were computed by averaging the distances between the origin and five points located not at the zone midpoint but at the upper third of the pharyngeal zone and at the leftmost third of the alveolar zone.

Separate Linear Mixed Model (LMM) statistical tests were run on the distance values gathered at the midpoint of C1 and C2 for each of the two clusters pairs [kt]-[gd] and [tk]-[dg] with speaker as random factor. The LMM tests had the fixed variables 'sequence' (with levels [kt] and [gd] for one test, and [tk] and [dg] for the other test), 'C place' (with levels 'velar' and 'dental'), and 'zone' (with levels ALV, PAL, VEL and PHAR). Additional LMM tests, one for each cluster pair, were carried out on the C#C duration values for all tokens with 'sequence' and 'C place' as fixed variables. Least Significant Difference (LSD) post-hoc tests were performed on all main effects and significant interactions in order to find out whether numerical differences between pairs of levels of a given statistical variable reached significance or not. Given the large number of tests involved in the LMM analyses, the Benjamini-Hochberg (BH) correction procedure for adjusting the false discovery rate was applied to those variable comparisons which were of relevance to the present investigation. The significance level was set at $p < 0.05$.

3. RESULTS

Statistical results for the main effects and factor interactions obtained from the LMM tests run on the tongue distance data are given for the two pairs of clusters in Table I.

Results for the main effects of C place and zone are of little interest for our purposes: distances between the lingual splines and the origin of the ultrasound field of view were generally larger for velars than for dentals, and larger at the velar (VEL) and palatal (PAL) zones than at the pharyngeal (PHAR) and alveolar (ALV) zones. More relevant results are those involving the sequence factor. Thus, the significant sequence x zone interactions reported in the table happened to be associated with larger distances for [kt] than for [gd] and for [tk] than for [dg] at the pharyngeal zone (PHAR) and for [gd] vs [kt] at the palatal zone (PAL), and the significant sequence x C place interactions with greater distances for [k] than for [g] in the case of the [kt]-[gd] pair. In agreement with these statistical results, the lingual splines plotted in Figure 1 show a clear trend for most or all subjects to produce

(a) higher and more anteriorly the voiced than the voiceless stops whether they be velar or dental in the case of the [kt]-[gd] sequence pair (left graphs),

(b) somewhat more anteriorly the voiced than the voiceless stops for the [tk]-[dg] sequence pair (right graphs).

A trend for the tongue dorsum to occupy a higher position for [d] than for [t] in the cluster pair [tk]-[dg], which did not achieve significance, may also be observed. In sum, data for Catalan two-stop sequences reveal that, in comparison to voiceless stops, voiced stops may be produced not only with a more anterior postdorsum but also a higher, not lower tongue front dorsum configuration.

Table I: Significant main effects and factor interactions for the spline-to-origin distance values for the two-stop sequence pairs. NS: non-significant. * p <.05, ** p < .01, *** p <.001.

	kt/gd	tk/dg
sequence	NS	NS
C place	10.52(1,43)*	53.16(1,47)***
zone	9.53(3, 43)**	7.05(3,47)**
sequence x zone	48.57(3,43)***	3.95(3, 47)*
C place x zone	94.58(3,43)***	46.11(3,47)***
sequence x C place	6.96(1,43)*	4.14(1,47)*
sequence x C place x zone	NS	NS

Table II. Significant main effects and factor interactions for the segmental duration data for the two-stop sequence pairs. NS: non-significant. * p <.05, ** p < .01, *** p <.001.

	kt/gd	tk/dg
sequence	NS	19.69(1.100)*
C place	9.21(1,100)*	NS
sequence x C place	NS	NS

The extent to which differences in tongue position occur along the time domain may be seen in Figure 2. The figure represents the distance values between the four articulatory zones and the origin of the ultrasound field of view at all six temporal points starting at C1 onset (left edge of the lines) and ending at C2 offset (right edge) for the two cluster pairs across speakers. For both stop sequence pairs and mostly so for [kt]-[gd], distances are clearly larger for the voiceless cluster than for the voiced cluster at the pharynx (PHAR), and, judging from a comparison between the first and second halves of the lines, this difference occurs not only during the C2 voicing trigger but also during C1 which is supposed to undergo regressive voicing assimilation in Catalan. Figure 2 also shows the existence of a higher tongue dorsum position at the palatal zone (PAL) for the voiced vs voiceless cluster, mostly during C2 in the case of the [kt]-[gd] pair and during C1 for the [tk]-[dg] pair.

According to Table II, the statistical tests run on segmental duration yielded a main effect of sequence for the [tk]-[dg] pair which turned out to be associated with a longer voiceless vs voiced cluster (132.5 ms vs 104.2 ms across speakers). As to the [kt]-[gd] pair, a main C place effect was related to longer velars than dentals, which resulted mainly from durational differences between [t] and [d] (90.8 vs 65.8 ms).

4. DISCUSSION

Ultrasound data for Catalan stop-stop sequences reported in this study reveal the presence of a more anterior tongue body position for voiced vs voiceless stops in velar + dental and dental + velar stop sequences. This finding agrees with data for English reported in the literature exhibiting a greater pharyngeal cavity expansion for voiced vs voiceless stops [1, 3, 5]. Moreover, the ultrasound data show that this C2-dependent difference in back lingual configuration may be transmitted to the preceding coda stop which is supposed to undergo regressive voicing assimilation in Catalan. There is no agreement regarding oral cavity volume: in contrast with English according to some reports [3, 5], in the Catalan clusters subject to analysis a more anterior tongue body position for voiced vs voiceless stops did not cooccur with a lower tongue predorsum position due presumably to differences in the way aspirated and unaspirated stops are produced and in line with the associated differences in prominence of the stop release. Thus, it could be hypothesized that a reduction in oral cavity size is not required to occur for unaspirated voiceless stops due to a lesser intraoral pressure buildup than for the aspirated stop cognates. These spline-to-origin distance data turned out to be consistent with the segmental duration values in that the stop sequences were generally longer when voiceless than voiced. The validity of these findings should be ascertained with a larger sample size in future studies.

5. REFERENCES

[1] Bell-Berti F. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *J. Acoust. Soc. Am.* 57, 456-461.

[2] Ohala, J.J. 1983. The origin of sound patterns in vocal tract constraints. In: McNeilage, P. (ed.), *The Production of Speech.* New York: Springer-Verlag, 189-216.

[3] Perkell, J.S. 1969. *Physiology of Speech Productio*n, Cambridge, MA: The MIT Press.

[4] Recasens, D. & Rodríguez, C. 2018. Contextual and

syllabic effects in heterosyllabic consonant sequences. An ultrasound study. *Speech Commun* 96, 150-167.
[5] Westbury J. R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.* 73, 1322-1336.

Figure 1: Lingual configurations at the C1 and C2 midpoints for the sequence pairs [kt]-[gd] (left) and [tk]-[dg] (right) according to five Catalan speakers. The front of the mouth is on the right of the graphs.

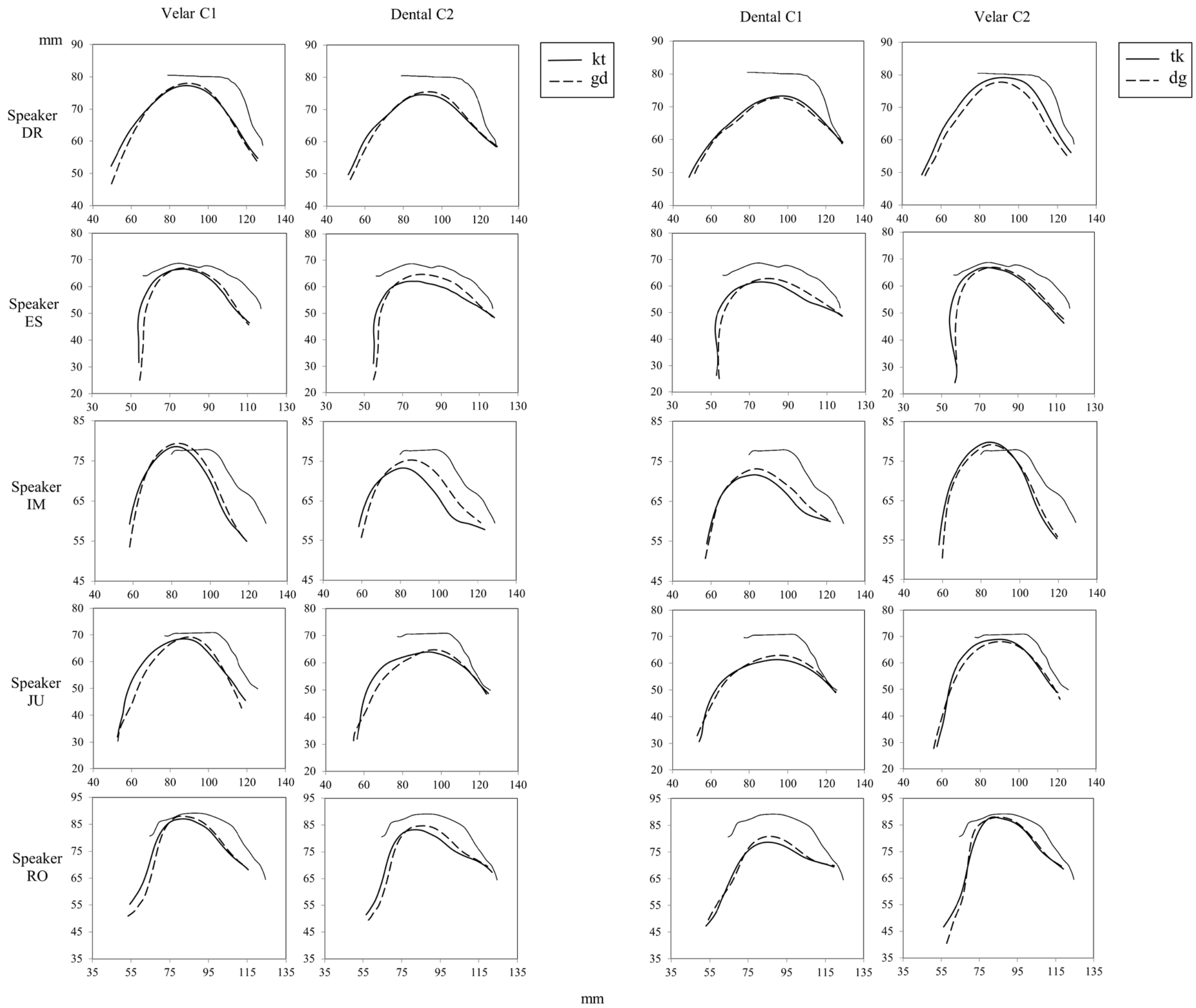

Figure 2: Cross-speaker spline-to-origin distances over time for [kt]-[gd] (left) and [tk]-[dg] (right) at each articulatory zone. The distance trajectories proceed through the time points 1 (C1 on), 2 (C1 mp), 3 (C1 off), 4 (C2 on), 5 (C2 mp), 6 (C2 off).