

AN EPG+UTI STUDY OF ITALIAN /r/

Lorenzo Spreafico^a, Chiara Celata^b, Alessandro Vietti^a, Chiara Bertini^b, Irene Ricci^b

^aALPs – Alpine Laboratory of Phonetics and Phonology – Free University of Bozen-Bolzano, Italy

^bLaboratorio di Linguistica 'Giovanni Nencioni' – Scuola Normale Superiore, Pisa, Italy
{lorenzo.spreafico, alessandro.vietti}@unibz.it; {chiara.celata, c.bertini, i.ricci}@sns.it

ABSTRACT

This paper describes a system for the acquisition, real-time synchronization and analysis of acoustic, electropalatographic (EPG) and ultrasonographic (UTI) data.

Simultaneous data on linguo-palatal contact and tongue sagittal profiles are captured for rhotic consonants produced by a native speaker of Italian. Three anterior variants of /r/ ([r], [ɹ] and [ɹ̥]) are shown to be realized with an apical tongue gesture, but different vowel-related coarticulation patterns. The paper discusses the implication of the proposed analysis for a coherent investigation of lingual and linguo-palatal dynamics.

Keywords: rhotics, electropalatography, ultrasound tongue imaging, multichannel database, Italian.

1. INTRODUCTION

Speech articulatory complexity poses challenging questions to both phonetic-phonological theories and speech technology research. The wide range of applications of articulatory studies explains the growing interest in the collection of multi-level corpora of speech data integrating data from diverse instrumental techniques [7,8,14,15,16].

Within the framework of multi-level articulatory approaches to speech variation, this paper describes an original system for the acquisition, real-time synchronization and analysis of acoustic, electropalatographic (EPG) and ultrasonographic (UTI) data.

The aim of the paper is twofold. The first aim is to introduce the implementation procedure and the obtained system for high-speed synchronous audio-EPG-UTI recording. The system is used to capture simultaneous data on linguo-palatal contact and tongue sagittal profiles of rhotic consonants produced by native speakers of a Tuscan variety of Italian. EPG and UTI instrumental techniques are to be seen as complementing each other inasmuch as the former observes linguo-palatal dynamics from the perspective of the realized contacts - with particular precision in the anterior vocal tract - while UTI focuses on linguo-palatal distances and

movements of the whole tongue profile, including tongue root and postdorsum.

The second goal of the paper is that of providing a multi-level articulatory definition for alveolar rhotics according to well-established linguo-palatal contact indexes and common procedures for lingual shape analysis. Innovative data reduction techniques allowing a coherent investigation of lingual and palatal dynamics are also discussed. Rhotic sounds are known for their exceptional degree of syntagmatic, cross-linguistic, and even cross-individual variability [5,13]; an integrated perspective on the dynamics of the vocal tract is expected to provide cues for a deeper understanding of their phonetic properties and systemic behaviour.

2. MULTICHANNEL ACOUSTIC-ARTICULATORY SYSTEM

2.1. Experiment design

The speech corpus has been collected through a multi-repetition sentence reading task performed by 2 adult female speakers of a Western Tuscan variety of Italian (Pisa and Livorno districts). 62 sentences of equal length (8 phonetic syllables) and consistent intonation structure were selected from the list in [3]. The /r/ appeared either in intervocalic position, or in a biphonemic consonant cluster. When embedded in clusters, the rhotics could be preceded or followed by obstruents of various places of articulation (bilabial, alveolar and velar) and laryngeal status (voiced and unvoiced). The vocalic context could also vary, with /i/, /a/ and /u/ symmetrically distributed before and after the consonantal interval. A total of 558 stimuli was collected for the two speakers. The present analysis is based on a subset of 237 stimuli produced by one of the speakers.

2.2. Data acquisition and annotation procedure

The data acquisition system was able to synchronously record EPG and UTI data at sufficiently high frame rates to correctly characterize the movements of the tongue, as well as the acoustic speech signals. The acquisition platform was developed using Articulate Instruments hardware

and the Articulate Assistant Advanced (AAA) software tool [17] in the ALPs lab in Bozen.

Electropalatographic data was captured via the WinEPG (SPI 1.0) unit recording EPG at 100Hz. Ultrasound data was captured via the Ultrasonix device used in conjunction with a head-mounted micro-convex probe, with depth set at 80mm and angle at 127°, capturing a mid-sagittal tongue image at a rate of 91Hz.

In order to synchronize EPG with UTI, the microphone signal was passed through the WinEPG unit. That way, the WinEPG added a synchronization tone to the microphone signal and that was used by AAA to synchronize the EPG with synchronization pulses from the SonixTablet.

The audio files were exported into Praat (5.3.62) for an acoustic and auditory analysis of /t/ realizations.

Each /t/ token was first categorized according to manner of articulation (i.e. trill, tap, approximant, fricative). The constriction interval was annotated as voiced or unvoiced according to spectrographic evidence.

The following analysis focuses on the three alveolar rhotic variants, namely, the tap [ɾ], the fricative [ʀ] and the approximant [ɹ] variant. While the tap was the most frequent realization in the dataset, approximants and fricatives were also present across phonetic contexts, and no specific distributional bias could be detected for any of the variants.

3. EPG+UTI ARTICULATORY ANALYSIS

The analysis focused first on the articulatory correlates of the three rhotic variants, with the aim of verifying to what extent the acoustic-auditory classification could be corroborated by coherent sets of articulatory characteristics. Within-category variation was also analysed by investigating the influence of the phonetic context on the realization of variants. In particular, the present contribution focuses on vowel-induced coarticulation.

The articulatory measurements that are described and discussed in the following sections were taken at the midpoint of all constriction intervals.

3.1. EPG analysis

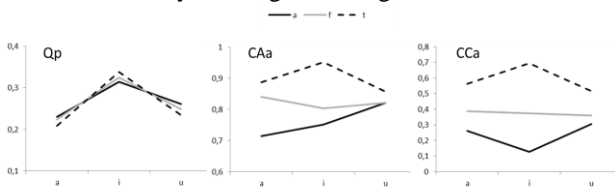
A traditional analysis based on the contact index method ([10,11]) was performed on the anterior (rows 1-4) and posterior (rows 5-8) parts of the palate. The Qp index (percentage of contact in the posterior palate) was extracted in order to determine the amount of activated electrodes in the velar-mediopalatal region. In addition, two indexes were extracted for the anterior palate. CAa (contact

anteriority in the anterior palate) served as a measure of the anteriority of contact along the sagittal dimension. CCa (contact centrality in the anterior palate) was calculated to determine the coronal extension of linguo-palatal contact; in fact, the activated electrodes in the four central columns of the palate receive a higher score than the activated electrodes on the four (two on the right and two on the left) peripheral columns.

We hypothesized that the linear combination of the three dependent variables Qp, CAa and CCa may account for the categorization of the tokens into taps, fricatives and approximants as well as for differences in their coarticulatory behaviour as a function of variations in the vocalic context. The interaction between the two factors ‘variant’ and ‘vowel place’ in multivariate tests confirmed the hypothesis (e.g. Pillai’s trace $F=2.311$, $df=12$, $p<.001$), thus indicating that the vocalic context influenced the realization of the three variants to a different extent. Univariate tests were run to ascertain the contribution of the individual dependent factors in the interaction and they showed that the two indexes concerning the anterior contact patterns did contribute significantly (CAa: $F=3.912$, $df=4$, $p<.01$; CCa: $F=3.629$, $df=4$, $p<.01$) while Qp gave non-significant results (Fig. 1). Between-subject tests and Bonferroni multiple comparisons showed in particular that, in the posterior palate, there was a significant effect of vowel place ($F=61.710$, $df=2$, $p=.000$), with highest contact values for the /i/ > /u/ > /a/ contexts, but a non-significant effect of variant, thus indicating a generalized increase of contact in the medio-palatal/velar region caused by coarticulation with the high front vowel. In the anterior palate, variant produced significant effects (CAa: $F=23.375$, $df=2$, $p=.000$; CCa: $F=40.883$, $df=2$, $p=.000$) while vowel place did not. Post-hoc Bonferroni tests revealed in particular that the three variants were consistently different for contact extension (tap > fricative > approximant) and less so for contact anteriority (tap > fricative, approximant) since in the /u/ context the approximant and the fricative were not significantly different for contact anteriority. Only two constriction locations for the rhotics were thus detectable when the adjacent vowels required a raised and retracted tongue dorsum. The tap consistently showed the anterior and most extended contact pattern across vocalic context and turned out to be the least resistant to V-induced coarticulation.

Figure 1: Average Qp, CAa and CCa values at constriction midpoint for the three rhotic variants as a function of vocalic context. Grey line: fricatives; black continuous: approximants; black

dashed: taps. High resolution images are available on the web by clicking on the figures.



The EPG analysis thus revealed that the three rhotic variants were all characterized by an anterior articulation with no active post-dorsum involvement, except for an increase in posterior contact due to the coarticulation with high back vowels. The linguo-palatal contact investigation also revealed that the most robust difference among variants referred to contact degree in the anterior palate, with taps consistently realizing more contact and an anterior configuration than both approximants and fricatives.

3.2. UTI analysis

Using the algorithm implemented in the AAA software [2], cubic splines were fitted to sagittal tongue curves at each mid-constriction for /r/. Fitted splines were exported to a workspace to calculate an average tongue contour for each variant, based on means at each of 42 fan lines (Fig. 2).

Figure 2: Means for the tongue shape during the constriction phase for /r/. Green line: palate; blue lines tap and fricative; red line: approximant. Teeth on the right, pharynx on the left.

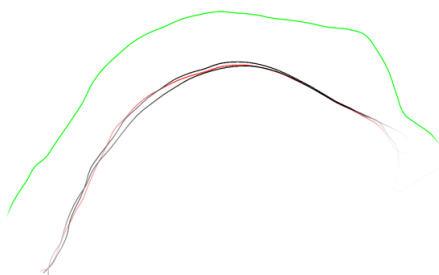
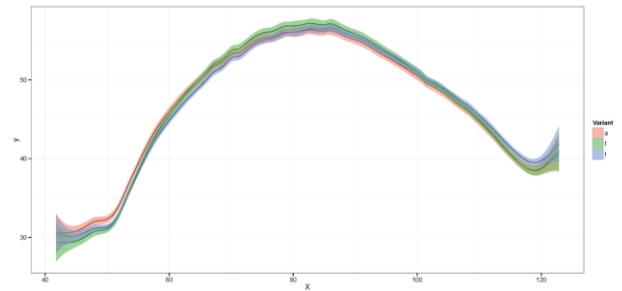


Figure 2 shows palate and mean splines for the alveolar tap, the fricative and the approximant. Overall, the mean profiles describe a mid-bunched contour, where the tongue tip forms the primary constriction; the front of the tongue is lowered while the dorsum is raised towards the hard palate. The visual comparison of the contours for each variant showed that the tongue post-dorsum and root were lower for [r] than for [ɹ] and [ɻ].

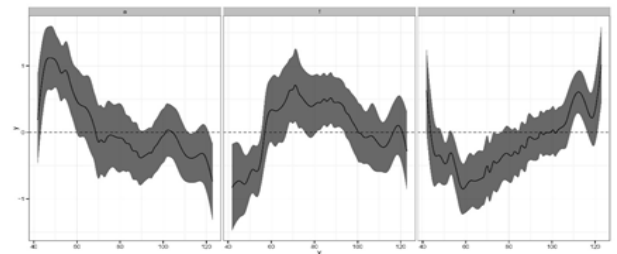
SS ANOVA quantitative analysis [4] showed differences of greater magnitude (Fig. 3).

Figure 3: Smoothed splines for the /r/ variants.



Though being very close to one another, the curves appeared to diverge for the utmost posterior and anterior parts of the tongue. Bayesian confidence intervals of the interaction effects were calculated to determine whether and where the curves were significantly different (Fig. 4).

Figure 4: Interaction effects with Bayesian confidence intervals for [ɹ], [ɻ] and [r].



As regards approximants, the smoothed spline was not significantly different from that of the spline that best fits all data (= Bayesian intervals encompassed zero) for about two-thirds of the total length of the tongue. The most significant difference corresponded to the posterior part of the tongue, thus suggesting slight root retraction in the articulation of approximants. As for fricatives, the significant difference corresponded again to the post-dorsum, displaying tongue lowering. As for taps, the difference mostly involved the anterior part of the tongue, presenting a tip-raising gesture.

To further investigate such apparent dissimilarity of the underlying splines, we extracted smoothed splines for the tongue shape during the closure phase of /r/ in the /a/, /i/ and /u/ environment respectively (Fig. 5, 6). Data clearly confirm V-induced coarticulation found in the EPG analysis.

Figure 5: Smoothed splines for /r/ when surrounded by [a] (red), [i] (green) and [u] (blue).

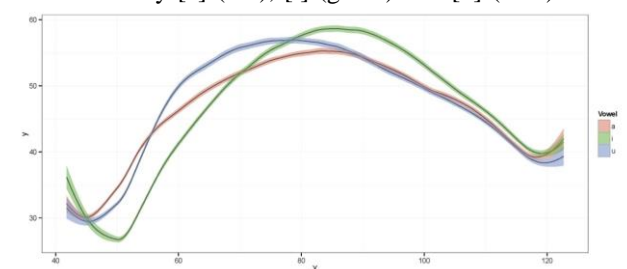
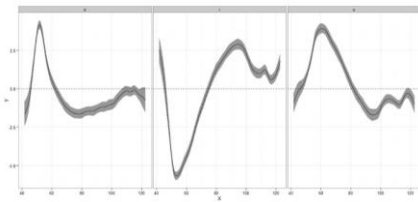


Figure 6: Interaction effects with Bayesian confidence intervals for [a], [i] and [u].



An original scalar measure was finally elaborated and the areas enclosed by the tongue spline, the roof spline, and the front line 3 and the front line 5 (FrontalArea1); or the front line 5 and the front line 7 (FrontalArea2); or the front line 7 and the dorsal line 3 (DorsoFrontalArea) in the AAA workspace were calculated (Fig. 7).

Figure 7: Red box: FrontalArea1; blue box: FrontalArea2; yellow box: DorsoFrontalArea.

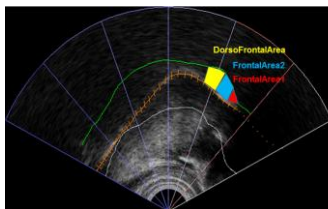
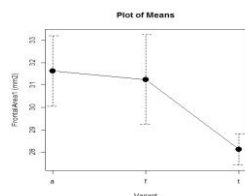


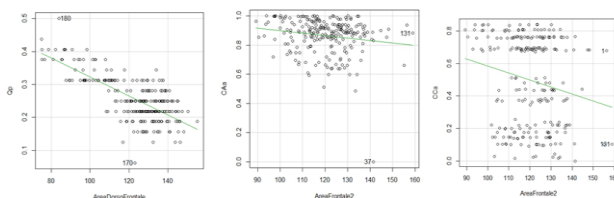
Figure 8 shows that the mean measure for the FrontalArea1 increases from the tap, to the fricative, to the approximant, thus mirroring EPG data on constriction degree shown in §3.1.

Figure 8: Plot of mean FrontalArea1 for [ɹ], [ɹ̥] and [r].



As a last step, in order to establish correlations between the EPG indexes and the UTI area measures, a stepwise linear regression was computed. Results (Fig. 9) showed a significant but rather mild interaction between Qp and DorsoFrontalArea, as well as an even slighter interaction between FrontalArea2 and both CAa and CCa.

Figure 9: Correlations among EPG indexes and UTI scalar measures of areas.



4. DISCUSSION

This study has shown the implementation and analytic reliability of a multi-level acoustic-articulatory system allowing real-time alignment of acoustic, UTI and EPG data on lingual movements during speech articulation.

The system proved useful to characterize the apical rhotics of a selected Tuscan Italian variety by showing that the observed articulatory variability is to be referred to both linguo-palatal contact patterns and tongue body and root synergistic behaviour with respect to vowel-dependent articulatory movements.

In particular, the study has shown that [ɹ], [ɹ̥] and [ɹ̥] are realized by means of substantially similar lingual gestures, i.e. involving apico-predorsal coupling, predorsum raising but no tongue-body retraction, as typically reported for apico-alveolar non-trilling rhotics ([9,11]). However, the three variants differ for constriction degree and coarticulatory resistance to adjacent vowel, with taps showing the most constrained apical configuration and the most /i/-like tongue dorsum configuration.

Position in the syllable and consonantal context may account for other production events that are left to future investigation.

Additional implications of the current study refer to the theoretical and methodological challenges of combining electropalatographic and ultrasonographic information in the description of the articulatory events. The adoption of innovative scalar measures for UTI investigation such as the area measures will allow establishing direct correlations between the palatal contact indices and the distance measures calculated for selected portions of the lingual profile. The prosecution of this study will thus highlight the existing relationships between the palatographic / sonographic representations of lingual movements and the resulting acoustic events (formant structures) during the production of contextually different rhotic realizations.

5. ACKNOWLEDGEMENTS

Thanks to Alan Wrench for his helpfulness. Financial support from: Provincia Autonoma di Bolzano – Alto Adige, Ripartizione allo studio, Università e ricerca scientifica 2013-16 “The articulatory sociophonetics of bilinguals in South-Tyrol: The Ultrasound Tongue Imaging potential”; Scuola Normale Superiore, Laboratorio di Linguistica and project 367-GR13Celata “Modeling speech variation in the socio-communicative context” 2013-2015.

6. REFERENCES

- [1] Articulate Instruments Ltd 2010. *WinEPG Installation and User's Manual: Revision 1.18*. Edinburgh, UK: Articulate Instruments Ltd.
- [2] Articulate Instruments Ltd 2014. *Articulate Assistant Advanced User Guide: Version 2.15*. Edinburgh, UK: Articulate Instruments Ltd.
- [3] Celata, C., Bertini, C., Ricci, I. 2014. Proprietà acustiche e articolatorie di /r/ nella Toscana occidentale. *X Convegno Nazionale AISV Torino*, 22-24 January, 2014.
- [4] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.* 120/1, 407-415.
- [5] Docherty, G., Foulkes, P. 2001. Variability in /r/ production: Instrumental perspectives. In: Van de Velde, H., van Hout, R. (eds) *r-atics: sociolinguistic, phonetic and phonological characteristics of /r/*. Bruxelles: ILVP/ULB, 173-184.
- [6] Lawson, E., Scobbie, J., Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *J. Sociolinguistics* 15/2, 256-268.
- [7] Meister, E., Meister, L. 2012. Multimodal Corpus of Speech Production: Work in Progress. In: Tavast, A. et al. (eds.), *Human Language Technologies. The Baltic Perspective*. Amsterdam: IOS Press, 146-153.
- [8] Narayanan, S. et al. 2014. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research, *J. Acoust. Soc. Am.* 136/3, 1307-1311.
- [9] Nicolaidis, K., Baltazani, M. 2011. An electropalatographic and acoustic study of the Greek rhotic in /Cr/ clusters. *Proc. 17th ICPhS Hong-Kong*, 1474-1478.
- [10] Recasens, D., Pallarès, M.D. 1999. A study of /r/ and /r/ in the light of the "DAC" coarticulation model. *Journal of Phonetics* 19, 267-280.
- [11] Recasens, D., Espinosa, A. 2007. Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica* 64, 1-28.
- [12] Schabus, D. 2014. The MMASCS multi-modal annotated synchronous corpus of audio, video, facial motion and tongue motion data of normal, fast and slow speech. *Proc. 9th LREC Reykjavik* 3411-3416.
- [13] Spreafico, L., Vietti, A. 2013. On rhotics in a bilingual community: A preliminary UTI research. In: Spreafico, L., Vietti, A. (eds), *Rhotics. New data and perspectives*. Bolzano: BU Press, 57-79.
- [14] Steiner, I., Richmond, K., Marshall, I., Gray, C. 2012. The magnetic resonance imaging subset of the mngu0 articulatory corpus. *J. Acoust. Soc. Am.* 131/2, 106-111.
- [15] Steiner, I., Knopp, P., Musche, P., Schmiedel, A., Braun, A., Ouni, S. 2014. Investigating the effects of posture and noise on speech production. *Proc. 10th ISSP Cologne*, 413-415.
- [16] Wrench, A.A. 1999. *The MOCHA-TIMIT articulatory database*. <http://www.cstr.ed.ac.uk/research/projects/artic/mocha.htm>
- [17] Wrench, A.A., Scobbie, J.M. 2008. High-speed cineloop ultrasound vs. video ultrasound tongue imaging: comparison of front and back lingual gesture location and relative timing. *Proc. 8th ISSP Strasbourg*, 57-60.