## Automatic Analysis of Geminate Consonant Articulation using Real-time Magnetic Resonance Imaging

Using real time magnetic resonance imaging (rtMRI) to study speech production offers advantages over other methods of articulometry. It allows the entire vocal tract to be viewed at once, and provides dynamic information about all components of the vocal tract, rather than specific flesh points. While methods exist for automatic segmentation of tissue boundaries in MRI data (Bresch & Narayanan 2009), such approaches are computationally intensive, and the examination of articulatory phenomena in segmented MRI data requires additional processing techniques. In this paper, we present a method of direct image analysis of MRI data especially applicable to the study of consonant articulation, and assess its utility in the analysis of Italian geminate production.

Although it has been established that Italian geminate and singleton stops differ in their temporal (Bertinetto 1981) and kinematic properties (Smith 1995; Esposito & Di Benedetto 1999), it is unclear whether these differences result from intrinsic durational differences, or differences in constriction target (Löfqvist, 2005). A proper understanding of the basis of the singleton-geminate contrast requires better knowledge of the coarticulatory relationship between consonants and their surrounding segments (Gili Fivela et al. 2007), yet although Italian consonant production has been studied extensively using palatography (Payne 2006) and other techniques, no current method allows examination of global articulatory dynamics at the level of detail provided by rtMRI.

A native speaker of Standard Italian was imaged while producing lexical items contrasting singleton and geminate stops, using a custom MR Imaging protocol (Narayanan et al. 2004; Bresch et al. 2006). The subject repeated phrases containing one member of a minimal (or near-minimal) pair, e.g. [paga]-[pagga]. Carrier phrases were designed to minimize co-articulation effects on the consonants of interest. Given the well established temporal differences between these singleton-geminate pairs, these tokens provide benchmarks with which to determine whether rtMRI is able to capture these durational (and, perhaps spatial) differences.

A method was developed to estimate articulator movement and constriction duration directly from pixel intensity variation in the MR Image sequences, exploiting the fact that in a given image frame, high pixel intensity typically corresponds to concentrations of soft tissue, while low pixel intensity corresponds to regions of less tissue density, or air (Lammert et al. 2010).

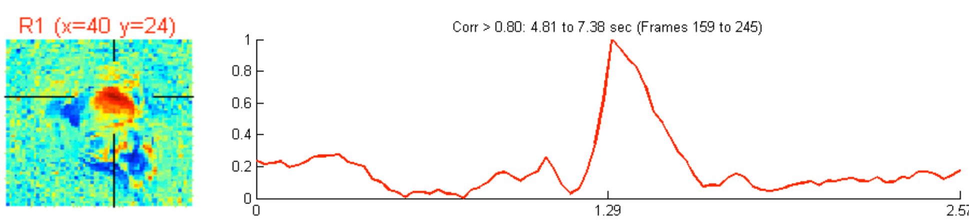


Figure 1: **Estimating dorsal articulation from correlated pixel intensity. Left**: correlation of all pixels in a single image frame with pixel of interest, located on tongue dorsum (red: highly correlated; blue: low correlation); **Right**: change in mean intensity over time of all pixels in correlated cohort; function peak corresponds to mid-consonantal production of dorsal stop [k].

For each consonant under consideration, a single pixel was selected in a region of interest in the vocal tract (interlabial, alveolar, or velar) at target achievement. The mean of this pixel's intensity and the intensities of all pixels whose intensity values were correlated over the frames of interest (the target consonant and the context vowels) was calculated, generating a time function (Fig. 1). Pixels pertaining to the area over which a single constriction organ moves during an utterance will exhibit similar intensity values and thus meet the 'cohort' threshold of 80% intensity correlation. Intensity variation in the correlated pixel cohort is observed as linguistically significant vocal tract constrictions are made in the area of the selected pixel. The constriction duration using the time function generated is calculated as the point in time at which 25% maximum intensity is initially reached to the point in time at which intensity decreases to 25% maximum intensity.

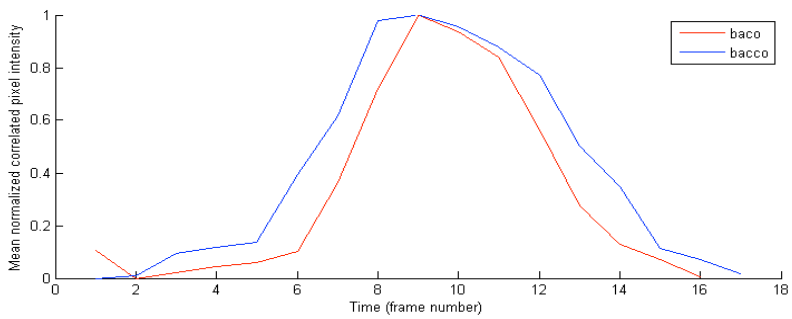


Figure 2: **Time functions of mean correlated pixel intensities** in velar region of vocal tract during production of [baco] and [bacco].

Constriction durations for geminates estimated using this method were found to be longer than those for singletons in all regions of the vocal tract: labial geminates were found to be 18.4% longer; coronals 49.9% longer, and dorsals 22.9% than their singleton counterparts.

It has been observed (Esposito & Di Benedetto 1999) that the acoustic closure duration of geminate stops in Italian is roughly twice as long as that of lexically contrasting singleton stops. The acoustic data collected during this study is consistent with this finding: labial geminates were found to be 77% longer; coronals 107% longer, and dorsals 128% longer than their singleton counterparts.

The results of this study indicate that rtMRI is, indeed, able to capture differences in constriction duration and that these differences may be quantified by tracking pixel intensity variation over time, as a constriction is formed and released. Given that it is now possible to estimate the temporal properties of vocal tract constrictions, it is our aim to develop a method of quantifying differences in the degree of constriction (compression) in singletons and geminates (as noted in Payne 2006) by measuring maximum pixel intensity in a specified region corresponding to the area of constriction. The ability to reliably quantify both the temporal and spatial properties of constriction gestures will ultimately allow for a more complete understanding of the articulatory dynamics of gestures of all types.

(Research supported in part by NIH grants DC 008780 and DC 007124.)

**References:**


Bertinetto, P.M. (1981). Strutture prosodiche dell'italiano. Firenze: Accademia della Crusca.

Bresch E. & Narayanan S. (2009). *Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images*. IEEE Trans. Med. Imag., 28(3): 323-338

Bresch E., Nielsen J., Nayak K. & Narayanan S. (2006). *Synchronized and noise-robust audio recordings during real-time MR Imaging scans*. JASA 120(4): 1791-1794

Esposito, A. & Di Benedetto M.G. (1999). *Acoustical and perceptual study of gemination in italian stops*. JASA, 106(4): 2051-2062

Gili Fivela B., Zmarich C., Perrier P., Savariaux C., Tisato G. (2007). *Acoustic and kinematic correlates of phonological length contrast in Italian consonants*. Proc. International Conference of Phonetic Sciences, Saarbrücken, Germany: 469-472

Lammert, A., Proctor, M., and Narayanan, S. (2010). *Data-Driven Analysis of Realtime Vocal Tract MRI using Correlated Image Regions*. Proc. Interspeech 2010, Makuhari Messe, Japan, 26-30 Sept.

Löfqvist, A. (2005). *Lip kinematics in long and short stop and fricative consonants*. JASA, 117(2): 858-878

Narayanan S., Nayak, K., Lee S., Sethy A. & Byrd D. (2004). *An approach to real-time MRI for speech production*. JASA 115(4): 1771-1776

Payne, E. (2006) *Non-durational indices of Italian geminates: an EPG study*. Journal of the International Phonetic Association, 36(1): 83-95

Smith, C.L. (1995). *Prosodic patterns in the coordination of vowel and consonant gestures*. In: B.Connell & A.Arvaniti (eds), Papers in Laboratory Phonology IV, Phonology and Phonetic evidence. Cambridge, CUP: 205-222