Vocal Tract Shaping and Dark Tone Quality - An Investigation Using Real-Time MRI

Elisabeth Lynn[a], Shrikanth S. Narayanan[b] and Adam C. Lammert[a]

[a]Worcester Polytechnic Institute, [b]University of Southern California

Introduction: Dating at least to the work of Manuel Garcia[2], vocal pedagogy has identified a tone quality termed "dark" as desirable for human voice performance. Darkness is often associated with the additional terms "warm", "round" and "covered", and is a key component of the coveted "chiaroscuro", or "bright-dark", tone quality, which is an aesthetically important property of Western classical voice performance[6]. Darkness is associated acoustically with lowered formant frequencies[8], and may stem from specific vocal tract shaping, including a lowered larynx and widening of the pharynx[3]. Evidence further suggests that vocal tract shaping counter to that associated with darkness (i.e., larynx raising, pharynx widening) results in bright tone quality commonly described as "twang"[9], which is possibly also accompanied by a wider lip opening and a constricted oral cavity[7].

Despite what is known about the origins of dark tone quality in vocal tract shaping, technological limitations have made it challenging, until recently, to observe the production of dark tone quality dynamically, during natural voice performance, and with a relatively complete view of the vocal tract. Several open questions remain. First, can it be confirmed that aspects of vocal tract shaping noted by prior studies as being associated with dark tone quality are present in a dynamic, more ecologically valid production context? Second, are there other additional aspects of vocal tract shaping, currently unknown, that contribute to the production of dark tone quality? The present study aims to address these questions using real-time magnetic resonance imaging (rtMRI) quantify and analyze vocal tract shaping during voice performance of several trained sopranos displaying a range of tone qualities.

Method: The subjects for the present study were four (4) female native speakers of American English (subject identifiers: H5, S2, M1 and L4). All subjects were sopranos trained in the Western classical tradition. Subjects sang portions of "Susanna's aria" from Mozart's
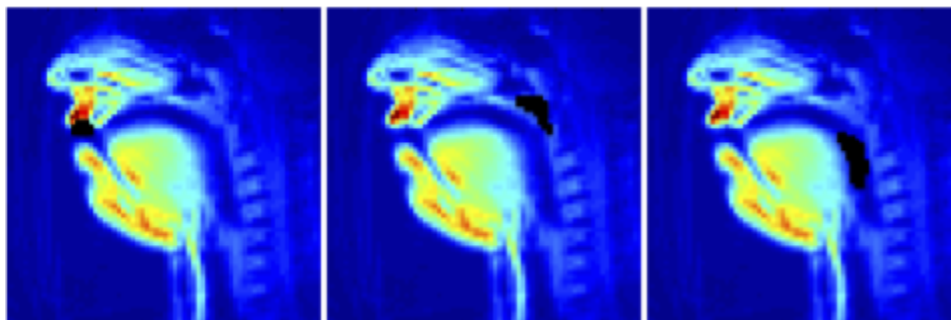

Figure 1: *M1's mean images with ROIs for labial narrowing, velum raising and pharynx narrowing superimposed.*

Marriage of Figaro. Subjects were instructed to sing the specified portion in their **normal** performance style (in instructions identified as "ringing"), especially **dark** (identified as "covered") and especially **non-dark** (identified as "bright") tone qualities, with separate acquisitions made of each of the three qualities. During data collection the subjects were in a supine position. The rtMRI data were collected on a GE Signa 1.5T scanner, using a 13-interleaf spiral gradient echo pulse sequence with a repetition time of 6.5 ms[5]. Images were reconstructed using a sliding-window procedure, creating an image sequence with an effective rate of 22.41 frames per second. Spatial resolution of the reconstructed images was 68×68 pixels over a field of view of 20×20 cm. Audio was simultaneously recorded at 20 kHz using an MRI-safe optical microphone, and subsequently denoised[1]. All image sequences were cropped to the aria portion corresponding to the lyrics "Deh vieni, non tardar, o gioja bella/ Vieni ove amore per goder t'appella". Cropping was performed by listening, and manual identification of frames corresponding to the first and last frames during which the voice was audible.

Region of Interest (ROI) analysis[4] was used to extract articulator-specific kinematic time series associated with (1) labial narrowing, (2) velum raising and (3) pharyngeal narrowing during vocal performance using three corresponding, automatically-determined regions (Fig. 1). Within-subject,

articulator-specific differences associated with dark and non-dark tone quality types were tested by performing two-sample t-tests on corresponding time series values. Across-subject, articulator-specific differences associated with tone quality in terms of average vocal tract shape were tested using paired t-tests on time series means. Means for the dark and non-dark conditions were first normalized to the percent difference from the "normal" condition.

Results: All sopranos exhibited within-subjects significant differences in mean labial narrowing in the dark and non-dark conditions ($p \ll 0.001$). The majority of sopranos exhibited such within-subjects differences in mean velum raising (H5, S2, L4), and half of sopranos exhibited such differences in mean pharynx narrowing (M1, L4). Across subjects (see Table 1), there was a significant difference in mean labial narrowing in the dark condition (M = 0.56, SD = 0.42) versus the non-dark condition (M = -0.26, SD = 0.08) such that the lips were narrower during the former (t(3) = 4.32, p = 0.023). There was no such significant difference across subjects in mean velum raising (dark: M = 0.01, SD = 0.07; non-dark: M = 0.04, SD = 0.03; t(3) = -0.96, p = 0.41), nor was there such a significant differences in mean pharynx narrowing (dark: M = 0.00, SD = 0.13; non-dark: M = -0.03, SD = 0.12; t(3) = 1.46, p = 0.24).

Discussion: When looking at individual subjects, clear differences in vocal tract shaping can be observed when comparing dark and non-dark tone qualities in terms of the average position of all three articulators considered - labial narrowing, velum raising and pharynx narrowing. Moreover, the direction and magnitude of the differences is consistent across subjects when considering labial narrowing, such that labial aperture is narrower during the production of dark tone quality. Differences across subjects are not consistent with regard to velum raising and pharynx narrowing, however. These results shed new light on the production of dark tone quality, in that they are somewhat conflicting with traditional accounts, which emphasize the importance of the pharynx. Although the pharynx does seem to play a role, that role is variable across individuals, perhaps reflecting individual differences in training. The present results suggest that labial aperture may be the primary and most consistent driver of dark tone quality, which may also be related to earlier suggestions that labial widening is important for the production of "twang"[7].

Table 1: *Mean values for labial narrowing, velum raising and pharynx narrowing by tone quality and subject.*

Subject	Labial Narrowing			Velum Raising			Pharynx Narrowing		
	Dark	Normal	Bright	Dark	Normal	Bright	Dark	Normal	Bright
H5	73.6	38.9	30.2	52.6	57.4	57.3	39.9	48.8	38.6
S2	64.4	34.1	28.1	67.4	66.5	71.2	57.8	56.6	59.5
M1	63.0	42.8	26.9	59.0	56.7	59.2	53.1	51.0	49.3
L4	34.9	34.9	25.1	57.9	54.2	55.8	53.4	47.4	49.9

1. Bresch, E., Nielsen, J., Nayak, K., & Narayanan, S. (2006). Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *The Journal of the Acoustical Society of America*, *120*(4), 1791-1794.
2. Garcia, M. (1840). *traité complet de l'art du chant*. Mayence, Paris: Schott.
3. Hertegård, S., Gauffin, J., & Sundberg, J. (1990). Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. *Journal of Voice*, *4*(3), 220-230.
4. Lammert, A. C., Proctor, M. I., & Narayanan, S. S. (2010). Data-driven analysis of realtime vocal tract MRI using correlated image regions. In *Eleventh Annual Conference of the International Speech Communication Association*.
5. Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America*, *115*(4), 1771-1776.
6. Stark, J. (1999). *Bel canto: a history of vocal pedagogy*. University of Toronto Press.
7. Story, B. H., Titze, I. R., & Hoffman, E. A. (2001). The relationship of vocal tract shape to three voice qualities. *The Journal of the Acoustical Society of America*, *109*(4), 1651-1667.
8. Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica et Logopaedica*, *22*(1), 28-48.
9. Titze, I. R., Bergan, C. C., Hunter, E. J., & Story, B. (2003). Source and filter adjustments affecting the perception of the vocal qualities twang and yawn. *Logopedics Phoniatrics Vocology*, *28*(4), 147-155.