Time to target attainment in singing and speech

Reed Blaylock, University of Southern California, reed.blaylock@gmail.com

Introduction While most of the speech production literature focuses on conversational speech (or the closest approximation of laboratory speech that can be recorded in a laboratory setting), a large part of human behavior is based on artistic speech like singing. Singing uses the same fundamental sounds as conversational speech (which is why singing is usually intelligible); however, singing notably differs from speech in pitch, amplitude, and duration to satisfy melodic and rhythmic constraints.

For example, Ramanarayanan et al. (2011) showed that sung articulations do not participate in speechlike phrase-boundary strengthening or lengthening effects. This could be because phrase-boundary effects are restricted to laboratory speech or because the articulator movements were constrained by musical meter. Under either interpretation, the implication is that different types of speech have different patterns of articulator movement. Furthermore, those differences likely stem from larger goals of the behavior—that is, because is more isochronous than conversational speech, articulator duration may be controlled more heavily and may therefore be less malleable. To understand the cognition that underlies speech production, it is important to consider how speech sounds manifest in a variety of speech-based behaviors, not just laboratory speech.

This project compares sung and spoken articulation by examining whether the tongue body takes the same amount of time to reach vowel constriction targets in singing as it does in speech. On the one hand, singers may attempt to reach vowel targets as quickly as possible to rapidly achieve an aesthetic vowel quality goal; if so, the time it takes for the tongue to travel to its target should be as brief as in speech, or perhaps even shorter. On the other hand, the tongue might move more sluggishly because sung vowel durations can last for thousands of milliseconds without interference from other gestures, leaving the tongue with ample time to reach its target.

Method Singing and speech video data were collected from a trained female soprano at the University of Southern California (USC) by USC's Speech Production and Articulation kNowledge group (SPAN) (Bresch & Narayanan 2010, Ramanarayanan et al. 2011). The videos were acquired using real-time MRI of the midsagittal plane with a reconstructed frame-rate of 22.41 frames per second. Using a region of interest analysis, measurements of tongue body movement time series were taken for each vowel in two sung English passages and three read-aloud TIMIT sentences. Articulatory landmarks of tongue body movement were manually annotated in MviewRT (Tiede 2010). The annotated targets included: 1) the beginning of the tongue movement towards its vowel constriction target, and 2) the end of tongue movement towards its constriction target. Time to target attainment was then defined as the difference between landmarks 2) and 1).

Results A Welch's two-sample t-test found that time to target was significantly longer for singing than for speech (Figure 1; $t = 6.934$, $df = 49.922$, $p < 0.001$). If singers do have an aesthetic task to reach vowel targets quickly, other forces prevent sung vowel targets from being attained as quickly as possible.

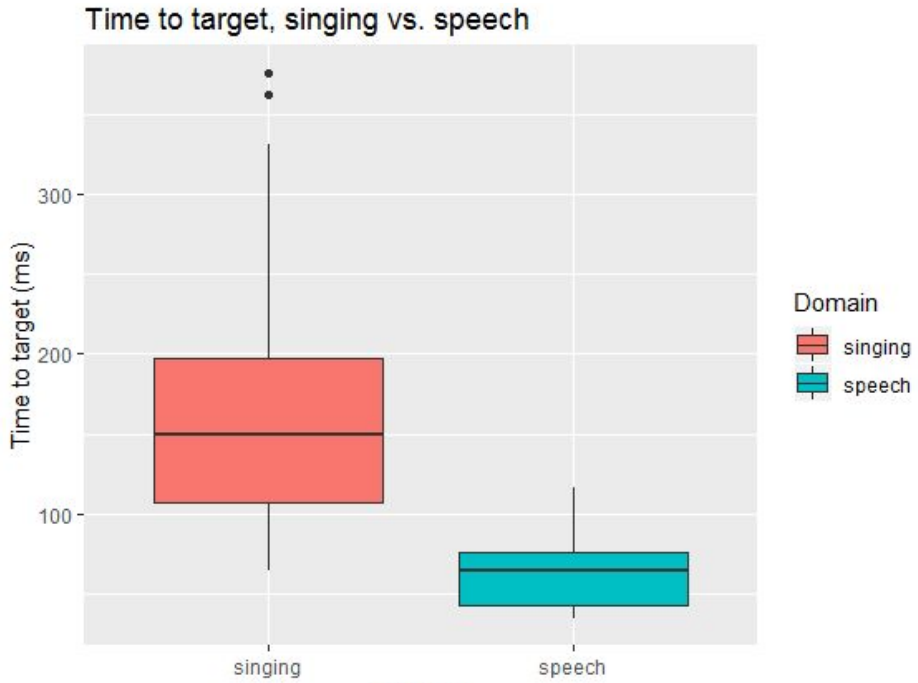

Figure 1. Time to target in singing domain (left) is greater than in the speech domain (right).

Implications Future work should consider the speed of sung consonants as well to determine whether the slower articulator movement observed here is due to vowel duration or a more global slowing of sung articulation.

To fully characterize the mental units of speech and their organization, it is important to understand both articulatory constraints for different types of speech behavior. The current project offers a small taste of the types of questions that can be tested with studies of singing, such as what the kinematics of vowel lengthening are under various conditions or how aesthetic goals influence gestural coordination. [Supported by NIH.]

References

1. Bresch, E., & Narayanan, S. (2010). Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. The Journal of the Acoustical Society of America, 128(5), EL335-EL341.
2. Ramanarayanan, V., Lammert, A., Byrd, D., Goldstein, L., Narayanan, S. (2011). "Planning and execution in soprano singing and speaking behavior: An acoustic/articulatory study using real-time MRI." International Seminar on Speech Production (ISSP), Montreal, Canada.
3. Tiede, M. (2010). MVIEW: Multi-channel visualization application for displaying dynamic sensor movements. development.