Ingo Titze, Associate Editor

The Combination of Imaging and Computer Simulation: A Likely New Wave In Studio Voice Technology

Ingo Titze

Ingo R. Titze

Journal of Singing, January/February 2014
Volume 70, No. 3, pp. 329–330

The tape recorder, the video camera, and now a variety of sound processing applications on digital systems have become standard aids in studio teaching. They give information to the singer and the teacher, either in real time or with delayed feedback, about the complex relations between sound production and sound perception. A new door is opening that may accelerate voice training with technological assistance. It begins with large-scale imaging of the entire airway system. As magnetic resonance and low dose x-ray scans of the respiratory anatomy become faster and more economical, it is not difficult to imagine that an entry-level voice student could receive a full MRI scan at the beginning of vocal training. This would go along with currently available video stroboscopic imaging of the vocal folds. Rather than looking at the body from the outside and inferring what the instrument looks like on the inside, we would have a view of all the major airspaces and tissue interfaces. We would look at specific anatomic details of a given singer rather than generic charts of male, female, and child anatomy. Once the neutral shape of the airway is in front of us, specific phonatory, respiratory, and articulatory gestures can be overlaid from faster two dimensional scans.

But what will the visual scans obtained in a few imaging sessions do to facilitate exploration in weekly lessons or daily practice? Dreaming a little bit more, a practice room or studio with a computer and a wide screen will be able to serve up a virtual clone of the vocalist's instrument. A sound simulator will have been constructed from the singer's airway system. Controls will be provided to the singer and teacher for changing lung pressure, individual muscle activations, and articulatory gestures.

So, what is the value of this for human sound production? Why not simply put the computer clone on stage? Well, I tried this nearly 25 years ago with Pavarobotti, our singing robot. It took enormous technical skill behind stage to operate the robot, even when all segments of the sounds were simulated several months in advance. Instantaneous control and sound generation would have kept at least two humans (two brains, four arms, and four feet) busy just to change pitch, loudness, voice onset, and a few articulatory gestures. While man-made instruments are made to be played with hands, lips,

and feet, the larynx is made to be played with respiratory and articulatory musculatures. We can play a violin with five fingers on the strings and one hand on the bow. We can play a piano with two hands and one foot. We cannot, however, play our airways easily with hand and foot gestures.

Then why have a robot as a virtual clone? Two answers come to mind immediately: (1) we can change one thing at a time in simulation, and we know exactly what we are changing; and (2) we can push the limits without breaking anything. Simply said, exploration becomes more direct and less risky. What will a given voice sound like with more or less vibrato, with more or less ring, with or without roughness or hoarseness, when a muscle is weak or paralyzed? For those who question their voice category or gender, we might be able to demonstrate what a given voice will sound like with a classification or gender transformation.

In summary, I don't have much fear that singing robots will take over the live and spontaneous vocalizations of humans and other vertebrates, but skill acquisition and exploration with new vocal dimensions may be enhanced with future studio clones derived from imaging in combination with computer simulation.