Annual Report of the Yorkshire Philosophical Society, 2022, pp64-66.

Human Voice Synthesis and the Vocal Tract Organ

Professor David M Howard FREng
Founding Head of Department of Electronic Engineering, Royal Holloway, University of London
david.howard@rhul.ac.uk

Every day we use our voices without giving their operation any thought. Our voices are basic to our ability to communicate with one another and in this digital communication age where electronic devices communicate with us vocally, electronic voice synthesis has come of age.

A healthy vocal instrument is essentially taken for granted and typically, speakers and singers are basically unaware of their voice production system or how it works until something goes wrong. We learn to speak during the first years of by observing and listening to the world around us and usually this needs no special training.

Human voice production has three main elements:

1. a *power source* from breathing and the movement of air to and from the lungs

2. a *sound source* from the larynx and/or constrictions within the mouth/throat, and

3. the *sound modifiers* due to the shaping of the mouth and throat (the vocal tract).

The first known attempt at creating speech sounds synthetically was the Speaking Machine of Baron von Kempelen created in the 1790s. Figure 1 shows the author holding a copy of the von Kempelen Speaking Machine made by Geoffrey Coffin of Principal Pipe Organs in York. The bellows under the right arm act as the lungs (the *power source*), the darker box in the middle contains a reed that can vibrate (the *sound* source) and the leather tube in the left hand act as the vocal tract (the *sound* modifiers). In addition, it has 'ss' and 'sh' 'whistles' underneath the darker box in the middle (the tube and cone) which are controlled by the two buttons on the opposite side. When these are used, and the user can stop the reed vibrating by pressing the third button. A demonstration can be heard at [1].

Figure 1: The author holding a replica von Kempelen speaking Machine made by Geoffrey Coffin of Principal Pipe Organs in York.

Whilst the von Kempelen speaking machine demonstrates the principles of human speech production as a mechanical model, which for its time is really remarkable, it is limited in terms of its overall speech sound repertoire, the lack of a pitch control and its playing awkwardness.

The advent of electronics enabled models of human voice production to be created that consisted of:

(a) a ***sound source*** – either a 'buzz' with a pitch that can be varied for pitched sounds such as all the vowels and consonants such as those in 'judge', 'zoom', 'money' and 'ring' and/or a noise-like sound for speech sounds that are not pitched (try singing them!) such as the consonants in 'sea', 'she', and 'earth'.

(b) ***sound modifiers*** – which change the perceived nature of the sound source buzz or noise-like sound by removing or filtering out parts of it.

Providing the buzz and/or the noise-like sound sources and the sound modifier filtering are set up to match very closely those generated within the human vocal system (these have been studied and modelled over many decades), the resulting output can be highly natural sounding and nowadays, indistinguishable from natural human sounds.

The Vocal Tract Organ

My own more recent approach to creating vocal sounds has been to make a three-dimensional scan of the vocal tract for different vowel sounds and from it, to create 3-D printed models. This model is designed to couple to a loudspeaker via which is played an appropriate larynx buzz sound source as described above. The set-up is shown in figure 2 where the original scan is shown on the left and the 3-D printed tract in front of an artificial head for clarity atop its loudspeaker is shown on the right. The larynx source which models the output from the vibrating vocal folds makes use of the Liljencrants/Fant (LF) model of the larynx output that is commonly used in speech synthesis. This system synthesises fixed vowel sounds – not running speech and it was conceived as a new musical instrument with commonalities to a pipe organ; hence it is called the *Vocal Tract Organ*. There is a direct link with large pipe organs, some of which have a stop named *Vox Humana*, but the sound is not convincing as being a human voice whereas the output from the Vocal Tract Organ is clearly vowel-like.

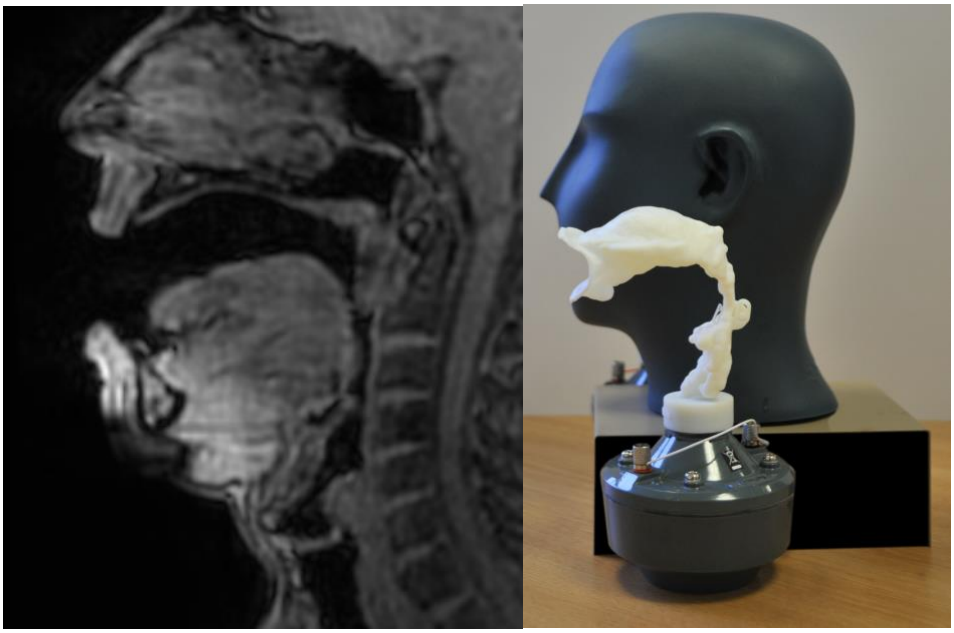

Figure 2: Magnetic resonance image for an adult male vowel in 'card' (left) and the 3-D printed version atop a loudspeaker placed in front of an artificial head for clarity (right).

The first version of the Vocal Tract Organ made use of an LF larynx model that was programmed using an Arduino microcomputer controlled by two joysticks, one for pitch and volume and the other for vibrato rate and extent. The second version makes use of an electronic piano or synthesiser MIDI keyboard on which chords of up to 8 notes can be played, and these can be played through a single vocal tract since the system is linear – it is like having up to 8 larynxes driving one vocal tract! Since no music exists for the Vocal Tract Organ, the author wrote a demonstration 4-part piece for two singers and two parts played on the Vocal Tract Organ for which the musical score can be seen at [2] and heard at [3].

The Vocal Tract Organ was first demonstrated in the presence of HRH the Princess Royal at the 2013 Royal Academy of Engineering Summer Soiree at the Ron Cooke Hub, University of York, on 27th June 2013 as an after dinner flash-mob performance of *O mio babbino caro* by Puccini sung by soprano Dr Helena Daffern accompanied by the author. This was repeated at the Graduands' Dinner in the Merchant Adventurers' Hall, York which was video recorded [4].

Figure 3: 3-D printed vocal tract of Nesyamun.

The most recent application of the Vocal Tract Organ has been the creation of a sound of the 3,000-year-old Egyptian Mummy, Nesyamun whose remains are displayed at Leeds Museum [5]. An MRI scan was made from which a 3-D printed vocal tract was created as shown in figure 5 which was placed atop a Vocal Tract Organ to create the vocal sound [6]. This work attracted huge world-wide media interest and the paper itself has had over a third of a million downloads to date. To our knowledge, this is the first vocal sound created in this way and it offers a new way to explore voices from the past, given that vocal tract models can be edited digitally with the potential to create a set of likely vocal sounds and perhaps even spoken words and sentences.

Links

1. https://www.youtube.com/watch?v=Lw1MsSHHlQs
2. http://davidmhoward.com/pdf/Music4Downloading/VocalVisionII-20June2013.pdf
3. https://www.youtube.com/watch?v=pUryWk-s9Ig
4. https://www.youtube.com/watch?v=SX-f1oU0_Kk
5. https://www.nature.com/articles/s41598-019-56316-y/
6. https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-019-56316-y/MediaObjects/41598_2019_56316_MOESM2_ESM.wav