KES Transactions on Innovation in Music: Vol 1 No 1
Special Edition - Innovation in Music 2013 : pp.7-19 : Paper im13bk-001

The Vocal Tract Organ

David M Howard

York Centre for Singing Science, Department of Electronics, University of York, UK
david.howard@york.ac.uk

1. Abstract

This paper describes an innovative musical instrument known as the Vocal Tract Organ that has arisen out of research into the acoustics of singing to promote understanding and enable a new performance paradigm of the human singing voice in performance. A fundamental aspect of the research side of this endeavour is to investigate individual elements of the singer's craft, such as intonation, vibrato, voice quality, vocal fold vibration, tongue position, jaw position, setting of the lips, vowel production, consonant production, posture, breath control, in terms of the physiological settings, acoustic outputs and their uses and roles in the context of musical performance. The Vocal Tract Organ consists of 3-D printed life-size vocal tracts that are excited by an electronically synthesised larynx source via small loudspeakers and the instrument is played via a standard piano-style keyboard. The Vocal Tract Organ offers the potential for a novel approach to both performance and composition as it brings together the performance opportunities offered by on of the oldest musical instruments, the pipe organ, in the context of one of the most expressive musical instruments that is basic to human life, the human singing voice.

2. Introduction

The human vocal tract is very flexible and able to produce a wide range of sounds that humans use for sophisticated speech communication, attracting attention, providing warnings to others, expressing emotion, making sound effects and perhaps the oldest of all; singing. Almost everyone is affected in some way by one or other style of singing, and most people engage in some sort of singing activity even if it is only humming tunes gently to oneself.

Overall understanding of the acoustics of the singing voice has advanced greatly over recent years in terms of the observed outputs for different singing styles, such as opera, belt, pop, rock, country and western, overtone and choral, and the various change in voice production associated with them. Knowledge of the physiological gestures made during singing is less well known in detail since it cannot be directly observed. Singing teachers work in the physiological domain and infer the gestures that contribute to an acoustic output that is appropriate to the desired signing style and interpretation of the music being sung. Some instructions from singing teachers are not directly related to what is actually happening in terms of articulation but are more related to a frame of mind or associated gesture in the form of a *psychological hook*. Part of the inspiration for this work is to understand better exactly what is happening within the vocal tract during singing and speech through direct observation using magnetic resonant imaging (MRI) techniques and to promulgate this to enable a fuller understanding of voice production.

Vocal tract images in 3-D from MRI imaging have been used to enable electronic synthesis of natural sounding spoken and sung sounds such as vowels through computer simulation (Speed et al., 2013, 2014). Having such vocal tract shapes has enabled 3-D prints to be created for use as lecture/workshop demonstration artefacts for furthering understanding of voice production. The visual similarity between these tracts and the pipes of a pipe organ, lead to the notion that the tracts could be used to create a special implementation of a pipe organ, albeit with an electronic sound source rather than a vibrating reed sound source at present, and thus the Vocal Tract Organ idea was formulated.

3. Human voice production

The human voice production system consists of three elements: the power source (breathing), the sound source (the vibrating vocal folds in the larynx for pitched sung sounds) and the sound modifiers (the varying resonant acoustic properties of the tubes of the throat, mouth and nose above the larynx). For the purposes of considering voice production and its acoustics, only the sound source and sound modifiers are considered.

The singing sound source is the output from the vibrating vocal folds vibrating in the larynx and this is activated and sustained due to a controlled flow of air from the lungs. The vocal folds are two small muscles that are aligned such that they can act as a valve to completely and securely shut the airway to the lungs, thereby protecting the lungs from ingress of particles, insects, dirt or any liquids. The vocal folds can be set in vibration to create a tone due to a regular synchronous vibration of both vocal folds such that they meet together at the midline once per cycle. The instant at which they meet is the main acoustic excitation to the sound modifiers or *vocal tract*, and healthy vocal folds will meet together and move apart, opening and closing the valve in a cyclic fashion during a sung note. If a soprano sings the A above middle C, the international tuning reference that has a frequency of 440 Hz, her vocal folds will open and close 440 times a second. Singers can change the pitch of a sung note by changing the number of vibrations per second, and they achieve this by altering the tension in the fold muscle tissue (rather like turning the tuning peg on a violin) or by change the portion of each vocal fold that is vibrating thereby altering the mass of the vibrating folds (rather like using thicker and thinner strings in a piano for lower and higher notes respectively). A trained singing can achieve a pitch range of two and a half octaves or more; in normal speech the pitch range averages around one octave.

No human can produce a note that remains exactly fixed at a single pitch; there will always be some variation. Everyone exhibits a *flutter*, which manifests itself as small changes in fundamental frequency suggesting that when synthesizing sung sounds, some variation is essential for a natural result. More importantly for singing is vibrato which is present to a great extent in Western trained opera singers, and sung fundamental frequency with vibrato typically will vary at a rate of approximately 5.5Hz to 7.5Hz with a variation range of between ± 0.5 and ± 2 semitones (Sundberg, 1987). Other singing style make use of vibrato to a greater or lesser degree, and it is common to include vibrato rate and depth controls in singing synthesis systems.

The sound modifiers are the throat, mouth and nose, known collectively as the *vocal tract*, in singing and speech. The nose can be switched in or out using the *soft palate* or *velum*, which acts as a valve to couple in or shut out the nose from the rest of the vocal tract. The shape of the vocal tract can be altered by moving the *articulators*, which are mainly the tongue, lips and jaw, and this has the effect of altering the volume of the mouth and throat. Note that the volume of the nose cannot be changed. During speech the articulators are moving almost continually; there is not much communication advantage to be gained from a non-changing sound. During singing there is somewhat less movement of the articulators since sounds such as vowels are often sustained for much longer times that in speech as a function of the performance requirements of the musical score.

As the articulators move they change the volume of the vocal tract, and this has a consequential effect on the acoustic resonant properties of the tract, which move in frequency. These resonances are called *formants* in the context of speech and singing and they will be at different frequencies depending on the vowel being articulated. The centre frequency of each formant is what is used to characterize vowels acoustically and they are usually from the lowest formant upwards in frequency as *F1* or the *first formant* for the peak that is lowest in frequency, *F2* (*second formant*) and *F3* (*third formant*). It is only the first three or even the first two formants that contribute to the perception of vowels in practice; the higher formants tend not to vary during running speech or singing. Six or more formants can often be found during acoustic analysis of the output, and these higher formants provide a contribution to the identity of an individual voice, whether speaking or singing.

4. Vocal Tract Organ implementation

A pipe organ is an acoustic harmonic synthesiser in which the stops enable the player to create different overall sounds depending on which are selected or *drawn*. In addition to the stops that enable harmonic synthesis, there are solo reed stops (e.g. serpent, trombone, tuba, trumpet, orchestral trumpet, krumhorn, bassoon, oboe, horn, clarinet) and for many years there has been a reed stop available known as a *vox humana* (human voice), which is typically only found on large organs. The *vox humana* is a reed stop that it has a metal reed at the base of each of its pipes to provide a rich harmonic sound source and above is a resonant tube that has been shaped to produce an output sound that purports to be that of a human voice. In the author's organ playing experience, *vox humana* stops rarely if ever has much in common with a human voice. For organists, given the close relationship between the organ and the singing voice in the context of accompanying choral music, a true *vox humana* stop would be very beneficial, and the 3-D vocal tracts provide a route to obtaining far more natural sounding vowel sounds than can be achieved with *vox humana* pipes.

Figure 1: MRI image of a sung “ah” vowel by a tenor on the A below middle C (A3, 220 Hz).

These vocal tract models are created from MRI images, taken in the GE 3 Tesla HDx Excite MRI scanner in the University of York’s Neuroimaging Centre, of a subject’s vocal tract producing a steady vowel. The data capture involves lying down in the MRI machine, which is a claustrophobic and uncomfortable environment with the subject being encased in a tube where there is a high level of ambient noise that rises when the MRI machine is gathering data (in-ear foam inserts are obligatory for hearing protection but these compromise vocal output monitoring). The data capture process itself involves sustaining a steady sound, for this work a vowel, with minimum tract or head movement to avoid blurring the images for over 16 seconds in order to allow the capturing of a full set of 81 pictures (512 by 512 pixels resolution) to cover the vocal tract. An example is shown in figure 1 for the midline of a sung “ah” vowel by a tenor on the A below middle C (A3, 220 Hz). In these MRI images the airway is shown in black and skin, bone and other tissue are non-black. Notice that the teeth do not show at all. The vocal folds are at the lower end of the image where the airway ends. The lips and nose are to the right of the figure and the bulk of the tongue can be seen, in this case for an “ah” vowel the tongue sits low in the mouth to create a large mouth cavity; the main consequence of this is that the throat is narrow since the bulk of the tongue is pushed backwards to enable it to be low in the mouth.

In practice, the use of MRI is not the most ideal way to gather vocal tract shape data due to the 16 second exposure time required, which is long compared to the time of articulation of a vowel in speech, the presence of local ambient acoustic noise, the effect of gravity on tissue due to the need for the subject to be in a supine position where the head is in a horizontal position rather than its usual vertical position in speech and the expense and availability of the MRI machine itself. However, at the present time there is no alternative approach that enables the level of detail needed to create realistic lifelike vocal tract shapes for this work.

From the set of MRI pictures across the vocal tract a 3-D image can be obtained, which can be converted into the '.obj' file format required for 3-D printing of the vocal tract. Some manual intervention is usually required around the boundaries of the airway to ensure that these are not breached; typically the mouth-nose boundary of the hard palate is not too clearly defined (this is the case in the image shown in figure 1). A 3-D print of the set of MRI images for which that shown in figure 1 is the central midline image is shown in figure 2. Here, the tract is shown sitting on its loudspeaker drive unit (Adastra model 952.210, 16 ohm, 60 Watt). The lower end of the vocal tract is specially shaped to connect tightly onto the loudspeaker drive unit to ensure the acoustic sound source is properly coupled to the tract.

Figure 2: A 3-D print of the set of MRI images for which that shown in figure 1 is the central midline image, sitting atop its loudspeaker drive unit as used in the Vocal Tract Organ.

Each printed vocal tract can be made to sound if a suitable voice source signal is applied at the position of the larynx in the neck (Howard and Murphy, 2008). The stimulus waveform that is used for the voice source is commonly used in speech synthesis and it is a close approximation to the observed output from the vibrating vocal folds. Its implementation is based on the Liljencrants/Fant (LF) glottal source model (Fant and Liljencrants, 1985), which has been synthesised in practice using Pure Data, or *PD*, (WWW-1). PD is well suited to this because it enables a wavetable synthesiser to be implemented that is based on either one cycle that is either (a) calculated from a set of harmonic amplitudes (a pulse and a sawtooth waveform is available in the system that is based on these; the user can switch between them), or (b) drawn by hand using the mouse (this is how the LF model is implemented enabling changes to its shape to be easily tested). The implementation of this glottal source for the Vocal Tract Organ for multi-part, or *polyphonic*, synthesis is described in Howard et al., (2013). In order that the result is perceived as being close to a natural output, each channel has a separate setting for vibrato rate, vibrato depth and volume. An overall volume control is also included which can be set using the mouse and a slider or externally manipulated via a MIDI control parameter. These can be set independently either via an on-screen slider with the mouse or over MIDI (Musical Instrument Digital Interface) via any programmable MIDI controller device.

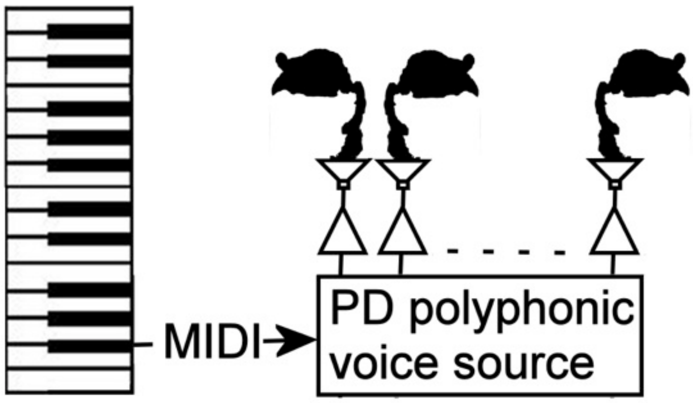

Figure 3: Block diagram of the Vocal Tract Organ.

A block diagram of the components of the Vocal Tract organ is shown in figure 3. Any MIDI keyboard can be used to play the instrument, and its MIDI output is linked to the MIDI input of a computer on which the polyphonic PD voice source synthesis system is running. In practice, six channels have been implemented (six-note polyphony) to enable four-part chorale style music to be played (quasi soprano, alto, tenor, bass, or SATB) with two spare channels to 'catch' moments in *legato* playing between chords when up to two new notes are selected before the old notes have been released. The use of six-note polyphony has been found to be sufficient in practice, but it does depend on playing technique and it could readily be increased to eight-channel polyphony in terms of the PD patch depending on computer speed and availability of sufficient audio output channels. The six outputs are routed via a multi-channel digital to analogue converter, an RME Fireface 400 [www-2], to an amplifier and loudspeaker atop of which sits the 3-D printed vocal tract.

The Vocal Tract Organ has been set up with just one "stop", which is an "ah" vowel. It is envisaged that other vowels will be added in the future when tracts are available. In order to ensure that the individual outputs from each of the vocal tracts is not identical, the set of six tracts for an "ah" vowel have been printed such that their lengths are very slightly different. Changing the overall length of the tract will have an effect on all the resonances of the tract, or *formants*, which characterise the sound of a vowel. Lengthening or shortening the tract will lower or raise all the formants respectively. Provided the length variation is small relative to the whole length, the effect will be to slightly change the vowel quality without change it to sound like another vowel. In practice for the Vocal Tract Organ, the lengths of the six tracts have been set to: 17.00 cm, 17.25 cm, 17.50 cm, 17.75 cm, and 18.00 cm.

5. First musical performances with the Vocal Tract Organ

The prototype Vocal Tract Organ was designed and implemented based on an initial requirement to play four-part chorale-like music to enable it to be used for multi-part vocalise-like ensemble music. As there is no specific repertoire for the Vocal Tract Organ itself, the author composed two pieces were composed to demonstrate the instrument and to enable its output to be heard firstly alongside and compared directly with, live singers, and in the second as an accompanying instrument for a solo singer.

In the first example, the author's *Vocal Vision II*, two male singers sing two parts and two other parts are played on the Vocal Tract Organ. The score for *Vocal Vision II* is shown in figure 4. *Vocal Vision II* is written as a four-part barbershop-style vocalise in which the 2nd bass and 2nd tenor parts are sung by male voices and the 1st bass and 1st tenor parts are performed on the Vocal Tract Organ. Notice that the entries of the voices are staggered to enable each of them to be heard clearly. The ordering of the voices on entering is bottom up (2nd bass up to 1st tenor) which results in a alternation between the human and synthesised vowels. The singers are asked to use same vowel sound that is being produced by the Vocal Tract Organ, which so far is that in “spa”. Part of the challenge for the performers is to blend their outputs with those from the Vocal Tract Organ to ensure that an overall good timbral ensemble is established and maintained. As with all singing performance, careful listening is paramount. The direction in bar 1 of the score to use “different vowels on repeat if possible” has so far not been a possibility due to non-availability of multiple vowel tracts which would be implemented as additional stops. Dynamics have not been provided in this version since the existing version of the organ does not offer a swell pedal for volume control. A performance of *Vocal Vision II* can be seen and heard on YouTube [WWW-3].

Figure 4: The score for the author's *Vocal Vision II*; a four-part barbershop-style vocalise in which the 2nd bass and 2nd tenor parts are sung by male

voices and the 1st bass and 1st tenor parts are performed on the Vocal Tract Organ.

In the second example, the Vocal Tract Organ has been used to accompany a solo soprano singing an opera aria. The original event for which this was organised was a black tie after-dinner short *flashmob* opera aria entertainment to highlight engineering in a musical context, and a member of the British Royal Family was in attendance. The piece was *O mio babbino caro* from the opera *Gianni Schicchi* (1918) by Giacomo Puccini. The usual keyboard accompaniment for this aria is an orchestral reduction and whilst it works well as a piano accompaniment, it is not well suited for accompanying on the Vocal Tract Organ. In order to facilitate this performance, the author created a new keyboard part in the spirit of a four-part chorale; the score is shown in figure 5.

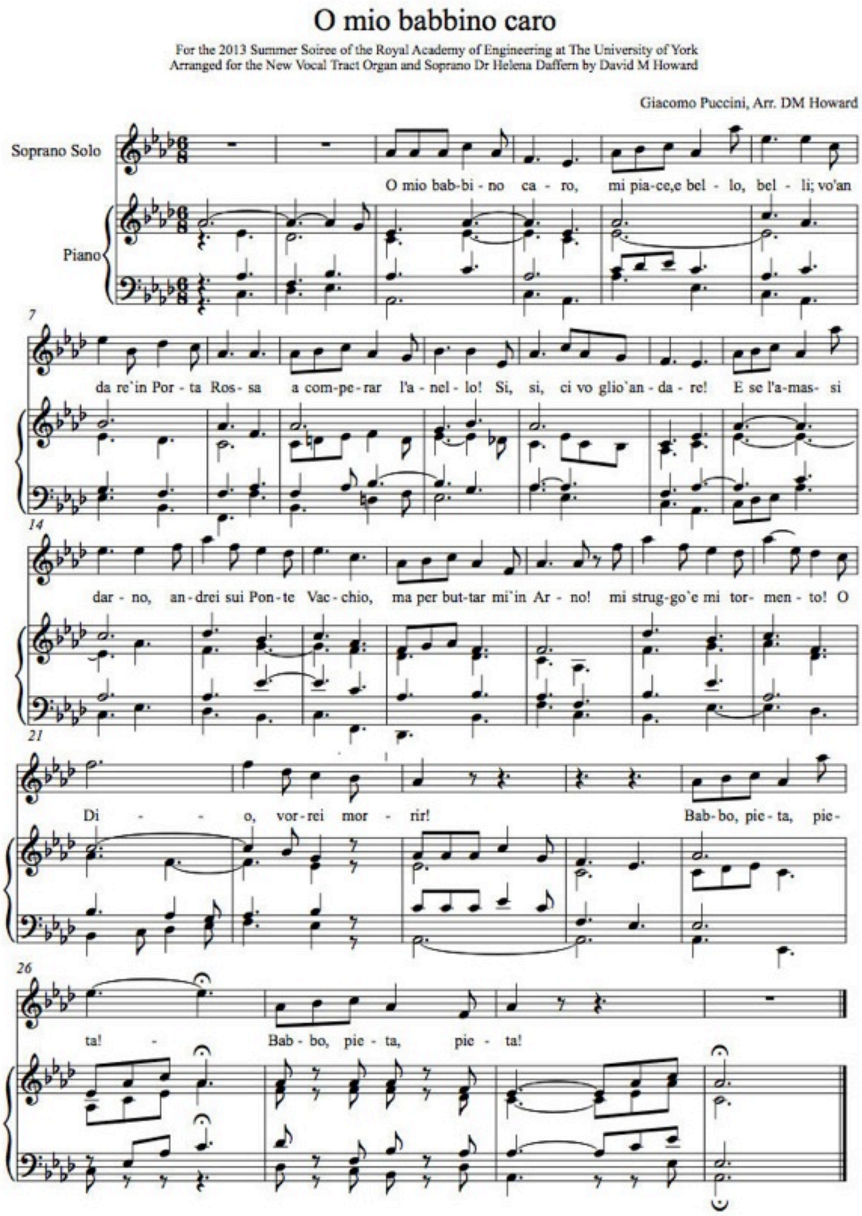

Figure 5: The author's four-part chorale-like realisation of the accompaniment for *O mio babbino caro* from the opera *Gianni Schicchi* (1918) by Giacomo Puccini to be performed on the Vocal Tract Organ, originally in *flashmob* style.

The score was arranged to enable the *flashmob* aspect of the performance such that it starts with a single note, which builds up into a chord to enable the audience to hear what an individual tract sounds like without giving any clue as to the music that follows. In addition, the soloist, who is waiting in the audience, is given her first note right from the start of the introduction. In addition, the score does not double the tune; it serves to provide a harmonic context to support the soloist. Note that the Vocal Tract Organ accompaniment does have one brief moment where it takes the tune in bars 28 and 29, which was a deliberate ploy to ensure that listeners have a chance to hear the instrument itself. In performance, the chord at the end of bar 25 was cut short to enable the soloist to sing the top note alone with gusto, crescendo and placing a pause on it for artistic and dramatic effect.

There was a second performance of this piece in the Merchant Adventurers' Hall, York for the University of York's 50th Anniversary Graduands' Dinner in July 2013, and this can be viewed on YouTube [WWW-4 from 2m50s following a brief presentation about the organ]. The musical performance interpretations of the score described above are evident in this performance.

6. Conclusions

The Vocal Tract Organ is a new musical instrument that enables vocal sounds to be employed in an innovative manner in music making. The design of the Vocal Tract Organ brings together the pie organ and the singing voice in a unique manner. The vocal tracts themselves function as a much truer representation of the human voice than the classic *vox humana* reed stop that is found on a number of large pipe organs, offering new possibilities for solo organ works and accompaniment. Two pieces have been described, both have audio files available on-line for audition and the scores have been provided herein.

The future potential for this instrument in terms of development is considerable, including the additions of the following.

- Varying vocal tract lengths with keyboard pitch (e.g. man, woman, child).
- Different stops for different vowels (e.g. "ee", "er", "uu").
- Adding a second manual (keyboard) to perform vowel-based counterpoint.
- Dynamic vibrato rate and depth controls (e.g. via knee levers).
- Volume control (swell pedal).

Adding other vowel stops and perhaps a second manual offers the possibility of creating the sound of a virtual choir that has a number of different vowels mixed together. This would be a departure from typical choral music offering new timbral possibilities for exploitation in musical performances. There are many pieces for manuals only in the pipe organ repertoire, and these could be performed directly on the Vocal Tract Organ, thereby offering new timbral possibilities for their interpretation. The role of temperament in tuning will be explored in the future (changing this within the PD synthesiser is straight forward), and this has the potential to support research into tuning strategies in choirs.

7. Acknowledgements

The author is indebted to the singers who have taken part in the performances, Helena Daffern (mezzo soprano), and Ben Lindley (tenor) Bertrand Delvaux (bass), as well as Pete Turner for help with the 3-D prints and the electronic implementation of the Vocal Tract Organ.

8. References

[1] Fant G, Liljencrants J, Lin QG. (1985). A four parameter model of glottal flow, STL-QPSR, 2**,** (3), 119-156.
[2] Howard, D.M., and Murphy, D.T. (2008). *Voice science, acoustics and recording*, SanDiego: Plural Press.
[3] Howard, D.M., Daffern, H., and Brereton, J. (2013). Four-part choral synthesis system for investigating intonation in a cappella choral singing, *Logopedics Phoniatrics Vocology*, **38**, (3).
[4] Speed, M., Murphy, D.T., and Howard, D.M. (2013). Three-dimensional digital waveguide mesh simulation of cylindrical vocal tract analogs, *IEEE Transactions in Audio, Speech and Language Processing,* **21**, (2), 449-455.
[5] Speed, M., Murphy, D.T., and Howard, D.M. (2014). Modeling the vocal tract transfer function using a 3D digital waveguide mesh, *IEEE Transactions in Audio, Speech and Language Processing,* **22**, (2), 453-464.
[6] Sundberg, J. (1987). The science of the singing voice, Dekalb: Illinois University Press.
[7] WWW-1: http://www.puredata.info (last accessed 2nd June 2014).
[8] WWW-2: http://www.rme-audio.de (last accessed 2nd June 2014).
[9] WWW-3: http://www.youtube.com/watch?v=pUryWk-s9lg (last accessed 2nd June 2014).
[10] WWW-4: http://www.youtube.com/watch?v=SX-f1oU0_Kk (last accessed 2nd June 2014).