The Vocal Tract Organ:
a new musical instrument using 3-D printed vocal tracts*

http://dx.doi.org/10.1016/j.jvoice.2017.09.014

David M Howard FREng
Department of Electronic Engineering, Royal Holloway, University of London, UK

Abstract
The advent and now increasingly widespread availability of 3-D printers is transforming our understanding of the natural world by enabling observations to be made in a tangible manner. This paper describes the use of 3-D printed models of the vocal tract for different vowels that are used to create an acoustic output when stimulated with an appropriate sound source in a new musical instrument: the Vocal Tract Organ. The shape of each printed vocal tract is recovered from magnetic resonance imaging. It sits atop a loudspeaker to which is provided an acoustic L-F model larynx input signal that is controlled by the notes played on a MIDI (musical instrument digital interface) device such as a keyboard. The larynx input is subject to vibrato with extent and frequency adjustable as desired within the ranges usually found for human singing. Polyphonic inputs for choral singing textures can be applied via a single loudspeaker and vocal tract, invoking the approximation of linearity in the voice production system, thereby making multiple vowel stops a possibility while keeping the complexity of the instrument in reasonable check. The vocal tract organ offers a much more human and natural sounding result than the traditional Vox Humana stops found in larger pipe organs, offering the possibility of enhancing pipe organs of the future as well as becoming the basis for a 'multi-vowel' chamber organ in its own right.

* This paper was presented at the 46th Annual Symposium: Care of the professional voice

Keywords: vocal tract; 3-D printing; MRI; vowels; vox humana; pipe organ

1. Introduction

Artificial computer-based synthesis of human speech and singing are commonplace today. Various strategies exist for modelling human speech and singing production that have enabled electronic speech and singing synthesis for a number of applications such as telephone answering services, text reading systems, timetable announcements, on-board transport announcements, announcements in lifts, the provision of warning messages and backing and solo sung tracks in music studio recordings.

MC	**Voice synthesis method**	FCP
	Manipulation of recorded natural speech waveforms *no knowledge of speech production mechanism*	
	Linear predictive synthesis (LPC) *all-pole acoustic model of the vocal tract*	
	Formant synthesis *direct control of formant centre frequencies and bandwidths*	
	Articulatory physical modelling synthesis *control of articulation in a vocal tract model*	

Table 1: Overview of four common methods used for voice synthesis in terms of their *model complexity (MC)* and the *flow of control parameters (FCP)* indicated by the light gray to black columns (data from [1]).

A useful breakdown of four commonly used and readily available synthesis methods is provided by [1] to indicate the key differences to consider between these methods in terms of (a) the *model complexity (MC)* which indicates the degree of knowledge relating to the voice production process required in the method, and (b) the *flow of control parameters (FCP)* which indicates the amount of control data that are required to perform the synthesis usually expressed as the control data sampling rate. This is illustrated in table 1. For the manipulation of recorded natural speech waveforms, FCP is highest (highest data sampling rate), requiring updates once at every fundamental period and it becomes progressively lower moving down the table to a minimum for articulatory physical modelling synthesis requiring updates once per articulatory gesture. (This is summarised in the FCP column in table 1 by the shades of gray with black indicating the maximum update rate.) Conversely, the MC in terms of the knowledge of speech production required for the synthesis techniques listed is the converse; a maximum for articulatory physical modelling synthesis and a minimum for the manipulation of recorded natural speech waveforms. (This is summarised in the MC column in table 1 by the shades of gray with black indicating the greatest amount of speech production knowledge required.)

Styger and Keller [1] argue that the choice between different synthesis techniques involves a trade-off between the FCP and MC depending on the overall computation required and the naturalness of the final output. If one wishes to create a highly natural result that provides a close approximation to the speech of a specific speaker, then articulatory physical modelling synthesis is most appropriate, but it requires in-depth detailed knowledge of articulation in natural speech. In the case of the Vocal Tract Organ, this implies that the use of measured shapes of the human vocal tract should provide the best basis for achieving highly natural speech/signing synthesis, especially since detailed knowledge of articulation in natural speech or singing is not needed since the vowels are static for vocalise-style choral singing.

In terms of articulation and gaining a greater understanding of the detailed shape within the vocal tract during speech and singing, magnetic resonance imaging (MRI) can provide

accurate and major insights [2]. Indeed, MRI is being explored in the context of vocal tract articulation during singing to enable acoustic properties observed during singing performance to be related directly to articulation itself [e.g. 3-6]. There are consequences and therefore caveats during MRI data collection in terms of how natural the recording experience is of being: (a) supine [7, 8], (b) enclosed in a potentially claustrophobic space, (c) exposed to the MRI machine's acoustic noise, and (d) less able to hear oneself for auditory feedback purposes when producing utterances. However, it has been shown that subjects are able to replicate closely the utterances they produce in the MRI machine when the experience within the machine is replicated beforehand and afterwards lying on a foam mattress in an anechoic chamber listening MRI machine noise over closed back headphones [9].

It was during the collection of MRI data in relation to singing voice production that the idea raised itself of the possibility of making a new musical instrument based on a 3-D printed life-size model of a vocal tract with a suitable loudspeaker drivers, a technique demonstrated by Fujita and Honda [10]. This is very much in keeping with the work of von Kempelen [11] and others [12-14] on speaking machines. However, a key difference is that they were manipulating the shape of the vocal tract as mechanical analogues of the human speech production system to create dynamic variations to simulate running speech. The author has a modern replica of a von Kempelen speaking machine (see figure 1) which is used for demonstrating voice production; the visual, tangible and rather unusual nature of such demonstrations is an important aspect for engaging those new to the field of speech and singing science [15].

A major part of the inspiration for this work is: (1) to understand better exactly what is happening within the vocal tract during singing and speech through direct observation using MRI, (2) to promulgate this to enable a fuller understanding of speech and singing production and vocal tract shape differences in individual singers and speakers, and (3) to create a novel music performance instrument. The latter is possible through the use of static life-like 3-D printed vocal tracts of specific sung vowels which provides a direct link with vocalise singing (music sung by one or more singers just on vowels). This provided the catalyst for the creation of the Vocal Tract Organ as a new musical instrument. The Organ has potential as a new 'natural' sounding musical performance instrument as well as providing opportunities for engineering and voice science outreach to raise awareness of vocal function with specialist and non-specialist audiences alike.

Given the visual similarity between 3-D printed vocal tracts and the pipes of a pipe organ and the possibility of exciting them musically via a keyboard to create a choral vocalise, the instrument has become known as the 'Vocal Tract Organ'. There is a pipe organ stop called 'vox humana' (human voice) but Howard [16] notes that '*typically a vox humana stop rarely sounds anything like a human voice and tends to be rather nasal and harsh sounding*'. In addition, a major voice in organ building, George Ashdown Audsley (designer of the vast Wanamaker organ in Philadelphia, USA) notes of the vox humana stop [17] that *"even the best results that have hitherto been obtained fall far short of what is to be desired. ... Of all stops of the organ, the Vox Humana is the one to which distance lends the greatest charm."* [13, Vol. 1, page 574]. and that *"such stops when heard in their immediate neighborhood are coarse and vulgar in the extreme."* [13, Vol 2, p 609].

Brackhane, F. and Trouvain [18] note that the vox humana pipe organ stop was once a substitute for boy's choirs in church and that the measured formant frequencies of vox humana pipes: (a) varied between organ builders, and (b) were, on average, higher than those found in a natural human voice. Howard [19] notes that the comparative rarity of the vox humana stop on church organs could be due to it being a poor substitute for a choir, probably due to the raised formant frequencies.

Thus the Vocal Tract Organ has potential as a musical instrument both in its own right and as a basis for providing natural-sounding vox humana stops to a pipe organ.

2. Materials and Methods

2.1 Principles of operation of the Vocal Tract Organ

The Vocal Tract Organ models human sung vowel production in keeping with the source/filter model described by Fant [20] that has formed the basis for describing the acoustics of speech and singing production and for their formant synthesis over many decades. A commonly used and well established source model in format synthesis systems is the L/F model [21] that defines mathematically a method for calculating individual cycles that approximate the derivative of glottal flow as observed in human speech and singing.

During speech and singing the fundamental frequency of the vibrating vocal folds is varied to convey pitch to listener(s) as intonation in speech and notes in singing. There is always some variation in the perceived pitch, even if the singer or speaker attempts to hold a fixed pitch output, due to the presence of *flutter* which results in small changes in fundamental frequency [22]. Additionally in singing, there is vibrato which is particularly obvious in Western opera singing but present in other genres also, where the fundamental frequency is varied at a rate of approximately 5.5Hz to 7.5Hz with a variation range of between ± 0.5 and ± 2 semitones [23]. The Vocal Tract Organ incorporates controls for vibrato rate and depth across these ranges. It is noteworthy that the pipe organ vox humana stop is almost invariably used with a tremulant stop which modulates the flow of air to the pipes producing a tremolo.

The filter section of the source/filter model describes the acoustic resonances of the vocal tract (throat, mouth and nose), which vary as different sounds are articulated which changes the three dimensional shape of the vocal tract. In the case of the Vocal Tract Organ, these are an inherent aspect of the 3-D printed models themselves, since the individual tract shapes for specific sounds are captured and therefore the acoustic properties should be maintained in direct relation to those shapes. However, successful synthesis of speech sounds based on 3-D printed models relies on the resonances of the 3-D models having similar acoustic properties in terms of their centre frequencies and bandwidths to the natural resonances of the human vocal tract. The centre frequencies of the vocal tract resonances are primarily related to the shape of the vocal tract and since this is based on direct measurement of a human vocal tract articulating the relevant speech sound, these should be well defined.

The bandwidths of the vocal tract resonances relate to the acoustic properties of the walls of the vocal tract and since the materials available for 3-D printing are very different to human tissue, this is where the greatest potential acoustic output difference lies. In the frequency and time domains, the bandwidths relate to the widths of the vocal tract resonances and the ringing decay after an excitation pulse respectively. The bandwidths are functions of the acoustic absorbing/reflecting properties of the tract wall itself; today's 3-D printing materials and human tissue are clearly very different. The material used for printing the vocal tracts is governed by what is available in practice with 3-D printers (in this case it was Verowhiteplus RGD835 which is stable with time and unaffected by moisture), therefore leaving the wall thickness as the only variable that can be altered. Tests with the printed tracts revealed that a wall thickness of 2 mm gave the closest approximation to vocal tract resonance bandwidths observed in practice.

2.2 Vocal Tract Organ implementation

The Vocal Tract Organ currently exists in two forms: (1) as a polyphonic (currently 6 note) instrument playable from a MIDI (musical instrument digital interface) keyboard, and (2) as an Arduino-based embedded systems implementation playable from two joysticks. In both formats, the sound source is implemented as a synthesised LF glottal source model [21].

In the MIDI keyboard version of the Vocal Tract Organ, the sound source is synthesised in real-time using the freeware system Pure Data, or Pd, [24]. Pd enables the use of a wavetable synthesiser whose stored cycle can be hand-drawn as a representation of the LF model and modified live in real-time during synthesis. In this way, changes in the LF waveshape can be easily implemented and their effect can be heard immediately (in practice the audible variation is very modest for different hand-drawn waveshapes). In addition, aliasing effects related to the more normal mathematical implementation of the LF model as piecewise functions with derivative discontinuities [25] are likely avoided since discontinuities are les likely in a hand-drawn version. Two standard waveforms used in music synthesis are also available to enable comparison with other source waveforms: a pulse and a sawtooth. In practice, these are calculated in Pd as a harmonic series, but they can also be hand drawn as desired. Switching between the three sound sources can be accomplished via on-screen buttons controlled with the mouse in real-time during synthesis. Details of the Pd implementation for the Vocal Tract Organ is given in [19, 26]. In order that the result be 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.

The sound modifiers within the Vocal Tract Organ are based on vocal tract 3-D volume measurements gained from MRI images taken in the Neuroimaging Centre at the University of York with the General Electric 3.0 Tesla Excite MRI scanner. A 3-D fast gradient echo sequence was employed that created eighty 512 by 512 pixel midsagittal plane images over 16 s (full details can be found in [27]). During the 16 second scan the subject is asked to sustain a steady vowel sound on a given pitch and to be aware of maintaining her/his vocal tract shape in a static configuration to the best of their ability. Acoustic monitoring by subjects of their vocal outputs is severely compromised due to the high ambient acoustic noise created by the scanner and the consequent Health and Safety requirement to wear in-ear foam inserts under plastic closed-back ear defenders for hearing protection. In addition, the subject is lying down in a potentially highly claustrophobic tube in the MRI machine and asked to remain as still as possible during the 16 second scanning time to avoid image blur. It should be noted that the effect of gravity on the supine rather than upright vocal tract could influence to a small degree the surrounding positioning of the tract walls [6-8].

An audio benchmark recording was established by repeating the scanning procedure in an anechoic room in both supine and standing positions, wearing earplug hearing protection but with a pair of Audio-Technica ATH-M30 closed-back headphones over which MRI scanning noise was played [9]. Speed et al. monitored vocal fold vibration with an Electrolaryngograph [28] to enable a pitch-matched source to be used for resynthesis. They found that there were only slight differences between the formant frequencies and amplitudes when supine and standing, indicating that MRI (supine) tract shapes are very close but not exactly comparable with standing tract shapes. They also found a very close match between synthesised vowels based on the MRI images and the supine anechoic recordings, confirming accurate repetition abilities of subjects. This serves to confirm the accuracy of vocal tract shapes gained from MRI scanning which are the sound modifiers in the Vocal Tract Organ.

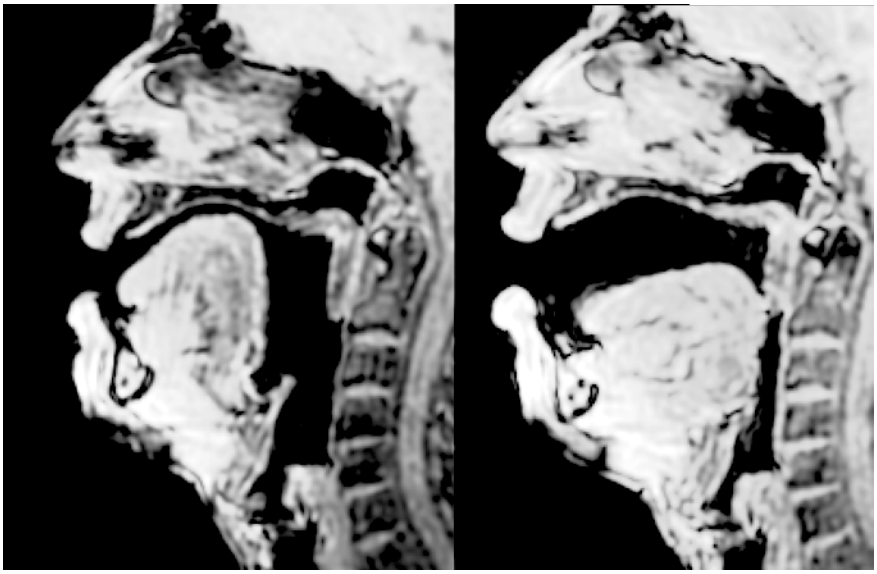

Figure 1: MRI image of the sung vowels /i:/ (left) and /ɔ:/ (right) by a tenor on the A below middle C (A3, 220 Hz).

An example MRI frame in the midline for the vowels /i:/ and /ɔ:/ sung by a tenor on A3 (220 Hz) are shown in figure 1. The airway is represented in black; non-black areas are skin, bone and other tissue. The teeth are absent in these scans which does have acoustic consequences in relation to the presence of possible side-branch cavities [29], but for the purposes of the organ the absence of teeth is not considered to be major given its overall output sound quality, but is something to be considered for the future when appropriate scans can be obtained. Here for the vowel in 'Pete', the bulk of the tongue is raised creating a narrow mouth cavity and consequentially an enlarged throat region characteristic of an /i:/ vowel production. For the vowel in 'port' the mouth is larger than the throat and the lips are rounded (this is not visible in this plane). In total there are 80 images and taken together these enable a 3-D representation of the vocal tract to be realised. Establishing the shape of the airway was achieved using ITK-SNAP [30] that joins the airway across the complete set of 80 frames together in three dimensions. A certain amount of hand editing (<1 mm) is required to ensure that the airway is appropriately selected mainly due to poor image contrast at air/tissue boundaries due to blurring or tissue movement during the scans.

The larynx end of the tract has to be coupled to a loudspeaker to enable the sound source to be the acoustic input to the 3-D printed vocal tract. This requires a small aperture loudspeaker drive unit with some ready means of coupling for its output. After patient investigation, a suitable loudspeaker was found: the Adastra model 952.210 (16 ohm, 60 Watt). This loudspeaker drive unit is commonly used for public address systems in the open air, for example on top of vans to play music or address crowds. A tight acoustic coupling between the loudspeaker drive unit and the 3-D printed vocal tract is achieved via an extension that is added to the vocal tract which fits over the threaded end of the loudspeaker driver. An approximately average glottal opening (~10 mm * 4 mm) is cut into the end of the vocal tract to provide an air path for the sound source to enter the tract.

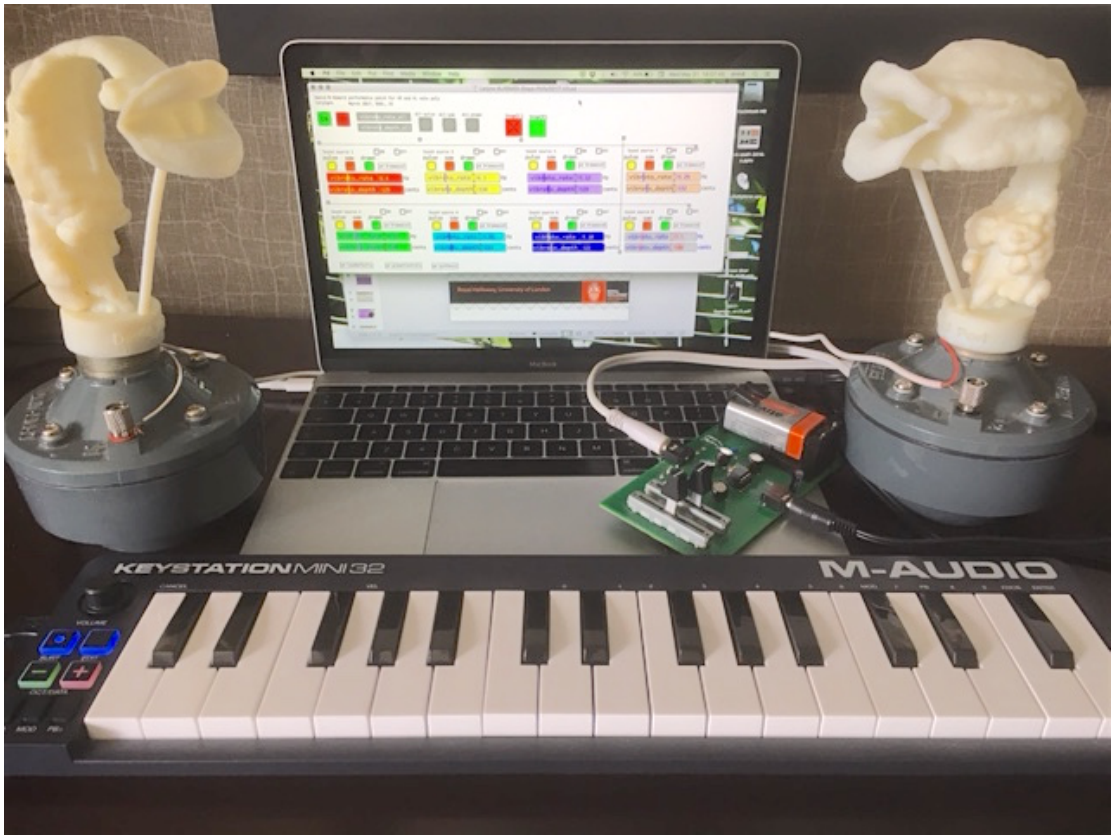

Figure 2: The portable two-stop Vocal Tract Organ consisting of a MacBook running Pure Data (Pd) that generate the source waveforms for the notes played on the MIDI keyboard, a stereo battery-powered power amplifier board and two Adastra model 952.210 loudspeaker drivers, one for each stop, with a 3-D printed vocal tract for /i:/ (left) and /ɔ:/ (right).

Figure 2 shows a complete version of a portable Vocal Tract Organ in which there are two 3-D printed tracts sitting atop Adastra 952.210 loudspeaker drivers, an /i:/ tract on the left and an /ɔ:/ tract on the right. The laptop screen shows the highest level of the Pd patch and the small key MIDI keyboard is played to provide the musical notes. The small circuit board is a battery-powered stereo power amplifier circuit which drives the two loudspeaker drivers and included individual slider volume controls for each channel.

The original Vocal Tract Organ was six-note polyphonic (six notes could be played at the same time) and this was implemented using six tracts (with slightly different lengths to provide some subtle acoustic differences between them) with their loudspeaker drivers. Invoking the principle of linearity in the system, in keeping with the source/filter model of speech production [20], enables the practical step in the Vocal Tract Organ of mixing all the larynx signals together in Pd and sending them to one loudspeaker driver with its vocal tract. Whilst it is now known that there are non-linear interactions in natural human voice production between the vibrating vocal folds of the larynx and the vocal tract [e.g. 31], their acoustic influences mostly affect whether the vocal folds are able to vibrate or not depending on the vocal tract resonance settings. In the Vocal Tract Organ there are no vocal folds and the source itself is generated electronically and therefore there is no non-linear coupling between the source and the resonances of the 3-D printed Vocal Tract. Thus multiple sound sources can be sent via the loudspeaker through a single 3-D printed vocal tract. This saves on electronic resources and enables a portable version, but more importantly, it allows for stops to be provided for different vowels. Hence the presence of two vowel tracts (/i:/ and /ɔ:/) in the Vocal Tract Organ shown in figure 2, the left and right channels of the stereo audio output

from the laptop providing the sound source for the /i:/ and /ɔ:/ vowel tract respectively. In the future it will be possible to make use of a soundcard with greater than 2 channels to provide many more stops (e.g. a number of different vowels for an adult female, adult male and child).

3. Results

3.1 Vocal Tract Organ output

Vocal tract shapes based on MRI outputs for five different vowels (/i:/, /ɜ:/, /a:/, /ɔ:/ and /u:/) have been produced for the Vocal Tract Organ to date. Invoking linearity and sending all the separate source waveforms to a single vocal tract is a key feature of the Vocal Tract Organ that will enable its development as a multi-stop instrument that is portable. The individual excitation waveforms are generated with each note having its own unique vibrato rate and depth which ensures source independence between the notes.

In order to demonstrate the acoustic frequency response of these tracts, comparisons are presented between the vowels spoken by the original informant in an acoustically anechoic room immediately after he had been in the MRI machine. This was part of the recording protocol adopted [9] where recordings made in the anechoic room before and after the MRI recording were directly compared with acoustic simulations of the outputs from the measured vocal tract shapes and the match was deemed to be good. Acoustic data recorded in an anechoic room before and after the MRI recording for the subject whose vocal tracts are presented with the Vocal Tract Organ enabled natural and Vocal Tract Organ formant data to be compared.

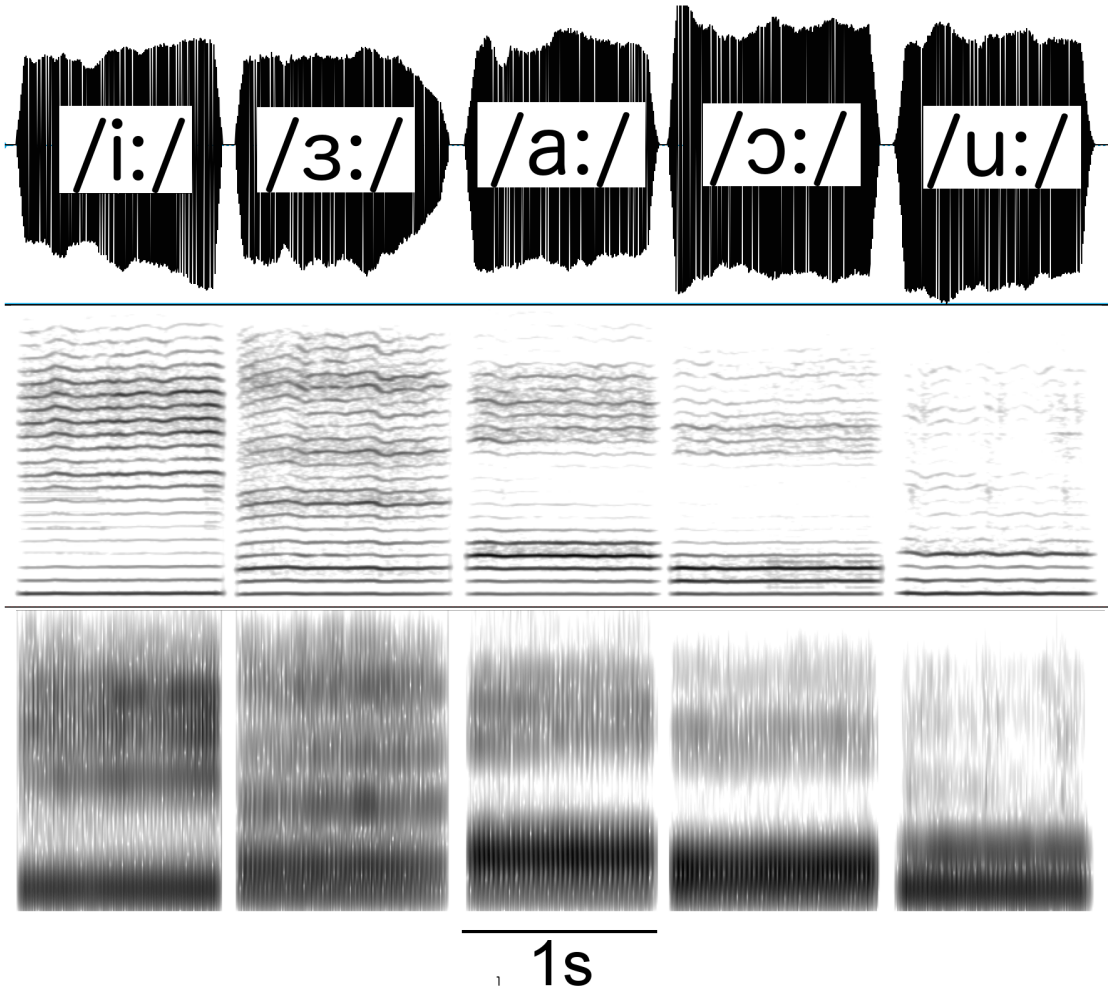

Figure 3: Time waveforms (upper), narrow-band 60 ms analysis window 0-5 kHz spectrograms (centre) and wide-band 3 ms analysis window 0-5 kHz spectrograms (lower) from Praat for the five vowels (/i:/, /ɜ:/, /a:/, /ɔ:/ and /u:/) spoken by the original adult male subject from whose MRI data the 3-D printed vocal tracts were made. Measured formant frequencies (F1, F2, F3) from Praat are given in Table 2.

Figure 3 shows time waveforms and narrow-band (60 ms analysis window) and wide-band (3 ms analysis window) 0-5 kHz spectrograms from Praat [32] for the five vowels (/i:/, /ɜ:/, /a:/, /ɔ:/ and /u:/) spoken by the adult male subject from whose MRI data the 3-D printed vocal tracts were made. Figure 4 shows the equivalent plots for the same five vowels from the Vocal Tract Organ with the excitation being an LF model with vibrato provided by the Arduino system. Measured formant frequencies (F1, F2, F3) gained from Praat for each of these sets of vowels are given in Table 2.

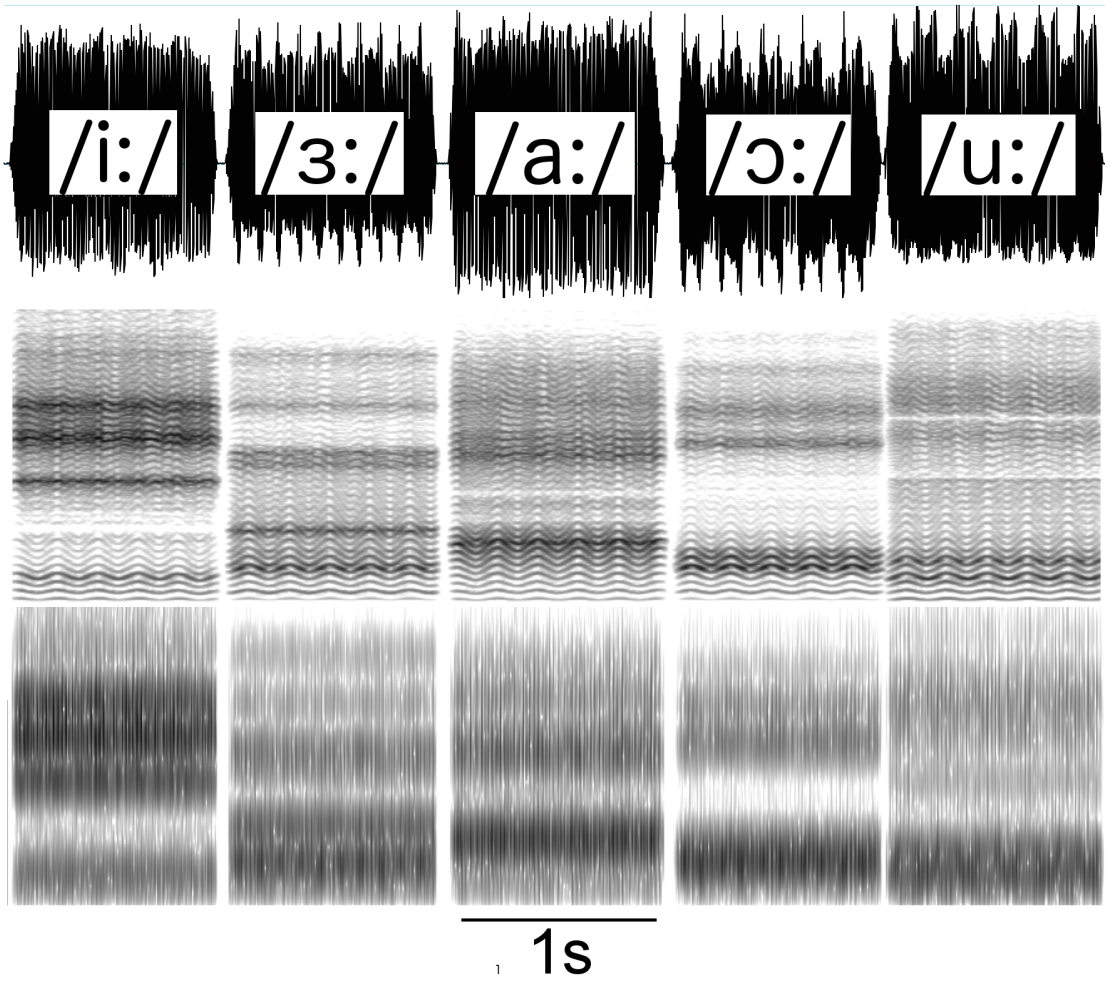

Figure 4: Time waveforms (upper), narrow-band 60 ms analysis window 0-5 kHz spectrograms (centre) and wide-band 3 ms analysis window 0-5 kHz spectrograms (lower) from Praat for the five vowels (/i:/, /ɜ:/, /a:/, /ɔ:/ and /u:/) produced by the Vocal Tract Organ driven by the Arduino with a fundamental frequency of 127 Hz. Measured formant frequencies (F1, F2, F3) from Praat are given in Table 2.

vowel		*F1 (Hz)*	*% diff*	*F2 (Hz)*	*% diff*	*F3 (Hz)*	*% diff*
/i:/	VT Organ	446	78	2091	-3	2831	-7
	MRI informant	251		2151		3022	
/3:/	VT Organ	619	19	1250	25	2434	-9
	MRI informant	521		1000		2665	
/a:/	VT Organ	849	32	1037	1	2579	-9
	MRI informant	642		1030		2812	
/ɔ:/	VT Organ	700	31	820	25	2722	-3
	MRI informant	533		653		2802	
/u:/	VT Organ	527	108	818	-12	1866	-14
	MRI informant	253		925		2154	

Table 2: Measured formant frequency data in Hz for 4 vowels uttered by an adult male speaker of American English from [26] along with average formant values (in brackets) measured in Hz using Praat for the vowels shown in figure 2. No data were provided in [26] for the vowel /ɔ:/.

Table 2 shows the measured formant frequency values for the human informant whose MRI data provided the vocal tract shapes for these measurements along with a percentage difference for the measured frequencies of the Vocal Tract Organ compared with those from the human informant. All the Vocal Tract Organ vowels have considerably higher F1 values than the original informant (large positive percentage values), and these differences are particularly high for /i:/ and /u:/. The matches for the second formant are much closer, and this is most notable for the vowels /i:/ and /a:/ which are very close. The third formant values for the Vocal Tract Organ are all under those from the human informant, but the percentage differences are rather small. Overall, it appears that there are good matches for F3, slightly less good matches for three of the vowels for F2 and poor matches for F1 which are all too high.

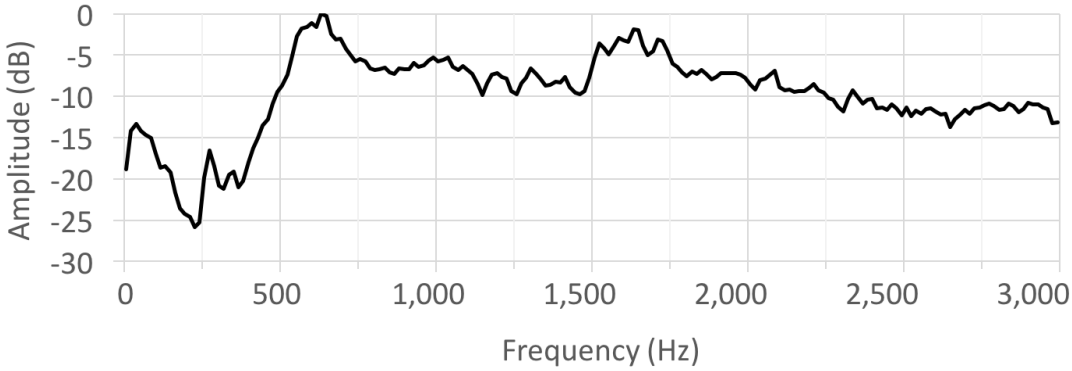

Figure 5: Output spectrum from the Adastra model 952.210 loudspeaker driver for a white noise (flat spectrum) input across the frequency range of interest for the first three formants on the five vowels (0-3 kHz).

The excitation signal used in the Vocal Tract organ is fed to the Vocal Tracts via the Adastra loudspeaker driver. This has the frequency response shown in figure 5 for a white noise (flat spectrum) input across the frequency range of interest (0-3 kHz) for the first three formants on the five vowels. It can be seen that while there is a reasonably flat response from around 750 Hz upwards, albeit with a small peak around 1600 Hz, there is a very marked drop in response below 750 Hz. This explains why the first formant values measured from the Vocal Tract Organ are too high as it can be seen from table 2 that the reference values from the human informant are all well below 750 Hz. In the future it may be possible to compensate

for this roll-off assuming that the loudspeaker driver is able to handle higher levels in its low frequency range without distortion.

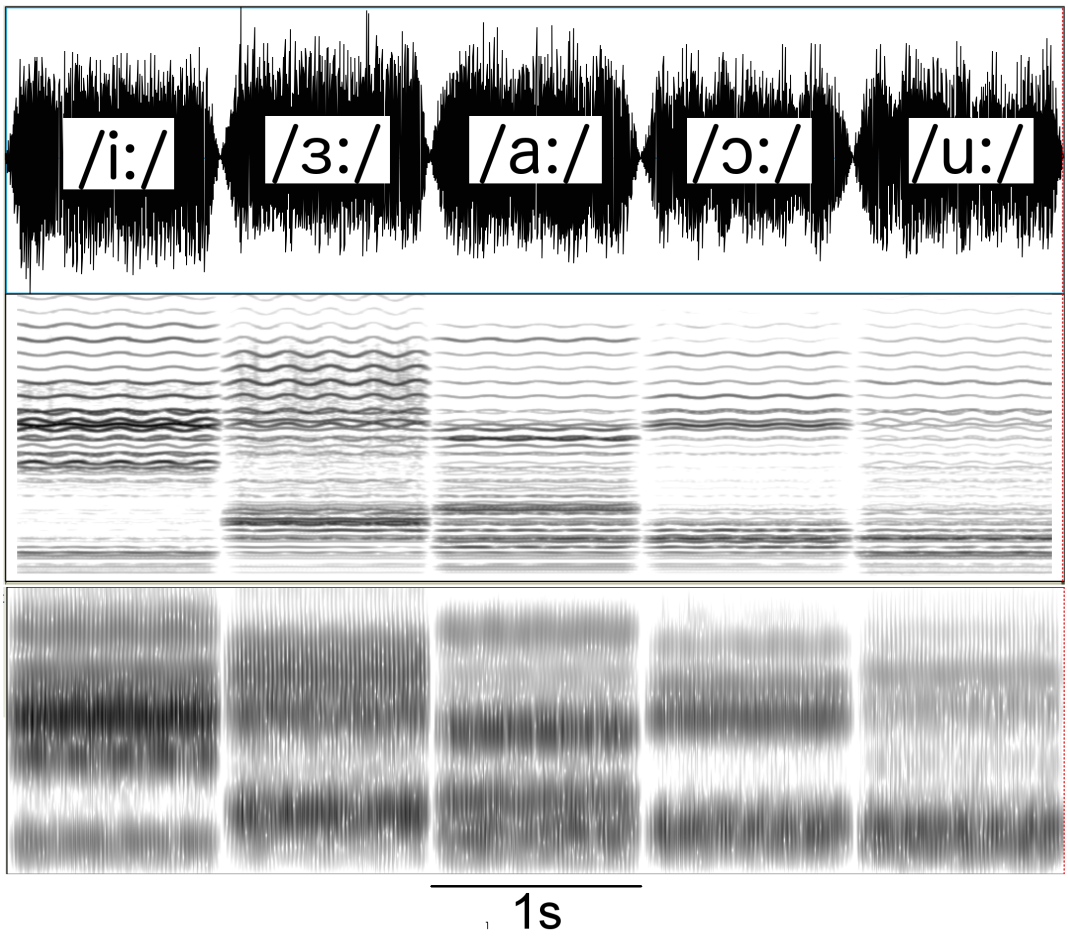

Figure 6: Time waveforms (upper), narrow-band 60 ms analysis window 0-5 kHz spectrograms (centre) and wide-band 3 ms analysis window 0-5 kHz spectrograms (lower) from Praat for the five 3-D printed vowel vocal tracts (/i:/, /ɜ:/, /a:/, /ɔ:/ and /u:/) excited with a G major chord in the Vocal Tract Organ in the Barbershop pitch range (G2, D3, G3, B3) sent to a single vocal tract loudspeaker driver for each vowel.

When it functions as an organ, the Vocal Tract Organ can be played polyphonically. To explore the acoustics of this, a 4-note G major chord in the male Barbershop pitch range (G2, D3, G3, B3) is considered with reference to figure 6 (a chord appropriate to a male barbershop pitch range is considered since the vowels are all for a male vocal tract). This excitation signal consists of the four L/F waves, one for each note of the chord, which have been summed together and sent to a single loudspeaker driver coupled to the appropriate vocal tract, invoking the principle of linearity as described above. The effect of the synthesis of different vibrato rates and depths for each note of the chord can be clearly seen in the spectrogram as phase differences in the vibrato traces on the harmonics, especially in the formant regions. It is worth noting that this difference can be most clearly observed at higher frequencies due to the use of a linear amplitude scale, because the overall vibrato extent is a function of the harmonic number. Harmonics that belong to the same note are evident where the vibrato remains in phase.

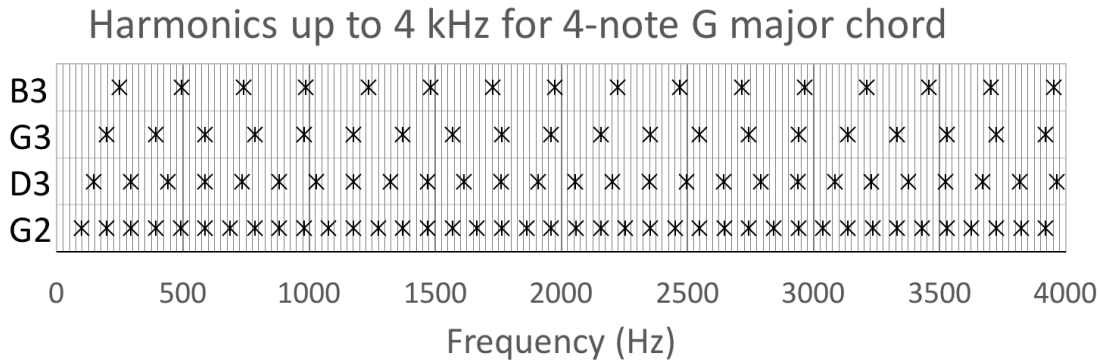

Figure 7: Frequency positions of the harmonics up to 4 kHz for the four notes of the G major chord (G2, D3, G3, B3) used as the excitation as excitation for the spectrograms shown in figure 4.

Since these harmonics originate from a four-note major chord (G2, D3, G3, B3), a number of coincident harmonics will be apparent between the four notes which are shown in figure 7, which indicates the frequencies of the harmonics for each of the four notes of the chord used. There are regions in the narrow band spectrogram in figure 6 where this can be observed directly due to the fact that the vibrato phases are different but the mid frequency of the harmonics are the same. This is particularly evident for example just below 3000 Hz (three fifths the way up the narrow band spectrogram plots), which corresponds to coincident harmonics of the lower three notes (the 30th of G2; the 20th of D3; and the 15th of G3) and the very close coincidence with a harmonic of the upper note (the 12th of B3).

3.2 The Vocal Tract Organ as a musical instrument

To demonstrate the musical possibilities of the Vocal Tract Organ, the author composed *Vocal Vision II* for a two channel Vocal Tract organ and two male singers. *Vocal Vision II* [33], a four-part barbershop-style vocalise, was conceived as a piece to demonstrate the naturalness of the acoustic output from the Vocal Tract Organ when heard alongside singers. The Vocal Tract Organ plays the 1st bass and 1st tenor parts and the two male singers sing the 2nd bass and 2nd tenor parts. Some of the chords are build up note by note enabling the output from the Vocal Tract Organ to be compared directly with the singer's notes. All four parts sing on the vowel /a:/. In performance, there is an expectation that the singers will aim to blend in with each other and the Vocal Tract Organ; of course, nothing reciprocal can be returned. A performance of *Vocal Vision II* can be seen and heard on YouTube [34].

The first performance was accompanying a soprano singing Puccini's *O mio babbino caro* (from *Gianni Schicchi*, 1918) originally for a flashmob after black tie dinner presentation. For this, the author created a new keyboard part in a chorale-style [35] because the orchestral reduction was musically unsuitable for the Vocal Tract Organ (see the last few minutes of YouTube [31]).

4. Further development

Next steps in the development of the Vocal Tract Organ include reviewing its possibilities for use as a musical instrument in its own right with composers writing music especially for it. In addition, the Vocal Tract Organ itself will be developed further as a musical instrument in a number of ways. Stops will be added as 3-D printed vocal tracts for women and children are

available to be added in the future. Different vowels stops will be added as they become available. In addition, it would be possible to add 3-D printed tracts for animals when available.

harmonic	*adult male*	*adult female*	*child*
1	male /a:/ 8'	female /a:/ 8'	child /a:/ 8'
2	male /a:/ 4'	female /a:/ 4'	child /a:/ 4'
3	male /a:/ 2 2/3'	female /a:/ 2 2/3'	child /a:/ 2 2/3'
4	male /a:/ 2'	female /a:/ 2'	child /a:/ 2'
5	male /a:/ 1 3/5'	female /a:/ 1 3/5'	child /a:/ 1 3/5'
6	male /a:/ 1 1/3'	female /a:/ 1 1/3'	child /a:/ 1 1/3'
7	male /a:/ 1 1/7'	female /a:/ 1 1/7'	child /a:/ 1 1/7'
8	male /a:/ 1'	female /a:/ 1'	child /a:/ 1'

Table 3: Hypothetical stop list for a large Vocal Tract Organ containing stops for the first 8 harmonics of the vowel /a:/ (for an explanation of stop footages and their relation to the harmonic series, see [32]).

Individual stops will therefore exist for different informants (men, women, children, animals) and each has the potential to provide various vowels. The pipe organ itself is an acoustic harmonic synthesiser with stops that put pipes on the harmonic series. This practice will be taken advantage of in the future enabling a huge array of stops potentially to be provided on the Vocal Tract Organ. For example, individual vowels from individual informants could be provided with stops on the harmonic series (denoted as footages on a pipe organ) resulting in a stop list such as that shown in table 3 for just one vowel across the first 8 harmonics. In practice, pipe organs do not cover all these harmonics for individual stop types, but until they have been explored acoustically with listeners in terms of their musical possibilities it would seem sensible to keep all options open. This could be repeated for a number of vowels, probably restricted to phonetically long vowels only, such as /ɜ:/, /a:/, /ɔ:/ and /u:/.

Extending the principle of linearity discussed above, providing these would be done within the Pd synthesis software and would not require any additional 3-D printed vocal tracts, making the implementation of multiple harmonic stops relatively trivial in practice. In terms of overall tuning, different temperaments could be made available (equal and just temperaments are available on the Arduino version of the Vocal Tract Organ). New temperaments could be readily explored providing a means for comparing the result of and trying out new choral tuning strategies, including micro tuning, for example for future choral musical offerings.

5. Conclusions

The Vocal Tract Organ creates the sound of the human voice through the use of 3-D printed vocal tracts whose shapes have been derived from magnetic resonance imaging. These are excited acoustically with a synthesised L/F glottal source model with a different but appropriate vibrato applied to each individual note of polyphonic chords. The acoustic model behind the Vocal Tract Organ invokes linearity due to the nature of the sound source and the 3-D printed tracts where non-linear effects found in the human voice cannot exist. This has enabled a single 3-D tract to be excited with multiple sound sources that are summed together, making polyphony and in the future, multiple stop footages simple to implement in practice.

The Vocal Tract Organ is a chamber instrument that implements a far more human-sounding vowel sound than the *Vox Humana* pipe organ reed stop that has been available on a number of large pipe organs for centuries. Perhaps one day, *Vox Humana* stops on large instruments will be replaced by or implemented as 3-D printed vocal tracts of human vowels. The Vocal Tract Organ has the potential to become a highly complex instrument in its own right. Of course, before such developments were set up as a functioning musical instrument, it would have to be ascertained how interesting and potentially musically useful the resulting sounds might be. After all, when humans sing together in choirs, they do not create a fixed additive harmonic series of vowels as their corporate output, but perhaps this would be in itself a tantalising synthesis with its own musical possibilities.

Acknowledgements

The author is indebted to the staff of the York Neuroimaging Centre for their help and professional attentiveness in acquiring the MRI images, to Pete Turner for help with the 3-D printing, to Matt Speed and Amelia Gully for advice on editing the MRI images, as well as to the singers who have taken part in the cited YouTube video performances: Esme Smith (mezzo soprano), Ben Lindley (tenor) and Bertrand Delvaux (bass).

References

1. Styger, T. and Keller, E., Formant synthesis, In: *Fundamentals of speech synthesis and speech recognition*, Keller, E. (Ed.), Chichester: John Wiley and Sons, 1994 pp. 109-128.
2. B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," J.Acoust. Soc. Amer. **100** (1) (1996) 537–554.
3. Echternach, M., Sundberg, J., Arndt, S., Marki, M., Schumacher, M., and Richter, B. Vocal tract in female registers--a dynamic real-time MRI study, J. Voice, 24, (2) (2010) 133-139.
4. Echternach, M., Burk, F., Burdumy, M., Traser, L., Richter, B. Morphometric Differences of Vocal Tract Articulators in Different Loudness Conditions in Singing. PLoS ONE 11 (4) (2016) e0153792. https://doi.org/10.1371/journal.pone.0153792
5. Delvaux B, Howard D.M. A New Method to Explore the Spectral Impact of the Piriform Fossae on the Singing Voice: Benchmarking Using MRI-Based 3D-Printed Vocal Tracts. PLoS ONE 9 (7) (2014): e102680. https://doi.org/10.1371/journal.pone.0102680
6. Vos, R.R., Murphy, D.T., Howard, D.M. and Daffern, H. Determining the Relevant Criteria for Three-dimensional Vocal Tract Characterization, J. Voice. (2017) https://doi.org/10.1016/j.jvoice.2017.04.001
7. Shiller, D.M., Ostry, D.J., and Gribble, P.L. Effects of gravitational load on jaw movements in speech, J. Neurosci. 19 (1999) 9073–9080.
8. Tiede, M.K., Masaki, S., Wakumoto, M. et al. Magnetometer observation of articulation in sitting and supine conditions. J. Acoust. Soc. Amer. 102 (1997) 3166.
9. Speed, M., Murphy, D.T., and Howard, D.M. Modeling the vocal tract transfer function using a 3D digital waveguide mesh, IEEE Trans. Aud. Sp. Lang. Proc. **22** (2) (2014) 453-464.
10. Fujita, S., and Honda, K. An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models, Acoust. sci. and tech. **26** (4) (2005) 353-357.
11. von Kempelen, W. *Le mechanisme de la pavola, suivi de la Description d'une machine parlante*, Vienna: Degen, J.V, 1791.

12. Sawada, H., and Hashimoto, S. Mechanical construction of a human vocal system for singing voice production, Adv. Robotics. **13** (7) (1998) 647–661.
13. Kitani, M and Sawada, H., Mechanical Reproduction of Human-Like Expressive Speech Using a Talking Robot, Proc. Int. Conf. Biomet. Kansei Eng. doi: 10.1109/ICBAKE.2013.45 (2013) 229-234.
14. Birkholz, D., Jackel, D., and Kröger, B.J. Construction and control of a three-dimensional vocal tract model. Proc. Int. Conf. Acoust. Spe. and Sig. Proc. 2006 Vol. 1. 873-876.
15. Howard, D.M. Raising public awareness of acoustic principles using voice and speech production, J. Acoust. Soc. Amer. **131**, (3) (2012) 2405-2412.
16. Howard, D. M. The Vocal Tract Organ and the Vox Humana organ stop, J. Mus. Tech. Ed., 7, (3), (2015) 265-277.
17. Audsley, G. A. *The art of organ building*, in two volumes, New York: Dover Publications inc. 1965.
18. Brackhane, F. and Trouvain, J. On the similarity of tones of the organ stop *Vox Humana* to human vowels, The Phonetician, **107-108** (2013) 7-20.
19. Howard, D. M. Developing the Vocal Tract Organ, J. Creat. Mus. Sys. 1 (1) (2016) Available at https://openmusiclibrary.org/article/986358/.
20. Fant, G. The acoustic theory of speech production, The Hague: Mouton, 1960.
21. Fant G, Liljencrants J, Lin QG. A four parameter model of glottal flow, STL-QPSR, **2** (3) (1985) 119-156.
22. Ternström, S., Friberg, A., and Sundberg, J. Synthesizing choir singing. J. of Voice, 1(4), (1988) 332-335.
23. Sundberg, J. The science of the singing voice, Dekalb: Illinois University Press, 1987.
24. Puckette, M.K. Pure Data: Another Integrated Computer Music Environment, Proceedings, Second Intercollege Computer Music Concerts, Tachikawa, (1997) 37-41. [Computer program] Version 0.47-1, retrieved http://www.puredata.info
25. Kawahara, H., Sakakibara, K., Banno, H., Morise, M., Toda, T., and Irino, T., Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation, Proc. An. Summit and Conf., (2015), 520-529
26. Howard, D.M., Daffern, H., and Brereton, J. Four-part choral synthesis system for investigating intonation in a cappella choral singing, Log. Phon. Voc. **38** (3) (2013) 135-142.
27. Speed, M., Murphy, D.T., and Howard, D.M. Three-dimensional digital waveguide mesh simulation of cylindrical vocal tract analogs, IEEE Trans. Aud. Sp. Lang. Proc, **21** (2) (2013) 449-455.
28. Abberton, E.R.M., Howard, D.M., and Fourcin, A.J. Laryngographic assessment of normal voice: A tutorial, Clin. Ling. Phon. **3** (3) (1989) 281-296.
29. Traser, L., Birkholz, P., Flügge, T.V., Kamberger, R., Burdumy, M., Richter, B., and Echternach, M. Relevance of the Implementation of Teeth in Three-Dimensional Vocal Tract Models, J. Sp. Hear. Lang. Res., 12, (2017) 1-15.
30. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Sean Ho, Gee, J.C. and Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 31(3) (2006) 1116-28. [Computer program] Version 3.6.0, retrieved 15 July 2017 from www.itksnap.org
31. Titze IR Nonlinear Source-Filter Coupling In Phonation: Theory. J. Acoust. Soc. Am. 123 (2008) 2733–2749.
32. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program] Version 6.0.14, retrieved 13 May 2017 from http://www.fon.hum.uva.nl/praat/
33. Howard, D.M. (2013). Vocal Vision II [Music score] http://www.davidmhoward.com/pdf/Music4Downloading/VocalVisionII-20June2013.pdf (last accessed 24th August 2017)
34. Howard, D.M. (2013). Vocal Vision II [Performance] https://www.youtube.com/watch?v=pUryWk-s9Ig (last accessed 24th August 2017)

35. Howard, D.M. (2013). Music score of Vocal Tract Organ arrangement of 'O mio babbino caro' by Puccini http://www.davidmhoward.com/pdf/Music4Downloading/OMioBabbinoCaro.pdf (last accessed 24th August 2017)
31. Howard, D.M. (2013). Vocal Tract Organ arrangement of 'O mio babbino caro' by Puccini [Performance] https://www.youtube.com/watch?v=svZNaiBIKMU&feature=autoshare (last accessed 24th August 2017)
32. Howard, D.M. and Angus, J.A.S. Acoustics and psychoacoustics, 5th Ed., London: Routledge, 2017.