Yulia Draginda

“CHANT” for the Vox Humana: formantic waveform synthesis

Introduction.

Vox Humana is one of the most mysterious organ stops. Invented in the early 1500s, it is considered to be the oldest reed stop in the organ building. Its name is literally translated from Latin as the “human voice”, that is why it could be logically associated with the sound of human singing. As wrote J. Goebl, “In one of the large organs Vox Humana was built in the Fernwerk[1]... and its sound immediately reminded me the tender high sopran voice” (Goebl 1967). However, very often Vox Humana was perceived in a completely different way. For example, very remarkable is the Burney's description of the Vox Humana in the St. Bavo church in Haarlem: “It does not at all resemble a human voice, though a very good stop of the kind... I must confess that, of all the stops I have yet heard which have been honoured by the appellation of Vox Humana, no one in the treble part has ever yet reminded me of anything human so much as of the cracked voice of an old woman of ninety, or in the lowest parts of Punch singing through a comb” (Wedgwood 2018). The characteristics cited above are so dissimilar, that it is hard to believe, that they were written about the same organ stop. My project thus aims to investigate the original sounds of Vox Humana, find the possible way to explain its peculiar perceptional properties and synthesize its sound which will clearly reflect them.

I. Construction of Vox Humana and the human vocal tract.

As the typical reed stop, the pipe of Vox Humana produces sound by the oscillation of a reed against a hollow tube (usually called a “shallot”) and has a similar construction to the stopped diapason flue pipe (s. Fig. 1).

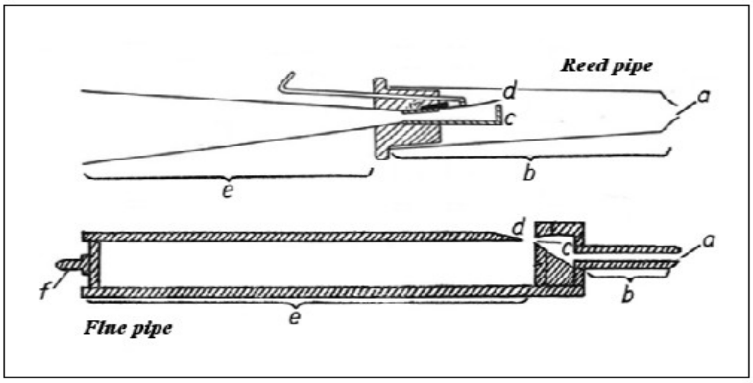


Fig. 1 Schematic drawing of a reed pipe and a flue pipe of the type stopped diapason. The air flows into the pipes (a) passing the socket or boot (b). The air in the reed pipe (top) will be excited by the reed tongue (d) that lies on the shallot (c). The excitation of the air in the flue pipe (bottom) is possible by increased air pressure at the windway (c) and the continuation towards the upper lip (d). The resonator (top e) and the body (bottom e) act as acoustic filters. (f) represents the cap needed for stopped flue pipes. (Brackhane and Trouvain 2013).

[1] Fernwerk – the Echo organ, when the pipes are enclosed, at a distance from the main case.

But, unlike other reed stops, the resonator of Vox Humana is relatively short (approx. 14 cm.) and, what is the most important, has a constant size independent of the pitch of the pipe (s. Fig.2). It is a unique feature of Vox Humana, because for almost every other organ stop consisting of reed pipes, the length of the resonator increases successively with the decreasing pitch of the pipes.

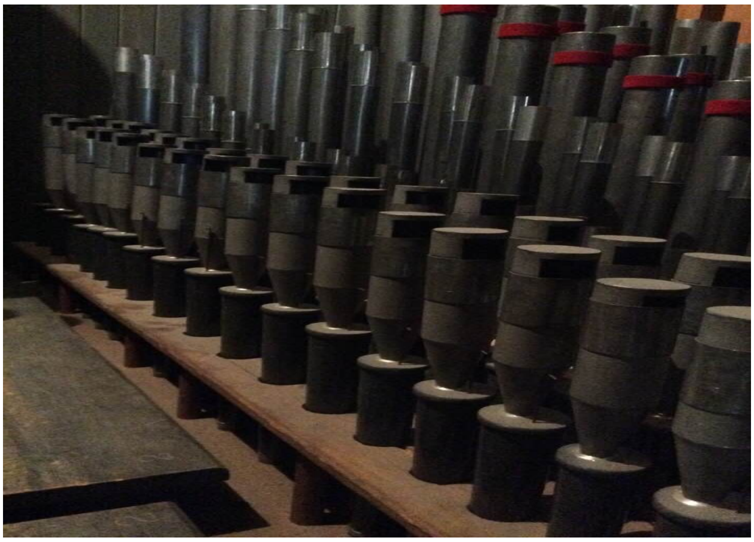


Fig.2 Vox Humana stop of St. Andrew&Paul church, Montreal, Canada

Because of this construction detail, as it was described in many studies (e.g., Howard 2015, Brackhane 2015, etc.), the resonators might act as a filter in such a way, that formants can be observed that are similar to those found in human vowels. This analogy is evident when we consider the human voice as a musical instrument combining a non-uniform tube resonator (the vocal tract for the human voice vs. the tube of the Vox Humana) with a forced excitation mechanism (the vocal folds for the human voice vs. the vibrating reed for the Vox Humana). A characteristic of voice is the formants: energy peaks in the spectrum that arise from the resonances of the source (s. Fig. 3).

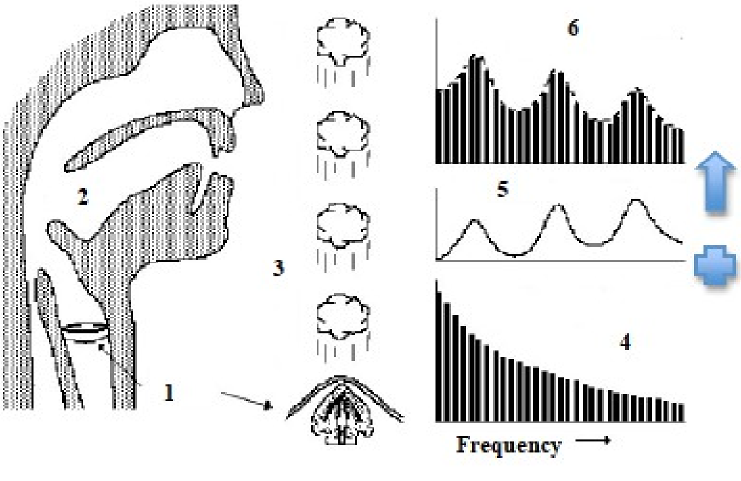


Fig 3. Filtering of the sound produced by the impulses of vocal folds (1). (2) – resonator (vocal tract), (3) – a sequence of pulses, (4) – harmonic spectrum of the sound produced by the pulsation of vocal folds, (5) – resonances of the vocal tract (formants), (6) – resulting spectrum. (Michon 2010)

Based on this similarity, I decided to apply the formant analysis and synthesis methods, which are widely used for a singing voice, to Vox Humana. This is how the idea of this project was born.

II. CHANT and formantic wave-functions synthesis

One of the easiest and elegant methods is the formantic wave-functions synthesis, which is commonly abbreviated as FOF, originated from its translation from French: “Fonction d'onde formantique” (Rodet 1984). The main idea behind this technique is the reconstruction of a sound by superimposing the contributions of each formant. In this case, the formantic wave-function is a “grain”, or elementary function, representing a fundamental period of a signal, which corresponds to each formant. This method makes possible to obtain the result similar to the traditional source-filter model, but requires a smaller number of calculations, as well as allows the intuitively understandable control over the parameters. The elementary function, in general, is a sinusoid, multiplied by an exponential envelope: $\mathrm{e}^{-\alpha t}\sin(\omega t+\varphi)$, it corresponds to the unitary impulse response to the second-order filter, and its parameters (amplitude, frequency, decay) correspond to the parameters of a formant (amplitude, central frequency, bandwidth).

Formantic wave-functions synthesis is the main principle of CHANT, which was originally the singing voice synthesizer, but was expanded to synthesize also other sounds (Rodet et al. 1984). CHANT is based on subtractive synthesis, where the source signal with a wide spectrum is passing through a complex filter. The main assumption here is, that the transfer function of the vocal tract can be modelled by simulating formant frequencies and formant amplitudes. The synthesis process, consisting of the artificial reconstruction of the formant characteristics, is done by exciting a set of resonators by a voicing source, which simulates the vocal fold vibration.

CHANT could be implemented in two different ways, as a filter bank, or as a set of the wave-functions. In the first interpretation, the resulting signal consists from the sum of elementary signals corresponding to the impulse response of the second-order resonance filters with the transfer function $H(z)=\sum_{i=1}^{J}\frac{1+d_i z^{-1}}{1+a_i z^{-1}+b_i z^{-2}}$, where the parameters a, b and d control the central frequency, the bandwidth and the slope respectively. *J* is the number of the FOF, equal to the *J* parallel second-order filter sections. Hence it is possible to divide the excitation signal *E(k)* to *J* formantic wave-functions, which correspond to the *J* spectral regions. And if the excitation is just a sequence of pulses, $E(k)=\sum_{-\infty}^{+\infty}p_n(k)$, where n is the pulse index, then the filter response *S* is the sum of partial responses $S_n(k)$, offset by the one period of the fundamental frequency F_0. But, the same response

$S_n(k)$ is also the sum of *J* partial responses of parallel sections $s(k)=\sum_{i=1}^{J} S_{n,i}(k)$, where *$S_{n,\,i}(k)$* are the formantic wave-functions corresponding to formants of the system.

An alternative implementation replaces the filters with a bank of damped sine wave generators and realizes a time-domain wave-functions synthesis. In the CHANT method, the formantic wave-functions contain also a modulation by an asymmetric Hanning window, what allows to model more precisely the attack of the waveform. They have the following structure:

$$s(k)=0, k<0; \qquad (1)$$

$$s(k)=1/2(1-\cos(\beta k))e^{-\alpha k}\sin(\omega k+\varphi), 0\leq k\leq \pi/\beta \qquad (2)$$

$$s(k)=e^{-\alpha k}\sin(\omega k+\varphi), \pi/\beta<k \qquad (3)$$,

where ω is the central frequency of the maximum (in Herz), α is the bandwidth (in Herz as well), and the parameter π/β, which corresponds to the attack duration and controls the width of the "skirts" of the formant peak.

. The amplitude envelope used in CHANT presents no first- or second-order discontinuity, and could be obtained by table lookup for *(1/2)(1 - cos[(βk])* and s*in(ωk),* and by successive multiplications by $e^{-\alpha}$ for $e^{-\alpha k}$. As it was shown in (Rodet 1984, or Spanier 1999), assuming that $\beta^2 >> \alpha^2$, the shape of the power spectrum of the FOF *s(k)* in the neighborhood of the center frequency $\omega_c/2\pi$ has the form $K/[\ \alpha^2+ (\omega_c-\omega)^2]$, where $K= ((e^{-\alpha\pi/\beta} +1)/2)$*^2.* It is independent of *β* and almost identical to the transfer function of a second-order filter section, the central pulsation of which is ω_c, and the bandwidth is α/π. Thus, the appropriate parameters provided, it is possible to synthesize the sound *S(k)* from the sum of the q partial FOFs (s. Fig. 4). For each formantic wave-function, the required for synthesis parameters are amplitude, bandwidth, central frequency and attack time.

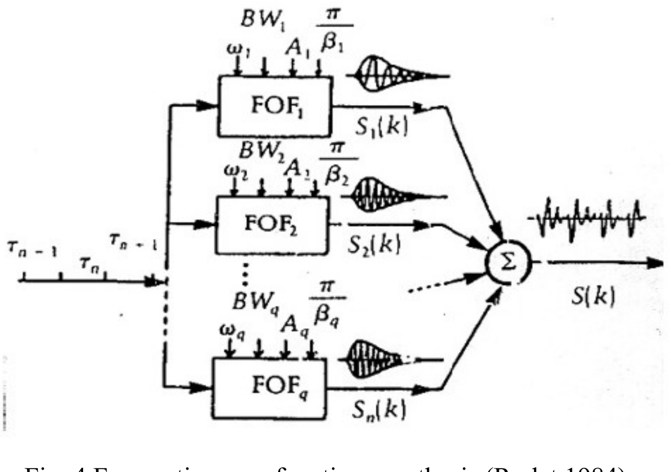


Fig. 4 Formantic wave-functions synthesis (Rodet 1984).

III. Implementation: analysis.

For my experiment, I made a recording of the pure Vox Humana tones, pure Prestant 4'[2] tones, as well as their combination with and without tremulant in different frequency ranges at the Wolff French classical organ in the Redpath Hall[3]. In order to get the parameters for synthesis, I implemented in MATLAB a simple formant analyzer, based on the linear prediction estimating the parameters of the filter (Snell 1993). The estimation of the formants relies on the fact that a pair of poles

$$\frac{1}{(1-r\,e^{j\theta}z^{-1})(1-r\,e^{-j\theta}z^{-1})}=\frac{1}{1-2\mathrm{r}\cos(\theta)z^{-1}+r^{2}z^{-2}}$$

has a maximum peak at angular frequency *θ,* i.e. at the frequency $F_s\,\theta/2\pi$, where F_s is a sampling rate. The bandwidth of a pair of poles in the LP filter depends on the distance of the pole from the origin, and is defined as $-F_s\frac{\ln r}{2\pi}$, what equals twice the formant bandwidth where the amplitude response of the pole reaches the level -3dB below its maximum value. The linear prediction coefficients were obtained with the built-in MATLAB function `a=lpc(y,ncoeff),` where *y* is the previously pre-emphasized and windowed signal, and *ncoeff* is a filter order, which could be set as twice the number of formants plus two. The roots of the prediction polynomial are returned by the function `r=roots(a).` The amplitudes values were obtained by the peak-picking for the peaks of the spectral envelope. I have also set the (relatively weak) additional constraints for bandwidth to be less than an experimentally defined threshold value, in order to avoid the inappropriate peaks. This criteria, as well as the number of coefficients and the analysis range, must be adjusted according to the analyzed sound (are marked with the %check&adjust comment, s. Appendix 1). My first analyzed sound was the note A, corresponding to the 415 Vox Humana tone. I compared the output of my analyzer with the formant analysis in Praat for the respective signal frame. The results are represented inTable 1.

Frequency, Hz	Praat frequency, Hz	Bandwidth, Hz	Praat bandwidth, Hz	Amplitudes, dB
776.7	855	134,5	196	-40.2
2244.1	2266	253.5	239	-43.6
3245.5	3263	227.9	253	-44.4
4325.4	4339	289.9	370	-48.9
6105.6	6009	697.6	2372.94	-55.7
7573.4	7703	141.4	1835	-47.2

Table 1. Analysis results for the A sound.

The preliminary number of formants were defined as 6 (s. Fig. 5). The best correlation of my results

[2] The 4' sign means, that the tones speak one octave higher, than for the 8' stop: e.g., A for the concert pitch tuning will be not 440 Hz, but 880 Hz, etc.

[3] The organ description and the stop list: http://www.musiqueorguequebec.ca/orgues/quebec/redpath.html

with Praat was for the middle formants, which were stable and easy to define. The least correlation was for the first formant because of the probable influence of the DC component, as well as for the highest formants, which were rapidly changing (s. Fig. 6).

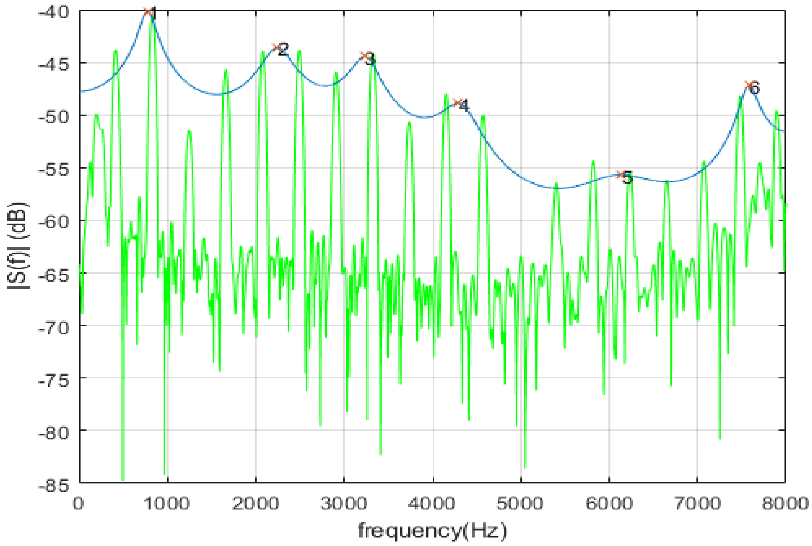


Fig. 5 Frame LP envelope and magnitude spectrum for the sound A (415 Hz)

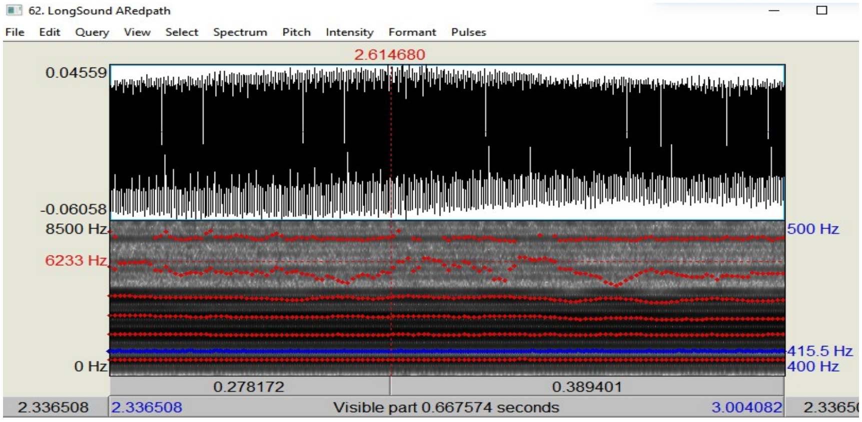


Fig. 6 Formant analysis in Praat for the sound A (415 Hz)

Thus I decided to use for the synthesis only first four formants which seem to be defined properly in both systems, and also correspond to the most meaningful spectral energy regions (s. Fig.7).

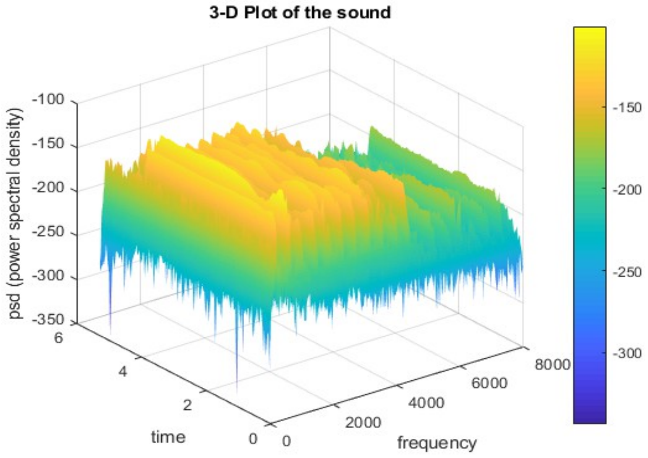


Fig. 7. Power spectral density for the sound A (415 Hz)

In order to test my analysis for the low frequencies range, I analyzed the sound C0 (124 Hz). I increased the prediction filter order to 18, and 7 formants passed the preliminary bandwidth criteria (s. Fig.8). The comparison of my results and the Praat analysis are represented in Table 2.

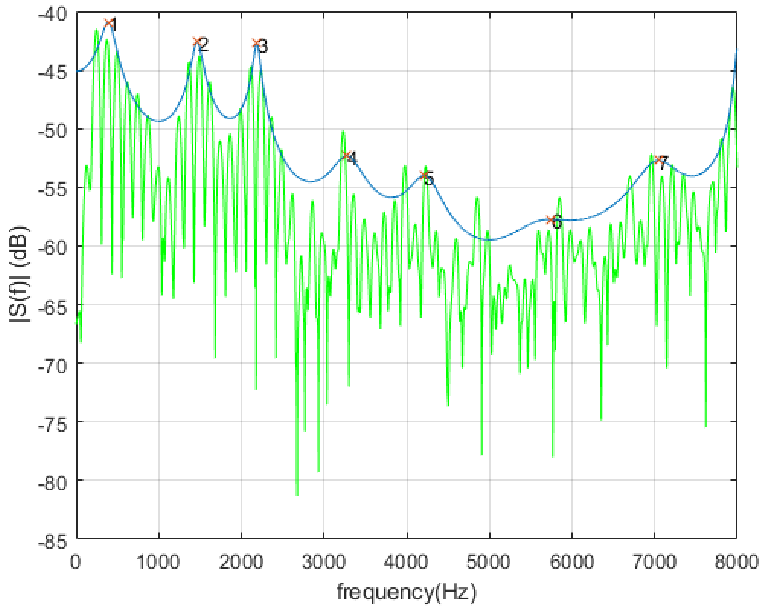
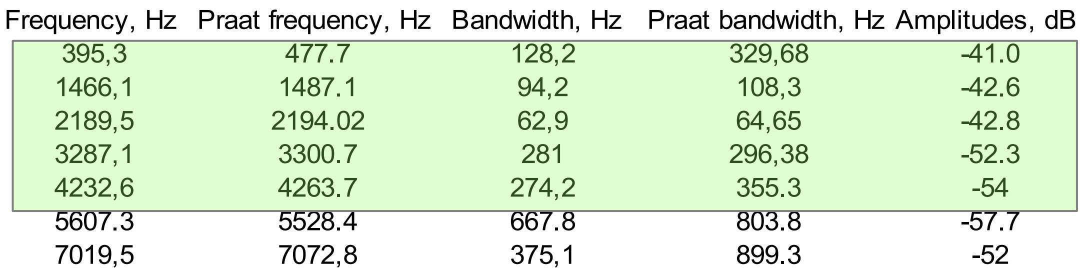


Fig. 8 Frame LP envelope and magnitude spectrum for the sound C0 (124 Hz)



Frequency, Hz	Praat frequency, Hz	Bandwidth, Hz	Praat bandwidth, Hz	Amplitudes, dB
395,3	477.7	128,2	329,68	-41.0
1466,1	1487.1	94,2	108,3	-42.6
2189,5	2194.02	62,9	64,65	-42.8
3287,1	3300.7	281	296,38	-52.3
4232,6	4263.7	274,2	355.3	-54
5607.3	5528.4	667.8	803.8	-57.7
7019,5	7072,8	375,1	899.3	-52

Table 2. Analysis results for the C0 sound.

As in the previous case, I decided to take for the synthesis the first five formants, which could be definitely distinguished, and correspond to the most prominent regions in the psd plot (s. Fig. 9).

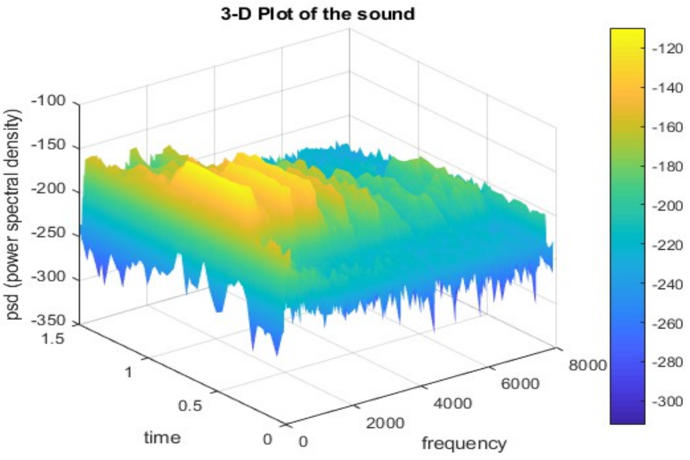


Fig. 9. Power spectral density for the sound C0 (124 Hz)

It is remarkable, that even despite the low fundamental frequency, a low degree polynomial is still able to model all the peaks in the spectrum, what in general could be problematic (Koppinen 2006).

For the high-frequency range test, I chose to analyze the tone c3 (this choice was also predefined by the perceptional data, which I will describe in the next section). The analysis results for this sound (992 Hz) were slightly different from the previous ones: instead of modelling the formants, the poles were moved towards the harmonics of the fundamental frequency, and the LP spectrum caught the harmonic peaks (s. Fig. 10). Thus, strictly speaking, this method is not appropriate for the high frequency range, but it was still possible to synthesize a convenient sound, based on the first six detected frequencies (993.9 Hz, 1978 Hz, 2981.7 Hz, 3974.7 Hz, 4973.8 Hz and 5980.7 Hz), because in fact these values are close to formants. This sound is discussed in the next section (Table 3).

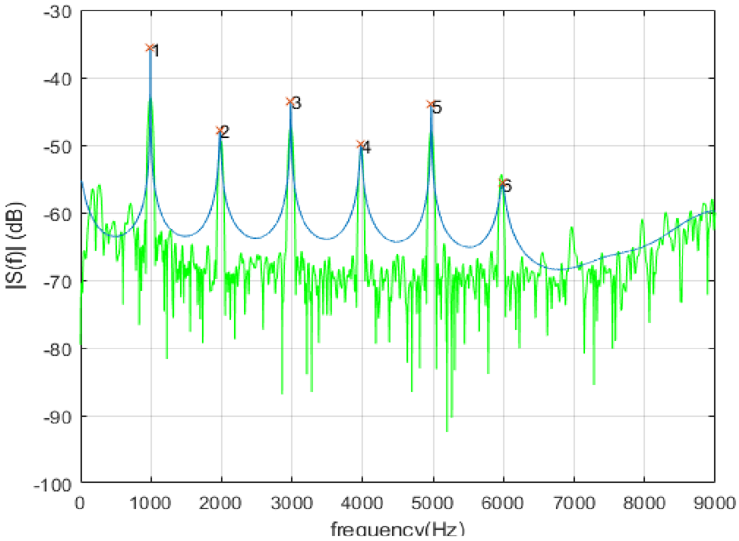


Fig. 10 Frame LP envelope and magnitude spectrum for the sound C3 (992 Hz)

III. Implementation: synthesis.

Formantic wave-function synthesis was originally implemented in the program CHANT (Rodet 1984), then was also available as the additional library for the OpenMusic (Michon 2010). Unfortunately, this *chant-lib* library seems not to work anymore. But the very similar algorithm is easy to build in CSOUND, where the special *fof* object can be used for generation of the wave-functions. The amplitude envelope for each pulse is generated by the GEN19, which can build the waveform from sinusoids. The envelope used in CSOUND has the following structure:

$x(t)=\frac{\sin(t+\varphi)+b}{a}$, where φ =3π/2 in the phase, and the parameters b=1 and a=2 are used to keep the *x(t)* between 0 and 1: $0\leq x(t)\leq 1$. This *x(t)* is used first for the attack, and then is read inversely to produce the decay. This envelope has almost the same structure, as the envelope proposed in

CHANT, with the only difference, that the duration of the decay in CSOUND is slightly shorter. Fig.11 shows the MATLAB plot of the elementary FOF, calculated according to the formulae in CHANT, and the test wave-function generated in CSOUND with the GEN19 object as described above.

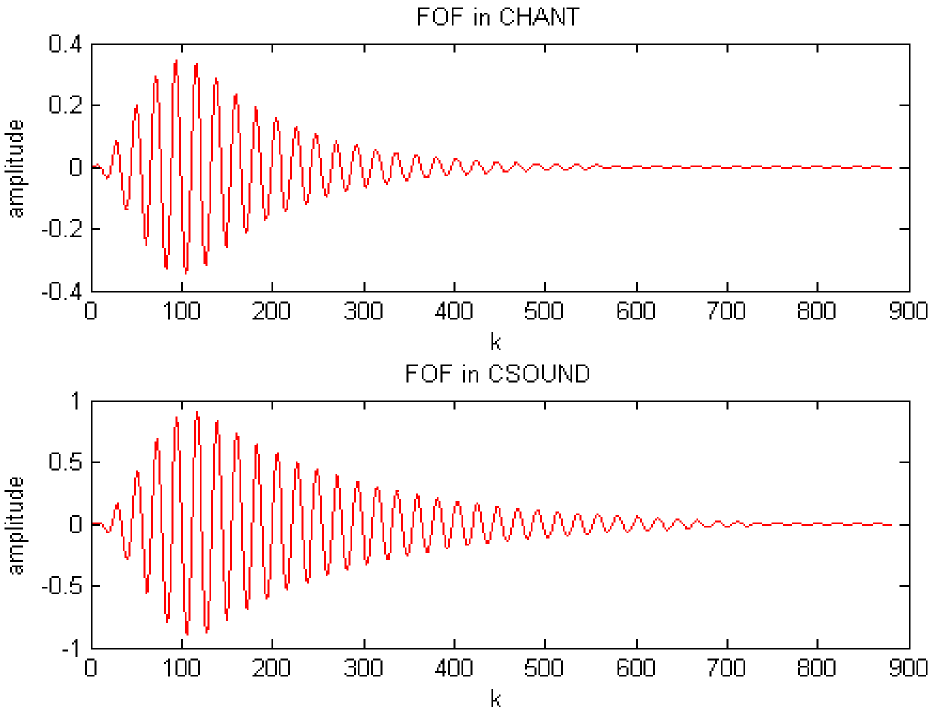


Fig. 11. Comparison of the FOF proposed in CHANT, and FOF used in CSOUND.

Hence, provided the input parameters, it is possible to synthesize the resulting signal *s(k)* from by summation of such elementary FOF "grains" in SCOUND.

Based on my analysis formant parameters, I synthesized the A, C0 and c3 tones for the Vox Humana (the example SCOUND score for the tone A is in Appendix 2). The waveform of the synthetic A sound is represented at the Fig. 12. It is important to notice, that all resulting sounds have the easily recognizable "reed" timbre very close to the original sound.

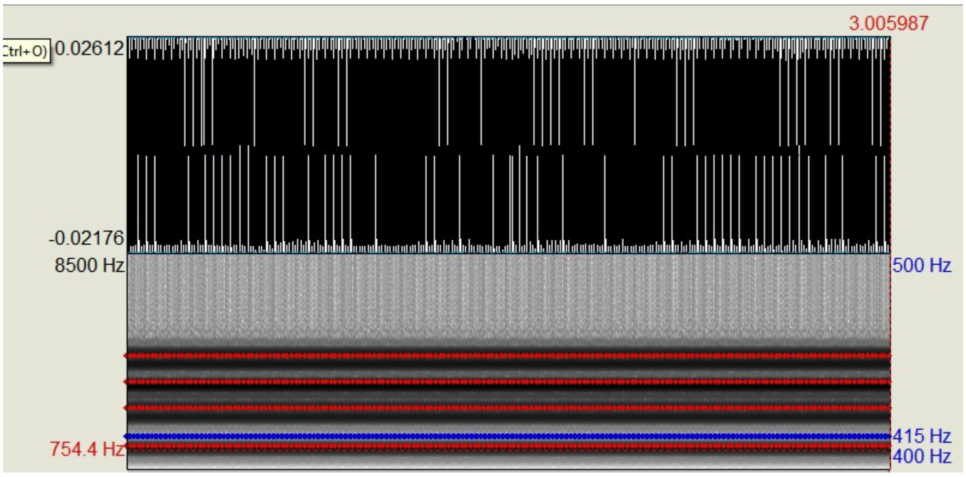
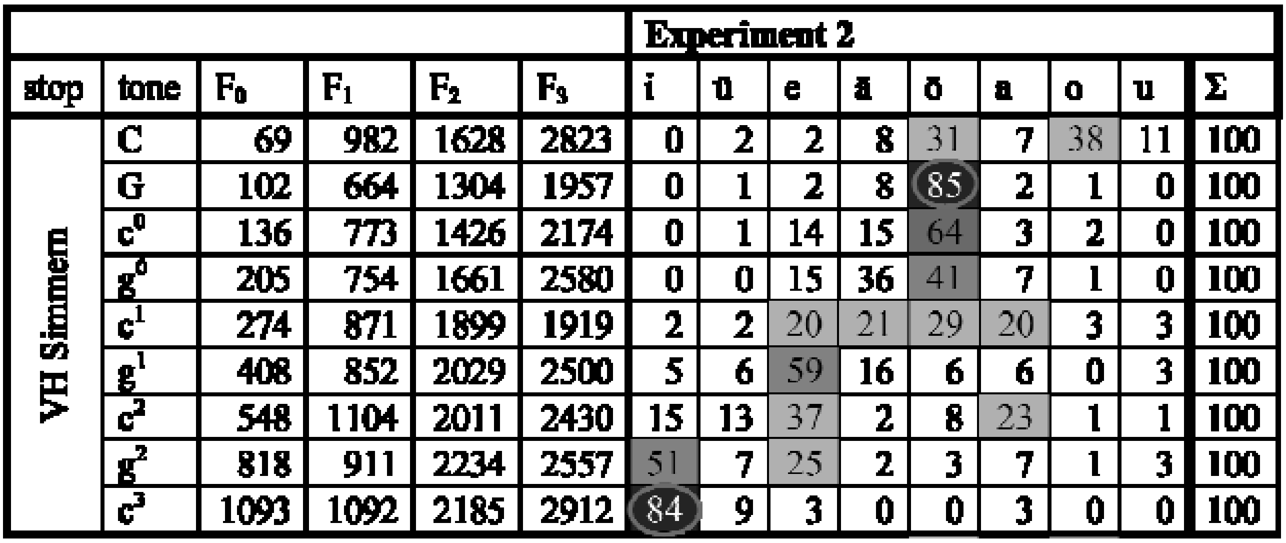


Fig. 12 Formant analysis in Praat for the synthesized sound A (415 Hz)

Having thus this set of original and synthesized sounds, I made an additional research in order to investigate, how they are correlated with the human voice, and why they could be perceived so differently. A large number of appropriate perception experiments are described in (Brackhane 2015), and the short summary could be found in (Brackhane and Trouvain 2013). As stated in their experiments, the lowest correlation was in the middle frequency range, and the highest for the high frequency range (s. Fig. 13, black and dark-grey areas). [4]



						Experiment 2								
stop	tone	F_0	F_1	F_2	F_3	i	ü	e	ä	ö	a	o	u	Σ
VH Simmern	C	69	982	1628	2823	0	2	2	8	31	7	38	11	100
	G	102	664	1304	1957	0	1	2	8	85	2	1	0	100
	c^0	136	773	1426	2174	0	1	14	15	64	3	2	0	100
	g^0	205	754	1661	2580	0	0	15	36	41	7	1	0	100
	c^1	274	871	1899	1919	2	2	20	21	29	20	3	3	100
	g^1	408	852	2029	2500	5	6	59	16	6	6	0	3	100
	c^2	548	1104	2011	2430	15	13	37	2	8	23	1	1	100
	g^2	818	911	2234	2557	51	7	25	2	3	7	1	3	100
	c^3	1093	1092	2185	2912	84	9	3	0	0	3	0	0	100

Fig. 13 Percentage of answers for Vox Humana and German vowels (Brackhane and Trouvain 2013).

For this study, it was necessary to slightly modify the pure synthesized sounds according to the basic organ performance rules (so that we could obtain the sound, which could be heard in the church or the concert hall). In fact, Vox Humana must be used in combination with strong or soft tremulant, and with the 4' or 8' flute stop. In order to obtain these mix, I merged my synthesized sounds with the recorded sounds of the Prestant 4' for the respective pitch and the *tremblant doux* (soft tremulant), as well as added a reverberation effect. These sounds (named in the sound folder as AsynthPrTr.wav and c3synthPrTr.wav) could be compared against the original sounds combinations (AredpathPrTr.wav and c3RedpathPrTr.wav respectively).

If we look closely at the Fig.13, we can see, that the middle A could probably correspond to the *e* vowel, and the high c3 might have a good similarity to the *i.* Hence, for this analysis, I recorded the human (*my own*) voice trying to produce (*sing*) these vowels at the respective pitches, then analyzed them with the same algorithm as described in the previous section, and synthesized with the obtained parameters in CSOUND. I examined the original and synthesized sound combinations against the original and synthesized human voice.

The results of my study are quite impressive. If we look at the sound A, we see, that there is no

[4] These experiments were made with the recordings of Vox Humana at the Stumm organ in Stephanskirche Simmern, Germany, which is tuned in a choir pitch (A=462 Hz), http://organindex.de/index.php?title=Simmern,_Stephanskirche , while the organ in Redpath Hall has a chamber pitch tuning (A=415 Hz). This explains the difference in the fundamental frequencies for the respective tones.

correlation between the formant regions for the Vox Humana sound and singing voice (Fig. 14). It is even more clear, if to look to the synthetic sounds (Fig. 15). That was more than expected, because the sounds have a really different timbre, and as shown in (Brackhane and Trouvain 2013), the biggest percentage of the answers in the middle frequency range was 59%, which is still below the statistical significance. If we add the required combination to the pure tone, it distorts the relatively stable formantic structure for the single tone, and tries to move it towards the human voice formants; however, in this case it does not succeed.

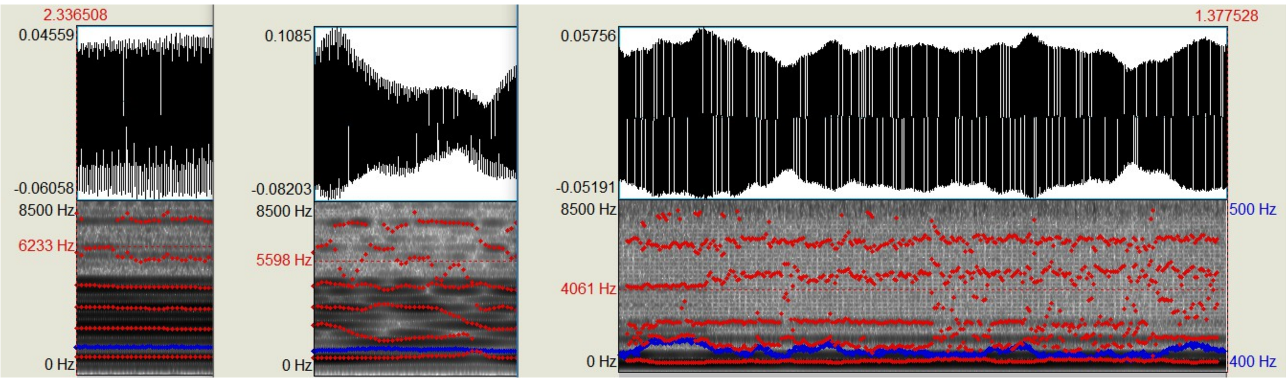


Fig. 14. Original sounds A, combination and voice. The sequence shows (left to right order) the original A sound, sound A with Prestant 4' and Tremolo, human voice singing the vowel *e*.

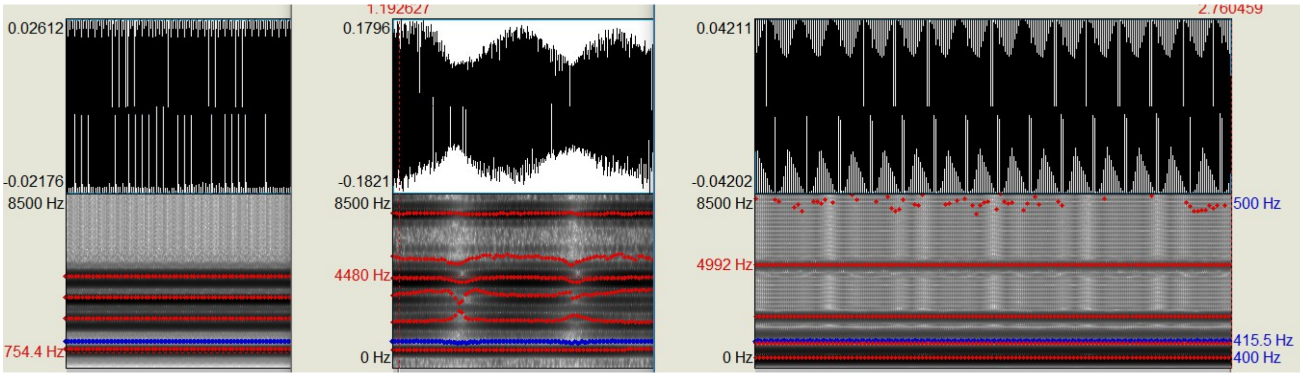


Fig. 15. Synthesized sounds A, combination and voice. The sequence shows (left to right order) the synthesized A sound, sound A with Prestant 4' and Tremolo, synthesized human voice singing the vowel *e*.

The sound c3 is a different case: for this sound, the formantic structure is almost similar to the human voice (s. Fig. 16).

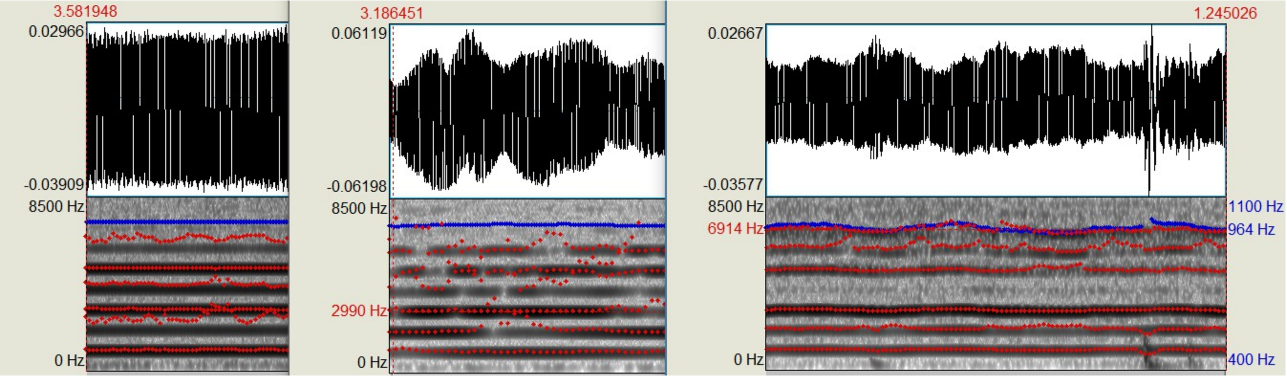


Fig. 16. Original sounds c3, combination and voice. The sequence shows (left to right order) the original c3 sound, sound c3 with Prestant 4' and Tremolo, human voice singing the vowel *i*.

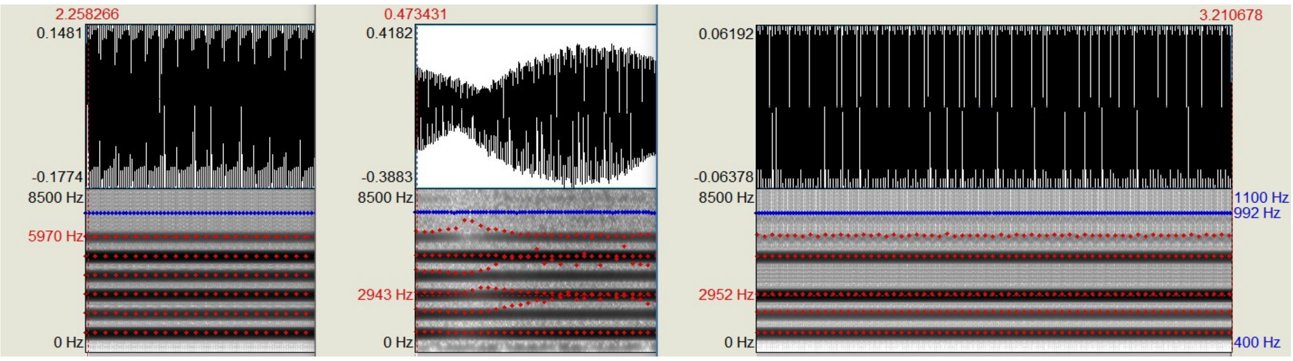


Fig. 17. Synthesized sounds c3, combination and voice. The sequence shows (left to right order) the synthesized c3 sound, sound c3 with Prestant 4' and Tremolo, synthesized human voice singing the vowel *i*.

The resulting comparison of formant frequencies given by my MATLAB analysis, Praat, as well as the data for the vowel *i* from the CSOUND manual (Vercoe 2005) is represented in Table 3.

c3 (MATLAB)	C3 (Praat)	Vowel i (MATLAB)	Vowel i (Praat)	Soprano i (CSOUND manual)
993,9	994,04	983.2	953.37	270
1978	1987,58	1972.3	2046.22	2140
2981,7	2983,39	2953.0	2886.17	2950
3974,7	3978,81	x	x	3900
4973,8	4973,7	4918.1	4873.09	4950
5980,7	5966,23	5949.2	5943.48	x

Table 3: Formant frequencies (in Hz) for the pure tone Vox Humana c3 and vowel i.

Except the first formant, there is some correlation between the formant frequencies for this data. The difference in the first formant could be understood, if to consider that the data provided in CSOUND was given for the lower fundamental frequency. As shown in (Joliveau et al. 2004), for the high pitch the first resonance is approximately equal to the fundamental frequency, which in my case in 992 Hz. The presence of the 4th formant at 3900 Hz in CSOUND data is less clear; I can only suppose, that it is because the "formantic gap", which occurs for the low pitch between the first and second formant, moves higher when the pitch increases, and thus could happen between 2953 and 4918 Hz in my case.

From the Fig.16-17 we also see, that adding the combination helps to fix the "walking" second formant near the second harmonic (the Prestant 4' does his job), as well as distorts the fourth formant, which is missing for the human voice. The combination sound is really close to the voice, what just confirms the perceptional results from (Brackhane and Trouvain 2013), where the similarity between c3 and the vowel *i* was noticed by 84% responders.

As a by-product of this research, we can also see, why the adding the 4' stop is in this case better than the 8' stop (what is recommended by some dictionaries): the former will amplify the second harmonic, while the latter will mostly contribute to the fundamental frequency. The disadvantage of it, however, is that they are usually out of tune in the winter season, and if this is the case, the 4' stop still should be avoided.

To sum up my experimental results, it is evident, that the Vox Humana, in fact, has very different sound quality in terms of correlation to the real human voice depending on the frequency range. It could be now explained by looking at the history of this stop. It was originated in Italy, during the High Renaissance era at the beginning of the 16th century. It is not widely known, that its Italian prototype, *Voce Humana,* was a totally different stop: it consisted from labial pipes and was the so-called *Schwebung*[5], specified by its soft shimmering, which occurs when two pipes are tuned slightly off pitch from each other and speak together. It was one of the most beautiful organ stops in the organ history, which was a reflection of humanistic principles of the Renaissance. But after being imported later to Germany, it was significantly transformed. According to the German Lutheran church tradition of this time, only God was almighty, and everything on the Earth was imperfect (human included). The Vox Humana might be a reflection of this idea: we could see, how the beautiful original Italian *Voce Humana* was re-branded at the end of the 16th century into the Vox Celestis ("Heaven Voice"); the Vox Humana got its present Regal-like construction and moved to the reed stops family. This construction allows it to produce the sounds close to the human voice in the high frequency range, which could be an attempt to model the boy's or angel's choir. That explains also its specification mentioned in the first section of my project, namely its relatively short length of the resonator: as suggested in (Howard 2015), it might correspond to the length of the child's vocal tract (approx. 14 cm.). In some rare organs, there is also an unusual stop called Vox Angelica ("Angel's voice"), which in fact is nothing else than the 4' Vox Humana. The empiric performance rule prescribing the usage of tremulant might be a reminiscence of the original Schwebung's shimmering; the analysis results show, that it helps to make the sound closer to the singing voice. And in the middle frequency range, its ridiculous, cracked sounding might be seen as the reminder about the frailty of the earthy life and the human imperfection.

IV. Conclusion

The analysis-synthesis proposed here is not the most accurate scheme. The more relevant way would be staying consistently in the one approach: either to use the filter-bank analysis, and then to synthesize sound by the inverse filtering of the prediction error; or first decompose the original sound to the sequence of the wave-functions (e.g., as described in D'Alessandro and Rodet 1989), and then use the appropriate parameters for the FOF synthesis: both ways are relatively complex. The disadvantage of the "mismatch" in my process is that in my analysis it is not possible to find some additional parameters required for synthesis. For example, the attack parameter β could not be defined, that is why it was set to the standard value for the all FOFs. The other problem was, that CSOUND used a slightly different amplitude envelope, as it was proposed in CHANT, but it did not affect significantly the general structure of the wave-functions.

[5] The full description of Schwebung stops: http://www.organstops.org/c/celeste.html

But then my approach, based on the accordance of the formant and FOF basic parameters, led to the more simple and intuitively understandable implementation. It still succeeded to synthesize the different sounds with the rich spectrum, had low computational costs, and provided a good illustration for all important features of the Vox Humana.

Bibliography

Adelung, W. 1982. “*Einführung in den Orgelbau.* “ Wiesbaden: Breitkopf.

Brackhane, Fabian. 2015. *“Kann was natürlicher, als Vox humana, klingen? Ein Beitrag zur Geschichte der mechanischen Sprachsynthese.”* Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie der Philosophischen Fakultäten der Universität des Saarlandes. Published in *PHONUS: Berichte zur Phonetik, Universität des Saarlandes*: 18(2015).

Brackhane, Fabian, and Jurgen Trouvain. 2013. “On the similarity of tones of the organ stop vox humana to human vowels.” *The Phonetician*: 107 (7).

Bohn, Tamaz, and Geza Nemeth. 2007. “Algorithm for formant tracking, modification and synthesis.” *Hiradastechnika*: LXII (1).

Clarke, Michael. 2000.“*FOF and FOG synthesis in Csound*” in “The Csound book: Perspectives in software synthesis, sound design, signal processing and programming.”MIT Press, 293–306.

D'Alessandro, Christophe, and Xavier Rodet. 1989. “Synthèse et analyse-synthèse par fonctions d'ondes formantiques.” *Journal Acoustique*: 2, 163-168.

Goebl, Joseph. 1967. *“Theorie und Praxis des Orgelpfeifenklanges: Intonieren und Stimmen: ein Handbuch für Orgelbauer und Organisten.”* Frankfurt a.M : Verlag Das Musikinstrument, 59.

Hillenbrand, James, and Robert Houde. 2002. “Speech Synthesis Using Damped Sinusoids.” *Journal of Speech, Language, and Hearing Research*: 45

Howard, D.M. (2015). “The Vocal Tract Organ and the Vox Humana organ stop.” *The Journal of Music, Technology and Education:*7, (3), 265-277.

Joliveau, Elodie, Smith, John and Joe Wolfe. 2004. “Vocal tract resonances in singing: The soprano voice.” *The Journal of the Acoustical Society of America:* 116(4), 2434–2439.

Koppinen, Konsta. 2006. Application of Linear Prediction. Published online, http://www.cs.tut.fi/kurssit/SGN-4010/LPsovellus_2004_en.pdf, last accessed 19.12.2018.

Michon, Romain. 2010. “*La Synthèse de la Voix Chantee par Fonctions d'Ondes Formantiques – Techniques, Outils Existants, Exemple d'implementation et Utilisation.*” Master Thesis, Université Jean Monnet, Saint-Etienne, France.

Rodet, Xavier. 1984. “Time-Domain Formant-Wave-Function Synthesis.”*Computer Music Journal*: 8(3), 9-14.

Rodet, Xavier, Potard, Yves, and Jean-Baptiste Barrière. 1984. “The CHANT Project: From the

Synthesis of the Singing Voice to Synthesis in General." *Computer Music Journal*: 8(3), 15-31.

Schwarz, Diemo, and Xavier Rodet. 1999. "Spectral envelope estimation, representation, and morphing for sound analysis, transformation, and synthesis.".ICMC: *International Computer Music Conference*, Oct 1999, Pekin, China.

Shneider, Thekla. 1958. "*Die Namen der Orgelregister. Kompendium aller Registerbezeichnungen aus alter und neuer Zeit mit Hinweisen auf die Entstehung der Namen und ihre Bedeutung.*" Bärenreiter: Kassel-Basel-London.

Snell, Roy C., and Fausto Milinazzo. 1993. "Formant location from LPC analysis data." *IEEE Transactions on Speech and Audio Processing:* 1(2), 129-134.

Spanier, Jonathan Robert. 1999. "Algorithms and VLSI architectures for parametric additive synthesis." Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/4536/, last accessed 01.12.2018.

Sundberg, Johan. 2002. "The KTH Synthesis of Singing." *Advances in Cognitive Psychology*: 2(2-3), 131-143.

Vercoe, Barry. 2005. "*FOF, The canonical CSOUND Reference Manual.*" Cambridge: MIT, 749.

Wedgwood, James. 2018. *"A comprehensive dictionary of organ stops."* Franklin Classics Trade Press. https://archive.org/stream/cu31924022450831/cu31924022450831_djvu.txt, last accessed 20.12.2018.

Software manuals and help articles:

CSOUND MATLAB
http://www.csounds.com/resources/documentation/

MATLAB
https://www.mathworks.com/help/matlab/

Praat
http://wstyler.ucsd.edu/praat//

Appendix 1. MATLAB (R2017b) formant analyzer.

```
close all;
clear all;
[x, fs] = audioread('ARedpath.wav');

x=resample(x,16000,fs); %resample
fs=16000; %check&adjust
ts=(0:length(x)-1)/fs; % times of sampling instants
figure(1);
plot(ts,x); % plot waveform
legend('Waveform');
xlabel('Time (s)');
ylabel('Amplitude');

%take a frame (compare to praat)
start1 = 2.6;
finish1 = 2.63;
start = start1*fs;
finish = finish1*fs;

% FFT
bins=16384;% Number of FFT bins
y = x(start:finish);
y1 = y.*hamming(length(y));
preemph = [1 0.96];
y2 = filter(1,preemph,y1);
z = fft(y2,bins);
z=abs(z(1:bins/2));
ff=(0:fs/bins:fs/2-fs/bins)';
figure(2);
plot(ff, pow2db((z/bins)),'g'); hold on; grid on;

% get Linear prediction filter
ncoeff = 14; %adjust
[a,g]=lpc(y2,ncoeff);
lspec = freqz(g,a,ff,fs);
amp = pow2db(abs(lspec));
figure (2);
[pks,locs] = findpeaks(amp);
ff_peaks = ff(locs);
text(ff_peaks+0.5,pks,num2str((1:numel(pks))'));
plot(ff,amp,  ff_peaks,pks,'x');
xlabel('frequency(Hz)');
ylabel('|S(f)| (dB)');


% find formant frequencies by root-solving
r=roots(a); % find roots of polynomial a
r=r(imag(r)>0.01);
angz = atan2(imag(r),real(r));
[ffreq,i]=sort(atan2(imag(r),real(r))*fs/(2*pi)); % convert to Hz and sort
bw = (-1/2*(fs/(2*pi))*log(abs(r(i))))*pi; % find bandwidth*pi

% determine the formants after the bandwidth and amplitudes restriction
nn = 1;
for kk=1:length(ffreq)
    if  bw(kk)<800 %check&adjust
        formants(nn) = ffreq(kk);
        alpha(nn)=bw(kk);
        nn = nn+1;
    end
end
```

```matlab
nnn=1;
for kkk=1:length(pks)
    if  pks(kkk)>-60 %check&adjust
        ampls(nnn)=pks(kkk);
        nnn = nnn+1;
    end
end

% print results
for i=1:length(formants)
    fprintf('Formant %d Frequency %.1f\n',i,formants(i));
    fprintf('Formant %d Alpha %.1f\n',i,alpha(i));
    fprintf('Formant %d Amplitude %.1f\n',i,ampls(i));
end


%plot 3D spectra
figure
N = floor(.05*fs);
freq = 0:10:8000;
freq = freq';
[S,F,T,P] = spectrogram(x,hamming(N),floor(N/9),freq,fs);
[x,y]=meshgrid(-2:1:2,-2:1:2);
surf(freq,T,20*log10(abs(P')));
h = surf(freq,T,20*log10(abs(P')));
set(h,'edgecolor','none')
title('3-D Plot of the sound')
xlabel('frequency')
ylabel('time')
zlabel('psd (power spectral density)')
colorbar
shading interp
```

Appendix 2. CSOUND synthesis score for the tone A (415 Hz)

```
<CsoundSynthesizer>
<CsOptions>
-odac -d
</CsOptions>
<CsInstruments>
/* fof.orc */
sr = 44100
kr = 4410
ksmps = 10
nchnls = 1

; Instrument #1.
instr 1
kfund init 415 ; fundamental frequency
koct init 0 ; no octaviation
kris init 0.003 ;attack duration (pi/beta)
kdur init 0.02 ;total pulse duration
kdec init 0.007 ; decay duration
iolaps = 14850 ;reserve memory
ifna = 1 ;use GEN10 for sinusoids
ifnb = 2 ;use GEN19 for envelope
itotdur = p3 ;total sound duration, defined in score by p3
; First formant.
k1amp = ampdb(-40.2)
k1form init 776.7
k1band init 134.5
; Second formant.
k2amp = ampdb(-43.6)
k2form init 2244.1
k2band init 253.5
; Third formant.
k3amp =ampdb(-44.4)
k3form init 3245.5
k3band init 227.9
; Fourth formant.
k4amp = ampdb(-48.9)
k4form init 4325.4
k4band init 289.9

a1 fof k1amp, kfund, k1form, koct, k1band, kris, \
kdur, kdec, iolaps, ifna, ifnb, itotdur
a2 fof k2amp, kfund, k2form, koct, k2band, kris, \
kdur, kdec, iolaps, ifna, ifnb, itotdur
a3 fof k3amp, kfund, k3form, koct, k3band, kris, \
kdur, kdec, iolaps, ifna, ifnb, itotdur
a4 fof k4amp, kfund, k4form, koct, k4band, kris, \
kdur, kdec, iolaps, ifna, ifnb, itotdur

; Combine all of the formants together.
asig sum (a1+a2+a3+a4)* 16384
```

```
outs asig, asig

endin
/* fof.orc */

</CsInstruments>
<CsScore>
/* fof.sco */
; Table #1, a sine wave.
f 1 0 4096 10 1
; Table #2, an envelope
f 2 0 1024 19 0.5 0.5 270 0.5
; Play Instrument #1 for five seconds.
i 1 0 5
e
/* fof.sco */
</CsScore>
</CsoundSynthesizer>
```