Synthesizer sings

About the difference between child and adult voice, the distinction between male and female voice, about soprano and tenor, about overtone singing ...
About the fantastic cybernetic system: the trinity of brain, speech organ and hearing.

When acoustic feedback fails
Once I was confronted with the fact that this cybernetic system was disrupted. The feedback mechanism did not or hardly work . On a Friday afternoon, back on my way home from the conservatory. A whole group of youth entered the train.

Looks like teenagers, on weekend home. Soon it turned out that it was a group of deaf-mute youths. Mutually communicating with sign language, occasionally (unintended?) voice sounds were heard.

And they were pretty confronting. The sounds had something inhuman. Something that did not spur at all with those nice sweet kids. Then I suddenly realized what it means when the acoustic feedback does not work when creating speech sounds.

Spectral envelope and pitch modulation
In another article it became clear how fantastically sensitive the human hearing is for spectral fingerprints, such a spectrum envelope or formant characteristic.

But likewise sensitive to the transition from one form to the other. And certainly not less sensitive to what happens under such a spectral envelope. For example, pitch modulation. We will deepen these matters on the basis of the patches, `VocFolds&Tract.pch2`, `Age&Gender.pch2`, `OvertoneSinger.pch2` and `SubtrSinger.pch2`.

Where do these formants actually come from?
We will look at that by means of a simplification of the human speech organ. We therefore make an abstraction, a model. As a result, reality becomes less complex and we soon have the idea that we understand what is happening. The thought that you understand it always gives satisfaction.

Presented in simplified form, we can dissect the speech organ into two parts: an excitator and a resonator. These are consecutively the vocal cords and linked to the vocal tract.

The latter can be compared to a cylindrical tube of approx. 17 to 18 cm long. This tube is connected to the vocal cords, the excitation oscillator. Acoustically speaking, this is broadly similar to a clarinet.

Such a system behaves like a resonator with resonant frequencies that relate as 1: 3: 5 ... All odd multiples of the fundamental frequency ('1'), the first resonance mode of the tube.

For a tube length of 17 to 18 cm we then arrive at about 500 Hz for the lowest resonant frequency. The next resonances we find at 1500, 2500, 3500 and 4500 Hz and so on.

Formant simulation through delay line with negative feedback
Speak for yourself the sound 'uh'. Your mouth position is then relaxed, your tongue lies flat in the mouth. In this position, your organ of speech is most similar to the simplified model. This simple speech model can be found in the patch VocFolds & Tract.

Load this patch and listen to Variation 1. You will hear a tune in the blues scale with a pulse waveform of which each pulse lasts equally, 0.32 milliseconds. Also called a one shot.

The pitch is now varied by varying the 'out time' of this pulse signal. This is done by the Excitation oscillator, which triggers the Pulse module. This basic sound has a fixed spectrum envelope due to its fixed pulse duration.

In the X-Fade module, gradually change the potentiometer position from -64 to 0. You will now hear the resonances of the voice channel.

The voice channel is simplified by a delay line with negative feedback. In this way we have built a resonator that indeed corresponds to the resonance characteristic of the idealized vocal tract.

If you experiment with the delay time, you can change the length of the vocal tract, as it were. Gradually change the delay time to 0.5 millisecond. Indeed, the formants shift an octave upwards and you hear the well-known smurf sound, known and notorious of the older generation of samplers.
By varying the pulse duration you can give the basic signal, the 'vocal cord sound', another spectrum envelope, another timbre.

Formants and speaking and singing pitch
Load the Age & Gender patch. The sound generation as follows: vocal cord signal, voice channel followed by the two parallel bandpass filters.

An attempt with only two formants and an adequate input signal to give an impression of various singing voices of different age and gender.

If you listen to the eight variations in succession you should get a sounding impression of: a singing little child, boy / girl, soprano, mezzo-soprano, alto, tenor, baritone and bass.

Age and gender impression are mainly determined by the pitch of the vocal cord signal. As an average, the male voice sounds one octave lower than the female voice and the child's voice a fifth higher than a female voice, respectively about 100, 200 and 300 Hz.
These are obviously averages. Everyone knows examples from his environment that form an exception. For example, men with a remarkably high voice and, conversely, women with a very low voice.

Overtone singing
You speak every day and you hear voices every day. You therefore no longer think about what a fantastic flexible musical instrument the human speech organ is.
In overtone singing you can hear how wonderful independent vocal chords and different mouth and tongue positions can be controlled.

For example, in wind instruments there is a strong feedback from resonator to excitation oscillator. This is so strong that the excitation oscillator can only vibrate at one of the resonance frequencies of the resonator.

Different pitches are realized by shortening the resonator tube by opening holes, which reduces the effective acoustic length. Or by literally shortening or lengthening the tube physically by means of a sliding mechanism similar to the trombone, or by valves like the trumpet.

Our own biosynth
Based on what we hear when speaking or singing, we can control the breath, vocal cords and resonator through the acoustic feedback mechanism extremely accurately and independently of each other.

If you listen to overtone singing then you suddenly realize that our simple resonator model based on a comb filter with negative feedback is a very simple representation of things.

In essence, the speech channel ends in a very flexible oral cavity. It can function as one band filter for the vowels oo, oh, ò and aa. But also as two simultaneous band filters. This is done by curving your tongue and by varying your mouth cavity, for example with the vowels, uh, ü, a, ay and ee.

The oral cavity is then divided into two separate compartments. By spouting your mouth you can extend the effective length of the voice channel a bit and thus lower the resonance frequencies. You can also reduce those resonance frequencies by making the mouth opening very small.

Go tweaking with especially fine structure modulation. If you switch off all fine structure modulation and also deactivate the Glide module you will notice that little or nothing remains of the singing impression. (To make tweaking a little clearer, in the `SubtrSinger.pch2` patch I colored the most important experimental modules purple.)

Synthesis model and control

You've probably already discovered that a proper controlling of whatever synthesis model is very decisive for the sound impression you experience.
So every simple hardware or software synth is capable of much more than what you normally hear.

That you do not hear that is mainly due to the rough and poor controlling of pitch, loudness and timbre. Timbre: the dynamic modulation of spectral envelope. Go and listen to Perry Cook's 'Spasm', a physical model of song synth that is a much more accurate approach to reality than Age & Gender and OvertoneSinger.

Load the `SubtrSinger.pch2` patch. I have stripped the vocal synth model so far that with each subtractive synthesizer, which has an oscillator with various waveforms, a VCA with simple envelope generator and a filter with various choice characteristics such as highpass, lowpass and bandpass, the same results can be achieved. Provided ..., yes the right precise nuanced controlling. This stripped example lacks the so-called singer's formant (ca 2500Hz), which is very characteristic of classic male voices. The simulations tenor, baritone and bass are therefore less convincing than the other examples.

The moral of this story: the synthesis power of today is no problem, the sadder it is with the controlling. Even when compared to an acoustic instrument, controlling on the average synth is poor and often very coarse. Let alone in comparison with the human voice: for sure, best synth ever ...

Ernst Bonis

This article was previously published in Dutch as 'Synth met meer noten op z'n zang' in Interface 112 October 2007.

internet
Perry Cook’s "Spasm", singing synthesizer
www.cs.princeton.edu/~prc/SingingSynth.html

http://www.cs.princeton.edu/~prc/PRCThesis.pdf

literature
The Science of the Singing Voice
Johan Sundberg
1989 Northern Illinois University Press

The Science of Musical Sound
John R. Pierce
Scientific American Books
1983 New York and Oxford