

Synth zingt

Over het verschil tussen kinder- en volwassenenstem, het onderscheid tussen mannen- en vrouwenstem, over sopraan en tenor, over bventoonzingen. Vooruit, leer je synthesizer zingen...

de drie-eenheid in klanksynthese

Nog altijd ben ik verbaasd en verbijsterd over dat fantastische cybernetische systeem: de drie-eenheid van menselijk brein, spraakorgaan en gehoor.

Feilloos herken je stemmen uit duizenden. Hoe komt het dat ik direct een geboren Tilburgs-spreker herken. Hoe doe ik dat nou: Vlaams spreken. Heel handig overigens als je op bezoek bent bij de zuiderburen. Ze hebben dan niet in de gaten *da kik ei'lijk n'n 'Ollandse kaaskop ben* en dat heeft zo zijn voordelen.

Maar toch blijft het frustreren dat ik maar gedeeltelijk onder woorden kan brengen wat nou het klinkende verschil is tussen het Vlaams gesproken zoals in de provincies Antwerpen en Vlaams-Brabant enerzijds en bijvoorbeeld ons eigen Brabantse Nederlands.

Enkele verschillen zijn gemakkelijk te duiden. Bijvoorbeeld hebben wij Nederlanders de neiging om bepaalde klinkers te diftongeren. De *ee* in *beer* en *eten* uit te spreken als *eej*. De Vlaams-Brabantse / spreekt men uit met de tongpunt wat meer naar achteren in de mond. En zo kan ik nog wel een paar verschillen min of meer objectief duiden.

Maar toch, voor het overgrote deel weet ik het niet. Dat is af en toe knap frustrerend voor een geluidsnerd als ik. Wel direct horen dat het Vlaams dat ik hoor uit het pajottenland stamt, maar dat ik niet precies kan uitleggen hoe dat komt; waaraan het ligt *dát* ik dat hoor.

Dat iedereen zomaar zonder enige oefening onderscheid kan maken in kinder-, vrouwen- en mannenstemmen. Dat er mensen zijn als Mr. Yamaha Nederland, Ruud Veenstra:

met een fantastisch stem-imitatietalent. Vaak denk ik nog terug aan de autoritjes samen met Ruud op weg naar een VL-1 demo. Moeiteloos deed 'ie dan een Duitser met Engels accent, of een Japanner die probeert Nederlands te spreken. Of hij imiteerde mijzelf. Wat mij dan bewust maakte van m'n eigen typerende spraakeigenaardigheden. En, als je Ruud dan vroeg hoe 'ie dat deed, wist 'ie het uiteraard niet. Kortom een fantastisch cybernetisch systeem die drie-eenheid van brein, spraakorgaan en gehoor: staat garant voor *best synth ever!* Geheel en al onbewust werkt dit regelsysteem. Vanuit je grijze massa produceer je bij benadering de klank die je in gedachten had. Terwijl je dat doet hoor je dat de output niet helemaal klopt. Ogenblikkelijk regel je dan je ademdruk, stembanden en mondstand bij tot het gewenste resultaat. Daar sta je eigenlijk nooit bij stil. Zonder dit *real time feedback* systeem geen 'menselijk' klinkende stem of zang.

een verstoorde terugkoppeling

Eens werd ik ermee geconfronteerd dat dit terugkoppelingsmechanisme niet of nauwelijks werkte. Op een vrijdagmiddag, terug van het conservatorium naar huis. In de trein stapte een hele groep jeugd in.

Zo te zien aankomende pubers, lekker met weekend naar huis. Al snel bleek dat het een groepje doofstomme jongeren was. Onderling werd met gebarentaal gecommuniceerd, waarbij zo nu en dan ook (onbedoelde?) stemgeluiden te horen waren.

En die waren behoorlijk confronterend. De klank had iets onmenselijks. Iets dat helemaal niet spoorde met die aardige lieve kinderkoppies. Toen realiseerde ik me opeens wat het betekent als de akoestische feedback niet werkt bij het vormen van spraakklanken.

spectrale omhullende(n) en toonhoogtemodulatie

In een andere aflevering van synthworkshop werd al wel duidelijk hoe fantastisch gevoelig het menselijk gehoor is voor spectrale vingerafdrukken als zo'n spectrum envelope of formantkarakteristiek.

Maar evenzo gevoelig voor het overgaan van de ene naar de andere formantstructuur. En zeker niet minder sensitief voor wat er onder zo'n spectrale omhullende gebeurt. Bijvoorbeeld toonhoogtemodulatie. Beide zaken gaan we wat uitdiepen aan de hand van de patches, *VocFolds&Tract*, *Age&Gender*, *OvertoneSinger* en *SubtrSinger*.

Waar komen die formanten eigenlijk vandaan?

Dat gaan we bekijken aan de hand van een vereenvoudiging van het menselijk spraakorgaan. We maken daarom een abstractie, een model. Zo'n vereenvoudiging doet een mens goed. De werkelijkheid wordt daardoor minder complex en we hebben dan al gauw het idee dat we begrijpen wat er gebeurt. De gedachte dat je het begrijpt stemt altijd tot voldoening. Je weet dan waar Abraham de mosterd haalt.

Zo vereenvoudigd voorgesteld kunnen we het spraakorgaan ontleden in twee delen: een excitator en een resonator. Dat zijn opeenvolgend de stembanden en daaraan gekoppeld het spraakkanaal.

Dat laatste is –wat kort door de bocht– te vergelijken met een cilindrische buis van zo'n 17 à 18 cm lang. Aan het begin is die buis verbonden met de stembanden, de excitatie-oscillator. Akoestisch gezien is dat in grote lijnen vergelijkbaar met een klarinet.

Zo'n systeem gedraagt zich als een resonator met resonantiefrequenties die zich verhouden als 1 : 3 : 5 ... Allemaal oneven veelvouden van de grondfrequentie ('1'), de eerste resonatiemodus van de buis.

Voor een buislengte van 17 à 18 cm komen we dan uit op ongeveer 500 Hz voor de laagste resonantiefrequentie. De volgende resonanties vinden we dan bij 1500, 2500, 3500 en 4500 Hz enzovoorts.

formantsimulatie met een delay line met negatieve feedback

Spreek voor je zelf eens de klank 'uh' uit. Je mondstand is

dan ontspannen, je tong ligt plat in de mond. In deze stand lijkt je spraakorgaan nog het meest op het vereenvoudigde model. Dit simpele spraakmodel vind je in de patch *VocFolds&Tract*.

Laad deze patch en luister naar *Variation 1*. Je hoort nu een zwalkend deuntje in de bluestoonladder met als basisklank een pulsgolfvorm waarvan elk pulsje even lang duurt, 0,32 milliseconde. Ook wel een *one shot* genaamd.

De toonhoogte wordt nu gevarieerd door de '*uittijd*' van dit pulssignaal te variëren. Dit gebeurt door de *Excitatie-oscillator*, die de *Pulse module* triggert. Deze basisklank heeft door z'n gefixeerde pulsduur een vaste spectrum-omhullende verkregen.

Laat het riedeltje gezellig doorzeuren en verander in de *X-Fade module* de potmeterstand geleidelijk van -64 naar 0. Je hoort nu de resonanties van het spraakkanaal ontstaan.

Het spraakkanaal is vereenvoudigd voorgesteld door een delay line met negatieve terugkoppeling. We hebben op deze manier een resonator gebouwd die inderdaad overeenkomt met de resonantiekarateristiek van het geïdealiseerde spraakkanaal.

Als je nu experimenteert met de delay-tijd verander je als het ware de lengte van het spraakkanaal. Verander geleidelijk de delay-tijd naar 0.5 milliseconde. Inderdaad ja, de formanten verschuiven een octaaf naar boven en je hoort de welbekende *smurfensound* ontstaan, bekend en berucht van de oudere generatie samplers.

Door de pulsduur te variëren kun je het basissignaal, het 'stembandgeluid', een andere spectrum-omhullende, een andere kleur geven.

formanten en spreek- en zangtoonhoogte

Laad de *Age&Gender patch*. Deze lijkt als twee druppels water op de patch uit een andere synthworkshop, *Singer*. Het virtuele model is nu echter wat verfijnd.

In de *Singer patch* werden de formanten opgewekt door

direct achter het imitatiestembandsignaal twee bandpass filters (FltNord) te plaatsen. Nu gaat de klankopwekking als volgt: stembandsignaal, spraakkanaal gevolgd door de twee parallelle bandpass filters.

Een poging met slechts twee formanten en een adequaat inputsignaal een indruk te geven van diverse zangstemmen van onderscheiden leeftijd en geslacht.

Als je achtereenvolgens de acht *variations* luistert zou je een klinkende impressie moeten krijgen van respectievelijk: een zingend klein kind, jongen/meisje, sopraan, mezzosopraan, alt, tenor, bariton en bas.

Leeftijds- en geslachtsindruk worden voornamelijk bepaald door de toonhoogte van het stembandsignaal. Als gemiddelde geldt dat de mannestem één octaaf lager klinkt dan de vrouwenstem en de kinderstem een kwint hoger dan een vrouwenstem, respectievelijk ongeveer 100, 200 en 300 Hz.

Dit zijn uiteraard gemiddelden. Iedereen kent wel voorbeelden uit z'n omgeving die een uitzondering vormen. Bijvoorbeeld mannen met een opvallend hoge stem en omgekeerd vrouwen met een heel lage stem.

overtone singing

Ofwel in goed Nederlands, boventoonzingen. Spreken doe je elke dag en evenzo hoor je elke dag stemgeluiden. Je staat er daarom helemaal niet meer bij stil wat een fantastisch flexibel geluids/muziekinstrument het menselijk spraakorgaan is.

In boventoonzingen hoor je hoe geweldig onafhankelijk stembanden en onderscheiden mond- en tongstanden kunnen worden aangestuurd.

Zo bestaat er bijvoorbeeld bij blaasinstrumenten een sterke terugkoppeling van resonator naar exciter-oscillator. Zodanig sterk zelfs dat de exciter-oscillator slechts kan trillen op één van de resonantiefrequenties van de resonator.

Verschillende toonhoogtes worden gerealiseerd door de resonatorbuis te verkorten door gaten te openen, wat de effectieve akoestische lengte verkleint. Of door de buis letterlijk fysiek te verkorten of te verlengen door een schuifmechanisme als bij de trombone, of d.m.v. ventielen zoals bij de trompet.

Zo niet bij onze eigen *biosynth*. Op basis van wat we horen bij spreken of zingen, kunnen we door het akoestisch feedbackmechanisme uiterst nauwkeurig en onafhankelijk van elkaar ademspanning, stembanden en resonator aansturen.

Als je naar boventoonzang luistert dan wordt je opeens duidelijk dat ons simpele resonatormodel gebaseerd op een kamfilter met negatieve terugkoppeling wel een erg eenvoudige voorstelling van zaken is.

In wezen eindigt het spraakkanaal in een zeer flexibele mondholte. Die kan functioneren als één bandfilter voor de klinkers *oe*, *oo*, *ò* en *aa*. Maar eveneens als twee simultane bandfilters. Dat doe door enerzijds je tong te krommen en anderzijds je mondstand te variëren zoals bijvoorbeeld bij de klinkers, *eu*, *uu*, *èè*, *ee* en *ie*.

De mondholte wordt dan als het ware verdeeld in twee aparte compartimenten. Door je mond te tuiten kun je de effectieve lengte van het spraakkanaal een beetje verlengen en zo de resonantiefrequenties wat verlagen. Ook door de mondopening heel klein te maken verlaag je die resonantiefrequenties.

Ga tweaken met vooral de fijnstructuurmodulatie. Als je alle fijnstructuurmodulatie uitschakelt en ook de *Glide module* deactiveert zul je bemerken dat van de zangimpressie weinig of niets overblijft. (Om het tweaken wat overzichtelijker te maken heb ik in de patch *SubtrSinger* de belangrijkste experimenteerm modules paars gekleurd.)

synthesis model en control

Je hebt waarschijnlijk al ontdekt dat een juiste controlling van whatever synthesis model zeer bepalend is voor de

klankimpressie die je ervaart.

Zo is elke eenvoudige hardware of software synth tot veel meer in staat dan wat je er doorgaans uit hoort komen.

Dat je dat niet hoort is voornamelijk te wijten aan de grove en ongenueanceerde controlling van toonhoogte, luidheid en timbre. Timbre: de dynamische modulatie van spectrale omhullende. Ga eens luisteren naar Perry Cooks '*Spasm*', een physical model zangsynth die een veel nauwkeuriger benadering van de werkelijkheid is dan *Age&Gender* en *OvertoneSinger*.

Luister en vergelijk. Laad dan ook nog de *SubtrSinger patch*. Ik heb het zangsynthmodel hier zover gestript dat je met elke subtractieve synthesizer, die beschikt over een oscillator met diverse golfvormen, een vca met eenvoudige envelope generator en een filter met diverse keuzekarakteristieken als highpass, lowpass en bandpass, dezelfde resultaten kan bereiken. Mits..., jawel de juiste precieze genuanceerde controlling. In dit gestripte voorbeeld ontbreekt de zogenoemde *singer's formant* (ca 2500Hz), die heel kenmerkend is voor klassieke mannenstemmen. De simulaties tenor, bariton en bas zijn dan ook beduidend minder overtuigend, dan de andere voorbeelden.

De moraal van dit verhaal: aan de synthesekracht van vandaag ligt het niet, des te droeviger is het gesteld met de aansturing. Zelfs in vergelijking met een akoestisch instrument is de controlling op de doorsnee synth maar armetierig en veelal heel grof. Laat staan in vergelijking met de menselijke stem: *for sure, best synth ever!* Vooral als ze zo mooi verleidelijk Vlaams-Brabants spreekt; Dan, smélt ik ...

Ernst Bonis

Dit artikel werd eerder gepubliceerd (als 'Synth met meer noten op z'n zang') in Interface 112 oktober 2007.

luisteren
Stimmung (boventoonzang)

Karlheinz Stockhausen
HYPERION
CD 34571161150
CODAEX Deutschland GmbH

Mark van Tongeren, boventoonzang
www.fusica.nl/index.htm

Perry Cooks "Spasm", singing synthesizer
www.cs.princeton.edu/~prc/SingingSynth.html

meer weten
<http://hyperphysics.phy-astr.gsu.edu/hbase/music/vowel.html>
excellente tutorial over de menselijke stem

Klank en Muziek
Een combinatie van wetenschap en cultuur
John R. Pierce
Natuur en Techniek
1986 Maastricht
ISBN 90 70 157 55 1