

# Een spraaksynthesizer op UNIFACE

Klaas Robers, Valkenswaard

Al enige jaren staat er in Eindhoven op het hoogste gebouw van de Technische Universiteit een automatisch zendstation bestuurd door een P2000. Deze zender schakelt zich iedere avond zelf in en zendt dan een les uit van de morsecursus. Zendamateurs moeten morse leren voor hun hoogste zendmachtiging. Met iedere avond een les gaat dat het beste. Zendamateurs doen dat natuurlijk per radio.

## Inleiding

Voor deze morsecursus hadden wij de behoefte aan een schakeling waarmee spraak kon worden gegenereerd. Met deze spraak kondigt het morsestation zijn uitzendingen aan en noemt daarbij netjes zijn roepletters, zoals dat door de PTT geëist wordt.

De hele cursus loopt op een P2000 waarbij wij als voorwaarde hadden gesteld dat er niets mocht bewegen, dus geen cassettes, geen floppy's, ook geen hard disc. Bovendien staat alles de meeste tijd uit, ook de computer, dus moest alles in EPROM. Dat beperkt

de te gebruiken geheugencapaciteit sterk. Daarom kwam het domweg opslaan van gesampled en gedigitaliseerd geluid zoals bij de CD niet in aanmerking. De enige zuinige oplossing was een spraaksynthesizer.

## Fonemen

Er bestaan verschillende soorten speech synthesizers. Het oudst bekende is de "VOTRAX". Deze synthesizer sluit aan bij de fonetische schrijfwijze zoals die gebruikt wordt in woordenboeken. Per taal onderscheidt men een aantal verschillende klanken. Hiervoor ontwikkelde men

Fig 1: Blokschema van de formanten synthesizer in de PCF8200. Duidelijk is te zien dat het inputsignaal ruis kan zijn of een pulsvormige toon (uit databoeken Philips Components).

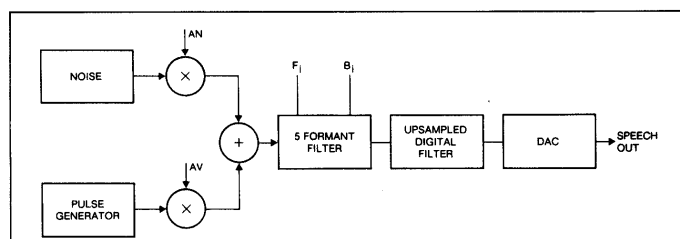
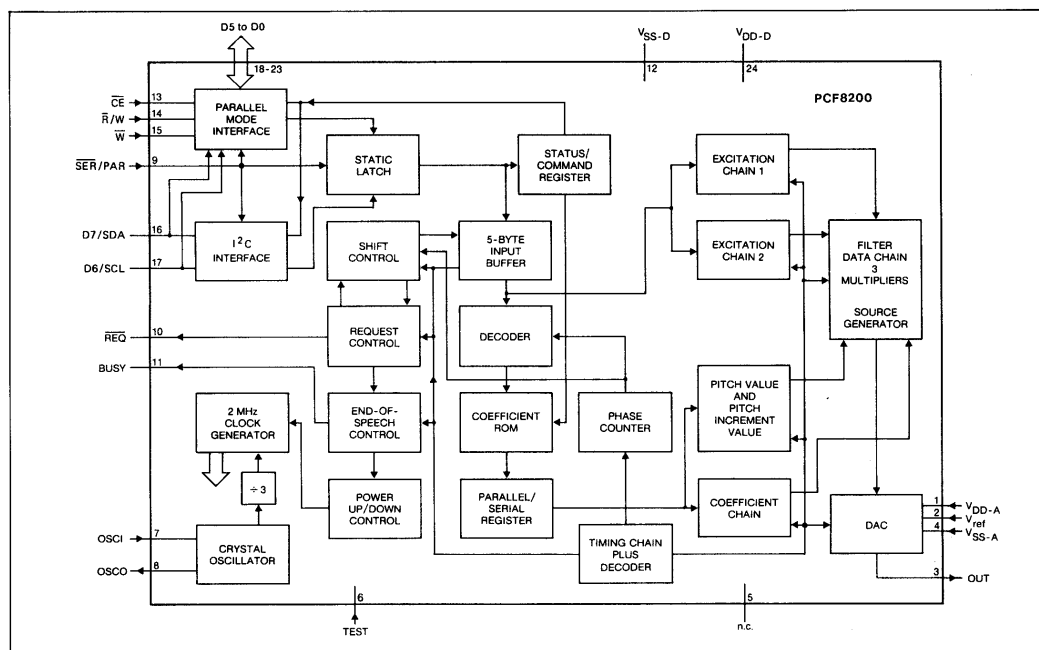


Fig 2: Compleet blokschema van de PCF8200. Alleen het deel rechts van VSS-D is de eigenlijke synthesizer, de rest is besturing en timing (uit databoeken Philips Components).



in vroeger jaren reeds een soort uitgebreid alfabet. Elke klank die in de taal voorkomt wordt met een tekenje aangegeven, meestal letters, schuine letters en letters met accentjes. Zo'n klank heet een foneem. Wie dit goed beheerst kan nog nooit gehoorde woorden aan de hand van het fonetisch geschrevene vrij correct uitspreken.

Zo werkt ook de VOTRAX. De met een speciaal fonetisch toetsenbord of in daarvoor aangepast ASCII ingegeven tekst wordt door de VOTRAX omgezet in een klankenstroom. Elk teken en andere klank. De overgangen worden enigszins "afgerond". Normale spraak vereist een datastroom van ongeveer 150 bits per seconde. De output is dan een typisch monotone computerstem, stem, meestal in een overduidelijk buitenlands (Engels) accent. Dit laatste omdat alleen de klanken worden opgewekt die karakteristiek zijn voor de taal waarvoor de chip is bedoeld. Een Nederlandse VOTRAX bestaat niet. Paul Ruysendaal heeft jaren geleden een P2000 insteekdoos ontworpen met daarin een VOTRAX-chip. Zo nu en dan zie je die nog op open dagen.

Hoe de klanken van de VOTRAX worden opgewekt is een soort fabrieksgeheim. In het eerste model zat alles in dichtgegoten blokken met daaraan een printconnector, later werd het een enkele chip. Het zelf toevoegen of wijzigen van klanken is niet mogelijk.

Wel kan van de lopende tekst de toonhoogte worden gevarieerd waardoor het kenmerkende monotone geluid wat wordt verlevendigd, maar echt mooi wordt het nooit. De verstaanbaarheid is zodanig, dat, als je er enige tijd aan gewend bent of weet wat er gezegd wordt, dan kun je het verstaan.

Het hangt bovendien sterk af van het karakter van de tekst. Als er plotseling iets onverwachts gezegd wordt dan versta je het toch weer niet. Dit leek ons niet goed genoeg voor de aankondigingen in de morsecursus.

#### Formanten

Een wat andere benadering volgt het Philips IC PCF8200. De achtergrond hiervoor is ontwikkeld op de Technische Universiteit Eindhoven door het Instituut voor Perceptie Onderzoek (IPO). Men beschouwt het menselijk spraakorgaan als een pulsgenerator met regelbare sterkte en toonhoogte, het strottehoofd dus, met daarachter een aantal regelbare filters in de vorm van de mondholte, tong, tanden en lippen. In plaats van pulsen kan ook ruis een ingangssignaal zijn voor de filterketen, dat is voor sissende spraakuitingen.

De chip PCF8200, er zijn ook al voorgangers geweest zoals de MEA8000, zitten geheel digitaal de pulsgenerator, de ruisgenerator, de amplitude regelingen, de filters en de DAC (zie fig 1). Het variëren van de generator

en de filters geschiedt om de 10 msec. Daarvoor zijn blokjes data nodig (frames) van 40 bits (5 bytes). Het IC vraagt zelf met een REQUEST signaal om het volgende byte; de computer hoeft bij het spreken dus niet zelf de tijd en de blok grootte in de gaten te houden.

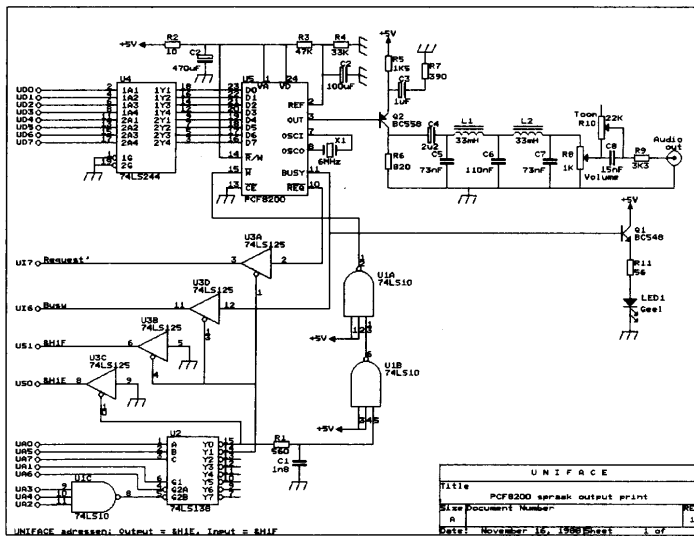
Het IPO heeft zeer uitgebreid onderzoek gedaan aan de spectrale inhoud van spraak en omdat een mensensmond redelijk uniform is heeft de chip geen echte voorkeur voor een bepaalde taal. Tenslotte, om een andere taal te leren hoeft je niet je mond te laten verbouwen. Twijfelaars moeten maar eens luisteren naar in Nederland geboren of jeugdig geadopteerde Vietnameesjes en Koreaantjes. Hier hoor ik ze met een onvervalst Brabants accent spreken.

Het IPO systeem heeft bewezen uitstekend verstaanbaar te kunnen zijn. Veel hangt natuurlijk af van de juiste besturing van de generatoren en de filters. Maar als dat goed gedaan wordt dan is de verstaanbaarheid geen punt. Men kan er zelfs stemmen van personen in herkennen. Door de tijdsduur van de frames aan te passen aan de tekst, korte frames als er veel verandert en lange frames in aangehouden klanken en tussen woorden, kan de datastroom worden teruggebracht tot 1500 bits/sec. zonder dat de kwaliteit hoorbaar wordt aangetast. Dit is waar wij voor de morsecursus voor hebben gekozen.

#### Datavoorbereiding

Het is natuurlijk mogelijk om, net als bij de VOTRAX, een set fonemen te maken en de datafiles voor al deze fonemen in de gewenste volgorde af te spelen. De kwaliteit degradeert dan tot fonemenspraak. Al beter is het om in plaats van fonemen alle overgangen tussen fonemen op te slaan. Daarmee neemt wel de hoeveelheid informatie kwadratisch toe. Dit noemt men di-fonen. Voor de PCF8200 is een complete set difonen beschikbaar. Veel beter is het nog om de nodige tekst te laten inspreken en deze zinnen, of delen daarvan, te laten analyseren. Het analyse apparaat levert dan direct de datafiles waarmee de PCF8200 overweg kan.

Fig 3: Schakelschema spraaksynthesizer met de PCF8200 op UNIFACE. Links zien we de aansluitingen op de 34-polige UNIFACE kabel. De audio-output staat rechts. Een (gele) LED geeft aan dat de PCF8200 'busy' is.



De analyse wordt uitgevoerd door een PC met daaraan een 12 bits A/D converter. De zinsdelen worden vanaf tape gedigitaliseerd en in het computergeheugen opgeslagen. Dat zijn onwijs veel data, iets van 120 kbit per seconde. Direct daarna volgt de analyse, waarbij het programma steeds over korte tijdsfragmenten de optimale instelling van de oscillator of ruisgenerator en de instellingen van de filters uitzoekt. De hoeveelheid data slinkt daarmee tot 3,5 kbits per sec. Deze datafile kan op disc of op floppy bewaard worden. Het is opmerkelijk te zien dat het nogmaals digitaliseren en analyseren van hetzelfde tape fragment een heel ander resultaat kan geven. Blijkbaar is de toevallig wat andere indeling in tijdsfragmenten cruciaal. Als de spreker twee of driemaal dezelfde zin heeft uitgesproken geeft dat een nog groter onderscheid. Je neemt natuurlijk de beste analyse.

#### **Oppoetsen**

Daarna volgt een edit-sessie waarbij het resultaat met de hand kan worden verbeterd. Hierbij verschijnen de berekende instellingen als lijnen op het computerscherm. Voor de oscillator is dat een eenvoudige golflijn die de toonhoogte aangeeft. Van de 5 formanten filters wordt de resonantiefrequentie en de bandbreedte (Q) op het scherm aangegeven als een in breedte variërende kronkelende baan. Dit alles doet nogal kunstmatig aan, maar na enig oefenen krijg je er wat kijk op. Met het toetsenbord is alles op een nogal primitieve manier te wijzigen en na elke wijziging kun je beluisteren wat het resultaat is. Het blijkt dat de analysator zich nogal eens verrekent en daardoor plotseling even de toonhoogte een octaaf laat zakken of stijgen. Ja, wat is de grondtoon en wat de boventonen hè? Vanaf het toetsenbord is dit snel in het reine te brengen, waardoor de spraak behoorlijk opknapt.

Voor onze experimenten konden wij gebruik maken van de analysator van het CARIN-lab. Bij dit navigatiesysteem voor de auto worden de aanwijzingen aan de bestuurder ook door een PCF8200 uitgesproken. Men had bij CARIN al ontdekt dat een goede stem voor de ingesproken teksten van het grootste belang was. De vraag aan Wim van Putten (TROS radio) om de aankondigingen van de morsecursus in te spreken bleek een schot in de

roos. De eerste proef die ik op een cassette-recordertje naar een clubavond had meegenomen was nog niet helemaal vlekkeloos, maar het was voor iedereen uitstekend verstaanbaar.

Voor de morsecursus werden alle zinnen zodanig verknipt dat bijna alles slechts eenmaal behoefde te worden opgeslagen. Als iets dus tweemaal voorkomt, zij het in een wat andere context, dan is dat toch hetzelfde stukje data. Op deze manier kon alles ruim in de 8 kilobyte ROM die ik daarvoor had gereserveerd.

#### **Interfacing**

De PCF8200 is heel gemakkelijk aan te sluiten op een computer. Er is een 8 bits brede databus die gebruikt wordt om de databytes in de chip te schrijven. Bij lezen (read) komen hier drie statusbits uit. Het IC kan dus memory mapped op een geheugenadres worden ingevoegd. De timing loopt geheel los van de computer op een eigen 6 MHz kristal dat domweg op het IC wordt aangesloten. Geen gepriegel met externe C-tjes dus.

Het benodigde programma is heel eenvoudig: wachten tot het request-statusbit 0 wordt en dan het volgende byte uit de datafile sturen. En dat net zo lang tot het stop-statusbit 1 wordt. Aan het eind van de datafile is een stop instructie opgenomen, daardoor stopt de synthesizer vanzelf.

Ook voor "domme hardware" is de PCF8200 gemakkelijk te besturen. Dat komt omdat de statusbits ook op aparte aansluitpennen zijn uitgevoerd. De databus wordt dan op "write only" gezet en kan direct worden aangesloten op de datalijnen van een EPROM. Een tellertje dat loopt op de statusbits is nodig voor de adressering van de EPROM en dat is zo'n beetje alles.

#### **Op UNIFACE**

Voor ons morsestation is een tussenform gebruikt. In de P2000 zit een UNIFACE computerdeel. Dat was toch al zo, want de zender moest aan en uit gezet worden, er moeten morsepiepjes daarnaartoe worden doorgegeven en de teksten voor de morselessen zitten in EPROMs (192 K), ook weer op een UNIFACE print. Nou lijkt de UNIFACE-bus geweldig veel op domme hardware. Om de PCF8200 hierop aan te sluiten zijn 4 simpele TTL-blokjes gebruikt. De buffer in de databus bleek achteraf zelfs overbodig.



Foto : Wim van Putten. Zijn stem is vastgelegd in de datafiles voor de spraaksynthesizer (Foto TROS persdienst).

Ook de laagfrequent output is de eenvoud zelf. Door de toegepaste, uit de Compact Disc techniek overgenomen, oversampling en digitale filtering is het uitgangssignaal in feite al schoon. Toch is voor de zekerheid maar een laagdoorlatend filter ingebouwd, aanwezige HF reststoring zou in de 2-meter zender merkwaardige dingen kunnen doen. Een eenvoudige toonregeling verzwakt de lage tonen eventueel wat.

#### **Resultaat**

De morsecursus met als roepletters PI7CWE kondigt zich in de 2 meter amateurband elke avond netjes in spraak aan. Of je kunt horen dat het uit de computer komt? Wij vinden van wel. Maar een echte computerstem is het ook weer niet. In elk geval is iedereen het er over eens dat het uitstekend verstaanbaar is. En als je hem van de radio kent herken je zelfs de stem van Wim van Putten.