

Vocoding och frekvensskiftningsexperiment inom det audiologiska forskningsfältet

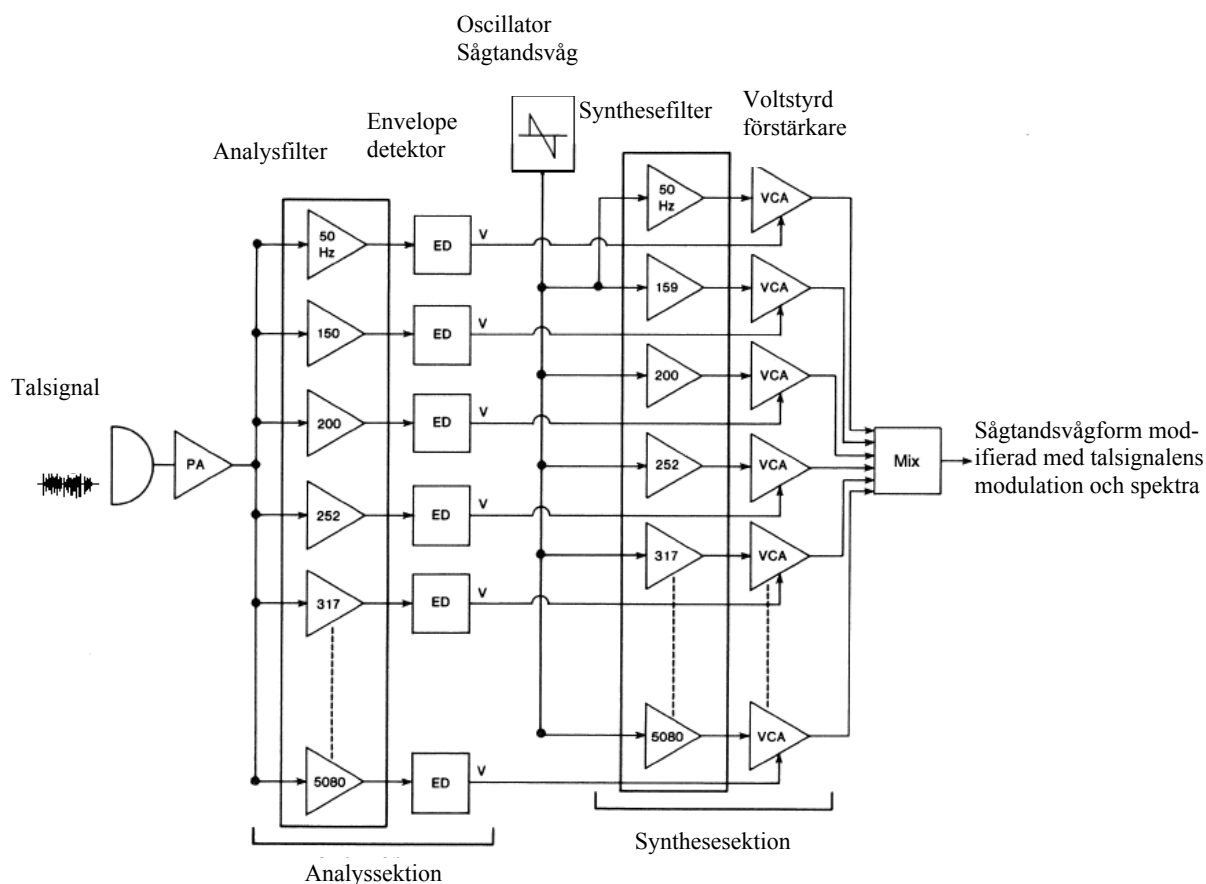
Av Morgan Karlsson

Vocoding

Några av de första försöken att återskapa tal elektroniskt gjordes på 30-talet av fysikern Homer Dudley som var verksam vid Bell Telephone Laboratories, New Jersey USA. Dudley arbetade med att konstruera en modell av talapparaten akustiska korrelerat vilka presenterades 1940 i artikeln "The Carrier Nature of Speech", som publicerades i *The Bell System Technical Journal*. Målsättningen var att reducera informationen i dåtidens telefonnät vid långdistanssamtal över atlanten utan att försämra taluppfattningen. Hans arbete som idag ligger till grund för formantkodning av tal kulminerade i en talanalysator som benämndes vocoder, samt en elektronisk talsynthesizer "The Voder" (Voice Operation Demonstrator) som kontrollerades via en klaviatur med tillhörande fotpedal. Tekniken demonstrerades vid 1939 års World's Fairs i New York och San Francisco. Voder/Vocodern återskapade flertalet akustiska särdrag i den mänskligt talapparaten, detta genom att såväl urskilja betydelsetunga formantregioner, som tonande och tonlösa språkljud, för att sedan återskapa dessa på syntetisk väg. Werner Meyer-Eppler som var verksam vid institutionen för Fonetik vid Bonns University 1948, var en föregångare som uppmärksammade vocodern och dess mångsidiga användningsområden.

Audiologiska experimenten med vocoding genomfördes på 60-talet av Ling, Denes, Takefuta, Minami och Pimonow. Slutet av 60-talet och början av 70-talet var den period då vocodern användes mest frekvent vid frekvensförflyttning inom det audiologiska forskningsfältet. Vocoding som signalbehandlande teknik är långt ifrån passe, vilket Posen och Reed så sent som 1993 visade med förbättrad taluppfattning av klusiler och frikativor genom användandet av en tonande/tonlös detektor med tillhörande vitt brus som bärvåg i vocoderns lågfrekventa synthesesedel. En vocoder är ett komplett analys-synthese system som bryter ned (analyserar) talljuden i ett antal närliggande frekvensband, och använder därefter amplituden i respektive frekvensband för att återskapa (synthesize) en signal som är mycket lik den ursprungliga analyserade talsignalen. Talljudens spektra och modulation överförs på så sätt till en s.k. bärvåg som kan vara allt från en sågtand, vitt brus, eller en komplex vågform med rik övertonsstruktur.

Inom Audiologin har diverse vocoders med olika antal filterband och specifikationer använts, men dess grundläggande mekanismer är desamma. Det är dock viktigt att poängtera att den tekniska konstruktionen i sig själv kan påverka kvaliteten på talet .



Figur 1 visar vocoderns olika beståndsdelar i block diagramform. I den vänstra sektionen processas talsignalen. I den högra delen processas vågformen som blir bärare av taljudsinformationen.

Till vänster i figur 1 syns vocoderns analysdel. Talsignalen filtreras med hjälp av ett antal branta bandpassfilter, med mittfrekvenser som optimerats för att täcka in det mänskliga talet och dess formanter. Filterbanken ”skivar” upp spektrumet av talsignalen. Varje ”skiva” av ljudet skickas därefter vidare till en envelopeföljare/detektor som läser av energin i det enskilda frekvensbandet. Envelopeföljarens utsignal består av en styrspänning som är proportionell mot styrkan i respektive frekvensband. Envelopeföljaren talar om hur mycket energi/amplitud som finns i varje frekvensband av talsignalen vid en viss tidpunkt. Analysdelens output kan med andra ord beskrivas som en uppsättning långsamt varierande styrspänningar som utgör en kod eller analys av talsignalens spektrum. Envelopeföljarens styrspänning som nu har taljudets duration och intensitet kontrollerar nu förstärkarna i synthesesektionen .

Synthesedelen av vocodern syns till höger i figur 1. En uppsättning bandpassfilter, identiska med dem i analysdelen, filtrerar ljudsignal/vågform nummer två, som benämns bärvåg eller carrier. Dessa filter ”skivar” upp bärvågens spektrum i ett antal band på samma sätt som analysdelen ”skivade” upp talsignalen. Varje ”skiva” skickas sedan vidare till en spänningsstyrd förstärkare. Denna förstärkare styrs av envelopeföljaren/detektorn i analysdelen. Ljudet i synthesedelens alla ingående förstärkare skickas därefter till en mixersektion som slutligen utgör vocoderns output. Talets spektrala variation kommer nu att överföras på bärvågen. Antag att det i analysdelens ingång finns en person som talar eller sjunger, och att det i synthesedelen finns en bärvåg med rik övertonsstruktur i form av en sågtandsvågform med konstant frekvens på 700 Hz. I vocoderns utgång kommer nu höras bärvågens tonhöjd på 700 Hz, men med talets spektrala variationer. Sågtandsvågen är nu bärare av talmodulationsinformation.

Detta är den grundläggande principen bakom vocoderns arbetssätt. Mer intressanta vocoderns inkluderar minst 16 filterband, tonande/tonlös detektor samt ton-till-styrspänningsomvandlare. Intressanta modeller i detta segment är Sennheiser VSM201, EMS 5000, Synton Syntovox 221 samt Bode's Vocoder är för att nämna några.

Denes gjorde taluppfattningstest på 10 normalhörande med en 11-bands vocoder. Talet sänktes med en faktor på 3, samtidigt som högfrekvent ljud flyttas proportionellt mer än basfrekvenser. Alla deltagarna fick 16 träningstillfällen på 20 minuter vardera. Vid varje träningstillfälle gjordes omfattande taluppfattningstester. En kontrollgrupp med normalhörande som enbart fått lyssna till oprocessat tal användes. Slutsatsen från studien var att taluppfattningen var likvärdig i båda grupperna och att vocodat tal går att lära sig. Taluppfattningen med vocodat tal förbättrades dessutom med 75 procentenheter efter 5.3 timmars träning.

Pimenow konstruerade en 7-bands vocoder som testade på gravt hörselskadade barn. Han använde uteslutande lågfrekvens ljud i synthesedelen under 300 Hz. Han rapporterade att 3 gravt hörselskadade barn lärde sig en vokabulär på 30 ord efter 3 veckors träning.

Takefuta utvärderade en vocoder som arbetade med 22 sinusvågor vars amplituder modulerades av en analysenhet med 1/3 oktavfilter. Talet presenterades till fyra olika grupper med en frekvensflyttningfaktor på 1.0, 1.4, 2.0 och 2.5. Alla grupper utom de som fått talet komprimerat med faktorn 2.5 fick ett förbättrat resultat efter träning. Den första gruppen med faktorn 1.0 (utan frekvensförflyttning) fick det högsta resultatet på 70 % rätt. Man noterade

även att konsonantljuden var lättare att identifiera än vokalljuden.

Posen och Reed utförde två experiment på normalhörande där bandbredden på det hörbara området reducerats till 800 Hz. Två lågpasfilter med en sammanlagd branthet på 300dB/oktav användes för att simulera en grav diskantnedsättning.

I experiment 1, där alla språkljud processades fann man att perceptionen av nasaler samt halv vokaler försämrades medan frikativor och klusiler förbättrades.

I experiment två modifierades därför konstruktionen till att endast vara aktiv då energin i talsignalen var som störst i diskantområdet. Denna konstruktion lyckades förbättra uppfattningen av frikativor och klusiler utan att försämra perceptionen av nasaler och vokaler.

Trots vocoderns flexibilitet och möjligheter har få positiva resultat rapporterats med denna teknik. Undantaget är Posen och Reeds modifierade vocoderschema som flyttade frikativor och klusiler till basområdet, detta när ett antal kriterier varit uppfyllda. Den förbättrade taluppfattningen av frikativor och klusiler var blygsamma 6 procentenheter .

Vid en kritisk granskning observeras att flertalet vocoderexperimenten från 60-70 talet inte skiljt på tonande respektive tonlöst talmaterial. Tonlösa konsonanter har sitt maxima i diskantområdet vilket kräver en annan typ av bärvåg (vitt brus) för att kunna återges på ett mer korrekt sätt. Vidare har flertalet av patienterna i dessa experiment haft mycket kraftig nedsättning i basområdet vilket försvårat identifieringen av flyttade taljudskomponenter .

Genomgående i dessa studier observeras dock att taluppfattningsresultatet förbättrats avsevärt med träning. Vocoderns potential inom frekvensförflyttning borde vara stor, detta beroende på att analysdelen och synthesdelen kan optimeras var för sig efter en specifik hörselsituation. Med digital signalprocessing har dessutom vocodertekniken utvecklats än mer, vilket ger fler möjligheter till forskning med modifierade vocoderscheman.

Frekvensskiftning

Frekvensskiftning är en annan variant av frekvensförflyttning som har utvärderats inom audiologin av Raymond, Proud och Fletcher. Denna forskning var som störst på 60-talet och innefattade såväl hörselskadade som normalhörande. Taluppfattningsexperiment utfördes med både nedåtskiftning/subtraktion och uppåtskiftning/addition. Därefter studerade man mer ingående nyttan av att flytta högfrekvent ljud till basområdet -subtraktiv frekvensskiftning- hos patienter med grav diskantnedsättning. Analog såväl som digital teknik har använts för att utvärdera den frekvensskiftande teknikens fördelar .

Tekniken som även gått under benämningen *klangumwandler*, är en linjär transponerande hörapparat som adderar/subtraherar ett fast värde, angivet i Hz, till alla spektrala komponenter i den aktuella signalen. T.ex. en sågtand med grundtonen 500 Hz har sin första överton vid 1 kHz, andra vid 1.5 kHz, tredje vid 2 kHz osv. Om vi nu skiftar denna sågtand 100Hz nedåt i frekvens erhåller vi en grundton på 400Hz, med första övertonen på 900Hz, andra på 1.4 och tredje på 1.9 kHz. I den utgående signalens spektrum har således 100 Hz subtraherats från alla ingående frekvenskomponenter i insignalen. Eftersom övertonerna inte är multiplar av grundtonens frekvens så innebär det att övertonerna inte är harmoniskt relaterade till grundtonen. Detta ger upphov till en dissonant ljudupplevelse hos lyssnaren. Däremot är det viktigt att poängtera om vi gör en subtraktion med ett värde som är harmoniskt relaterat med sågtandens grundfrekvens så kommer däremot slutresultatet fortfarande vara harmoniskt. I exemplet ovan skulle det bli att vi subtraherar 500 Hz från sågtandens alla spektrala komponenter .

Denna typ av frekvensskiftning ger inte upphov till någon bandbreddsreduktion, däremot kan spektrala komponenter överlagras vilket ger problem med vikningsdistorsion. Detta problem kan elimineras genom att man först filtrerar signalen innan frekvensskiftning appliceras. Om alla frekvenser under 5 kHz skiftas nedåt med 20 %, en skiftning med 1000 Hz, så måste alla ljudkomponenter under 1 kHz först filtreras bort. Denna typ av filtrering kommer dessvärre att försämra taluppfattningen av de språkljud som har sin energi i detta frekvensområde.

Fletcher studerade hur effekten av frekvensskiftning påverkar taluppfattningen hos normalhörande. Hans data visar på att frekvensskiftning såväl nedåt (subtraktion) som uppåt (addition) försämrar taluppfattningen.

Raymond and Proud utvärderade effekten av frekvensskiftning på 16 patienter med sluttande audiogram. Medelvärden på tontrösklarna var följande: 250 Hz=16dB, 500 Hz=14dB, 1 kHz=18dB, 2 kHz=39dB, 3 kHz=52dB, 4 kHz=64dB. Försökspersonerna fick 16 timmars träning utspridda över två månader tillsammans med en erfaren audionom. Träningen bestod dels av att lyssna på enkla berättelser och enstaka ord med frekvensskiftat tal, och dels en lyssningssituation där frekvensskiftat tal presenterades till ena örat och normalt tal till andra örat. Man använde såväl kvinnliga som manliga talare. Efter avslutad träning utförde man tre stycken olika taluppfattbarhetsmätningar: en manlig stämma frekvensskiftad med 400 Hz, en kvinnlig stämma frekvensskiftad med 750 Hz, och slutligen oprocessat tal hos en kvinnlig talare. All data visade på att försökspersonerna förbättrade sina resultat med hjälp av träning, dessutom var resultatet med den frekvensskiftade kvinnostämman bättre än med den manliga stämman. Ingen av deltagarna fick dock något signifikant bättre resultat med frekvensskiftat tal, oavsett hur det presenterats, jämfört med oskiftat tal före träningsperioden .

I slutet av 60-talet utförde Biondi och Biondi frekvensskiftande experiment med hjälp av digital teknologi. De rapporterade att ett par patienter med måttlig nedsättning var mycket negativt inställda till denna ljudbild och vägrade därför att fortsätta studien. Deltagarna med grav nedsättning lyckades däremot uppfatta en betydande samling ord, både i isolerad form, och ord presenterade i ett meningssammanhang. En av deltagarna med grav hörselnedsättning lyckades lära sig merparten av orden som presenterades. Problemet var att han endast kunde hålla ett 30-tal ord i minnet åt gången .

Experiment med frekvensskiftning har inte gett någon förbättrad taluppfattning hos normalhörande eller hos personer med lätt till måttlig hörselnedsättning. Detta har även gällt då man applicerat små portioner av frekvensskiftning. Problemet med frekvensskiftning kan dock vara dess spektrumförstörande inslag där inte deltonrelationen blir intakt efter genomförd frekvensförflyttning.

De positiva resultat som rapporterats kan relateras till omfattande talträning och till patienter med grav hörselnedsättning. Flertalet normalhörande och patienter med lätt till måttlig hörselnedsättning har dessutom uttryckt svårigheter att tolka det frekvensskiftade ljudet som talljud. Språkljudens akustiska korrelat har således starkt skiljt sig från dess ursprungliga utseende vilket troligtvis försvårat identifieringen.