KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK – ESAT
Kardinaal Mercierlaan 94, B-3001 Heverlee

in samenwerking met
LABO VOOR EXPERIMENTELE OTORHINOLARYNGOLOGIE
Kapucijnenvoer , B-3000 Leuven

CODING TEMPORAL INFORMATION IN SOUND PROCESSORS FOR COCHLEAR IMPLANTS

Jury:
Prof. Dr. ir. E. Aernoudt, voorzitter
Prof. Dr. J. Wouters, promotor
Prof. Dr. ir. M. Moonen, promotor
Prof. Dr. ir. J. Vandewalle
Prof. Dr. ir. D. Van Compernolle
Prof. Dr. G. Govaerts
Prof. Dr. L. Feenstra
Prof. Dr. ir. S. Peeters (UIA, Antwerpen)
Dr. R. P. Carlyon (MRC/CBU, Cambridge)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Luc GEURTS

U.D.C. 681.3*E4

Mei 2000

© Katholieke Universiteit Leuven
Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (België)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm, elektronische of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without the prior written permission from the publisher.

D/2000/7515/11
ISBN 90-5682-241-1

Voor Leentje en Matthijs

Voorwoord

Dit proefschrift is het resultaat van enkele jaren onderzoek in het domein van cochleaire implantaten of hoortoestellen voor dove mensen. De drie studies die centraal staan, hebben één ding gemeen: allen behandelen ze tijdsafhankelijke effecten in de geluidsverwerking voor cochleaire implantaten en de daarmee gepaard gaande invloed op de waarneming van de geïmplanteerde persoon.

Wetenschappelijk onderzoek doe je niet alleen. Elke beginnende onderzoeker heeft nood aan mensen die hem bijstaan, raad geven, opbouwende kritiek geven en oppeppen. Deze mensen wil ik hier even in de schijnwerpers plaatsen.

In de eerste plaats is dat prof. Jan Wouters. Enkele jaren geleden werd aan deze universiteit een document bedeeld, getiteld: "Profiel van de goede promotor"[1]. Ik heb dit altijd zorgvuldig bewaard met de bedoeling om het voor deze gelegenheid terug op te dissen. Met een kritische geest, die mij dankzij hem al vrij goed eigen is, heb ik zijn kenmerken getoetst aan de vermelde criteria. Het is niet aan mij om mijn promotor een quotering te geven, maar ik zou niet spaarzaam geweest zijn. Hij was altijd beschikbaar om over nieuwe opstellingen, experimenten en meetgegevens te discussiëren. Regelmatig stapte ik vol frustratie zijn bureau binnen met tegenvallende resultaten in de hand, maar vol goede moed en nieuwe ideeën voor verdere experimenten stapte ik weer buiten. Dankzij hem beschikte ik over de luxe om zo goed als voltijds aan mijn doctoraat te werken, wat niet iedereen gegeven is. Belangrijk is ook zijn bijdrage aan de goede sfeer op het lab. Er werd hard gewerkt, maar er waren tussendoor erg plezierige momenten.

De taak van tweede promotor, verbonden aan de faculteit toegepaste wetenschappen, heeft prof. Marc Moonen op zich genomen. Ik heb getwijfeld of deze thesis, het resultaat van multidisciplinair onderzoek, wel geschikt was als *ingenieurs*thesis, maar dankzij zijn morele steun en opbouwende kritiek op de eerste versie, weet ik dat ik, als ingenieur, trots mag zijn op dit werk.

De vier juryleden die ook deel uitmaakten van het begeleidingscomité, wil ik ook danken voor de opbouwende kritiek op de vorige versie van dit proefschrift. Elk van hen heeft echter nog een ander belangrijk aandeel in het tot stand komen ervan. Het implantaat dat gebruikt werd voor alle testen, is het Laura-implantaat, het geesteskind van prof. Stef Peeters. Dankzij zijn

[1] http://www.kuleuven.ac.be/admin/oc/niv3pbis/oc-i91.htm

enthousiasme en doorzetting was en is Vlaanderen een koploper op het gebied van onderzoek en ontwikkeling van cochleaire implantaten. Hij heeft mijn prestaties steeds van nabij gevolgd, en ook de nodige technische informatie over het implantaat verschaft om alle experimenten te kunnen uitvoeren. Prof. Georges Govaerts wil ik danken voor zijn kritische raadgevingen bij het uitvoeren van spraaktesten bij geïmplanteerden. Zulke testen houden inderdaad veel meer in dan enkele woordjes aanbieden en tellen hoeveel er correct geïdentificeerd zijn. Prof. Joos Vandewalle en prof. Dirk Van Compernolle heb ik mogen kennen als uitmuntende lesgevers tijdens mijn ingenieurs-opleiding. De kennis over signaal- en spraakverwerking die zij me hebben bijgebracht, bleek onmisbaar voor dit werk. De twee overige juryleden wil ik ook danken voor hun bijdrage. Prof. Louw Feenstra dank ik voor de ideale omstandigheden waarin ik mijn onderzoek heb mogen uitvoeren aan zijn afdeling. Finally, I would like to thank Dr. Bob Carlyon for his co-operation on the study, described in the fifth chapter, for which his experience and inspiration was indispensable.

Ik wil ik ook graag mijn uitzonderlijke waardering uitspreken voor de vele cochleair-implantgebruikers die hebben deelgenomen aan de vaak moeilijke en eentonige experimenten. Ik heb er steeds van versteld gestaan met welk enthousiasme en uithouding ze urenlang achter de computer hebben doorgebracht om vakjes op het scherm aan te duiden. Hun enige motivatie was vaak dat ze eens iets wilden terugdoen voor het feit dat ze weer konden horen, en dat ze ons daarmee kennis en inzicht konden verschaffen om het systeem nog te kunnen verbeteren.

Ik ben al die jaren in het lab omringd geweest door zeer fijne collega's, die mij rechtstreeks of onrechtstreeks geholpen hebben. Zo heb ik kunnen profiteren van de wetenschappelijke ervaring van Astrid van Wieringen, die bovendien enkele taken op zich heeft genomen die ik eigenlijk zelf had moeten doen, zoals het afnemen van testen en het schrijven van teksten. Valerie Vandenbroeck heeft mij zeker zoveel vooruitgeholpen met mijn onderzoek, als ik met haar licentiaatsthesis, waarin voor het eerst nieuwe algoritmen bij cochleair-implantgebruikers werden geëvalueerd. Jeff Vanden Berghe dank ik voor de vele discussies die we voerden, hetgeen het resultaat alleen maar ten goede kwam. Dennis Kemps stond altijd paraat om één van de vele technische problemen zo snel en zo goed mogelijk op te lossen. Dr. Christian Desloovere dank ik voor zijn deskundige uitleg tijdens de operaties, waarbij ik over zijn schouder mocht meekijken. Van Etjen Swannet heb ik vooral geleerd hoe je met slechthorenden, en dus ook met cochleair geïmplanteerden, moet communiceren. Lea Van Caneghem heeft mij steeds geholpen bij vele kleine, vervelende dingetjes die geregeld moesten worden. Allemaal, heel erg bedankt. Ik wil ook van de

gelegenheid gebruik maken om de nieuwe collega's Jean-Baptiste Maj, Ann Spriet en Els Gelders, succes te wensen met hun verdere onderzoekscarrière.

Een groot deel van de proefpersonen zijn geïmplanteerd door Dr. Erwin Offeciers in het Medisch Instituut Sint-Augustinus te Wilrijk. Ik wil hem en Dr. Paul Govaerts speciaal danken voor het bezorgen van alle nodige informatie over deze mensen. Ook wil ik Jan Janssen, Günther Van der Borght en Jef Snellings van Cochlear Technology Centre Europe, het vroegere Philips Hearing Implants, danken voor alle technische ondersteuning in verband met het Laura-implantaat.

De financiering van dit werk werd gedeeltelijk gedragen door het Vlaams Instituut voor de bevordering van het Wetenschappelijk Technologisch onderzoek in de industrie (I.W.T.) en door het Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (F.W.O.).

Mijn ouders stonden in voor de financiële steun en de coaching tijdens mijn studies. Dat dit niet zo evident is, heb ik pas nadien voldoende leren beseffen. Ik wil hen bij deze dan ook de dank betuigen die ze toen misschien een beetje meer mochten krijgen. Dan is er ook nog Matthijs, die waarschijnlijk nog geen benul heeft van wat papa nu doet, maar onbewust heeft hij me alle problemen helpen relativeren door zijn guitenstreken en lachsalvo's.

De laatste plaats is een bijzondere plaats, en die is voor Leentje. Zij heeft mij destijds ten volle gesteund toen ik nog twijfelde of ik aan een doctoraatsstudie zou beginnen. Vaak verloste ze mij van het tobben over tegenvallende resultaten. Zij heeft me ook geholpen bij het maken van de uitgebreide Nederlandse samenvatting, die bedoeld is om de leek duidelijk te maken wat ik zoal heb gedaan. Ik hoop dat ze mij nog lang kan steunen in alles wat ik doe. Deze thesis draag ik dan ook op aan haar en aan Matthijs.

Abstract

This thesis includes three studies in the field of cochlear implants, hearing aids for the deaf. In each study, a temporal effect in the neural representation of a sound is related to the sound processing strategy and the influence of this effect on the perception of the implantee is investigated. Prior to the experiments, a computer-based set-up is developed to conduct psychophysical experiments and to design and evaluate new speech processing algorithms with users of the Laura cochlear implant. In the first study, it is investigated how one phenomenon of the physiology of normal hearing, namely rapid neural adaptation, can be included in a CIS-like speech processor and whether it provides a higher speech intelligibility. Some phonemes are identified better, and although the increases of the scores are small, they are significant and obtained without training or feedback. The topic of the second study is the coding of the fundamental frequency of voiced phonemes. It is shown that the temporal cue, manifested as fluctuations in the speech envelope, is very important for pitch perception, and that optimal performance is already obtained for modulation depths of around 20 %, provided that many channels are stimulated. In the last study, it is demonstrated for the first time that implantees are sensitive to very small timing differences, from 200 µs, between two pulse trains on widely separated channels.

Korte Inhoud

Dit proefschrift bevat drie studies in het domein van cochleaire implantaten, hoorapparaten voor dove mensen. In elke studie wordt het verband onderzocht tussen een temporeel effect in de neurale representatie van een geluid en het geluidsverwerkingsalgoritme, en de invloed van dit effect op de waarneming van de geïmplanteerde persoon. Voor de aanvang van de experimenten werd een computer-gebaseerde opstelling gebouwd, waarmee men gemakkelijk psycho-fysische experimenten kan uitvoeren en nieuwe geluidsverwerkingsalgoritmen kan evalueren bij proefpersonen. Hierbij kan gebruik gemaakt worden van alle mogelijkheden van het Laura-implantaat. In de eerste studie wordt een nieuw algoritme voorgesteld waarin een fysiologisch temporeel verschijnsel van het normale gehoor, nl. neurale adaptatie, wordt nagebootst. Uit de resultaten van de spraakidentificatietesten blijkt dat de proefpersonen bepaalde klanken beter verstaan. In de tweede studie wordt de codering van de fundamentele frequentie van klinkers onderzocht. De resultaten tonen het grote belang aan van de temporele codering, die zich voordoet als amplitudevariaties van de klankomhullende. Optimale prestaties worden bekomen vanaf modulatiediepten van ongeveer 20%, op voorwaarde dat vele kanalen gelijktijdig gestimuleerd worden. In de laatste studie wordt voor het eerst aangetoond dat cochleair-implantgebruikers zeer kleine tijdsverschillen, vanaf 200 µs, tussen twee pulstreinen op ver van elkaar verwijderde kanalen kunnen waarnemen.

Coderen van temporele informatie in geluidsverwerkingsstrategieën voor cochleaire implantaten

Uitgebreide Nederlandse samenvatting

1 Inleiding

1.1 Een remedie tegen doofheid

De handicap van doofheid wordt vaak onderschat. Een dove kan dan wel probleemloos zelfstandig handelen, maar het onvermogen om gemakkelijk te communiceren met anderen bemoeilijkt alle sociale contacten. De dove *hoort* er niet bij. Er zijn enkele manieren om doofheid te omzeilen, maar deze hebben aanzienlijke nadelen: het *lipbeeld* bevat niet alle informatie van het gesproken woord, de meeste normaalhorenden spreken geen *gebarentaal*, en alles *opschrijven* gaat te traag. Bovendien zijn dove mensen vaak moeilijk te verstaan, omdat ze zelf niet horen wat ze zeggen. Een klassiek hoortoestel, dat het geluid met een microfoon opvangt, versterkt en dan terug aanbiedt aan het oor, volstaat niet voor de ernstig slechthorenden.

Als een oor doof is, functioneren één of meerdere fysiologische processen niet meer naar behoren. In een normaal werkend oor wordt geluid opgevangen door de oorschelp, omgezet in een mechanische trilling van het trommelvlies, de gehoorbeentjes en het basilair membraan in het slakkenhuis, en nadien in een reeks zenuwprikkels die naar de hersenen gaan om daar *geïnterpreteerd* te worden. Met een cochleair implant (CI) kan men gehoorsensaties opwekken door deze processen over te nemen. Eén van de onderdelen is een reeks van elektroden die in het slakkenhuis, of cochlea, zijn ingebracht tijdens een chirurgische ingreep. Door een elektrische stroom te laten vloeien tussen twee elektroden kunnen de gehoorzenuwvezels geactiveerd worden, en neemt de geïmplanteerde persoon geluid waar. Dit werd reeds vastgesteld op het einde van de 18e eeuw (Volta, 1800), maar verregaand onderzoek in dit domein begon pas na 1970. De eerste klinische toestellen waren beschikbaar in het begin van de

jaren 1980. De gebruikte technieken waren nog vrij experimenteel en de resultaten waren dan ook beperkt. Geluid waarnemen was mogelijk, maar spraak verstaan nauwelijks. Sindsdien vergrootte fysiologisch en psychofysisch onderzoek het inzicht in het effect van elektrische stimulatie van de gehoorzenuw. Hedendaagse implantaten stellen de gebruikers dan ook in staat om spraak te verstaan. Echter, slechts enkele CI-dragers verstaan alles wat gezegd wordt, en voor velen vergt het nog steeds een grote inspanning om een gesprek te volgen.

In dit proefschrift worden drie studies beschreven waarin enkele aspecten van CI technologie nader onderzocht worden. In elke studie wordt de invloed van een temporeel effect in het geluidsverwerkingsalgoritme op de waarneming van de geïmplanteerde persoon onderzocht. Zo een algoritme bepaalt hoe het geluid, dat opgevangen wordt met een microfoontje, wordt omgezet in een reeks elektrische pulsen die aangeboden worden aan de elektroden. Hierbij tracht men de zenuwvezels te laten reageren alsof het geluid akoestisch werd aangeboden aan een normaal werkend oor.

1.2 Enkele gegevens over de werking van het oor

Figuur 1 toont de drie onderdelen van het oor: het buitenoor, het middenoor en het slakkenhuis of de cochlea. In een normaal werkend oor wordt geluid als volgt verwerkt. De oorschelp vangt het geluidssignaal op en de gehoorgang leidt het naar het trommelvlies dat zal meetrillen. De gehoorbeentjes in het middenoor brengen deze mechanische trilling over naar het ovale venster van de cochlea, zodat de drukgolf zich voortplant in de met vloeistof gevulde cochlea. Deze is door twee membranen onderverdeeld in drie delen (fig. 2A). Op het onderste membraan, het basilair membraan, bevindt zich het orgaan van Corti, dat de haarcellen bevat (fig. 2B). De drukgolf in de cochlea doet het basilair membraan op en neer bewegen, zodat de haarcellen open en dicht gaan aan de stereocilia. Hierdoor verandert de chemische samenstelling in de haarcel, zodat de daaraan verbonden gehoorzenuwvezel zal afvuren. Ze zendt dan een puls naar de hersenen via de gehoorzenuw.

De verdeling van de zenuwactiviteit over de volledige lengte van de cochlea (± 35 mm) is afhankelijk van het geluid. Indien dit een lage toon is, zullen vooral de zenuwvezels aan het uiteinde van de cochlea, de apex, actief zijn. Een hoge, scherpe toon zal vooral de zenuwvezels aan de basis, dus dicht bij het ovale venster, activeren. Dit noemt men de *tonotopische*

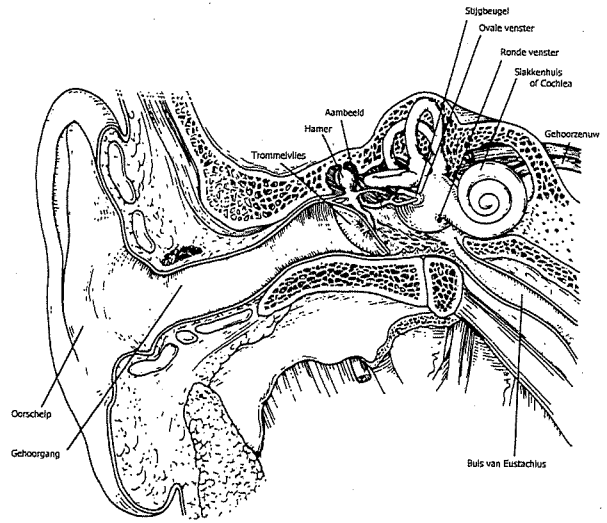


Fig. 1 Het buitenoor, het middenoor en het slakkenhuis van de mens. Uit Pickles (1988, Fig. 2.1)

organisatie van de cochlea. Een spraaksignaal is opgebouwd uit een mengeling van lage, hoge en middentonen, en het is juist de specifieke combinatie ervan die de klank bepaalt en de luisteraar in staat stelt om de klank te identificeren.

Het is ook belangrijk om te weten hoe de gehoorzenuw reageert bij verschillende geluidsintensiteiten. Zoals eerder vermeld, zal de gehoorzenuw pulsen naar de hersenen sturen. De intensiteit van één zo'n puls is altijd dezelfde, m.a.w. een luidere klank zal de puls niet groter maken. Wat dan wel gebeurt, is dat er meerdere pulsen gegenereerd worden. Een bepaalde zenuwvezel zal meer pulsen genereren en andere zenuwvezels, die bij een lagere intensiteit niet aktief waren, zullen ook beginnen afvuren. Kort gezegd, de totale afvuursnelheid stijgt met de geluidsintensiteit.

Bestaande cochleair-implantaat systemen houden rekening met bovenvermelde eigenschappen van het normaalwerkend oor. De volgende paragraaf beschrijft hoe dit gerealiseerd wordt.

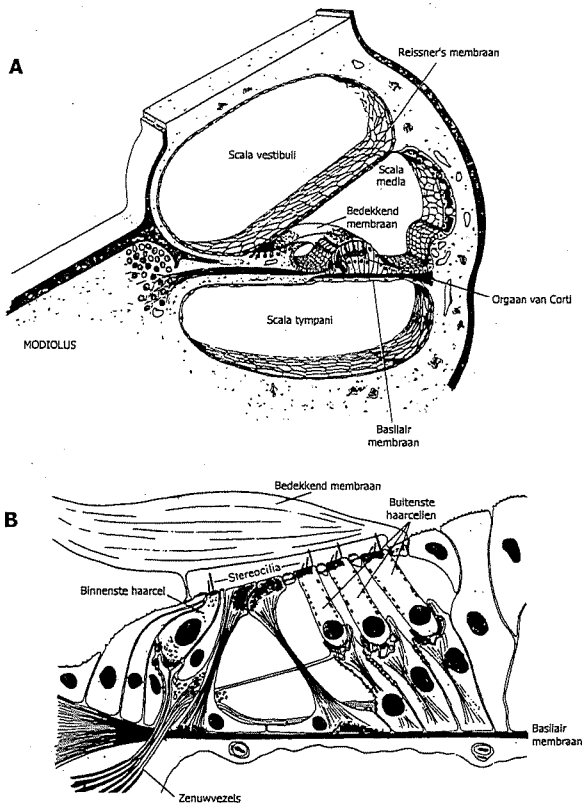


Fig. 2 (A) Een dwarsdoorsnede van de cochlea toont de drie delen gescheiden door twee membranen. (B) Het orgaan van Corti. De stereocilia buigen bij het bewegen van het basilair membraan. De zenuwvezels zijn verbonden met de haarcellen. Uit Pickles (1988, Fig. 3.1)

1.3 Cochleaire implantaten

Het apparaat bestaat altijd uit twee delen: een intern deel dat tijdens een chirurgische ingreep wordt ingebracht, en een externe processor (fig. 3). Bij de huidige klinische apparaten gebeurt de koppeling tussen beide delen transcutaan[2]. De processor stuurt dan informatie via een uitwendige zendspoel en een inwendige ontvangstspoel naar het intern deel. De essentiële component hiervan is een reeks van elektroden die in het slakkenhuis geplaatst wordt via het ronde venster (fig. 1). Een elektrische stroom tussen twee of meerdere elektroden kan gehoorzenuwvezels in de nabijheid activeren. De omzetting van geluid, dat opgevangen wordt met een microfoontje in de oorhanger, in een reeks elektrische signalen gebeurt in de externe processor. Dit is in feite een kleine computer die de berekeningen voor deze omzetting uitvoert volgens een bepaald programma of algoritme. De informatie over de elektrische signalen gaat via de twee spoelen naar een inwendige chip, die de signalen zal genereren.

Een stroom vloeit steeds van een bepaalde elektrode naar één of meerdere andere elektroden. De combinatie van deze elektroden noemt men een *kanaal*. Men spreekt van *monopolaire* stimulatie als elke elektrode gecombineerd wordt met dezelfde verafgelegen elektrode. Daarentegen, als kanalen telkens gevormd worden door twee dicht bij elkaar liggende elektroden, dan spreekt men van bipolaire stimulatie. Het eerste heeft als voordeel dat er minder stroom nodig is om een bepaald aantal zenuwen te activeren, het laatste heeft als voordeel dat de stimulatie selectiever is. De reagerende zenuwvezels liggen dan namelijk dichter bij elkaar. Merk hier echter op dat met het beperkt aantal elektroden (tot 32) in de hedendaagse implantaten nooit de selectiviteit van een gezond oor (15000 haarcellen) kan bereikt worden.

In de huidige apparaten wordt meestal gebruik gemaakt van pulsvormige niet-overlappende signalen. De pulsvorm is symmetrisch of bifazig: eerst wordt een stroom enkele tientallen microseconden in een bepaalde richting gestuurd, en dan wordt een even grote stroom in de andere richting gestuurd gedurende dezelfde tijd. De signalen zijn niet-overlappend, hetgeen wil zeggen dat er nooit twee of meerdere pulsen tegelijk uitgestuurd worden. Op deze manier voorkomt men dat het ene signaal het andere gaat beïnvloeden.

[2] transcutaan = over de huid, in tegenstelling tot percutaan = door de huid

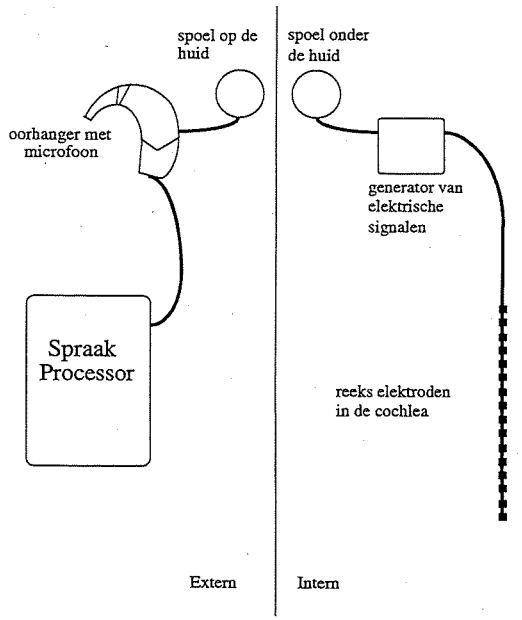


Fig. 3 Schematische weergave van een cochleair-implantaat systeem met trancutane koppeling. Het externe deel staat links, het interne rechts.

De taak van het geluidsverwerkingsprogramma is het opgevangen geluid om te zetten in een reeks pulsen voor elk kanaal. In de loop van de cochleaire-implantaatgeschiedenis zijn verschillende programma's ontwikkeld en getest. Allen hadden ze uiteraard tot doel om het spraakverstaan van de geïmplanteerde zo goed mogelijk te herstellen. Daarom wordt er getracht de fysiologie van het gezond oor na te bootsen.

In één van de eerste stappen van het programma wordt de tonotopische organisatie van het slakkenhuis gemodelleerd door het signaal op te splitsen in verschillende zogenaamde frequentiebanden. Zoals eerder vermeld, bestaat geluid in het algemeen uit een samenstelling van hoge, lage en middentonen, en deze worden in deze stap gescheiden. Dit kan vergeleken worden met een *equalizer* in een

stereo-installatie, waar de versterking van elke frequentieband apart kan geregeld worden. Het minimum aantal frequentiebanden wordt bepaald door het aantal bruikbare kanalen van de geïmplanteerde, het maximum is afhankelijk van het type toestel en varieert tussen 8 en 22. Het signaal in elke band wordt na nog enige verwerking aangeboden aan het overeenkomstige kanaal: de laagste tonen aan het diepst gelegen of meest apicale kanaal, de hoogste tonen aan het kanaal dat het dichtst bij het ronde venster ligt. Op deze manier wordt de tonotopie gesimuleerd.

De intensiteit kan men variëren door ofwel de pulsamplitude te verhogen, m.a.w. een grotere stroom te laten vloeien, ofwel door de pulsbreedte te verhogen, m.a.w. de stroom langere tijd te laten vloeien. Dus, als er een sterk signaal in een bepaalde frequentieband aanwezig is, dan zullen grote of brede pulsen gebruikt worden. Als gevolg zullen meer zenuwvezels reageren, en zullen ze dit ook vaker doen. De toon in dat kanaal zal luider klinken. Belangrijk is wel dat de pulsen niet te klein of te groot zijn. In het eerste geval zal de geïmplanteerde niets horen, in het laatste geval klinkt het signaal veel te luid. Daarom zal er steeds eerst een zogenaamde fitting gebeuren om het apparaat aan te passen aan de patiënt. Hierbij wordt voor elk kanaal de kleinste pulsgrootte gezocht die de geïmplanteerde nog kan waarnemen, en de grootste pulsgrootte die qua luidheid[3] nog kan worden verdragen. Het verschil tussen beide pulsen noemt men het *dynamisch bereik.*

Het algoritme dat als basis diende voor het hierna beschreven eigen onderzoek is het zogenaamde *Continuous Interleaved Sampling* algoritme (CIS), waarover eerst door Wilson *et al.* (1991) gerapporteerd werd. Het is geïmplementeerd in bijna alle klinisch gebruikte apparaten. Figuur 4 toont het blokschema.

Na de voorverwerking wordt het signaal opgesplitst in verschillende frequentiebanden met behulp van zogenaamde banddoorlaatfilters. Vervolgens wordt in elke band de *omhullende* van het gefilterde signaal bepaald. Deze is een maat voor de intensiteit van het signaal. In de compressiestap wordt het signaal aangepast aan het dynamisch bereik van het overeenkomstige kanaal, zodat het signaal nooit onhoorbaar of te luid zal klinken. In de laatste stap zal de gecomprimeerde omhullende een continue

[3] Merk op dat "luidheid" een psychische maat is. De waarde ervan wordt door een subjectieve luisteraar gegeven. "Geluidssterkte" daarentegen is een fysische maat, die met een apparaat kan worden gemeten.

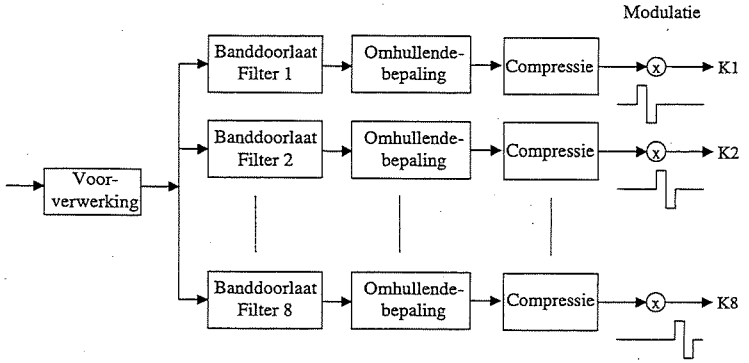


Fig. 4 Blokschema van het CIS algoritme met 8 kanalen.

reeks van pulsen moduleren. Dit wil zeggen dat de amplitude van elke puls overeenkomt met de grootte van de omhullende op dat moment. De *pulstreinen* in de verschillende kanalen zijn in elkaar genesteld (*interleaved*), wat betekent dat als een puls in een bepaald kanaal eindigt, er een nieuwe begint op een ander kanaal. Op die manier is de stimulatie ook continu: er is altijd wel ergens een puls bezig. De snelheid waarmee de pulsen elkaar opvolgen is afhankelijk van de pulsbreedte. Bij een brede puls zal het immers langer duren eer de volgende kan starten. Typische waarden voor de CIS strategie liggen rond de 1000 pulsen per seconde in elk kanaal. Figuur 5 toont een voorbeeld van het uitgangssignaal van een 8-kanaals CIS algoritme. De kanalen zijn genummerd van lage naar hoge frequentiebanden. Het diepst in de cochlea gelegen kanaal, dat dus tonotopisch overeenkomt met de laagste frequenties, heeft dus nummer 1. Het ingangssignaal is het woord "som". In het tijdsinterval 200-300 ms worden vooral kanalen 7 en 8 gestimuleerd. Dit komt overeen met de letter "s" die samengesteld is uit hoge, scherpe tonen. Na 350 ms begint de "o" die vooral kanalen 1 tot en met 5 activeert, en nadien (rond 500 ms) komt de "m" die bijna uitsluitend lage, doffe tonen bevat.

De proefpersonen die deelgenomen hebben aan de in deze thesis beschreven studies, zijn allen gebruikers van het Laura cochleair implant systeem (Peeters *et al.*, 1993), een product van Philips Hearing Implants. De elektrodenreeks bevat 8 bipolaire kanalen (fig. 6). In de meeste gevallen is de pulsbreedte 40 µs per faze en een nieuwe puls kan gegenereerd worden elke 100 µs. De maximale stroom bedraagt ongeveer 1,5 mA. In het

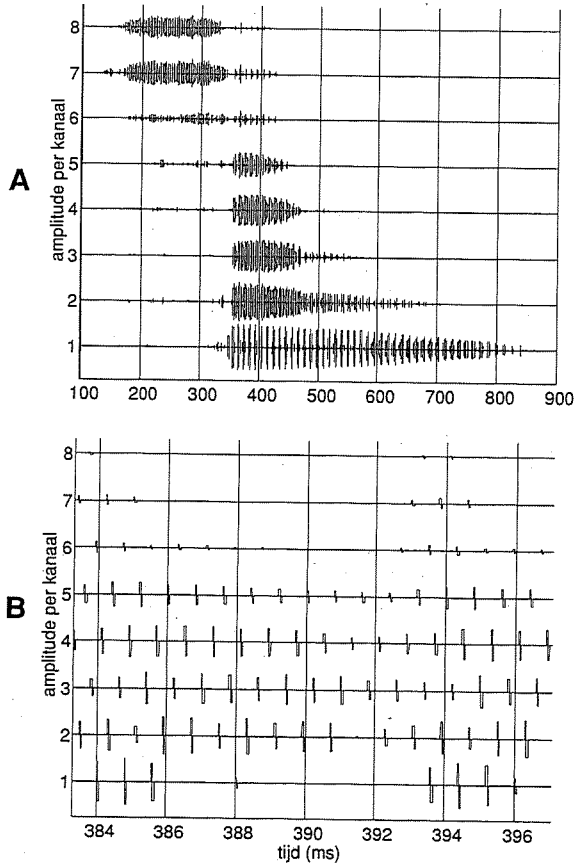


Fig. 5 (A) Uitgangssignaal van het CIS algoritme voor het woord "som". De pulsamplitudes in elk kanaal komen overeen met de omhullende in elke frequentieband. (B) Een detail van het zelfde uitgangssignaal, waarop de verweven, continue pulstreinen te zien zijn.

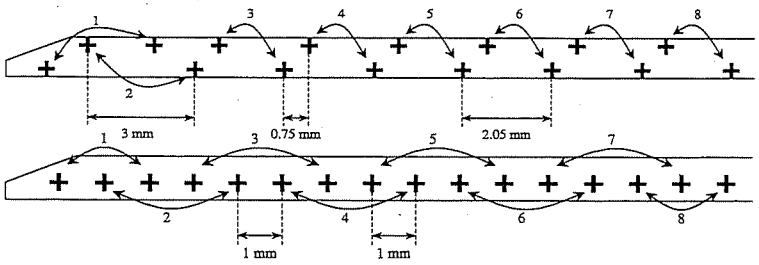


Fig. 6 Schematische weergave van twee versies van de Laura elektrodenreeks (niet op schaal getekend): type 5.6 boven en type 5.7/5.8 beneden. De kruisjes stellen de elektroden voor, en de pijltjes verbinden de elektroden die een bipolair kanaal vormen.

dagelijks leven gebruiken deze patiënten de Lauraflex processor, waarin een variante van het CIS algoritme geïmplementeerd is. Echter, voor de experimenten werd deze niet gebruikt. In de plaats daarvan werd een nieuwe computergestuurde opstelling gebruikt.

2. APEX

2.1 Overzicht

Om gemakkelijk allerlei experimenten te kunnen samenstellen en uitvoeren, werd een draagbare computergebaseerde opstelling gebouwd. APEX, een acroniem voor *computer Application for Psycho-Electrical eXperiments*, is een gebruiksvriendelijk hulpmiddel om testen af te nemen bij dragers van het Laura cochleair implant. Deze testen dienen ofwel om bepaalde nieuwe geluidsverwerkingsalgoritmen te evalueren, ofwel om na te gaan in welke mate de proefpersoon een verschil kan horen tussen stimuli waarvan één of meerdere kenmerken zijn gewijzigd. In het laatste geval spreekt men van psychofysische experimenten. Zonder enige kennis van computerprogrammatie kunnen nieuwe experimenten worden opgesteld en uitgevoerd en kunnen de nodige stimuli worden gecreëerd. Deze stimuli kunnen alle mogelijke vrijheidsgraden van het Laura implant benutten. Een schematische voorstelling geeft aan hoe alles in zijn werk gaat (fig. 7).

Alle gegevens die het experiment specifiëren, zitten in een eenvoudig tekstbestand, dat met eender welke tekstverwerker kan worden aangemaakt

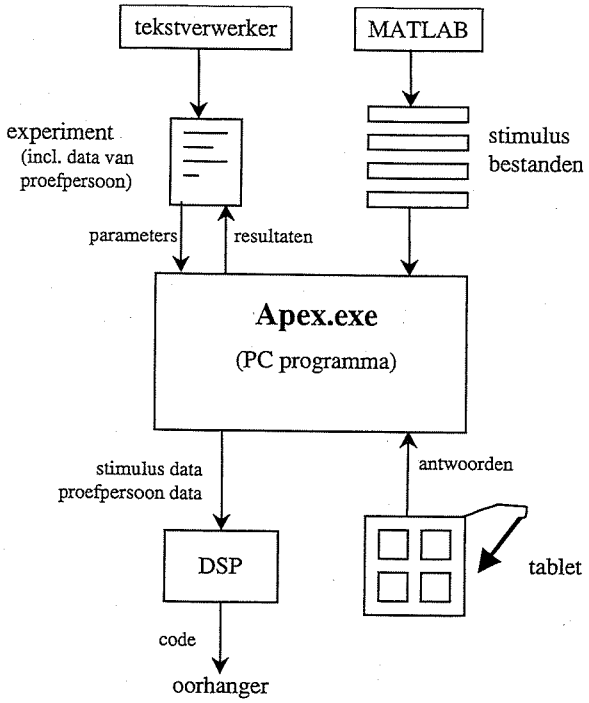


Fig. 7 Schema van APEX. De pijlen duiden de richting aan waarin de aangegeven informatie stroomt.

of aangepast. Deze gegevens bevatten o.a. de namen van de bestanden die de stimuli definiëren, enkele patiëntgebonden gegevens zoals het dynamisch bereik van de gebruikte kanalen, en parameters die de test karakteriseren (zie verder).

Om de elektrische stimuli te genereren kan de onderzoeker gebruik maken van het rekenpakket Matlab. Hiervoor zijn enkele routines geschreven die het mogelijk maken om op een snelle, eenvoudige manier deze stimuli samen te stellen. Het is ook mogelijk om een opgenomen en gedigitaliseerd geluidssignaal te verwerken tot een elektrische stimulus. Zo een routine vervangt dus de spraakprocessor die de patiënt normaal gebruikt. Het voordeel van deze werkwijze is dat

Matlabroutines gemakkelijk te programmeren en te *debuggen*[4] zijn. De stimulusgegevens worden opgeslagen in computerbestanden. Op dit niveau zijn deze gegevens nog onafhankelijk van de proefpersoon. Ze kunnen dus voor verschillende proefpersonen gebruikt worden.

Het PC programma *Apex.exe* controleert het volledige experiment. Het leest de gegevens uit het tekstbestand, bepaalt daaruit wanneer en welke stimulus naar de patiënt gestuurd wordt, leest het juiste stimulusbestand in en stuurt het naar een DSP-kaart[5] die voor de connectie tussen computer en proefpersoon zorgt (zie verder). De patiënt maakt een bepaald oordeel op basis van de aangeboden stimulus of stimuli, zoals bv. "welk woord is het?" en "welke van de twee stimuli klinkt luider?". Hij geeft vervolgens zijn antwoord in via een tabletje waarop de verschillende alternatieven staan. Het computerprogramma registreert alle antwoorden, doet er eventueel nog wat statistische bewerkingen mee, en slaat de resultaten op in het oorspronkelijke tekstbestand.

2.2 Apparatuur

Figuur 8 toont een foto van de gehele opstelling. Een schootcomputer wordt ingeschoven in een zogenaamd *docking station*, waarin zich een DSP-kaart bevindt. Aan de linkerkant is de oorhanger bevestigd aan een connector, die verbonden is met de DSP-kaart. Het grafisch tablet en de bijhorende pen bevinden zich rechts. De gehele opstelling is draagbaar, zodat de testen bij de proefpersonen thuis kunnen worden afgenomen.

2.3 Programma's

De software kan opgedeeld worden in drie delen: de Matlabroutines, het PC programma, en het DSP programma. De Matlabroutines dienen zoals eerder vermeld om stimuli te creëren. Het Laura implantaat kan elke 100 µs op eender welk kanaal een bifazige puls (40 µs per faze) met eender welke amplitude genereren. Het is wel zo dat de meeste stimuli een eenvoudige, logische structuur hebben, in welk geval een kleine set van korte routines al volstaat. Met de huidige set routines kan men bv. zeer gemakkelijk regelmatige pulstreinen genereren op verschillende kanalen, deze pulstreinen moduleren met andere signalen en stimuli achter elkaar plakken. Ook signaalverwerkingsalgoritmen zijn niet moeilijk te implementeren, aangezien

[4] debuggen is het opsporen en corrigeren van fouten in een computerprogramma.
[5] DSP = *Digital Signal Processing* of dus digitale signaalverwerking.

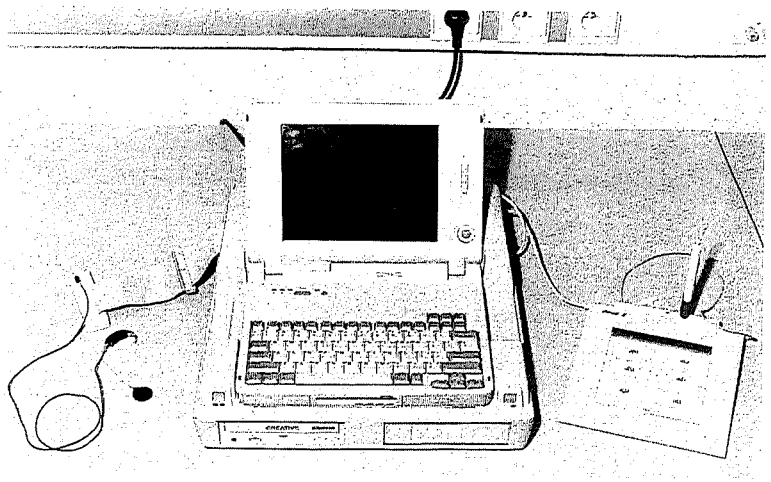


Fig. 8 De apparatuur van APEX

veel gebruikte typische functies voorhanden zijn in Matlab of in één van zijn zogenaamde *toolboxen*, zoals bv. het ontwerp en het uitvoeren van een filteroperatie (fig. 4).

Het PC programma controleert het verloop van de test. Psychofysische experimenten hebben een verschillend verloop naargelang de gebruikte procedure. De meest gangbare procedures kunnen met APEX uitgevoerd worden. De belangrijkste voor deze thesis zijn:

- Identificatie: de proefpersoon moet het woord dat hij gehoord heeft aanduiden of opschrijven. In het eerste geval komt het woord uit een set die de proefpersoon kent (gesloten-set), in het laatste geval kan het eender welk woord zijn (open-set).

- Balanceren: de proefpersoon moet de intensiteit van een stimulus aanpassen tot deze even luid klinkt als een andere stimulus.

- Adaptieve procedure: de proefpersoon moet het verschil waarnemen tussen twee of meerdere stimuli, bv. "duid de stimulus aan die het hoogste klinkt". Naarmate de test vordert en zolang de proefpersoon

het verschil kan horen, wordt dit verschil kleiner gemaakt, zodat de test dus moeilijker wordt.

- Constante stimuli: ook hier moet de proefpersoon het verschil trachten te horen tussen twee of meerdere stimuli, maar in dit geval is de volgende reeks van stimuli niet afhankelijk van het laatste antwoord van de patiënt. Moeilijk en gemakkelijk te onderscheiden stimuli wisselen elkaar willekeurig af.

Het DSP programma tenslotte zorgt voor de aansturing van het implantaat op basis van de gegevens die van het PC programma komen. Een belangrijke taak is het aanpassen van de stimuli aan de patiënt. De stimuli die in Matlab berekend zijn, zijn immers onafhankelijk van het dynamisch bereik van de gebruikte kanalen van de patiënt. Deze gegevens staan in het tekstbestand dat de test definieert. Het PC programma stuurt zowel de proefpersoongegevens als de stimulusgegevens naar het DSP programma, en deze combineert beide tot een stimulus die geschikt is voor de patiënt. Een tweede belangrijke taak van het DSP programma is het genereren van de code die het implantaat aanstuurt. Elk stukje code bevat informatie over de amplitude en het kanaal van één elektrische puls. De code wordt via oorhanger en de twee spoelen verzonden naar de interne chip.

3. Imiteren van neurale adaptatie

3.1 Neurale adaptatie

Zoals eerder vermeld, tracht een cochleair implant enkele functies van het oor over te nemen. Tot op heden is er echter weinig aandacht besteed aan het adaptatie-effect van de gehoorzenuw. Dit effect kan men kort omschrijven als de veranderende reactie van een gehoorzenuwvezel op een constante stimulus. Biedt men bv. een akoestische toon aan die niet verandert in geluidssterkte, dan zal de zenuwvezel in het begin aan een hoge snelheid afvuren, en daarna zal die snelheid afnemen. Dit effect is zichtbaar in figuur 9: de afzonderlijke grafieken tonen de afvuursnelheid (verticaal) in functie van de tijd (horizontaal). De geluidsintensiteit in dB SPL[6] is weergegeven bij elke grafiek. Reeds vanaf 10 dB SPL, hetgeen voor een

[6] dB SPL = *deciBell Sound Pressure Level* of deciBell Geluidsdrukniveau

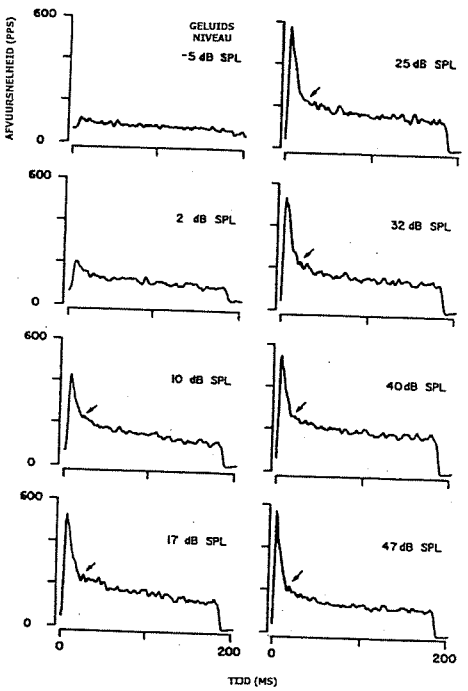


Fig. 9 Reactie van een gehoorzenuwvezel op een constante akoestische toon in functie van de tijd. De afvuursnelheid is weergegeven in aantal zenuwpulsen per seconde. De geluidssterkte is aangeduid boven elke grafiek. Uit Delgutte (1980, Fig. 1)

normaal oor nauwelijks of niet hoorbaar is[7], is er een adaptatie-effect aanwezig. De afvuursnelheid bij de aanzet van de toon bedraagt meer dan het dubbele van de snelheid nadien. Deze "adaptatiepiek" duurt 10 tot 20 milliseconden.

[7] Om een idee te geven over de dB-schaal: de geluidssterkte van een gewoon gesprek bedraagt ongeveer 60 dB SPL, een laag overvliegend gevechtsvliegtuig veroorzaakt een lawaai van 130 dB SPL.

Spraak is opgebouwd uit een relatief snelle opeenvolging van luide klinkers en stille medeklinkers, hetgeen een snel veranderende geluidssterkte tot gevolg heeft. Er zijn zelfs veel korte stiltes in spraak, zoals bv. vlak voor een *plosief* (p, t, k, b, d). Al deze snelle overgangen worden nog eens extra benadrukt door het adaptatie-effect, aangezien de zenuwvezel als het ware overdreven reageert bij elke snelle intensiteitsstijging. Ook uit verschillende spraaktesten bij normaalhorenden kan men afleiden dat dit effect een rol speelt bij het spraakverstaan (Drullman *et al.*, 1994; Shannon *et al.*, 1995).

Bij een cochleair implantaat wordt het geluid niet akoestisch, maar elektrisch aangeboden. Dus is het interessant om te weten hoe de gehoorzenuw reageert op een onveranderlijke elektrische toon. Figuur 10 illustreert de reactie op een elektrische pulstrein voor verschillende waarden van de constante amplitude. Enkel in de drie middelste grafieken is er een adaptatie-effect zichtbaar, en bovendien is de grootte van de adaptatiepiek veel kleiner dan deze bij akoestische stimulatie. Dus, om het gewenste effect te simuleren, moet het elektrisch signaal aangepast worden. In de volgende paragraaf wordt een algoritme voorgesteld dat het adaptatie-effect tracht te imiteren.

3.2 Geluidsverwerkingsalgoritme

Het standaard CIS algoritme houdt geen rekening met het adaptatiefenomeen. Een uitbreiding van dit algoritme kan hieraan tegemoet komen. Zoals in de inleidende paragraaf werd beschreven, berekent het CIS algoritme de omhullende van het signaal in elke frequentieband. In figuur 11A bv. is de omhullende van het nonsens woordje "aka" weergegeven. De eerste stap van de uitbreiding is het detecteren van snelle intensiteitsstijgingen. Telkens de omhullende van het standaard algoritme snel stijgt, wordt een nieuw signaal berekend dat enkel de aanzet van de opgekomen klank bevat (fig. 11B). Dit signaal komt overeen met de adaptatiepieken. Door dit nu weer op te tellen bij de oorspronkelijke omhullende bekomt men een signaal dat een verhoogde neurale activiteit zal veroorzaken bij plotse intensiteitsstijgingen (fig. 11C).

Vooraleer de adaptatiepiek kan opgeteld worden bij de oorspronkelijke omhullende, moet het aangepast worden aan de patiënt. Indien de piek te klein is, zal de proefpersoon immers geen verschil horen tussen beide algoritmen. Als de piek te groot is, zullen de woorden onaangenaam of vervormd klinken. Het pieksignaal zal eerst versterkt of vermenigvuldigd

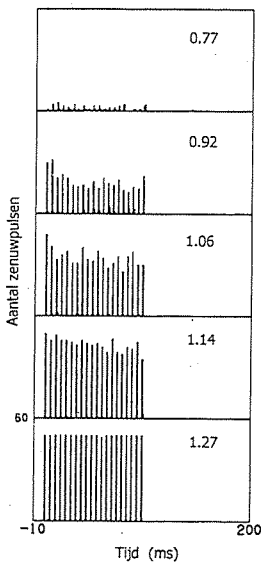


Fig. 10 Neurale adaptatie van de gehoorzenuw bij elektrische stimulatie. De verticale as geeft het aantal zenuwpulsen weer dat op elke elektrische puls volgde. De amplitude van de elektrische pulsen (in mA) is rechts van elke deelgrafiek weergegeven. Uit Javel *et al.* (1987, Fig. 3)

worden met een subjectafhankelijke factor. De waarde hiervan lag voor de proefpersonen die deelnamen aan de testen, tussen 1 en 2.

3.3 Experimenten

Acht Laura-gebruikers stemden in om mee te werken aan de evaluatie van het nieuwe algoritme. Alle identificatietesten werden afgenomen zowel voor het standaard CIS algoritme (CIS) als het nieuwe algoritme (EECIS). Zo kan men bepalen of de toevoeging van de adaptatiepieken enig effect heeft op het spraakverstaan. Twee proefpersonen echter werden uiteindelijk geweerd. Eén van beide kon nooit een verschil horen tussen de twee algoritmen, ook al werd het pieksignaal maximaal versterkt. Voor de andere klonk het nieuwe

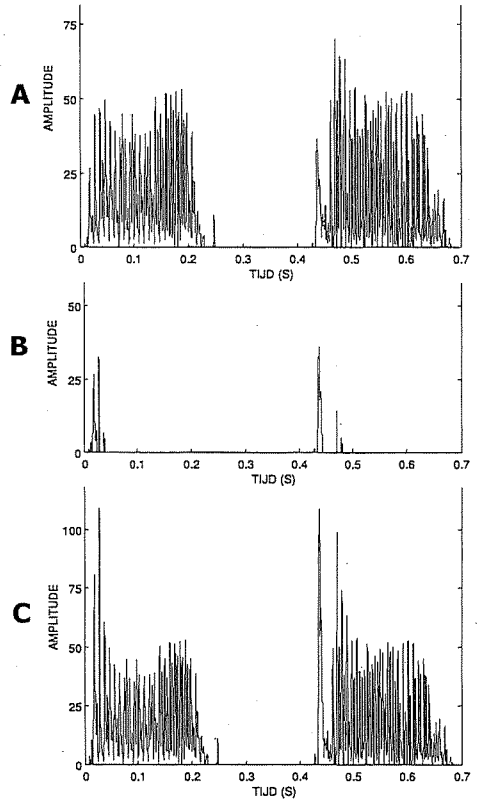


Fig. 11 (A) De omhullende van het spraaksignaal "aka" in één frequentieband. (B) De adaptatiepieken bekomt men door enkel de aanzetten te extraheren uit de omhullende. (C) Het nieuwe algoritme telt beide signalen op. De nieuwe omhullende simuleert het adaptatie-effect.

geluid altijd storend, tot het toegevoegde pieksignaal zo klein was, dat hij geen verschil meer hoorde. De overige zes proefpersonen vonden allen het nieuwe geluid beter verstaanbaar. De woorden klonken alsof ze duidelijker uitgesproken waren.

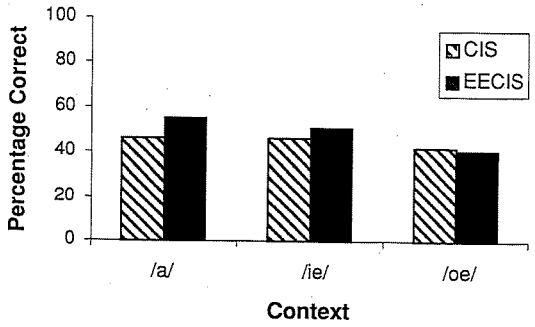


Fig. 12 Percentage correct geïdentificeerde woordjes voor de klinkercontexten "a", "ie" en "oe", en voor zowel het standaard algoritme CIS, als het nieuwe algoritme EECIS. De percentages zijn de gemiddelden over de zes proefpersonen.

In de eerste test moesten de proefpersonen woordjes identificeren van de vorm klinker-plosief-klinker. De klinker was "a", "ie" of "oe", en de plosief was "p", "t", "k", "b", "d" of "g"[8]. Deze klanken werden eerst getest, omdat deze de grootste pieksignalen tot gevolg hadden. Elk woordje werd 16 keer aangeboden, en de volgorde van aanbieding was willekeurig. Het percentage correct geïdentificeerde woordjes werd berekend voor elk algoritme en voor elke klinkercontext apart. De resultaten, gemiddeld over alle proefpersonen, zijn weergegeven in figuur 12.

Vooral de significante stijging van de scores voor de "a"-context valt op: van 46% tot 55% correct geïdentificeerde woorden. Voor de "i"-context is de stijging van 4% niet statistisch significant, en voor de "oe"-context is er een verwaarloosbare daling van 1%. Uit de analyse van de signalen achteraf blijkt dat de adaptatiepieken het grootst zijn voor de "a"-context, en het kleinst voor de "oe"-context. Een verbeterde versie van het EECIS algoritme dat hiermee rekening houdt, kan de resultaten gunstig beïnvloeden.

Om het nieuwe algoritme te evalueren in een bredere context, werden ook korte éénlettergrepige woorden aangeboden van de vorm medeklinker-klinker-medeklinker. Alle woorden waren bestaande Nederlandse woorden.

[8] Hiermee wordt de stemhebbende "k" bedoeld, zoals in "zakdoek" of in het Franse "garçon".

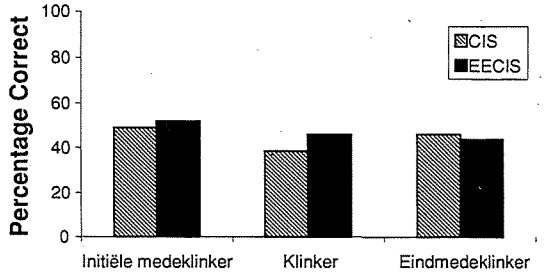


Fig. 13 Percentage correct geïdentificeerde klanken voor de éénlettergrepige woordjes, gemiddeld over de zes proefpersonen. Scores werden apart bepaald voor de initiële medeklinker, de klinker en de eindmedeklinker.

Hiermee kon nagegaan worden of er nog andere positieve of negatieve effecten op spraakverstaan konden zijn. Correct geïdentificeerde letters werden gescoord. De resultaten, gemiddeld over alle proefpersonen, zijn samengevat in figuur 13.

Het verschil in scores tussen beide algoritmen is niet significant voor de medeklinkers. Met het EECIS algoritme echter werden 7% meer klinkers correct geïdentificeerd. Deze stijging blijkt ook weer statistisch significant te zijn. Dit eerder onverwacht resultaat is waarschijnlijk te wijten aan het feit dat het adaptie-effect het contrast tussen opeenvolgende klanken versterkt.

3.4 Besluit

Het nieuwe spraakalgoritme imiteert het adaptatie-effect van de gehoorzenuw door een pieksignaal toe te voegen aan de normale omhullende bij een plotse intensiteitsstijging. Enkele spraaktesten bij zes proefpersonen tonen aan dat het nieuwe algoritme de woorden beter verstaanbaar maakt. De resultaten kunnen nog verbeteren indien het algoritme de zwakkere pieksignalen, zoals in de "ie"- en de "oe"-context, kan compenseren en indien de patiënten de mogelijkheid krijgen om het algoritme te gebruiken in het dagelijkse leven. De voorgestelde resultaten werden immers bereikt zonder dat de patiënt kon oefenen met het nieuwe algoritme. Het algoritme is zo ontworpen dat een implementatie in een draagbare processor mogelijk is.

4. Coderen van de fundamentele frequentie in het CIS algoritme

4.1 De fundamentele frequentie

De toonhoogte van een stem wordt bepaald door de snelheid waarmee de stembanden openen en sluiten. Op een vergelijkbare manier wordt de toonhoogte bij een snaarinstrument bepaald door de snelheid van de trillingen van de snaar. De fysische grootheid *frequentie* drukt het aantal trillingen per seconde uit. Zo heeft een typische mannenstem een frequentie van ongeveer 100 Hz, terwijl deze bij een vrouwenstem rond de 200 Hz ligt. Bijna elke klank is opgebouwd uit de grondtoon en zijn boventonen of harmonischen[9]. De frequentie van de grondtoon noemt men de fundamentele frequentie, en de frequentie van elke boventoon is daar een geheel veelvoud van. Dit heeft tot gevolg dat het verschil tussen de frequenties van twee opeenvolgende boventonen ook overeenkomt met de fundamentele frequentie. Hierdoor kan de fundamentele frequentie tot uiting komen in de reactie van alle zenuwvezels, dus ook bij zenuwvezels die niet reageren op de grondtoon zelf. Deze reactie wordt geïllustreerd in figuur 14. Elke deelfiguur toont, in functie van de tijd, de bovenste helft van de akoestische stimulus aan de rechterkant, en de afvuursnelheid van de zenuwvezel aan de linkerkant. De stimulus is een hoge toon die volgens een vast patroon, meer bepaald volgens een sinusfunctie, van amplitude verandert. Telkens worden twee op-en-neer-gaande bewegingen getoond. De snelheid waarmee deze verandering gepaard gaat komt overeen met de fundamentele frequentie. De grootte van de verandering noemt men de modulatiediepte. Het is duidelijk dat ook de zenuwvezel reageert volgens hetzelfde patroon. Deze toont ook pieken en dalen van aktiviteit die aan dezelfde frequentie elkaar opvolgen. Als de modulatiediepte vergroot, zal het verschil tussen pieken en dalen ook vergroten.

Hetzelfde effect kan ook bij elektrische stimulatie met een cochleair implantaat bereikt worden, indien de amplitude van het elektrisch signaal varieert met de snelheid van de fundamentele frequentie. In figuur 5A rond de 400-ms lijn (klinker "o") bv., is de variatie van de omhullende van de

[9] De klankkleur of het timbre wordt bepaald door de intensiteit van elke boventoon. Zo zullen bv. bij een "oe"-klank vooral de lagere boventonen overheersen, terwijl bij een "ie"-klank zowel enkele lagere boventonen als enkele hogere boventonen belangrijk zijn.

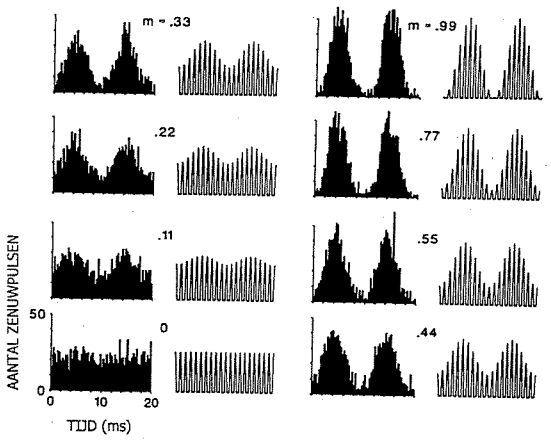


Fig. 14 Illustratie van de codering van de fundamentele frequentie in de neurale respons. De modulatiediepte van de akoestische stimulus is aangegeven boven elke deelgrafiek. Het positieve deel van de stimulus staat rechts, de zenuwactiviteit links. Uit Joris en Yin (1992, Fig. 1)

signalen in kanalen 1 tot 5 goed zichtbaar. Uit experimenten met cochleair geïmplanteerden blijkt dat deze gevoelig zijn voor het op deze manier weergeven van toonhoogtes. Indien de temporele variatie sneller gebeurt, dan zullen ze een hogere toon waarnemen. Vanaf een bepaalde frequentie, de exacte waarde varieert van persoon tot persoon, neemt deze gevoeligheid wel af, maar in de meeste gevallen ligt deze waarde boven het typische bereik van de fundamentele frequentie van de menselijke stem. Een vraag die tot op heden nog niet beantwoord is, is welke modulatiediepte nodig is om deze frequentie optimaal waar te nemen.

4.2 Experimenten

Vier Laura-gebruikers namen deel aan de twee experimenten van deze studie. In het eerste experiment werd nagegaan wat de invloed was van de modulatiediepte van een gemoduleerd signaal op het vermogen om toonhoogteverschillen waar te nemen. In het tweede experiment werden drie geluidsverwerkingsalgoritmen vergeleken in een vergelijkbare test.

In het eerste experiment kreeg de proefpersoon telkens twee toontjes op hetzelfde kanaal te horen. Elk toontje was een pulstrein waarvan de amplitude varieerde volgens een sinusfunctie (vergelijk met de stimuli in fig. 14). Ze waren identiek wat betreft kanaal, intensiteit, duur en modulatiediepte, maar de frequentie van de amplitudevariatie verschilde met 20%. De proefpersoon moest telkens aangeven welk van beide toontjes het hoogste klonk, en het antwoord was correct indien het toontje met de hoogste modulatiefrequentie gekozen werd. Verwacht werd dat dit moeilijker zou worden bij kleinere modulatiediepten, aangezien de frequentie in dit geval minder nadrukkelijk gecodeerd wordt in het elektrisch signaal. Elk onderdeel van figuur 15 geeft de resultaten van één proefpersoon weer (zie initialen rechtsboven). Er zijn opmerkelijke verschillen tussen de resultaten van de proefpersonen onderling, maar de belangrijkste overeenkomst is dat ze betere resultaten halen bij de grotere modulatiediepten. In sommige gevallen verbetert het resultaat wel niet meer vanaf een bepaalde waarde van deze diepte. Proefpersoon AE bv. scoort minstens even goed bij 20% als bij 99%. Proefpersonen JH en WM waren niet in staat om een verschil te horen tussen toontjes met modulatiefrequenties van 250 Hz en 300 Hz.

Voor het tweede experiment werden drie algoritmen gebruikt. Eén algoritme was het standaard CIS algoritme (CIS, fig. 16 midden). Het tweede algoritme, VLAK-CIS, was op 1 parameter na gelijk aan het standaard CIS algoritme. Deze parameter werd zodanig gekozen dat de fundamentele frequentie van een klank niet meer aanwezig is in de omhullende van de pulstreinen (fig. 16 boven). Het derde algoritme, F0-CIS, betrof een nieuw algoritme, dat een uitbreiding was van het standaard algoritme. Het was ontworpen om de modulatiediepte nog verder te verhogen t.o.v. het standaard algoritme (fig. 16 onder). Van elk algoritme werd bepaald welk het kleinst waarneembaar verschil in toonhoogte is. Telkens werden twee klanken, beiden een synthetische[10] "a" of een "ie", met een verschillende fundamentele frequentie aangeboden, en de proefpersoon moest de hoogste van de twee aanduiden. Zolang dit correct en consistent gebeurde, werd het verschil kleiner gemaakt.

De resultaten zijn weergegeven in figuur 17. Elke staaf geeft het frequentieverschil aan dat nog waargenomen kon worden, en is het gemiddelde van vier metingen. Hoe korter de staaf, hoe beter de

[10] *Synthetisch* wil hier zeggen dat het geen natuurlijke klanken zijn, door een mens uitgesproken, maar dat de computer deze signalen heeft gegenereerd.

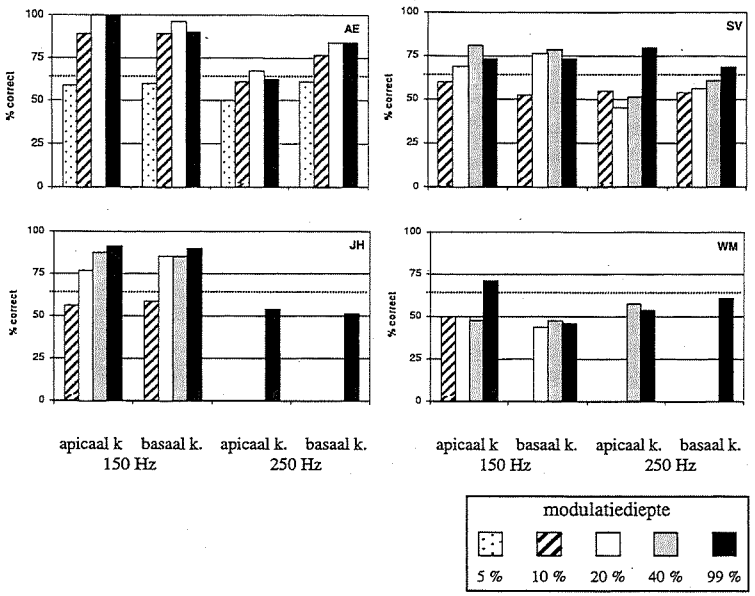


Fig. 15 Scores in percentage correcte antwoorden voor experiment 1. Elke deelgrafiek geldt voor een andere proefpersoon. Het kanaal en de laagste modulatiefrequentie zijn beneden weergegeven.

proefpersoon presteerde. De korte horizontale streepjes, verbonden met een vertical streep, geven aan in welke mate de vier metingen van elkaar afwijken. Soms kon zelfs het grootste frequentieverschil niet waargenomen worden. In dat geval wordt een staaf over de volledige hoogte van de grafiek weergegeven, met een vraagteken erin.

Een eerste opvallend feit is dat voor het VLAK-CIS algoritme, de proefpersonen soms nog steeds een onderscheid konden horen, zonder dat er een temporele variatie in de omhullende was. Uit de analyse van de uitgangssignalen blijkt dat de amplitude van de pulsen in elk kanaal verandert, als de fundamentele frequentie verandert. Waarschijnlijk hebben de proefpersonen ook deze verandering kunnen gebruiken om het onderscheid te maken. Een tweede belangrijke vaststelling is dat de twee andere algoritmen vaak wel tot een beter resultaat leiden. Dit is het geval

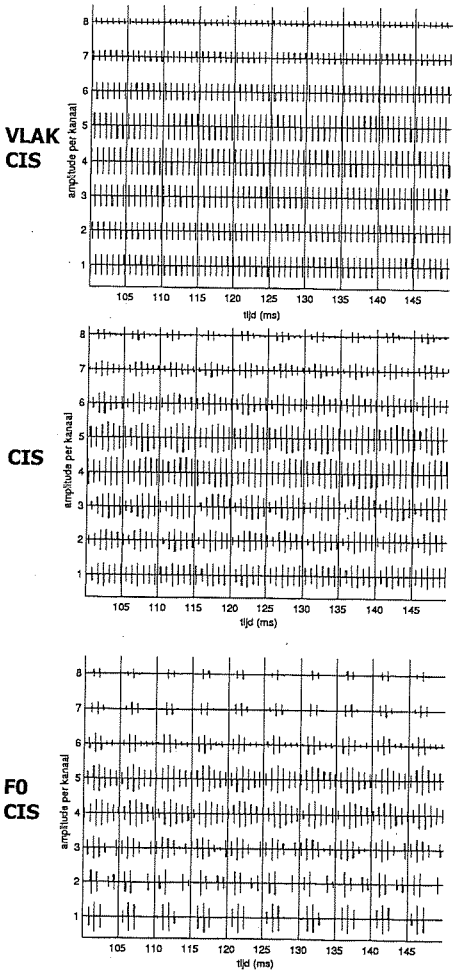


Fig. 16 De uitgangssignalen van de drie algoritmen voor de synthetische klinker "a" (fundamentele frequentie 200 Hz) in functie van de tijd: VLAK-CIS, standaard CIS en F0-CIS. Merk het verschil in modulatiediepte op.

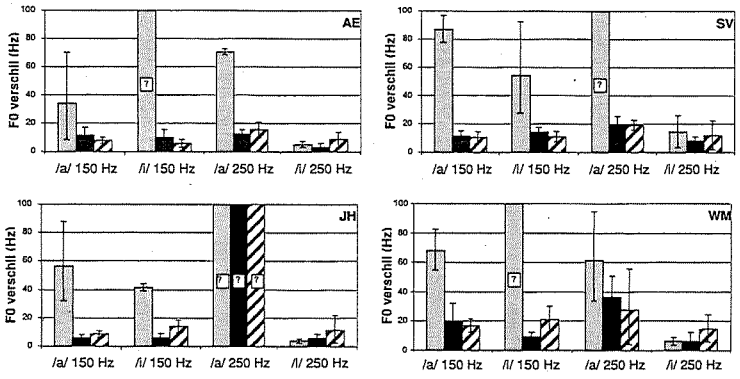


Fig. 17 Kleinst waarneembaar frequentieverschil, gemeten in experiment 2. Elke deelgrafiek geldt weer voor een andere patiënt. De klinker en de frequentie t.o.v. dewelke het verschil werd gemeten, zijn onderaan de deelgrafieken aangegeven. De algoritmen zijn VLAK-CIS (grijs), standaard CIS (zwart) en F0-CIS (gestreept). Als zelfs het grootst aangeboden verschil niet waarneembaar was, wordt een staaf over de volledige hoogte getoond, met een vraagteken erin.

voor alle proefpersonen bij de lage frequenties (150 Hz), en voor AE en SV bij de "aa" van 250 Hz. Dit toont aan dat de temporele variatie van de omhullende, die overeenkomt met de fundamentele frequentie, belangrijk en soms zelfs essentieel is om de toonhoogte correct te kunnen waarnemen. De uitzondering op de regel blijkt uit de resultaten voor de "ie"-klank bij 250 Hz. In dit geval zijn de resultaten voor het VLAK-CIS algoritme reeds zo goed, dat geen verbetering meer mogelijk is. Uit signaalanalyses blijkt dat de verandering van de amplituden in sommige kanalen een duidelijke en eenduidige relatie vertoont met de verandering van de fundamentele frequentie. Een derde belangrijke vaststelling is dat het nieuwe algoritme niet tot betere resultaten leidt t.o.v. het standaard algoritme. Nochtans, op basis van de signaalanalyses en de resultaten van het eerste experiment, kon dit toch verwacht worden. Een belangrijk verschil tussen de twee experimenten echter is dat in het eerste experiment altijd maar één kanaal gestimuleerd werd, terwijl in het tweede experiment alle beschikbare

kanalen gebruikt werden. Waarschijnlijk is in het laatste geval minder modulatiediepte nodig om tot eenzelfde resultaat te komen.

4.3 Besluit

De fundamentele frequentie bepaalt de toonhoogte van een klank. Bij het CIS algoritme komt deze tot uiting als een temporele variatie in de omhullende van het uitgangssignaal. In deze studie werd het belang onderzocht van de modulatiediepte, dus van de grootte van de variatie. In het eerste experiment werd gemeten hoe het vermogen om toonhoogtes te onderscheiden varieerde met de modulatiediepte. De modulatiediepte waarbij de hoogste scores behaald werden, varieerde tussen 20% en 99%. In het tweede experiment werden voor drie algoritmes het kleinst waarneembare verschil in fundamentele frequentie van een klank bepaald. Hieruit bleek dat de temporele variatie in de omhullende een belangrijke en soms zelfs essentiële rol speelt bij het waarnemen van de overeenkomstige toonhoogte. Een nieuw algoritme, dat de modulatiediepte nog kon vergroten t.o.v. het standaard algoritme, gaf geen betere resultaten, m.a.w. met de modulatiediepte verkregen in het standaard algoritme, zijn de resultaten reeds optimaal.

5. Detectie van kleine tijdsverschillen tussen twee kanalen

5.1 Motivatie

De laatste studie is meer fundamenteel van aard. Het was niet de bedoeling om een nieuw algoritme te evalueren, maar om te onderzoeken in welke mate cochleair-implantgebruikers bepaalde verschillen tussen twee stimuli kunnen horen. Elke stimulus bestond uit twee identieke pulstreinen, met een interval van 10 ms tussen de pulsen, die gelijktijdig op relatief ver van elkaar gelegen kanalen werden aangeboden. De stimuli verschilden van elkaar in de mate waarin de ene pulstrein later kwam dan de andere. Dit tijdsverschil varieerde tussen 0.1 ms en 2 ms. Het verband tussen deze studie en spraakalgoritmen als het CIS algoritme ligt in de volgorde van stimulatie van de kanalen in deze algoritmen. Zoals beschreven in de inleidende paragraaf, volgt er na elke puls op het ene kanaal een puls op een ander kanaal. De volgorde waarin de kanalen gestimuleerd worden, bepaalt het tijdsverschil tussen de pulsen op elk paar kanalen. Als bv. de kanaalvolgorde 1-2-3-4-5-6-7-8 is, en er elke 100 µs een puls gegenereerd

wordt, dan zal het tijdsverschil tussen naast elkaar liggende kanalen 0.1 ms bedragen, en tussen ver van elkaar liggende kanalen maximaal 0.4 ms bedragen[11]. Deze studie heeft als doel om na te gaan in welke mate dit verschil waarneembaar is voor de proefpersoon.

5.2 Experimenten

In het eerste experiment werden aan de vijf proefpersonen telkens twee stimuli aangeboden (fig. 18). In één van beide, de *standaard-* of *referentiestimulus*, was het tijdsverschil tussen de pulstreinen zo klein mogelijk[12], nl. 0.1 ms. In de andere stimulus, de *signaalstimulus*, was het tijdsverschil meestal groter en steeds in de andere richting. De stimulusparen die de proefpersonen moesten vergelijken, zijn aangegeven met de open cirkel of driehoek in figuur 18. Er waren twee testcondities. In de ene conditie kwam bij de referentiestimulus de pulstrein op het apicale kanaal[13] 0.1 ms later en had de signaalstimulus een vertraging op het basale kanaal (fig. 18 boven). In de andere conditie was dit andersom: de referentiestimulus had een minimale vertraging op het basale kanaal en de signaalstimulus had een vertraging op het apicale kanaal (fig. 18 onder). Als de tijdsvertraging verandert, zal de stimulus anders klinken. Dit verschilde naargelang de proefpersoon: ofwel klonken de signaalstimuli doffer dan de referentiestimuli, ofwel waren ze hoger in toonhoogte, ofwel galmden ze meer (vgl. echo).

Figuur 19 geeft voor elke proefpersoon afzonderlijk het percentage correcte antwoorden weer voor elke vertraging die getest werd (open symbolen). Merk op dat een score van 50 % wil zeggen dat de proefpersoon aan het gokken was. Immers, als er slechts twee mogelijke antwoorden zijn, dan is de slaagkans 50 % als een willekeurig antwoord gegeven wordt. Een score dicht bij 0 % wil zeggen dat de proefpersoon zeer duidelijk een verschil kon horen, maar dat hij altijd de verkeerde keuze maakte[14]. De curve voor de basale vertraging (cirkels) van proefpersoon SV komt het best overeen met het verwachte patroon. Voor de grootste vertraging (2 ms) haalt

[11] Merk op dat de sequentie opnieuw begint, als ze is afgelopen, dus na kanaal 8 wordt terug kanaal 1 gestimuleerd.

[12] Elke bifazige puls duurt 80 µs. Samen met een veiligheidsmarge van 20 µs tussen het einde van de ene en het begin van de volgende puls wordt het minimale verschil tussen het begin van elke puls 100 µs = 0.1 ms.

[13] Het apicale kanaal is het kanaal dat zich het diepst in de cochlea bevindt, dus het dichtst bij de apex.

[14] Merk op dat de proefpersoon niet kon weten of zijn antwoord goed of fout was.

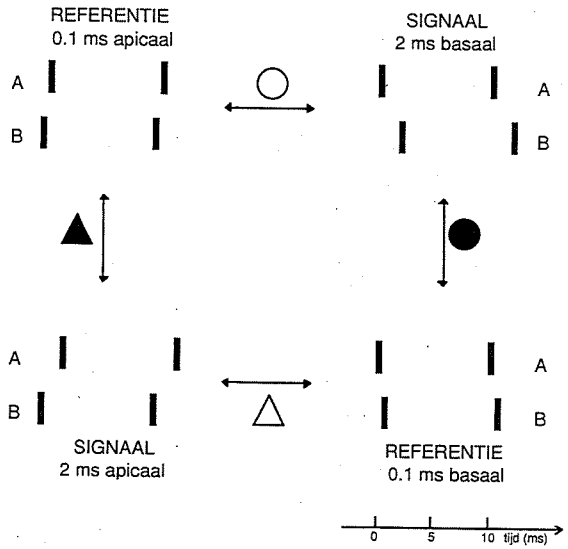


Fig. 18 Schematische weergave van de stimuli. Telkens worden twee opeenvolgende pulsen uit de pulstreinen op het apicale kanaal (A) en het basal kanaal (B) getoond. In de bovenste rij komt de signaalstimulus laatst en bevat deze een vertraging van 2 ms op het basale kanaal. In de onderste rij komt de signaalstimulus eerst en bij deze komt de pulstrein op het apicale kanaal 2 ms later. De open symbolen geven de stimulusparen aan die de proefpersonen moesten vergelijken in het eerste experiment. De gesloten symbolen komen overeen met een extra conditie die getest werd.

hij de hoogste score, voor de kleinste vertraging haalt hij 50 %. De scores voor de overige vertragingen liggen daartussen. Voor deze proefpersoon was het blijkbaar wel veel moeilijker om een vertraging op het apicale kanaal waar te nemen. Opmerkelijk voor de resultaten van DD is dat zijn scores voor de kleinste vertragingen significant verschillen van 50 %. Hij kon dus nog steeds een verschil horen, al had hij het daar zeer moeilijk mee. De overige drie proefpersonen konden dit verschil bijna perfect waarnemen. Zij haalden scores van meer dan 90 % (of minder dan 10 %) voor de kleinste vertraging. Zij hoorden dus het verschil tussen een stimulus waarin de pulstrein op het apicaal kanaal 0.1 ms later kwam, en

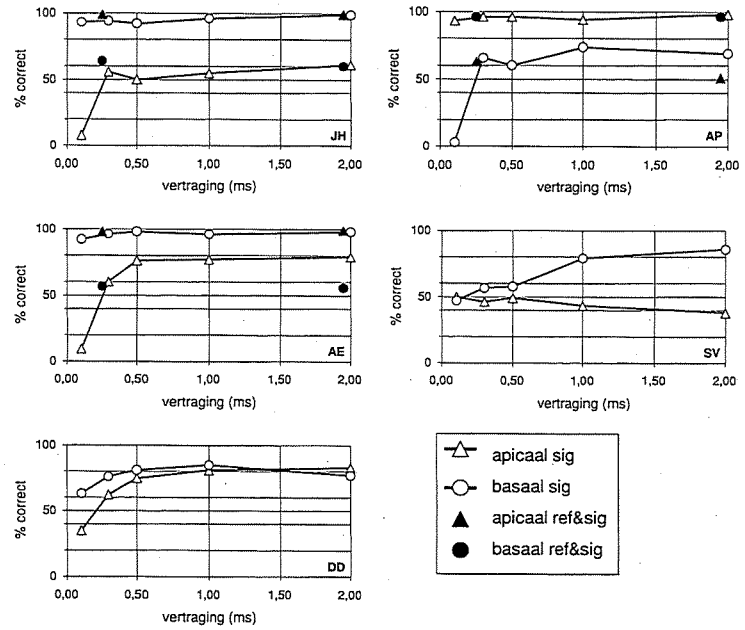


Fig. 19 Scores per proefpersoon voor het eerste experiment. Twee condities werden getest: resultaten voor de signaalstimuli met vertraging op het basale kanaal zijn aangegeven met cirkels, deze voor de signaalstimuli met apicale vertraging met driehoeken. De gesloten symbolen komen overeen met de resultaten van enkele extra metingen (zie tekst). Merk op dat elke score die ver afwijkt van 50 % aangeeft dat de proefpersoon een duidelijk verschil kon waarnemen.

een stimulus waarin de pulstrein op het basaal kanaal 0.1 ms later kwam. Proefpersonen JH en AE konden vooral de basale vertragingen goed waarnemen, voor proefpersoon AP was dit de apicale vertraging.

Een mogelijke verklaring voor de asymmetrie in de resultaten is dat de referentiestimuli van beide condities verschilden. Deze verklaring gaat niet op voor SV, aangezien voor hem de twee referentiestimuli niet onderscheidbaar waren. Voor een vertraging van 0.1 ms in de

signaalstimulus, en dus een vertraging van 0.1 ms op het andere kanaal in de referentiestimulus, scoorde hij in beide condities 50%. Bij de drie proefpersonen die zeer hoge scores behaalden voor de kleinste vertraging, werden extra metingen uitgevoerd. In de signaalstimulus was er een vertraging van 0.3 ms of 2 ms, en in de referentiestimulus een vertraging van 0.1 ms *op hetzelfde kanaal*. Deze vergelijking is visueel weergegeven in figuur 18 met de gesloten symbolen. De resultaten zijn met dezelfde symbolen weergegeven in figuur 19. Het merendeel van de scores valt samen met de scores van de oorspronkelijke metingen, maar waarbij de vertraging op het andere kanaal was. Dit wil dus zeggen dat dezelfde scores behaald worden als de referentiestimulus hetzelfde blijft. Daaruit volgt dat voor deze drie proefpersonen één van de referentiestimuli zeer duidelijk verschilde van al de andere stimuli.

Bij elke proefpersoon werden telkens twee kanalen geselecteerd die ver uit elkaar lagen, en dit met de bedoeling dat twee verschillende groepen zenuwvezels geactiveerd werden. Het is echter nog steeds mogelijk dat er een kleine groep zenuwvezels is die via beide kanalen gestimuleerd kunnen worden, en dat de proefpersonen oordelen op basis van de activiteit van die kleine groep. Om deze interactie uit te schakelen werd gebruik gemaakt van zogenaamde *maskeerders*. In dit geval werden twee extra kanalen geselecteerd die met pulstreinen gestimuleerd werden tijdens het verloop van de test. Deze twee kanalen liggen tussen de kanalen in die in het eerste experiment gebruikt werden, zodat de eventuele invloed van deze laatste kanalen op de voornoemde kleine groep zenuwvezels geneutraliseerd wordt. Enkele metingen van het eerste experiment werden herhaald, maar nu dus met pulstreinen op de maskeerkanalen. De scores lagen meestal iets lager, maar de proefpersonen konden allen nog duidelijk het verschil waarnemen. Dit bevestigt de stelling dat de proefpersonen oordeelden op basis van de neurale activiteit van twee verschillende groepen zenuwvezels.

5.3 Besluit

Het doel van deze studie was na te gaan of implantgebruikers in staat zijn om een verandering in de vertraging van een pulstrein op één kanaal t.o.v. deze in een ander kanaal waar te nemen. Eén van de vijf proefpersonen kon een vertraging van 1 ms waarnemen, terwijl drie andere bijna perfect een vertraging van 0.1 ms op het ene kanaal konden onderscheiden van een identieke vertraging op het andere kanaal. Dit overtreft met groot verschil de prestaties van normaalhorenden in een vergelijkbare akoestische taak. Met andere woorden is hier een gevoeligheid

van het gehoorsysteem aangetoond dat niet met akoestische stimulatie kan worden aangetoond. Indien dezelfde testen worden herhaald met maskeersignalen op tussenliggende kanalen, bekomt men vergelijkbare resultaten. Dit wil zeggen dat de proefpersonen oordeelden op basis van de neurale activiteit van twee verschillende groepen zenuwvezels. De resultaten hebben ook tot gevolg dat het veranderen van de stimulatievolgorde in algoritmen als de CIS strategie, een andere waarneming tot gevolg kan hebben. Uit de resultaten kan men echter niet besluiten welke volgorde optimaal is.

6. Algemene besluiten

Het voorgestelde werk bestaat uit één technische realisatie en drie studies met cochleaire-implantaatgebruikers. Eerst werd een computer-gebaseerde opstelling gebouwd, waarmee men gemakkelijk psycho-fysische experimenten kan uitvoeren en nieuwe geluidsverwerkings-algoritmen kan evalueren. Voor de eerste studie werd een nieuw algoritme ontwikkeld waarin neurale adaptatie werd nagebootst. Uit de resultaten van de spraakidentificatietesten blijkt dat de proefpersonen bepaalde klanken beter verstaan. In de tweede studie werd de codering van de fundamentele frequentie van klinkers onderzocht. De resultaten tonen het grote belang aan van de temporele codering, die zich uit als amplitudevariaties van de klankomhullende. In de laatste studie werd voor het eerst aangetoond dat cochleair-implantgebruikers kleine tijdsverschillen tussen twee pulstreinen op ver van elkaar verwijderde kanalen kunnen waarnemen. De drie studies bieden nieuwe inzichten in de waarneming van geluid via cochleaire implantaten en hopelijk brengt dit de wetenschap één of meerdere stappen dichter bij het ideale cochleaire implantaat.

Index

Voorwoord v

Abstract ix

Korte Inhoud xi

Coderen van temporele informatie in geluidsverwerkingsstrategieën voor cochleaire implantaten xiii

1 Inleiding xiii
1.1 Een remedie tegen doofheid xiii
1.2 Enkele gegevens over de werking van het oor xiv
1.3 Cochleaire implantaten xvii

2. APEX xxii
2.1 Overzicht xxii
2.2 Apparatuur xxiv
2.3 Programma's xxiv

3. Imiteren van neurale adaptatie xxvi
3.1 Neurale adaptatie xxvi
3.2 Geluidsverwerkingsalgoritme xxviii
3.3 Experimenten xxix
3.4 Besluit xxxii

4. Coderen van de fundamentele frequentie in het CIS algoritme xxxiii
4.1 De fundamentele frequentie xxxiii
4.2 Experimenten xxxiv
4.3 Besluit xxxix

5. Detectie van kleine tijdsverschillen tussen twee kanalen xxxix
5.1 Motivatie xxxix
5.2 Experimenten xl
5.3 Besluit xliii

6. Algemene besluiten xliv

Index xlv

Chapter 1 Introduction to cochlear implants 1

1.1 A potential solution for deafness 1

1.2 Basics of hearing 3

1.3 Cochlear implants 9
1.3.1 The hardware 9
1.3.2 The speech processing software 11
1.3.3 The Laura device 18

1.4 Speech tests and evaluation 19

1.5 The research questions 23
1.5.1 Mimicking neural adaptation 23
1.5.2 Coding of the fundamental frequency 23
1.5.3 Sensitivity to time delays between channels 24

1.6 Restrictions of time and means 25

1.7 Summary 26

Chapter 2 APEX: a research interface for psycho-physical and speech processing experiments with the Laura cochlear implant 27

2.1 Specifications 28

2.2 APEX 29
2.2.1 General description 29
2.2.2 The APEX hardware 29
2.2.3 The APEX software 31
2.2.3.1 The Matlab functions 31
2.2.3.2 The PC program: Apex.exe 34
2.2.3.3 The DSP program: 30electr.out 38
2.2.4 Two examples 40
2.2.4.1 Example.idn: an identification experiment 41
2.2.4.2 A discrimination experiment with adaptive procedure 44

2.3 Summary and conclusions 46

Chapter 3 Including neural adaptation in CIS processors 47

3.1 Neural adaptation 48
3.1.1 Neural response to acoustic stimuli 48
3.1.2 Existing models 48
3.1.3 Importance of adaptation 50
3.1.4 Neural response to electric stimuli 51
3.1.5 Foregoing work 53

3.2 Speech processing 53
3.2.1 Prefiltering and band-pass filtering 53
3.2.2 CIS: standard envelope extraction 55
3.2.3 EECIS: enhanced envelope extraction 56
3.2.4 Current mapping and pulse modulation 58

3.3 Identification experiments 60
3.3.1 Subjects 60
3.3.2 Fitting 61
3.3.3 Stimuli and procedure 62
3.3.4 Results 64
3.3.4.1. Stop consonants 64
3.3.4.2. CVC words 65

3.4 Signal analysis 67

3.5 Discussion 69

3.6 Summary and conclusions 71

Chapter 4 Coding of the fundamental frequency in CIS speech processors 73

4.1 Introduction 74
4.1.1 Fundamental frequency 74
4.1.2 Temporal coding of F0 in the auditory nerve 74
4.1.3 F0 in speech processors for cochlear implants 78

4.2 Speech processing 80

4.3 Experiment 1 85
4.3.1 Subjects 85
4.3.2 Stimuli and Procedure 86
4.3.3 Results 88
4.3.4 Discussion 90

4.4 Experiment 2 91
4.4.1 Stimuli and procedure 91
4.4.2 Results 97
4.4.3 Discussion 98

4.5 General discussion 99

4.6 Summary and conclusions 101

Chapter 5 Detection of small across-channel timing differences by cochlear implantees 103

5.1 Introduction 104

5.2 Experiment 1: Psychometric functions 106
5.2.1 Subjects 106
5.2.2 Stimuli and procedure 108
5.2.3 Results 112
5.2.4 Discussion 114

5.3 Experiment 2: Interspersed Maskers 116
5.3.1 Stimuli and procedure 116
5.3.2 Results 118

5.4 Experiment 3: Pitch judgements 118
5.4.1 Stimuli and procedure 118
5.4.2 Results 120

5.5 Experiment 4: Discriminating between 2-ms basal and apical delays 122
5.5.1 Stimuli and procedure 122
5.5.2 Results 122

5.6 General discussion 123
5.6.1 Comparison with previous measures of across-channel timing sensitivity 123
5.6.2 Comparison with sensitivity to timing differences between ears 126
5.6.3 Importance for speech processors 126

5.7 Summary and conclusions 127

Chapter 6 Conclusions and future directions 129

6.1 APEX 129

6.2 Mimicking neural adaptation in speech processors for cochlear implants 130

6.3 Temporal coding of the fundamental frequency in cochlear implants 132

6.4 Sensitivity to temporal delays between two channels of a cochlear implant 134

6.5 Final conclusions 136

References 139

Appendix Example of stimulus editing with the custom-made Matlab functions 151

List of publications 153

Curriculum vitae 155

Chapter 1

Introduction to cochlear implants

1.1 A potential solution for deafness

Deafness is one of the most severe handicaps. Loosing the ability to hear makes it very hard to communicate with other people, leading to social isolation. Although there are some alternative means of communication, they suffer from immense disadvantages: lip reading provides only a small part of the information of the spoken word, most normal hearing people do not 'speak' sign language and communication via written words is too time consuming and requires extra equipment. Also, in many cases it is very hard to understand what deaf people say, since they lack the feedback needed to produce highly intelligible speech. For most ethiologies, medical science can not yet cure deafness. A classical hearing aid, which picks up and amplifies the acoustical signal and presents it back to the ear, is not sufficient for the profoundly hearing impaired. Whatever the cause of deafness is, somewhere in the ear one of the physiological functions is not operational. In a normal ear, the acoustical signal is first converted into a mechanical vibration (ear drum, ossicles and basilar membrane), and next this information is transformed into neural spikes (hair cells and auditory nerve fibers) which are sent to the brain.

A cochlear implant is a device which attempts to restore the ability to hear. For the great majority the cause of deafness is located in the cochlea, more specifically at the hair cells. Therefore, the implant bypasses all the acoustical mechanisms and physiological functions between outer ear and cochlea. It consists of several electrodes which are placed in the cochlea, and a current flow between two or more of these electrodes can activate a group of auditory nerve fibers in their vicinity. An excellent introduction to the technique of electrical stimulation of nerve fibers in general is given by Ranck (1981). Although the idea of electrical stimulation of the ear is rather old - Allesandro Volta described first attempts in the previous century (Volta, 1800) -, real progress in this research area commenced in the 70s (House and Urban, 1973; Eddington *et al.*, 1978; Clark *et al.*, 1978; Shannon, 1983a; Shannon, 1983b). The first clinical devices were available

in the early 80s. In the beginning, the used techniques were of course very experimental, and successes were rather limited. Detection of sounds was feasible, but oral communication without lip reading was not. Since then, numerous physiological and psychophysical studies were conducted, leading to increased knowledge, improved technology and to very satisfying results. Nowadays, some cochlear implant users can communicate easily with other people, even without lip reading. However, many other users still obtain poor results. This large variability in the results is typical in the field of cochlear implant research, but unfortunately, there seems to be no clear relationship with pre-operative measures, except with duration of deafness and residual hearing (van Dijk *et al.*, 1999).

In this dissertation, some studies are described in which several aspects of cochlear implant technology are closely investigated. In each study, a temporal effect in the neural representation of a sound is related to the sound processing strategy and the influence of this effect on the perception of the implantee is investigated. Such a sound processing strategy converts the acoustical signal into one or more electrical signals to be applied to the electrodes in the cochlea. Obviously, the basis for every strategy is the physiology of the normal ear. Some of its essential components will shortly be described in section 1.2. Section 1.3 will give an overview of current cochlear implant technology with the focus on hardware, software and the link with physiology. More details will also be provided about the implant used for all studies in this thesis: the Laura device of Philips Hearing Implants. A few words about speech tests and the analysis of the results will be given in section 1.4. In section 1.5 the three major research questions will be introduced. The first deals with a temporal effect that emerges when the sound intensity suddenly increases, namely neural adaptation. The topic of the second study is the temporal representation of the fundamental frequency through the modulation in the sound envelope. In the third study, it is investigated whether implantees are sensitive to changes in the time delay between two electric pulse trains applied on widely separated electrodes. Some restrictions of time and means are inevitable in every research area. Those that were meaningful for the described studies are mentioned in the section 1.6.

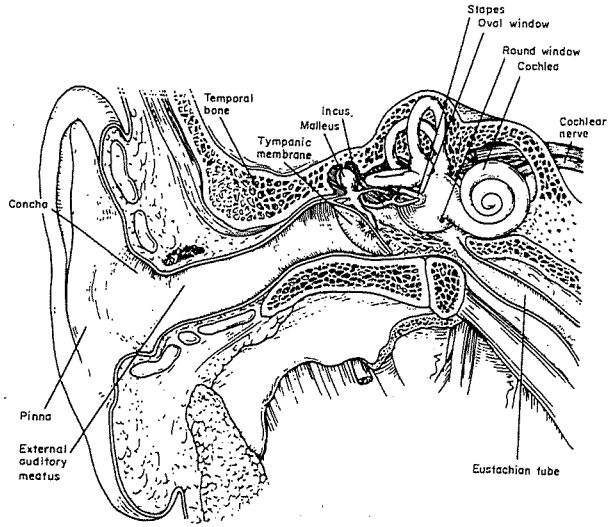


Fig. 1.1 The outer ear, middle ear and inner ear in man. From Pickles (1988, Fig. 2.1)

1.2 Basics of hearing

This section contains a very brief description of the physiology of hearing, with the emphasis on those phenomena that are important for current cochlear implant technology. More details can be found in Pickles (1988) and Handel (1989).

From pinna to nerve fiber

Prior to the auditory nerve, the ear consists of three parts: the outer ear, the middle ear and the inner ear or cochlea (figure 1.1). A sound wave is "captured" by the pinna, and is guided through the ear canal to the ear drum. Here, the acoustic wave is converted to a mechanical motion. One can model the outer ear as a band-pass filter, which emphasizes the frequencies between approximately 2 kHz and 7 kHz by an amount between 10 dB and 20 dB, the exact values depending on the direction of

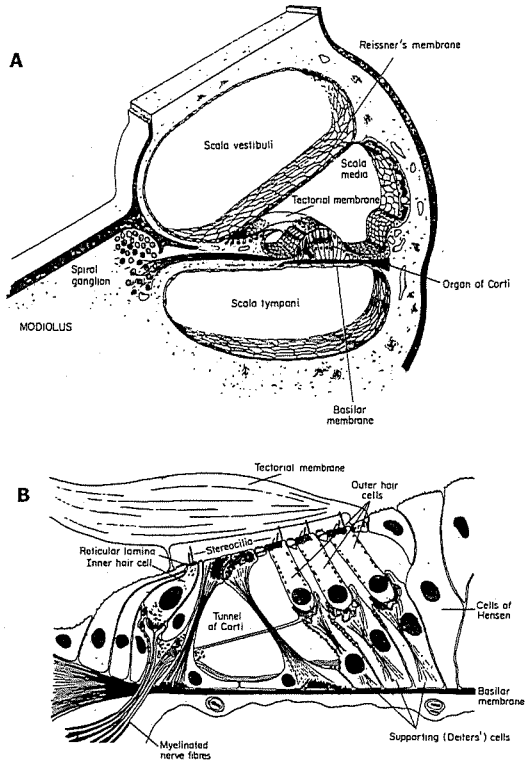


Fig. 1. 2 (A) A cross section of the cochlea, showing the three parts or scalae divided by Reissner's membrane and the basilar membrane. (B) A cross section of the organ of Corti. The stereocilia on the inner and outer hair cells will be bend if the basilar membrane moves. The nerve fibers are connected to the hair cells. From Pickles (1988, Fig. 3.1)

the sound wave and the shape of the outer ear. In the middle ear, three ossicles connect the ear drum to the oval window of the cochlea. The latter is filled with fluid, and the function of the middle ear is to match the impedance between air and fluid. Indeed, if the acoustic wave would arrive at the oval window directly, most of the energy (99.9 %) would be reflected.

A cross section shows that two membranes, the basilar membrane and Reissner's membrane, divide the cochlea in three parts or scalae (figure 1.2 A). At the end of the cochlea, the apex, the two outer parts are connected to each other. When the oval window vibrates, the fluid is displaced and the basilar membrane is forced to move up and down. The displacement of this membrane is first maximal at the base, and then progresses through the cochlea to the apex (see further). This causes shearing forces on the stereocilia of the hair cells. These are part of the organ of Corti, which is situated on the basilar membrane (figure 1.2 B). The deflection of the stereocilia determines the degree of coupling between the inner part of the cell and its surroundings. In rest, both parts contain chemical substances with a different ionic composition, resulting in a potential gradient: negative in the cell (between -45 and –70 mV) and positive around the cell+80 mV). When the basilar membrane is displaced, positive ions are driven into the cell. This intracellular depolarization causes the release of neurotransmitter, activating the auditory nerve, which is connected to the hair cell through the synapse. A good model of this mechanism is developed by Meddis *et al.* (1986; 1988; 1990) and further refined by Lopez-Poveda *et al.* (1997).

Tonotopy

The movement of the basilar membrane can be described as a wave, travelling from base to apex. The maximum displacement of this membrane differs for each position along the length of the cochlea, and depends on the frequency content of the stimulus (von Békésy, 1960; Sellick *et al.*, 1982). Figure 1.3 gives some examples of the maximum displacement of the travelling wave in response to pure tone stimuli as a function of the distance from the apex. Each plot illustrates the situation for an acoustic sinusoidal stimulus with a different frequency. It is clear that the position where the absolute maximum occurs depends on the frequency of the tone: the higher the frequency, the larger the distance of the absolute maximum from the apex. This relationship between frequency and absolute maximal displacement is referred to as the *tonotopic organization* of the cochlea.

Due to the haircell mechanism, this place-frequency coding is also reflected in the activity of the auditory nerve fibers. If the minimal sound level needed to achieve a significant increase in neural activity above the spontaneous activity, is represented as a function of the frequency of a pure tone, one obtains V-shaped curves, as shown in figure 1.4. Each plot shows the observed relation of two nerve fibers originating from the same tonotopic place. The curves were obtained at various positions along the length of the

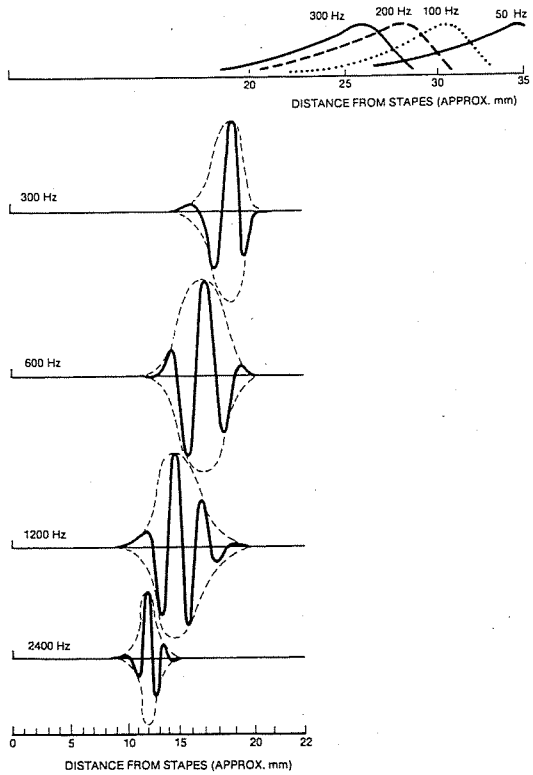


Fig. 1.3 The travelling wave on the basilar membrane for pure tones. The place of the maximum displacement depends on the frequency of the tone. From Handel (1989, Fig. 12.6)

basilar membrane. The frequency for which the smallest amount of energy is needed to activate the nerve fiber is called the *characteristic* or *best frequency*. Notice that this characteristic frequency depends on the position in the cochlea, again reflecting the tonotopic organization of the cochlea. The relation between frequency and position has been modeled by Greenwood (1990). According to his formula, the characteristic frequencies are approximately linearly distributed along the length of the

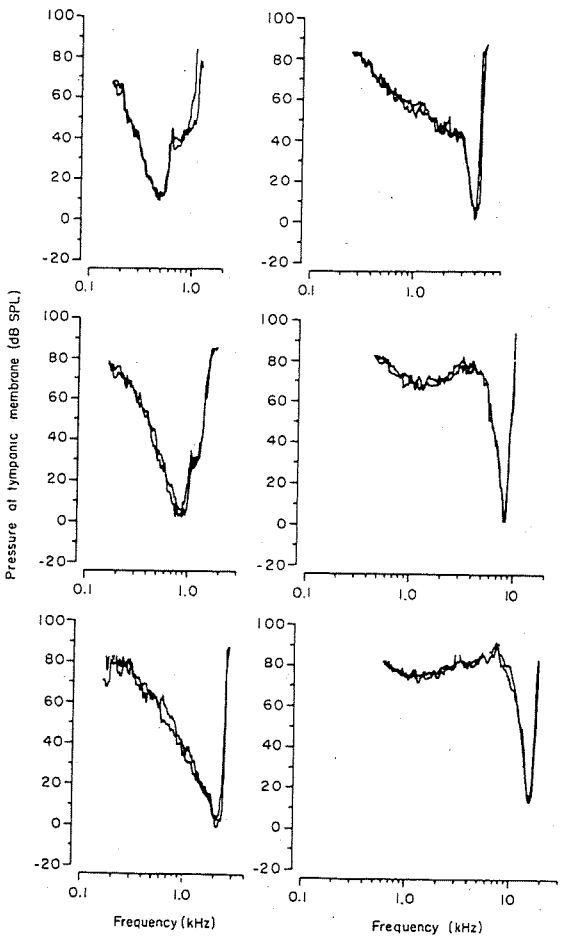


Fig. 1.4 Tuning curves of auditory nerve fibers at six different tonotopic places. Each plot shows the curve for two nerve fibers, having similar characteristic frequencies and thresholds. From Pickles (1988, Fig. 4.3)

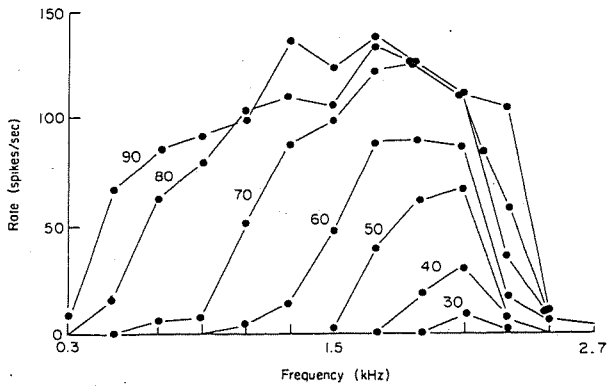


Fig. 1.5 The relation between discharge rate, intensity and frequency of a pure tone for a single nerve fiber. The characteristic frequency of the fiber is 1700 Hz. The numbers on the curves indicate the intensity in dB SPL. From Pickles (1988, Fig. 4.7)

basilar membrane for frequencies below 1 kHz, and logarithmically for higher frequencies. This distribution also explains why the tuning curves are broader for the lower frequencies. In conclusion, one can state that the cochlea acts as a spectrum analyzer.

Intensity coding

If the intensity of an incoming sound increases, the displacement of the basilar membrane will increase. Hence, the probability that a particular nerve fiber fires will increase too. It is generally assumed that the subjective measure loudness is proportional to spike count, although recent studies reveal that this relationship is not perfectly linear (Relkin and Doucet, 1997).

If intensity increases, not only the spike rate of nerve fibers in the vicinity of the corresponding tonotopic place will increase, but also more nerve fibers in a broader region will start firing. This is illustrated in figure 1.5. The discharge rate of an auditory nerve fiber was measured for a pure tone with constant intensity as a function of frequency. All the curves were obtained for a single nerve fiber with a characteristic frequency of 1700 Hz. Although the frequency seems to be coded less accurate according to the

tonotopy, the pitch of a loud pure tone is not less salient than that of a softer pure tone with the same frequency. This is probably due to the *phase-locking* effect: frequency is not only coded as the place in the cochlea, but also through the average time interval between successive spikes (Rose *et al.*, 1967). This time interval coding also correlates well with the pitch of periodic complex tones (Cariani and Delgutte, 1996). An overview article about periodicity coding in the auditory system is written by Langner (1992).

Although there are many other physiological phenomena in the ear, the concept of current cochlear implant systems is mainly based on the two mentioned above. The speech algorithms used today all attempt to mimic the tonotopic organization and the intensity coding. How this is accomplished, will be explained in the next paragraph.

1.3 Cochlear implants

1.3.1 The hardware

A cochlear implant system always consists of an internal part, implanted during surgery, and an external processor. This processor is mostly worn on the body, but very recently, some manufacturers implemented all external electronics in a BTE (behind the ear) hearing aid. In the early days, coupling between both parts was percutaneously, i.e. a wire from the external processor had to be plugged in a connector placed in the skull, to which all electrodes, inserted in the cochlea, were attached. At present, all clinical devices employ transcutaneous coupling using two inductively coupled coils, one placed beneath the skin during surgery, and one external coil placed on the skin. Usually, a magnet in each coil ensures that the external coil is kept in place. Figure 1.6 shows all parts of a cochlear implant device with transcutaneous coupling: a BTE with a microphone and transmitter electronics, a speech processor, two inductively coupled coils, an internal processor containing a receiver, a decoder and current sources, and finally the array of electrodes, which are inserted in the cochlea through the round window.

The hardware signal flow is as follows: an acoustic signal is captured by the microphone and sent to the speech processor, where it is converted from an analog to a digital signal. The speech processing algorithm is implemented in the digital signal processor. Here, the current amplitudes to be applied on the electrodes are calculated and coded. This information is

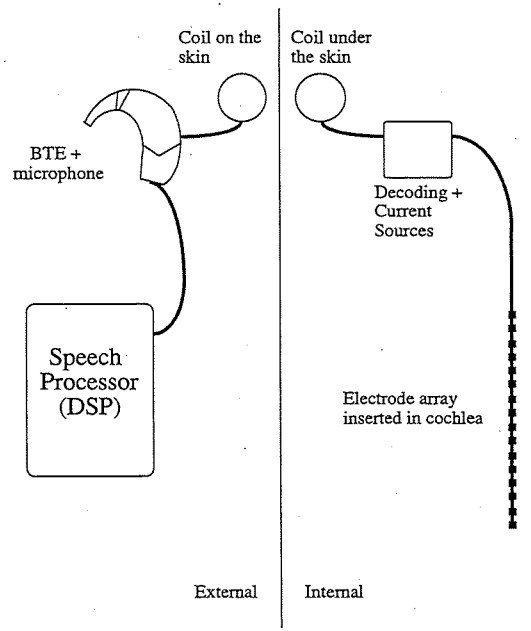


Fig. 1.6 A cochlear implant system with trancutaneous coupling. The external part is on the left, the internal part on the right.

sent to the transmitter in the BTE, which modulates this code with an RF carrier. This signal is sent through the two coils to the internal processor, where the current amplitudes are extracted. Finally, the current sources apply these currents on the electrode array.

The auditory nerves are commonly stimulated with biphasic current pulses. The width of each phase is typically within the range 10μs–300μs, and is either fixed or variable depending on the used algorithm. The net electric charges of both phases are balanced, in order to avoid DC current injection. Various stimulation modes exist to combine two or more electrodes. In monopolar mode, the current is applied between one electrode in the cochlea and one distant reference electrode which may as well be placed outside the cochlea. Typically, the same reference electrode is used for all other electrodes. In bipolar mode, two near electrodes are combined for the current flow, which is then limited to a

narrower region. This way, one aims to increase the selectivity of the stimulation. Finally, in common ground mode, there is one active electrode, with all the others connected to the electrical ground. Depending on the device, the mode and the electrode combinations are either fixed or can be altered. Each combination of electrodes between which a current will flow is called a channel. The maximum number of channels depends on the device: all commercial clinical devices at this moment contain at least 8 channels.

1.3.2 The speech processing software

Through the years, many speech algorithms have been used. It is not the purpose of this paragraph to give an historical overview. This can be found in Loizou (1998). The focus will be on the two currently most used strategies: the CIS strategy and the SPEAK strategy.

General

The software converts the speech signal into a series of electric pulses to be applied on the electrode channels. All technical degrees of freedom of the device can be employed: for each pulse the algorithm determines the amplitude, the width, the channel and the timing. Depending on the algorithm and the device, some of these parameters are fixed, others are varied as a function of the input signal. Obviously, all algorithms attempt to optimize the speech intelligibility. To recover the hearing sensations as natural as possible, the physiology of normal hearing is taken into account. As a result, to mimic the tonotopic organization of the cochlea, one of the first steps in all known algorithms is to perform a spectral analysis in some way by applying a bank of bandpass filters, by calculating the fast fourier transform or by extracting local maxima in the envelope of the instantaneous spectrum, such as e.g. the formant frequencies of vowels. The information in each frequency region will, after some more processing, be mapped on the corresponding channels. Low-frequency information will be mapped on deeply inserted electrodes, high-frequency information on the electrodes closer to the base. This way, the tonotopic organization is incorporated in the speech processing algorithms. Note however that the electrode array can not be fully inserted in the cochlea, which means that the apical nerve fibers can not be activated. Since it is undesirable to discard the sound information in the corresponding low-frequency regions, this information will be mapped on a location that is not a tonotopically exact match.

Intensity is coded by varying the pulse amplitude or the pulse width: if the energy in a particular frequency band is high, a large or wide current pulse will be applied on the corresponding channel. Indeed, if the amplitude or the width of a pulse is increased, the probability that a nerve fiber generates a spike increases too (Hodgkin and Huxley, 1952; Hartmann *et al.*, 1984; Javel *et al.*, 1987; van den Honert and Stypulkowski, 1987). Hence, if a series of large pulses is applied, a particular nerve fiber will generate more spikes, and the number of nerve fibers in the vicinity of the stimulated electrodes that will generate spikes increases too, together resulting in an increased spike count. The minimum and the maximum current level for a particular channel depend on the sensitivity of the subject for electrical stimulation on that channel. The threshold level is the smallest current level for which the subject can still detect the stimulus. The maximum level corresponds commonly to the highest current level that the subject can tolerate for a long time. The amplitudes of all current levels will be within these two limits, which are determined during a so-called fitting session. Note that the intensity coding is mimicked qualitatively rather than quantitatively. Current algorithms do not guarantee that the spike count resulting from the processed stimulus is the same as for the original acoustic stimulus.

In summary, a speech processing algorithm always contains at least the following steps: a frequency analysis to mimic the tonotopy, an energy extraction in each frequency band and a mapping within the dynamic range of the corresponding channel. The way these steps are implemented varies between algorithms. Some may also contain additional steps. Some commonly used strategies will now shortly be described.

CIS

The *Continuous Interleaved Sampling* (CIS) strategy was introduced by Wilson *et al.* (1991), and is now implemented in almost all clinical devices. It resulted in a substantial increase in speech intelligibility compared to the Compressed Analogue (CA) strategy that was commonly used until then. While in the latter case analogue signals are applied simultaneously on all channels, the CIS strategy makes use of brief pulses which are presented to the electrode channels at a high fixed rate and in a non-overlapping sequence. This way, electrical channel interaction is decreased dramatically. Figure 1.7 shows a block diagram of the CIS strategy with 6 channels.

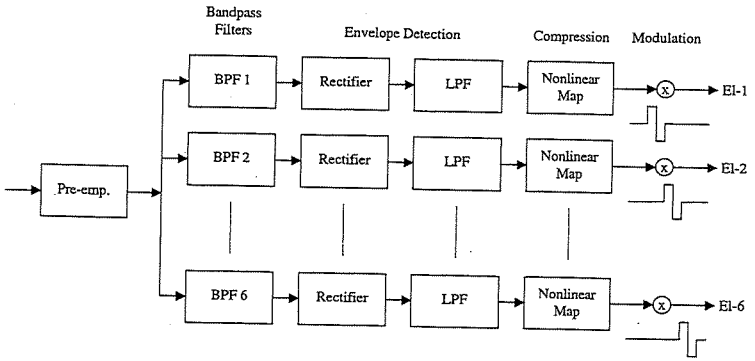


Fig. 1.7 Block diagram of the CIS strategy with 6 channels.

After some preprocessing, e.g. a pre-emphasis filter, a filter bank covering the speech frequency range is applied. It consists of as many band-pass filters as there are electrode channels available. In the original configuration six channels were used. The cut-off frequencies of successive filters overlap and are usually linearly distributed below 1 kHz and logarithmically above 1 kHz, corresponding to the place-frequency coding in man (Greenwood, 1990). As an example, the eight-channel filterbank used in the clinical Lauraflex processor is shown in figure 1.8. The next step is an envelope detection, realized by a full-wave rectifier and a smoothing low-pass filter, with a cut-off frequency of typically 400 Hz. Then, the envelope values are mapped within the dynamic range of the corresponding channel, using either a logarithmic function ($y = A.\log(x) + B$) or a power law function ($y = A.x^p + B$, with p typically between 0.2 and 0.5). These envelopes then modulate the amplitude of a pulse train, having a pulse rate equal to the overall pulse rate divided by the number of channels. The maximal overall pulse rate is the inverse of the total pulse duration. It is one of the essential features of CIS that none of the pulses on the various channels overlap, but that they are interleaved. This way, channel interaction is virtually nil. An example of the output of an eight-channelCIS algorithm is shown in figure 1.9 (A and B). The input was the Dutch word "som". The stimulation rate is 1250 pulses per second (pps) for each channel.

Many parameters of the CIS algorithm and their influence on speech reception scores have been investigated. For speech in quiet, these scores

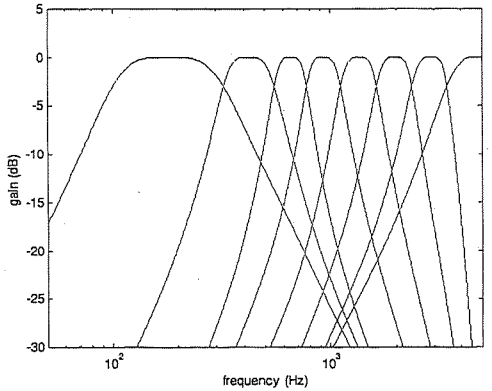


Fig. 1.8 The filter characteristics of the eight-channel filterbank used in the clinical Lauraflex processor. All filters are band-pass 4^{th} order Butterworth filters. Cut-off frequencies are linearly spread below 1 kHz, and logarithmically above 1 kHz0.

initially increase as the number of channels is increased, but then reach a plateau around eight channels (Lawson *et al.*, 1997). However, experiments with implantees and acoustic simulations with normal listeners show that these scores improve further for more than eight channels, if speech is presented in noise (Dorman *et al.*, 1998; Fu *et al.*, 1998). Other parameters that were investigated by Lawson *et al.* (1997) are the pulse rate, the pulse polarity, the channel order of presentation (from apex to base or staggered), the frequency range covered by the filter bank and the cut-off frequency of the smoothing filter. In general, best results were obtained for a high pulse rate (at least 833 pps per channel) and a high upper frequency of the filter bank (9500 Hz rather than 4500 Hz). Fu and Shannon (1998) measured the speech reception scores for various values of p of the power law compression function and found a rather broad maximum around p=0.2.

SPEAK / SMSP

The Spectral Peak strategy is implemented in the most used clinical device, the Nucleus cochlear implant, which holds 22 electrode channels in the cochlea. The algorithm is derived from the *Spectral Maxima Speech Processor* (SMSP), an experimental processor developed by McDermott *et*

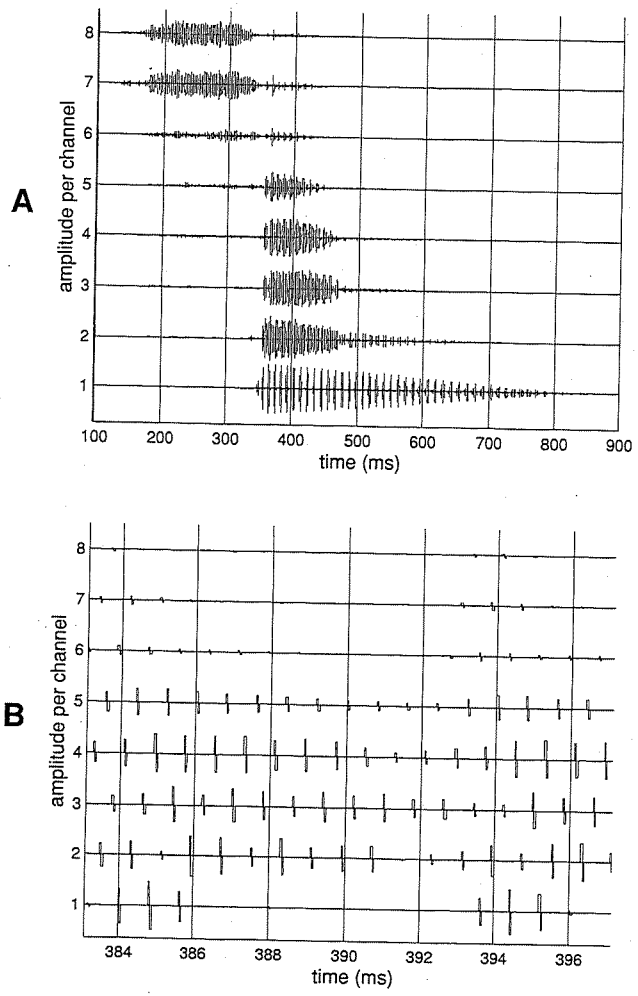


Fig. 1.9 (A) Output of the CIS algorithm for the word "som". The pulse amplitudes are shown for each channel as a function of time and reflect the envelopes of the bandpass filter outputs. (B) A more detailed view of the output, showing the interleaved pulse trains on the various channels.

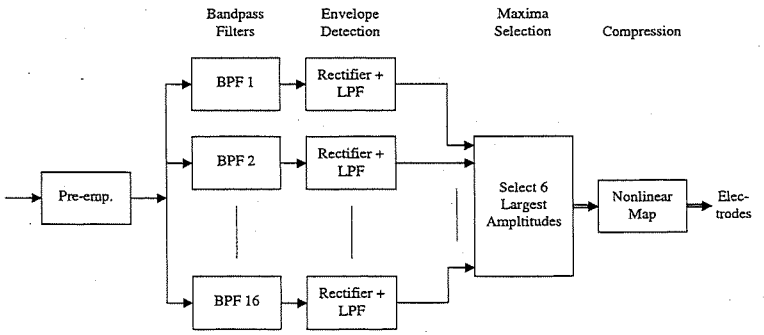


Fig. 1.10 Block diagram of the SMSP strategy

al. (1992). The major difference with the CIS strategy is that more bandpass filters are employed, and that only a limited number of channels are selected for stimulation, specifically those having the highest amplitudes. Another important difference is the substantially lower stimulation rate (for each channel, 300 pps or lower). A block diagram of the SMSP strategy is represented in figure 1.10.

After pre-emphasis, the signal is processed through a bank of 16 bandpass filters spanning the speech frequency range. Next, the envelopes are computed using a full-wave rectifier and a low-pass filter, with a cut-off frequency of 200 Hz. Each 4 ms, the six largest envelope outputs are selected for stimulation and compressed within the dynamic range of the corresponding channel. Finally, the six selected channels are stimulated consecutively. Since the maxima are selected each 4 ms, the overall stimulation rate is fixed at 1500 pps ($(4ms)^{-1}x6$). An example of the output of the SMSP algorithm is shown in figure 1.11. The input was the same as for the example shown in figure 1.9, i.e. the Dutch word "som".

The SPEAK strategy is a modified version of the SMSP. Instead of 16 filters, 20 filters are used, covering a wider frequency range. Also, the number of maxima selected is not fixed to 6, but varies from 5 to 10 depending on the spectral content of the input signal. This results in a variable stimulation rate, between 180 pps and 300 pps per channel.

The stimulation rate of the SPEAK strategy is lower than that of the CIS strategy, but more channels are used. So there is a trade-off between the

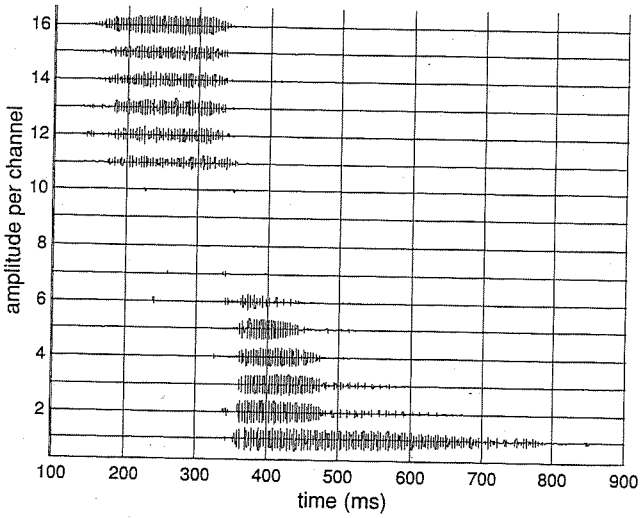


Fig. 1.11 Output of the SMSP algorithm for the word "som". The pulse amplitudes are shown for each channel as a function of time.

higher spectral resolution of SPEAK and the higher temporal resolution of CIS. Until today, speech perception results do not show clearly an advantage of one algorithm over the other. Moreover, it is difficult to compare both strategies "intra-subject", since there is no device in which both strategies can be implemented optimally.

Feature extraction strategies

Some older strategies are still used in the Nucleus implant. They are mentioned here because their basic concept differs from the above mentioned strategies. Instead of presenting a transformed waveform on the various electrodes, some features are extracted from the input signal. An example is the F0/F1/F2-strategy, in which the fundamental frequency (F0), and the amplitude and the frequency of the first two formants (F1 and F2) are calculated (Blamey *et al.*, 1987). Next, the electrodes that correspond tonotopically with the formant frequencies are stimulated with a pulse, the

amplitude of which is derived from the formant amplitude. For voiced phonemes, the stimulation rate is equal to the fundamental frequency; for unvoiced phonemes a quasi random rate of about 100 Hz is used. The MULTIPEAK or MPEAK strategy emerged from the F0/F1/F2-strategy. Here, some high-frequency information is added to the output, by stimulating three extra basal electrodes with a current derived from the envelope of three high-frequency bandpass filter outputs. This way, consonant information is better transmitted, as also revealed by identification tests (Skinner *et al.*, 1991). In fact, the MPEAK strategy can be considered as an intermediate step between the feature extraction strategies and the waveform presentation strategies, such as SMSP and CIS.

1.3.3 The Laura device

The subjects who participated in the studies described in this thesis, are all users of the Laura cochlear implant system (Peeters *et al.*, 1993), which is manufactured by Philips Hearing Implants. The electrode array consists of eight bipolar channels, as shown in figure 1.12. Through the years, three versions were developed, differing in the position of the electrodes on the strip, the size of the electrode surfaces and the way the bipolar channels are composed. The wider electrode separation for each bipolar pair in the later versions has the advantage that a lower current is needed to activate a certain number of nerve fibers. In general, the pulse width is 40 µs per phase, but 100 µs or 200 µs pulses can also be applied if the subjects' sensitivity to short pulses is too low. A single biphasic pulse with a fixed width and a variable amplitude can be applied to one of the eight channels each 100 µs if the pulse width equals 40 µs/phase. For larger pulse widths, a new biphasic pulse can be sent after 2 times the pulse width per phase plus 100 µs. Simultaneous or overlapping stimulation of two or more channels is not possible. The maximal value of the current amplitude is approximately 1500 µA.

In daily life, all the subjects use the clinical Lauraflex processor, in which a variation of the CIS strategy is implemented. The incoming speech signal is sampled at a rate of 13 kHz, and filtered by eight or less band-pass filters, the exact number corresponding to the number of available active channels. The filterbank spans the range from 100 Hz to 5 kHz and the cut-off frequencies are linearly distributed below 1 kHz and logarithmically above 1 kHz (see figure 1.8). The envelope detection consists of a method comparable to AM demodulation in radio broadcasting techniques. For the compression, a piecewise linear approximation to a square root curve is

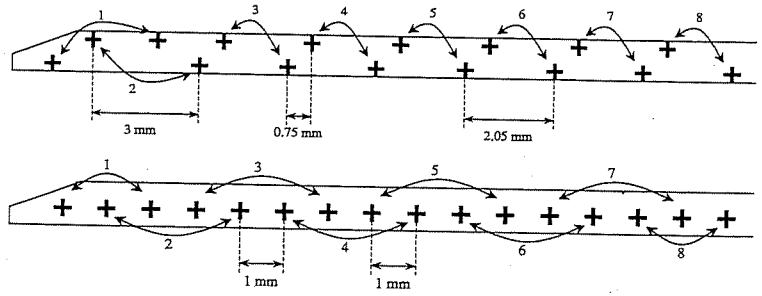


Fig. 1.12 Schematic representation of two versions of the Laura electrode array (not to scale): type 5.6 (top) and type 5.7/5.8 (bottom).

used. The overall stimulation rate is 10000 pps, so the stimulation rate per channel is this number divided by the number of active channels, e.g. 1250 pps if eight channels are used. For all the experiments however, this clinical speech processor is not used. Instead, a custom-made experimental set-up is used (see chapter 2).

1.4 Speech tests and evaluation

To evaluate a new speech processing algorithm, its benefit with respect to some known strategy is measured. Therefore, speech recognition scores on several speech tests, as obtained by listeners using the algorithms, are compared. Two broad classes exist: closed set and open set speech tests. In the former case, all response alternatives are known by the subject under test. Some examples are consonants in vowel context (asa, apa, ...), vowels in consonant context (heed, had, ...), or the Erber list (12 existing words, Erber and Alencewicz, 1976). In an open set speech test, the subject does not know the response alternatives, but usually some information about the speech material is given, such as the length or the composition of the words. Some examples are monosyllables of the form consonant-vowel-consonant (CVC), such as the NVA-list (Wouters et al., 1994), spondees having approximately the form CVCCVC, such as the BLU-list (Wouters et al., 1994), or sentence tests, such as the set composed by Plomp and Mimpen (1979). All these tests also differ in the degree of redundancy. A consonant or vowel test contains no context information, so that the subject can base his response only on what he has heard during the presentation of the

phoneme to be recognized. On the other hand, if some part of a sentence was not understood well, the brain can fill in the gap, based on the words that were understood and the possible meaning of the sentence. In between, the words of the NVA-list contain only little redundant information. Perceptual experiments revealed that the number of statistically independent phonemes varied between 2.4 and 2.7 for normal hearing and hearing impaired subjects (Bosman *et al.*, 1992). This means that, on the average, the subject has to identify more than 2 phonemes correctly in order to identify a word of this list. For sentences, usually the number of correctly identified keywords are counted. For spondees, the whole word must be correctly identified. There are two possibilities for monosyllables: either the words or the phonemes are scored. Obviously, for vowel and consonant tests, the correctly identified target phonemes are counted. Scores are mostly presented as percentages.

For these tests, and in some cases also for monosyllables, so called confusion matrices can be created in which element n_{ij} indicates the number of times response j was given for stimulus i. The diagonal elements correspond to the correct answers, the other elements to the confusions. It is often interesting to look for the largest off-diagonal elements, since these indicate which confusions were the most frequent. This provides some information about the major shortcomings of the algorithm. Such matrices can be further analyzed by calculating the relative feature information transmission (Miller and Nicely, 1955). This number indicates how well the several characteristics of the speech tokens are perceived. To do this, the stimulus set is first subdivided in several categories according to some feature. Common features for consonants are *voicing* (vocal chords vibrate during phoneme or not, e.g. "z" opposed to "s"), *frication* (phoneme is produced by forcing air through a narrow opening, e.g. "s" and "v"), *nasality* (sound energy comes out of the nose, e.g. "m"), *plosivity* (pressure is built up by closing nose and mouth and is then suddenly released, e.g. "p" and "d") and *place of articulation* (indicates the origin of the obstruction, e.g. in the front for the "p" and in the back for the "k"). For vowels, typical features are the formant frequencies, which correspond to the frequencies of the local maxima in the envelope of the vowel spectrum, and the duration of the vowel, e.g. the long vowel "oo" as in the Dutch word "boot" opposed to the short vowel "e" as in the Dutch word "bed".

As an example, consider the following confusion matrix for Dutch vowels. The rows correspond to the stimuli, the columns to the responses. Each word of the form /h/-vowel-/t/ was presented 12 times in randomized

order, so 72 words were presented in total. The sum of the diagonal elements is 40, and the score is 56%.

	hoet	hiet	haat	hot	hit	hat
hoet	8	2	0	2	0	0
hiet	4	4	0	1	3	0
haat	0	0	10	0	0	2
hot	2	0	0	7	0	3
hit	0	1	0	4	6	1
hat	0	0	0	7	0	5

Two features will be analyzed. F1, the first formant frequency, has two categories: *low* (hoet, hiet, hot, hit) and *high* (haat, hat). F2, the second formant frequency, has three categories: *low* (hoet, hot), *middle* (haat, hat) and *high* (hiet,hit). The categorization is based on the spectral analysis of the vowels. Now, a reduced matrix is calculated by summing the elements n_{ij} over all stimuli belonging to some category and the responses belonging to the same or some other category. This results in a 2x2 matrix for the feature F1, and in a 3x3 matrix for the feature F2:

F1	low	high	Σrow
low	44	4	48
high	7	17	24
Σcol.	51	21	72

F2	low	middle	high	Σrow
low	19	3	2	24
middle	7	17	0	24
high	9	1	14	24
Σcol.	35	21	16	72

Now, the *mean logarithmic probability MLP* or *entropy* can be used to calculate the amount of information at the input, the output and the transmission. For the input x, representing the stimuli, it can be calculated using the formula:

$$MLP(x) = -\sum_i p_i . \log_2(p_i) \qquad [1.1]$$

in which p_i is the probability of occurrence of stimulus i in the input. Since the logarithm base two is used, the formula gives the number of binary

decisions needed to specify the input. Similar formulas can be used for the information in the output y, representing the responses, and in the joint occurrence of input and output, which is represented in the (reduced) confusion matrix:

$$MLP(y) = -\sum_j p_j . \log_2(p_j) \qquad [1.2]$$

$$MLP(xy) = -\sum_{i,j} p_{ij} . \log_2(p_{ij}) \qquad [1.3]$$

The *mutual information* or *transmitted information* from input to output can now be calculated using:

$$T(x:y) = MLP(x) + MLP(y) - MLP(xy) \qquad [1.4]$$

The probabilities p_i, p_j and p_{ij} are estimated from the obtained matrix, respectively by using the row sum n_i, the column sum n_j and the matrix element n_{ij} divided by the total number of presentations n. This results in the following formula:

$$T(x:y) = -\sum_{i,j} \frac{n_{ij}}{n} . \log_2 \left(\frac{n_i . n_j}{n_{ij} . n} \right) \qquad [1.5]$$

Finally, the relative information transmission T_{rel} is calculated by dividing this result by the result that would be obtained if all information was transmitted correctly, i.e. if the reduced confusion matrix would be diagonal:

$$T_{rel} = \frac{T(obtained\ matrix)}{T(perfect\ matrix)} \qquad [1.6]$$

For the six-vowel confusion matrix shown above, this results in a relative information transmission of 33 % for F1, and 32 % for F2. A score of 100% means that the information for that feature is perfectly received: there are no confusions between stimuli belonging to different categories. A score of 0% means that the corresponding feature is not perceived at all, so confusion is maximal. In the third chapter, the obtained data will be analyzed using the presented method in order to compare two speech processing strategies.

1.5 The research questions

The three major research questions of this thesis will now be shortly introduced. Since these are rather independent of each other, a more detailed introduction will be given in the chapters dealing with the topic.

1.5.1 Mimicking neural adaptation

Since cochlear implants are intended to restore hearing, several physiological functions of normal hearing are incorporated in their design, such as the tonotopic organization and the intensity coding. However, other phenomena in normal hearing exist that are not adequately mimicked, such as neural adaptation. This effect can be described as the changing neural discharge rate while a stationary sound is presented. The discharge rate is maximal shortly after the onset of the sound and then decays gradually to a steady state level. The adaptation effect was observed, measured and modeled for pure tones (Smith, 1979; Westerman and Smith, 1984) and for speech (Delgutte, 1980; Delgutte, 1997). The "peaks" of neural discharge last about 15-20 ms and are more prominent, the more sudden the intensity of the stimulus increases. The temporal variations of the speech envelope seem to be emphasized by these peaks. Since the origin of the adaptation effect is at the synapses between inner hair-cells and auditory-nerves (Meddis, 1986; Meddis, 1988; Meddis *et al.*, 1990; Lopez-Poveda *et al.* 1997) and since a cochlear implant bypasses the haircells, no similar effect is seen in the nerve fiber's response to electrical stimulation (Parkins, 1989). Although there appears to be some form of adaptation, both the amplitude and the time constant of the effect differ substantially.

Two questions now arise: how can the adaptation phenomenon be included in a speech processing algorithm and what is the effect of the new algorithm on the speech intelligibility? In the third chapter a way to add the desired effect will be presented, by using a new signal processing strategy to calculate the speech envelope. Also, several speech tests were performed by cochlear implantees to evaluate the new algorithm (see also Geurts and Wouters, 1999).

1.5.2 Coding of the fundamental frequency

The fundamental frequency (F0) is the frequency of the first harmonic of a periodic sound. The pitch at which a person speaks is strongly related to

it. Yet, not all the harmonics – even not the first one – have to be present in the sound, in order to characterize the pitch. A more accurate definition would be that the fundamental frequency corresponds to the highest frequency for which all harmonics can be represented as an integer multiple of it. All vowels are periodic sounds, for which the frequency interval between the harmonics can be used to measure F0.

The response of a specific auditory nerve fiber to a vowel, or any harmonic complex, depends on its characteristic frequency and on the amplitudes of the harmonics in the vicinity of that frequency. In the case that a single harmonic dominates the response of the nerve fiber, the time representation of the neural response synchronizes to that harmonic, and nearly not to F0. On the other hand, the responses of many nerve fibers are influenced by several harmonics, and due to beating, the time representations of these responses show a clear modulation according to F0 (Miller and Sachs, 1984; Palmer, 1990). For the CIS algorithm, F0 is also apparent as a temporal variation in the envelope of the speech signal, because most frequency bands contain several harmonics (see fig. 1.9A around 400 ms). Obviously, the modulation depth of the F0 variations should be large enough in order to evoke the right pitch. Indeed, McKay *et al.* (1995) showed that if the modulation depth is too small, a pitch corresponding to the carrier frequency is perceived, instead of the modulating frequency.

The following questions arise: how can one control the modulation depth of the signal at the output of a speech processor, and does an increase of this depth improve the perception of the pitch corresponding to F0? In the fourth chapter, a new method to manipulate the modulation depth will be presented. Four cochlear implantees participated in several pitch discrimination experiments to evaluate the role of this parameter (see also Geurts and Wouters, 2000a).

1.5.3 Sensitivity to time delays between channels

The order of stimulation for the CIS strategy does not depend on the input signal, but is fixed. Several orders have been suggested: sequentially all channels from apex to base, or from base to apex, or a staggered order. An example for the last option is the default setting in the Lauraflex speech processor for eight channels: 1-4-7-2-5-8-3-6. It is reported that for one particular subject, speech reception scores were better for the apex-to-base order than for the staggered order (Lawson *et al.*, 1997). However, until now no extensive speech tests have been conducted to evaluate the suggested

orders thoroughly. Rather than experimenting with this parameter in a speech algorithm, it is meaningful to investigate first if it is feasible that there is any perceptual effect at all when pulse presentation order is changed. When this order is modified, the time interval between stimulation on either pair of channels changes.

So, the question is: can the brain compare different short time intervals between neural responses from different regions in the cochlea? In the fifth chapter a study is presented in which it was investigated to which extent a cochlear implantee can detect a difference in time delay between two pulse trains, concurrently presented on two widely separated channels. This study was carried out in collaboration with Dr. R. Carlyon of the MRC Cognition and Brain Sciences Unit, Cambridge, UK (see also Carlyon *et al.*, 2000).

1.6 Restrictions of time and means

The participation of cochlear implant users is essential for all the studies described in this thesis. The new speech algorithm which incorporates neural mimicking will be compared with the standard CIS speech algorithm. The results for several speech tests are needed in order to evaluate both algorithms in an objective manner. Also the speech algorithm which enhances the F0 variations in the envelope will be compared with the standard strategy. The just noticeable difference in F0 will be measured for each algorithm in several conditions. In the psychophysical study the subjects have to listen to and judge on hundreds of stimulus pairs. However, in Flanders, there are only about 15 adult Laura cochlear implant users that are willing to participate in such experiments, and several selection criteria reduce this number even more. Moreover, the time they can donate is limited too. This all means that all tests have to be carefully designed and optimized.

Also the finite degrees of freedom of the Laura cochlear implant have to be taken into account. A single current pulse can be sent to either channel every 100 µs. The amplitude can be set to any value between 0 and 1500 µA, but the width of the biphasic pulse is restricted either to 40 µs, 100 µs or 200 µs per phase. The combination of the 16 electrodes into eight bipolar pairs can not be changed. Another limitation is that it must be feasible for all new speech algorithms that they can be implemented in real time in a speech processor. So, complexity and necessary computing power and memory should not exceed certain limits.

1.7 Summary

The studies that will be described in the next chapters emanate from the current knowledge of physiology of hearing, psycho-acoustics and cochlear implant technology. The basics of these topics were summarized in this chapter. The three major research questions were shortly presented. The first deals with the question how neural adaptation can be incorporated in the speech processing for cochlear implants. The idea is that better mimicking of the physiology of normal hearing may improve speech intelligibility with a cochlear implant. The coding of the fundamental frequency in CIS processors is the topic of the second study. The focus will be on enhancing the F0 variations in the speech envelope at the processor's output. Several psycho-physical experiments are part of the third study, in which it is investigated to which extent cochlear implant users can detect time delays between stimulation on well separated channels. For all studies, the participation of implantees is indispensable, but unfortunately not infinite. Also the large but finite capacity of the Laura implant has to be taken into account. Before the above-mentioned studies are described, details will be given about the computer-based experimental set-up which was used for all experiments: APEX, a user friendly tool to create stimuli, program experiments, and deliver the stimuli to the subject under test. It can be used for psycho-physical studies and for the evaluation of new speech processing strategies.

Chapter 2

APEX: a research interface for psycho-physical and speech processing experiments with the Laura cochlear implant

Abstract

Until now, numerous psycho-physical experiments have been conducted with cochlear implant users. These provide a better understanding of the perceptual effect of varying one or more parameters of electrical stimulation such as current amplitude, stimulation mode, pulse rate or frequency, place of stimulation, pulse width, modulation depth of amplitude modulated pulse trains and duration of the silent gap between successive pulse trains. Various methods are used to obtain quantitative measures of the effect of these parameter variations. Some are the method of adjustment, method of limits, method of constant stimuli, rating scales and triad experiments (Stevens, 1951). APEX, an acronym for computer Application for Psycho-Electrical eXperiments, is a user friendly tool to conduct psycho-physical experiments and to investigate new speech coding algorithms with users of the Laura cochlear implant. Most common psycho-physical experiments can easily be designed and stimuli can easily be created without any knowledge of computer programming. The hardware consists of a PC, a DSP board, the Laura BTE (Behind The Ear) hearing aid and a graphics tablet. The pulsatile stimuli are composed using custom made Matlab functions and are stored on hard disc or CDROM. The functions convert either a speech signal into a pulse sequence or generate any sequence of pulses based on the parameters specified by the experimenter. In both cases, all technical degrees of freedom of the Laura device can be exploited. Since the stimuli at this level are not patient dependent, they can be used for all participating subjects. The APEX PC software reads a file which specifies the experiment and the stimuli, controls the experiment, delivers the stimuli to the subject through the DSP board, collects the responses from the graphics tablet and writes the results to the same file. The DSP program maps the amplitude of the stimulus pulses to values within the dynamic range of the subject under test and generates the sequence which drives the internal part of the Laura device. APEX is used for all research described in this thesis.

Geurts, L. and Wouters, J. (2000b). "A concept for a research tool for experiments with cochlear implant users", submitted to J.Acoust.Soc.Am.

2.1 Specifications

Several requisites are imposed on the hardware and the software of the experimental set-up. The first requisite of the hardware is that it will not induce additional limitations on the performance of the internal part of the cochlear implant. This implies that the hardware link between the computer and the internal part can transmit data at a sufficiently high rate. The parameters specifying one current pulse are coded as the time intervals between the falling edges within the code signal sent to the implant (see further). The resolution of this interval is in the order of 100 ns, which settles the minimal bandwidth of the hardware link. Also, a new code signal defining a current pulse should be able to start every 100 µs. A second requisite is that the whole set-up is portable, enabling participating subjects to be tested at their homes. At last, the set-up should contain as few custom made devices as possible, so that any part is generally available to other research groups. Another advantage is that long delays are prevented in case of a technical failure.

The experimenter who creates the stimuli and designs the experiment, should not have to be an expert in computer programming. The editor software for the electrical pulse stimuli should be simple and easy to handle. Another requirement is that all degrees of freedom of the internal part can be exploited. This means that every 100 µs a current pulse with any amplitude and any polarity can be delivered on any channel. However, there is a logic structure in most stimuli, which implies that a limited set of simple functions is sufficient. Some examples are stimuli consisting of pulse trains with constant amplitude and inter-pulse interval on one or more channels, amplitude modulated pulse trains or two consecutive pulse trains separated with a silent gap. It is also desirable that all parameters specifying the experiment can easily be set, preferably in a simple text file. The following experimental procedures should be available: a closed-set speech identification task with automatic calculation of the confusion matrix, the method of adjustment, including loudness balancing, the method of limits and the method of constant stimuli (Stevens, 1951). The software should be applicable to all patients, so it has to take account of possible technical differences between their devices. Also, not all subject dependent stimulus parameters are known in advance, such as the optimal current levels. Mostly, these are determined during a fitting procedure. The software should allow easy and fast manipulation of these subject data. In short, the generated stimuli should be independent of subject parameters, and the software should be able to fit these stimuli to the subject under test.

2.2 APEX

2.2.1 General description

A schematic representation of APEX is shown in figure 2.1. The experiment is controlled by the PC program *Apex.exe*, which runs under the Windows 95 operating system. The input of the program is a simple text file, which contains all parameters specifying the experiment. No specific programming skills are required to define a new experiment. Stimuli are created off-line in Matlab (version 5.1), enabling easy and flexible manipulation of all stimulus parameters. Several Matlab functions are available for stimulus generation: functions that convert sampled speech to pulse sequences and functions that generate stimuli specified by the input parameters. At this level, the amplitudes of the current pulses are not patient dependent, so the stimuli can be used for all subjects. The stimuli can be stored in files on hard disc or CDROM. The PC program loads the required files during the experiment and sends the data to a DSP board (TMS320C30), together with some subject and implant dependent parameters. The DSP program converts the subject independent amplitudes to values within the subject's dynamic range. A fast link, DSPLINK, connects the DSP to a BTE hearing aid of the same type as the one all Laura subjects use. The DSP program also generates the pulse sequence which codes each current pulse and sends it through the link. The subject can give his response on a graphics tablet, which usually is a stimulus number in a series of stimuli or a word out of a closed set of words. Alternatively, a mouse can be used. The PC program collects the responses, calculates some statistical values and writes the results to the input text file. The latter is done to keep all the conditions and the results of the experiment together in one file.

2.2.2 The APEX hardware

Figure 2.2 shows a picture of the whole set-up. A laptop computer (Compaq LTE Elite with 486DX processor) is inserted in a docking station (Compaq Smartstation) which contains the TMS320C30 DSP board[1]. On the left, a BTE is connected with the DSPLINK interface. The graphics

[1] The TMS320C30 DSP board is a product of Blue Wave Systems, Inc. (http://www.bluews.com)

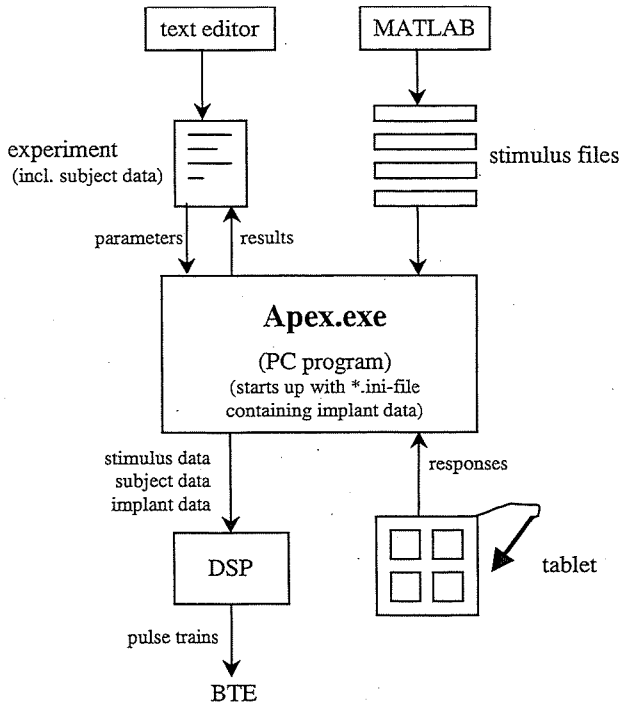


Fig. 2.1 Schematic representation of APEX.

tablet (Acecat II[2]) and the pen are on the right. The tablet is connected to the PC through the serial port.

Several *securities* of the Laura system limit the risks to the patient. First, there is no galvanic coupling between the computer and the implanted part. The information about the current pulses to be generated and the power for the electronic circuits are transmitted transcutaneously through two inductively coupled coils. Second, if there is an error in the code received by the internal processor, no current pulse will be generated. Third, there is a built-in technical limit on the highest current that can be generated of about

[2] The Acecat II graphics tablet is a product of Acecad, Inc. (http://www.acecad.com)

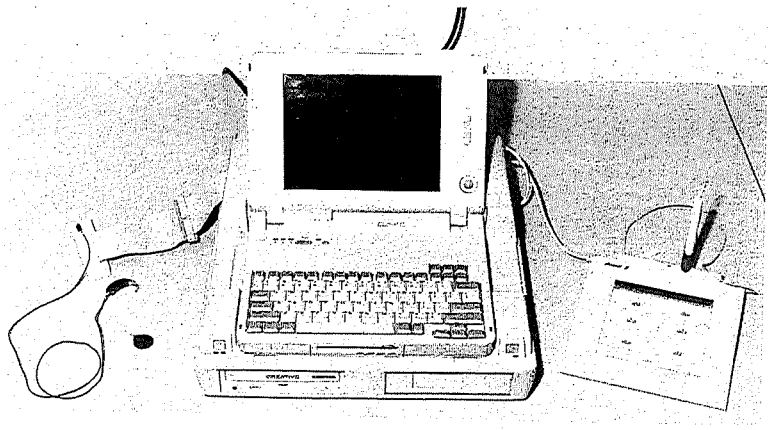


Fig. 2.2 The APEX hardware

1.5 mA. Fourth, each coil contains a permanent magnet which keeps the coils at their place. If a stimulus sounds very loud or irritating, the subject can easily remove the external coil.

Several additional parts are available to verify the stimuli generated by the APEX set-up, before an experiment is carried out. The external coil can be connected via a second coil to a box with a test implant, containing exactly the same electronics as those implanted in patients, except for the electrode array. The outputs of the current sources are connected to eight resistors. The voltages across these resistors are amplified using instrumentation amplifiers (Burr-Brown INA101HP) and four outputs of these amplifiers can be connected to a four-channel digital oscilloscope (Lecroy 9304AM) to vizualize the programmed pulsatile stimuli.

2.2.3 The APEX software

2.2.3.1 The Matlab functions

In an earlier version of the APEX platform, stimuli were created in real time on the DSP board (Geurts *et al.*, 1996). These stimuli were either the output of a speech processing algorithm or consisted of a pulse sequence of which the parameters could be varied by the experimenter. The current version uses stimuli that are created off-line with Matlab functions, which have several benefits compared to DSP programs: creating, debugging, changing and extending Matlab code is easier, the user interface is more

flexible, the output can be verified easily and several toolboxes, such as Matlab's signal processing toolbox, are available containing useful functions. The functions are either speech processing algorithms, converting an array of speech samples at the input into a pulse sequence, or synthesizers of more elementary stimuli as specified by the input parameters. Although each pulse within a stimulus can be applied on any channel with any amplitude, there is generally a logic structure in such a stimulus. So, a limited set of simple functions is sufficient for most desired stimuli.

The format of a pulsatile stimulus in Matlab is a matrix with twice as many columns as there are channels, and as many rows as ten times the length of the stimulus in milliseconds. Every row corresponds to a 100 µs interval and contains the pulses that will be applied during that interval. The amplitudes are specified in the first half of the columns, the modes in the second half. Every column in each half corresponds to an electrode channel. At this level, the channels and the amplitudes are not patient dependent. During the experiment, each channel is mapped to any of the eight implant channels and the value of the amplitude will be mapped within the subject's dynamic range on that channel in real time. Two points specify the mapping function, i.e. the currents that correspond to value "0.0" and to value "100.0" (see section 2.2.3.3). The maximum value of the amplitude is "204.6" ($(2^{11}-2)/10$). The value of the mode specifies the biphasic pulse:

"1"	40 µs/phase, apical electrode cathodic first
"2"	40 µs/phase, basal electrode cathodic first
"3"	100 µs/phase, apical electrode cathodic first
"4"	100 µs/phase, basal electrode cathodic first

If mode "1" or "2" is chosen, a new pulse can start after 100 µs. However, for mode "3" and "4" the resolution is 300 µs: 100 µs for the first phase, 100 µs for the second and 100 µs for the command that resets the amplitude. Table 2.1 lists the functions that can be used to create and edit a stimulus. All except the function *zeros.m* and the concatenation command, are custom made functions. An example of a stimulus synthesis is given in the appendix. Figure 2.3 displays a stimulus as it would appear on the computer screen by using the function *show.m*.

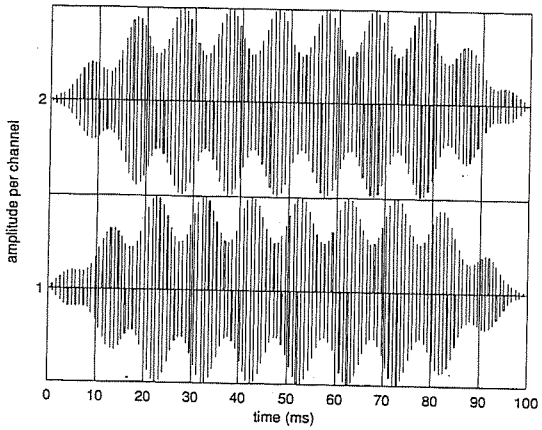


Fig. 2.3 Example of a stimulus as shown with the Matlab function *show.m*. The limited resolution does not permit that both phases of all pulses are visible.

Table 2.1: overview of the Matlab functions that can be used to create and edit pulsatile stimuli.

Function name	Description
zeros.m	generates an empty matrix
train.m	generates a pulse train with a given amplitude and inter-pulse interval on one channel
samtrain.m	applies amplitude modulation to a pulse train with a sine wave with a given modulation frequency and depth
gate.m	applies a window with linear ramps with a given rise/fall time
z=[x;y]	concatenates stimuli *x* and *y*
clearch.m	deletes the pulses on a given channel
show.m	displays the stimulus on a time scale (see figure 2.3)
compress.m	compresses and saves a synthesized stimulus

Before the stimulus is saved with the function *compress.m*, it is compressed in two steps. In the first step, it is checked if there are simultaneous pulses on two or more channels. Since the Laura device does not permit simultaneous stimulation, the pulse on the channel with the lowest priority, as specified with one of the function's input parameters, will be deleted. When there are simultaneous pulses on 2 channels with the same priority, one of them, randomly chosen, will be randomly shifted forwards or backwards. The output of the function *compress.m* is a symmetric square matrix with as many rows and columns as there are channels: the diagonal elements represent the number of pulses on each channel, the other elements represent the number of concurrent pulses on the channels corresponding to the row and the column of the element. In the second step, the stimulus is further compressed in order to utilize as few bits as possible. The information for a single pulse is coded in a 16 bit word: the three MSBs contain the channel (0-7), the next two bits the mode (0-3), and the 11 LSBs the current (0-2047). The floating point value of the current is multiplied by 10 and rounded to the nearest integer. The value "2047" is reserved: it indicates that another mapping function is valid (see section 2.2.3.3). This way it is possible to change the mapping function at any time within the stimulus. The experimenter has to specify the various options for this mapping in the experiment file (see section 2.2.3.3).

Finally, there are a few more functions that are part of the APEX system, but that are not intended to generate stimuli. In some experiments, stimuli are presented in a random order, e.g. in word or number identification experiments or when the method of constant stimuli is used. Several functions are available for the generation of a random list of numbers appropriate for the experiment (see example in paragraph 2.2.4). There is also one function that calculates the information transfer from a confusion matrix, as suggested by Miller & Nicely (1955) (see chapter 1, section 1.4).

2.2.3.2 The PC program: Apex.exe

The PC program is written in Visual C++ (Microsoft Developer Studio 4.0). The object oriented design contains several classes corresponding to different psycho-physical methods. The program recognizes the class by the extension of the input text file, and reads all the parameters specifying the experiment. In general, a series of stimuli is presented to the subject under test, who has to make a judgment about these stimuli. He gives a response using the graphics tablet or the mouse. The program collects and

summarizes all the responses and in some cases calculates one or more statistical measures. The results are always appended to the input text file, in order to keep the conditions and the results of the experiment together in one file. Table 2.2 gives an overview of all the experiments and an example of a typical application. Each class will be described shortly. A detailed description of the input text file for each class would be too lengthy, but two examples will be given in section 2.2.4.

Table 2.2: overview of all the experiments of Apex.exe

Class name	Task	Example of application
CIdentify	Identify a word or phoneme	Closed set identification
CDiscriminate	Indicate if 2 stimuli are the 'same' or 'different'	Phoneme or pitch discrimination
CBalance	Balance 2 stimuli in loudness	Typically used prior to a discrimination task
CAdjust	Find the best match to a given stimulus from a list of stimuli	Pitch matching
CAdaptive	Identify the 'signal' from a list of stimuli	Exploration of the limit of performance in a discrimination task
CConstant-Stimuli	Identify the 'signal' from a list of stimuli	Measurement of a psychometric curve
CTriade	Give the most similar or the most dissimilar pair out of 3 stimuli	Exploration of the perceptual most important features of different stimuli
CThreshold	Indicate the interval where the stimulus was presented	Determination of the threshold of hearing
CCount	Count the number of presented stimuli	Determination of the threshold of hearing

CIdentify (*.idn) This class is generally used for closed-set word identification tasks, e.g. consonants in vowel context or numbers. A randomized series of stimuli from a limited list will be presented to the subject under test. Next, the subject points with the pen to the word on the

tablet that he has heard. The program constructs a confusion matrix and appends it to the input file. Each entry of such a matrix indicates the number of times a particular stimulus (row) evoked a particular response (column). From these matrices the transmitted information for several speech features, such as duration, peak energy and voicing can be calculated (see chapter 1). Another way to analyze these data is to construct so called distance matrices which are used as an input to multi-dimensional scaling techniques. These reveal the perceptual dimensions which are used by the listener to distinguish the stimuli (e.g. van Wieringen and Wouters, 1999a).

CDiscriminate (*.dsc) In a discrimination task the subject under test is presented either with two different stimuli or two identical stimuli. These stimuli can either be speech tokens or more elementary stimuli for psycho-physical research. The subject has to indicate if the two stimuli were the same or were different. The program output is a list of all presented stimuli pairs with the number of times each of both possible responses was given.

CBalance (*.bln) In psycho-physical research, it is often necessary to balance a set of stimuli in loudness. This is to prevent that the loudness of a specific stimulus can be used as a cue to distinguish it from others. In a balancing task two stimuli are presented from which one has to be adjusted by the subject so that it sounds equally loud as the other stimulus. The parameter which is varied is the amplitude of the current pulses presented at the electrode array. The final amplitudes for each balanced pair are appended to the input file.

CAdjust (*.adj) An adjustment task is comparable with a balancing task, the difference being that not the amplitude, but some other parameter is varied. The criterion which has to be equalized is not restricted to loudness only, but can be any other cue, such as pitch or timbre. The input file contains a list of stimuli in which one or more parameters are varied, e.g. the pulse rate. A stimulus from this list and a fixed reference stimulus, which can have a totally different configuration, is presented. The subject can 'move' through the list and has to indicate which stimulus resembles the reference stimulus the most, following a given criterion. The output consists of the filename of the chosen stimulus.

CAdaptive (*.adp) An adaptive procedure is often used for a discrimination task: the subject is presented with a series of stimuli consisting of one so called 'signal' and one or more 'standards'. The signal differs from the standard stimuli in one or more parameters. This results in a cue that the

subject can use to differentiate between both stimuli. At the start of an adaptive procedure, the value of the parameter is made large resulting in a salient difference between the stimuli. The following presentation is determined by an X-down Y-up procedure (Levitt, 1971): the difference is made smaller after X correct responses, while Y incorrect responses will make it larger again. A *reversal* occurs when X or more correct responses are followed by Y incorrect responses, or vice versa. The values of X and Y can be chosen by the experimenter. The procedure stops after a predefined number of reversals. The program output is the followed sequence of the parameter values, the number of presentations and correct responses for each value and the mean of the values at the last *n* reversals, with *n* a number chosen by the experimenter.

CConstantStimuli (*.cst) The method of constant stimuli is also typically used for a discrimination task. The difference with an adaptive procedure is that the parameter value of the next presentation does not depend on the subject's response: a fixed number of presentations of a limited set of stimuli with different parameters will be presented in a random order. The output of the program leads to a psychometric curve, which represents the number or the percentage of correct responses for each parameter value.

CTriade (*.trd) In a triade experiment, the subject can listen to three distinct stimuli at a time simply by clicking the corresponding box on the screen or the graphics tablet. The task is to indicate which two of those three stimuli sounds the most similar and which two consists of the most unlike stimuli. The program output is a list of all possible stimuli pairs and the number of times that they were chosen to be the most similar and/or dissimilar pair. Also from these data distance matrices can be constructed which are used as an input to multi-dimensional scaling techniques.

CThreshold (*.thr) This class provides one way to determine the threshold of hearing. Two to four boxes appear on the screen, which will light up consecutively for several hundreds of milliseconds, according to the length of the used stimulus. Only during one interval, the stimulus will be presented, and the task for the subject is to indicate this interval. The amplitude of the stimulus is adaptively varied using an X-down Y-up procedure. The initial amplitude and the stepsize are set by the experimenter. The program output is a list of the amplitudes at the reversals and the mean amplitude at the last predefined number of reversals.

CCount (*.cnt) Another way to determine the threshold is the counted threshold estimate (Skinner, 1995). The stimulus is presented a random number of times, and the subject has to count the stimuli. The minimum and maximum number can be set by the experimenter, but are limited to two and five respectively. The amplitude is also adaptively varied using an X-down Y-up procedure. The program output is a list of the amplitudes at the reversals and the mean amplitude at the last *n* reversals.

2.2.3.3 The DSP program: 30electr.out

The DSP program receives three types of data from the PC program: subject data, stimulus data and sequencer timing data. The latter are stored in a simple text file *sequencer.ini*, and are sent to the DSP each time the PC program starts executing. Subject data are specified in the input text file (see further) and the data for a single stimulus are stored in a single file (*.*stm*). Both are sent to the DSP for each stimulus to be presented to the subject. The task of the DSP program is to combine and convert all these data to a pulse train which specifies each pulse of the stimulus. This pulse train is sent to the BTE where a 10 MHz carrier is amplitude modulated with the pulses. The RF signal is applied on the external coil, received through inductive coupling by the internal coil and demodulated and decoded by the internal circuitry, where current sources generate the pulse to be applied on the electrodes.

The stimulus data are stored in files, in which the information for a single pulse is coded in a 16 bit word: the three MSBs contain the Matlab channel (0-7), the next two bits the mode of stimulation (0-3), and the 11 LSBs the Matlab current (0-2047). The PC program sends the data in this format to the DSP. Two memory blocks of 1 kilobyte are used in the Dual Port Memory of the DSP board, which can easily and directly be accessed by both the PC and the DSP. While the PC sends the data to one block, the DSP processes the data in the other block. At this level, the channel and the current amplitude are not yet dependent on the subject. The DSP program converts both the channel and the current to subject dependent values. The mode is always left unchanged. The amplitude of the current pulse has to be mapped within the dynamic range of the subject. The mapping function is linear, so only two points are needed to define it. A Matlab amplitude of 0 is mapped to I0 and a Matlab amplitude of 1000 is mapped to I1000 (Fig. 2.4). In addition, if the Matlab amplitude is 0, no current pulse will be generated. I0 and I1000 are determined in a subjective manner, mostly at the start of a test session. In general, I0 and I1000 correspond to the subject's threshold

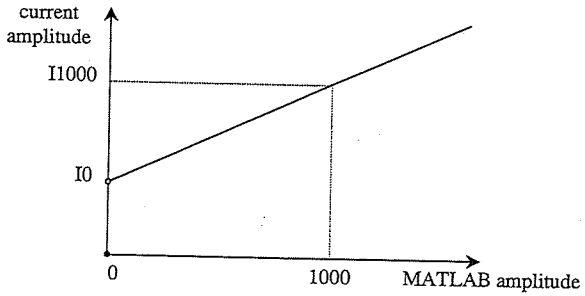


Fig. 2.4 Mapping function

and most comfortable level respectively. However, the experimenter is free to interpret both values in any way.

The number of the Matlab channel may also differ from the number of the corresponding stimulation channel of the subject's implant. This means that the DSP program also needs the information specifying which Matlab channel corresponds to which subject channel, in addition to the two levels for the current mapping. In the input files of the PC program, this mapping information is defined for every stimulus, e.g. as follows:

```
file=c:\apex\stimuli\example.stm
mapping=1,3,150,740,2,6,230,990
```

The file specified in the first line contains the stimulus. The next line contains the mapping information. Matlab channel 1 is mapped to subject channel 3, with I0=150µA and I1000=740µA. Next, the mapping information for Matlab channel 2 is given. It is possible to provide several mappings for each stimulus. The first mapping is valid from the beginning of the stimulus, and each time an amplitude of 2047 is received by the DSP program, the next mapping data in the list will be used (see section 2.2.3.1). Each group of mapping data is stored in a 8x3 table in the DSP's memory.

The last step is to convert the three parameters mode, channel and current to three timings which characterize the pulse train that is transmitted to the implant. The first pulse is a synchronization pulse, marking the start of a new pulse train. The width is typically around 4 µs. The next three pulses are so called data pulses and have a width of approximately 2 µs. The timings between the falling edges of subsequent pulses determine the

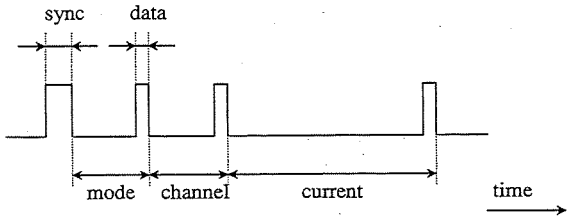


Fig. 2.5 Pulse train containing the information about a single current pulse.

stimulation parameters (figure 2.5): the first interval defines the mode of stimulation (3.5-8 µs), the second the channel (3.5-19µs), and the third the current (8-35µs).

The values of the widths of the sync and the data pulses and the timings corresponding to the various modes and channels vary somewhat between the devices, due to permitted technical tolerances. Therefore all these device dependent timings are sent to the DSP every time the PC program starts executing. The values are stored in the file *sequencer.ini* in the same directory as the executable file. Such a file should exist for each subject to be tested. The DSP program looks up each required value defining the channel and the mode in memory. The conversion from current amplitude to the corresponding timing is not device dependent and is also done by the DSP program: either a square root function is used, as in the Lauraflex, or a third order polynomial, which is based on the relation between the current and the time interval, as measured with the test implant of the laboratory.

The pulse train is generated with the processor's FLAGOUT bit, which is connected to the DSPLINK. This bit can be set high or low at processor speed, i.e. every 60ns its value can be changed. Refreshment is not needed: the bit holds its value after it is set or cleared. Also the ground (GND) and the +5V supply voltage (VCC) are connected to the BTE, to provide the necessary power.

2.2.4 Two examples

Each of the next two paragraphs contains an example of an input file and the corresponding output. The first is for an identification experiment, the second for a discrimination experiment with adaptive procedure. The

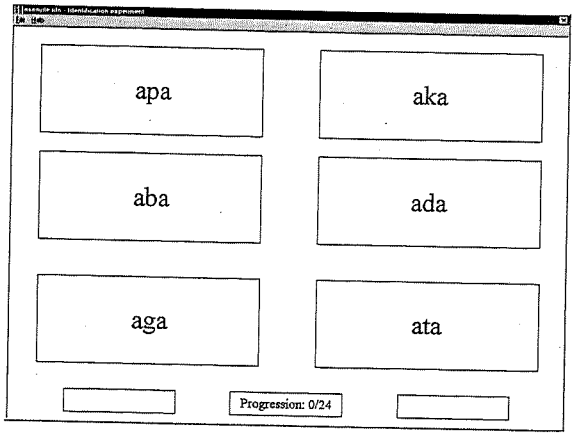


Fig. 2.6 The computer screen or the graphics tablet as it will appear during the experiment. The text in the boxes indicate the response alternatives. The boxes at the bottom are used for feedback (left box for correct response, "goed" will appear in the box, right box for incorrect response, "fout" will appear) or to indicate the progression of the experiment (first number corresponds to number of given responses, second number to total number of stimuli).

purpose is to illustrate the use of the software, and not to describe its full capabilities.

2.2.4.1 Example.idn: an identification experiment

The following experiment is a closed set stop consonant identification task, as used for the evaluation of a speech processing algorithm. The stimulus set consists of six stop consonants /p/, /t/, /k/, /b/, /d/, /g/ in intervocalic /a/ context (apa, ata, ...) of two male speakers. Each token will be presented twice to the subject under test in random order. The input file contains 4 sections. The first section specifies the size, the position and the text content of the six rectangles which will appear on the screen and/or on the graphics tablet (see figure 2.6). The text corresponds to the words that are presented during the experiment. The second section consists of the list of 12 stimulus files. First, the directory containing the stimulus files is specified. The second line includes the mapping data, which are valid for all stimuli. Then, for each stimulus the file name, the number of the

corresponding rectangle and a condition number corresponding to the speaker, is given. A confusion matrix will be constructed for all stimuli having the same condition number. The last line includes the file name of the DSP program, which will be downloaded to the DSP and executed when the experiment starts. The first parameter of the third section is a list of numbers, which specifies the order of presentation of the stimuli. The subsequent parameter indicates the number of initial stimulus/response pairs that will not be included in the confusion matrices. The last section contains some additional parameters for the experiment: the pause in milliseconds between the subject's response and the presentation of the next stimulus, the feedback parameter and the training parameter. In the case feedback is wanted, the correct answer will be shortly highlighted after the subject's response. Training means that the experiment is inverted: the subject first points to a word and subsequently is presented with the corresponding stimulus.

```
[screen]
box=5,5,45,30,apa
box=55,70,95,95,ata
box=55,5,95,30,aka
box=5,35,45,60,aba
box=55,35,95,60,ada
box=5,70,45,95,aga
font=28,Times

[input]
path=c:\tippex\stimuluspatterns\nonsens\cis\8ch\
mapping=1,1,10,1470,2,2,30,1470,3,3,140,630,4,4,90,680,…
5,5,80,910,6,6,220,640,7,7,430,1470,8,8,430,840
file=jwapa.stm
box=1
condition=1
file=jwata.stm
box=2
condition=1
file=jwaka.stm
box=3
condition=1
file=jwaba.stm
box=4
condition=1
file=jwada.stm
box=5
condition=1
file=jwakda.stm
box=6
condition=1
```

```
file=mdapa.stm
box=1
condition=2
file=mdata.stm
box=2
condition=2
file=mdaka.stm
box=3
condition=2
file=mdaba.stm
box=4
condition=2
file=mdada.stm
box=5
condition=2
file=mdakda.stm
box=6
condition=2
c30=c:\tippex\30electr.out

[randomlist]
list=12,5,7,12,5,1,5,11,9,8,3,7,7,4,4,2,6,9,10,12,11,6,…
8,2,1,3,10
skip=3

[procedure]
feedback=no
pause=1000
practice=no
```

Each time an experiment is completed, the results are appended to the input file. The first line specifies day and time of the completion of the experiment. The following section contains the two confusion matrices, one for each speaker. E.g., the first row of the first matrix shows that the word /apa/ was once identified correctly, and once as /ata/. The last section contains all stimulus response pairs, in the order of presentation.

```
Following results collected on Wed 22Apr98 at 15:10
[confusion matrices]
condition 1
apa     ata     aka     aba     ada     aga
1       1       0       0       0       0
0       0       2       0       0       0
0       0       2       0       0       0
0       0       0       1       1       0
0       0       0       1       1       0
0       0       0       2       0       0
```



```
condition 2
apa     ata     aka     aba     ada     aga
0       2       0       0       0       0
0       2       0       0       0       0
1       0       1       0       0       0
0       0       0       1       1       0
0       0       0       0       2       0
0       0       0       0       0       2

[stimulus/response pairs]
aga2/aga, ada1/aba, apa2/ata, aga2/aga, ada1/aba,…
apa1/apa, ada1/ada, ada2/ada, aka2/apa, ata2/ata,…
aka1/aka, apa2/ata, apa2/ata, aba1/aba, aba1/ada,…
ata1/aka, aga1/aba, aka2/aka, aba2/ada, aga2/aga,…
ada2/ada, aga1/aba, ata2/ata, ata1/aka, apa1/ata,…
aka1/aka, aba2/aba
```

2.2.4.2 A discrimination experiment with adaptive procedure

The following experiment is a so-called 2 interval – 2 alternative forced choice task (2I2AFC). The subject is presented with two stimuli, a *standard* and a *signal*, in random order, and has to point to the *signal*. As long as the subject correctly identifies the *signal*, the difference between both stimuli is made smaller. The difference arises from some parameter characterizing the stimuli, in this case the delay in milliseconds between two concurrent pulse trains on two different channels. The first section contains a list of *signal* or *variable* files, in which the parameter changes in discrete steps from file to file. The order of the files is such that discrimination varies from difficult to easy. For each stimulus, the file name, the mapping vector and a value are given. This is typically the value of the varying parameter. The second section contains the file name and the mapping data of the *standard* or *reference* stimulus. Details of the *method* or *procedure* of the experiment are given in the last section. The first line indicates that a 2I2AFC task will be used, with a silent interval of 500 ms between the stimuli. The first *signal* is the ninth file in the list (line 2). A 2-down 1-up procedure is used (line 3 and 4). At the start, the stepsize is 2 *signal* files, after four reversals it is 1 *signal* file (line 5). The experiment ends after 12 reversals (line 6), correct response feedback will be given (line 7), and the mean value of the parameter values at the last 8 reversal points will be calculated (line 8). The next presentation of stimuli starts 1500 ms after the subject's response (line 9), and there is no training (line 10). Training means in this case that the correct answer will be displayed on the screen before and during the presentation of the stimuli.

```
[varlist]
file=C:\Tippex\StimulusPatterns\Bob1Per100\del32.stm
mapping=1,2,400,600,2,7,450,723
value=3.2
file=C:\Tippex\StimulusPatterns\Bob1Per100\del34.stm
mapping=1,2,400,600,2,7,450,723
value=3.4
file=C:\Tippex\StimulusPatterns\Bob1Per100\del36.stm
mapping=1,2,400,600,2,7,450,723
value=3.6
file=C:\Tippex\StimulusPatterns\Bob1Per100\del38.stm
mapping=1,2,400,600,2,7,450,723
value=3.8
file=C:\Tippex\StimulusPatterns\Bob1Per100\del40.stm
mapping=1,2,400,600,2,7,450,723
value=4.0
file=C:\Tippex\StimulusPatterns\Bob1Per100\del42.stm
mapping=1,2,400,600,2,7,450,723
value=4.2
file=C:\Tippex\StimulusPatterns\Bob1Per100\del44.stm
mapping=1,2,400,600,2,7,450,723
value=4.4
file=C:\Tippex\StimulusPatterns\Bob1Per100\del46.stm
mapping=1,2,400,600,2,7,450,723
value=4.6
file=C:\Tippex\StimulusPatterns\Bob1Per100\del48.stm
mapping=1,2,400,600,2,7,450,723
value=4.8
file=C:\Tippex\StimulusPatterns\Bob1Per100\del50.stm
mapping=1,2,400,600,2,7,450,723
value=5.0

[reflist]
file=C:\Tippex\StimulusPatterns\Bob1Per100\del1.stm
mapping=1,2,400,600,2,7,450,723

[procedure]
sequence=afc,500,afc,-1
start=9
up=2
down=1
stepsize=2/0,1/4
stop=12
feedback=yes
numrevsformean=8
pause=1500
practice=no
```

The output of the program contains for each *visited* signal file the parameter value, the number of correct responses and the number of

presentations, in that order. Also the sequence of the parameter values of the visited signal files is given, and finally the mean value of the parameters at the last 8 reversal points.

```
Following results collected on Wed 04Jun97 at 14:47
Score for value 3.4: 1/3
Score for value 3.6: 6/8
Score for value 3.8: 7/11
Score for value 4: 4/5
Score for value 4.2: 7/8
Score for value 4.4: 3/4
Score for value 4.6: 2/2
Sequence=3.8,4.2,3.8,4.2,3.8,4,4.2,4.4,4.6,4.4,4.2,4,3.8,…
4,3.8,3.6,3.4,3.6,3.4,3.6,3.8,3.6
mean=3.775
```

2.3 Summary and conclusions

The APEX platform is developed to easily design and conduct experiments with Laura cochlear implantees. Various existing methods for classical psycho-physical tests and various speech intelligibility tests to investigate new speech processing algorithms are implemented. Stimuli are created off-line using Matlab functions, experiments are programmed off-line in simple text files. The PC program controls the experiment, while the DSP board serves as the interface to the implant. Expansion of the platform is also relatively easy. New Matlab functions can be created to generate stimuli, and new classes can be implemented in the PC program to conduct an experiment according to a new method or procedure. Also, the platform allows acoustical stimulation using a sound card.

Until now, many experiments have been conducted with the APEX platform. The concept and the implementation have proved to be suitable for all the experiments that were designed for the studies in this thesis (Geurts and Wouters, 1999; Geurts and Wouters, 2000a; Carlyon, Geurts and Wouters, 2000) and for several studies by other researchers (e.g. van Wieringen and Wouters, 1999a; van Wieringen and Wouters, 1999b).

Chapter 3

Including neural adaptation in CIS processors

Abstract

Speech processing algorithms for cochlear implants intend to mimic the physiological functions of the cochlea and the auditory nerve fibers. Tonotopic organization and intensity coding are incorporated in all existing multi-channel systems. However, one of the phenomena that is not included yet is the adaptation seen in the response of auditory nerve fibers to acoustic stimuli. In section 3.1, the effect and some existing models is described in more detail and the influence of the effect on speech perception is discussed shortly. Also the difference from adaptation to electric stimuli is explained, leading to the conclusion that the effect should indeed be implemented in speech processors for cochlear implants. A way to do this in the CIS strategy is presented in section 3.2. Both the standard strategy (CIS, without adaptation) and the new strategy (EECIS) are described in detail. The difference between both is the method used to calculate the speech envelope. The two strategies were tested perceptually with 6 post-lingually deafened users of the Laura cochlear implant (section 3.3). The tests included identification of stop consonants in 3 different vowel contexts and monosyllabic CVC (consonant-vowel-consonant) words. Significant improvements in correct identification scores were observed for stop consonants in intervocalic /a/ context (p=0.026): average results varied from 46 % correct for CIS to 55 % for EECIS. This improvement is mainly due to the better transmission of place of articulation. The differences in identification scores for stop consonants in /i/ and /u/ context were not significant. The identification scores for the medial vowels of the CVC words were significantly higher when the EECIS strategy was used: average results increased from 39% correct to 46% correct (p=0.018). No significant differences were observed between the results for initial and final consonants of the CVC words. An explanation of the results is proposed in section 3.4. The new algorithm and the results are discussed in section 3.5. In conclusion, the present results demonstrate that the inclusion of the rapid adaptation in the speech processing for cochlear implants can improve speech intelligibility.

Geurts, L. and Wouters, J. (1999). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J.Acoust.Soc.Am. 105, 2476-2484

3.1 Neural adaptation

3.1.1 Neural response to acoustic stimuli

In general, adaptation is the variation in response which occurs during a constant stimulus condition. In the case of auditory nerve fibers, it corresponds to the changing neural discharge rate or firing rate while a stationary sound is presented. Figure 3.1. shows the effect at several intensities as measured in the auditory nerve of a cat (Delgutte, 1980). Pure tones with constant intensity were presented acoustically and the firing rate was measured in function of time by counting the number of spikes within each small time interval for numerous presentations. In fact, the plots can be interpreted as firing probability functions. Already at very low intensities, there is a clear transition effect: the probability of discharge quickly rises to a maximum and then gradually decays to a steady-state level. Some authors suggested that the effect consists of at least two decaying components: rapid adaptation with a time constant of several milliseconds and a large amplitude, and short-term adaptation with a time constant of about 40 ms and a smaller amplitude (Smith, 1979; Westerman and Smith, 1984).

The effect is also seen in the response to speech signals (Delgutte, 1980; Delgutte, 1997). Figure 3.2 shows the response of several nerve fibers differing in characteristic frequency to the acoustically presented sentence indicated at the bottom. Several "peaks" of neural discharge, resulting from rapid adaptation, are prominent, and last about 15-20 ms. They occur each time the intensity increases in the frequency band for which the nerve fiber is most sensitive, e.g. for stop consonants, which are characterized by a sudden burst preceded by a silent period. Also, the more sudden the intensity increases, the larger the amplitude of the neural discharge peak is. Delgutte concluded that the rapid adaptation effect enhances the spectral contrast between successive phonemes.

3.1.2 Existing models

Several researchers included this effect in models describing the peripheral auditory processing. Westerman and Smith (1984) modeled the response function to stationary pure tones as the sum of three components: a constant term reflecting the steady state response and two decaying exponential functions reflecting the rapid and the short term adaptation. Meddis and co-workers developed a model predicting neural spikes from the

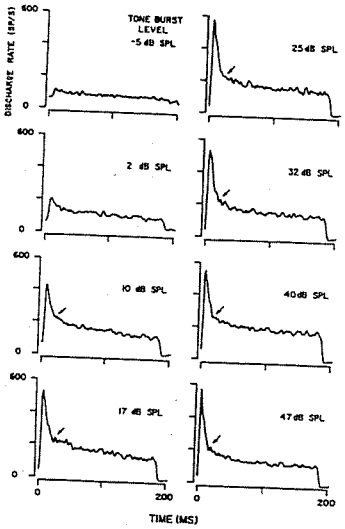


Fig. 3.1 Response of an auditory nerve fiber to acoustic pure tones in function of time. The intensity level is indicated in the top right of each plot. From Delgutte (1980, Fig. 1)

changing concentration of chemical components in the hair cell (Meddis, 1986; Meddis, 1988; Meddis *et al.*, 1990; Lopez-Poveda *et al.*, 1997). According to their model, adaptation is caused by the depletion of neurotransmitter at the synapses between hair cells and auditory nerves. This is very interesting, since a cochlear implant bypasses the hair cell. So, a different response to electric stimuli can be expected. Eggermont (1985) developed a model, based on existing stochastic models for birth and death processes. Patterson's time domain model of peripheral auditory processing includes adaptive thresholding (Patterson *et al.*, 1995; Holdsworth and Patterson, 1993). Such a threshold is calculated for each frequency channel and is a measure of the preceding sound energy in that channel. An output is generated only if the input exceeds this value. Delgutte measured and modeled the modulation transfer functions of several auditory nerves differing in characteristic frequency. Such a function expresses, as a function of frequency, the complex ratio of the modulation in the neural response to the amplitude modulation in the acoustic stimulus. He used this model to describe the neural encoding of

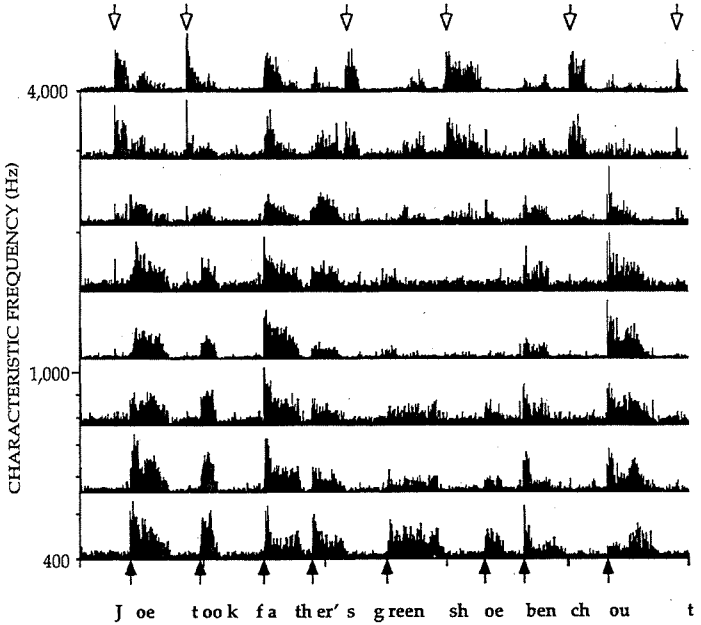


Fig. 3.2 Response of several auditory nerve fibers (characteristic frequency indicated at the left) to the sentence indicated at the bottom. The arrows indicate some discharge peaks occurring at sudden intensity increases. From Delgutte (1997, Fig. 16.1)

the temporal envelope of speech. His model could accurately predict the response as shown in figure 3.2 (Delgutte, 1997).

3.1.3 Importance of adaptation

Real life speech is characterized by the relatively quick succession, typically every 200-300 ms, of loud vowels and soft consonants, resulting in a rapidly varying intensity. Moreover, speech signals contain many silent intervals, e.g. prior to stop consonants. In the response of the auditory nerve, all these temporal variations at a syllabic rate seem to be emphasized by the adaptation "peaks", as illustrated in figure 3.2. Moreover, since the effect

depends on the intensity increases only in the frequency region for which the nerve fiber is most sensitive, it seems to increase the spectral contrast between successive phonemes. The importance of the typical low-frequency modulations for speech intelligibility was illustrated by Drullman *et al.* (1994). They reduced the slow temporal fluctuations of the speech by low-pass filtering the speech envelope - thereby reducing the adaptation effect -, but they preserved the spectral information. Speech recognition scores dropped fast for cut-off frequencies below 16 Hz. So, a loss of temporal information leads to poorer intelligibility. Furthermore, Shannon *et al.* (1995) showed that speech with reduced spectral information, but with the envelope information preserved, is still highly intelligible. They filtered the speech signals using four band-pass filters covering the speech frequency range, extracted and smoothed the envelopes with a low-pass filter, and used these envelopes to modulate white noise filtered by the corresponding band-pass filters. For cut-off frequencies of the envelope smoothing filter as low as 50 Hz, in which case the adaptation effect is preserved, speech recognition scores were still nearly perfect.

3.1.4 Neural response to electric stimuli

Since one of the origins of the adaptation effect is found at the synapses between inner hair-cells and auditory nerves, and since a cochlear implant bypasses the haircells, a different effect is seen in the nerve fiber's response to electrical stimulation. It is still true that the discharge rate may decay during the stimulus, which indicates that adaptation is not solely determined by the properties of the hair-cell / auditory nerve synapse (Wilson *et al.*, 1995; Killian *et al.*, 1994; Dynes and Delgutte, 1992; Parkins, 1989; Javel *et al.*, 1987; van den Honert and Stypulkowski, 1987). This is shown in figure 3.3: each plot shows the post-stimulus time histogram of a single nerve fiber in response to an electric pulse train (200 pps, 200 µs/phase), obtained at several intensities. It is clear that, at some intensities, there is a variation in the response, but the effect differs in several respects from the one observed for acoustic stimuli. First, the time constant is larger than that of short-term adaptation and much larger than that of rapid adaptation. Second, the ratio between the maximal discharge rate at the onset and the steady-state response is much smaller. Third, the effect is not consistent: not all fibers seem to show a decay in discharge rate (Dynes and Delgutte, 1992; Killian *et al.* 1994). Fourth, some fibers can cease to respond after a short time period of activity. This is the result of depolarization block and occurs mostly for stimuli at high rates and intensities (Killian *et al.*, 1994; Javel *et al.*, 1987). Fifth, as revealed by measurements of intracochlear evoked potentials (EPs),

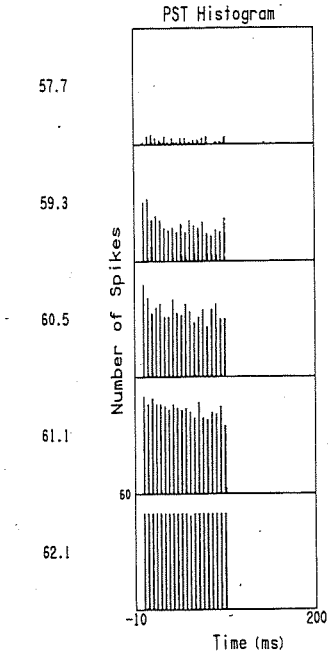


Fig. 3.3 Post-stimulus time (PST) histogram showing adaptation in the neural response to electric pulsatile stimuli. The amplitudes of the pulses in dB re 1µA are indicated on the left of each plot. From Javel *et al.* (1987, Fig. 3)

there is a large response on the first pulse of a high rate pulse train. Yet, this effect is due to the refractory properties[3] of the electrically stimulated nerve (Wilson *et al.*, 1995). In summary, there is no evidence of rapid adaptation of auditory nerves in response to electric stimulation, with the same time constant and amplitude as to acoustic stimulation. Hence, this effect should

[3] The absolute refractory period corresponds to the time interval after a nerve fiber discharge during which the nerve fiber can not discharge again. For electrically stimulated auditory nerve fibers, this interval lasts about 1 ms. During the succeeding relative refractory period, the nerve fiber can only discharge if the amplitude of the stimulus is high enough.

be added to a cochlear prosthesis processor to better mimic the neural response patterns to normal acoustic stimulation (Parkins, 1989).

3.1.5 Foregoing work

Two methods to incorporate rapid neural adaptation in a sound processor were suggested by Van der Borght (1994). In the first method, a sudden intensity increase is detected by comparing two subsequent local maxima in the speech envelope. If the difference exceeds a predetermined threshold, an "adaptation peak signal" is synthesized and added to the original envelope. It was difficult though to find a suitable value for this threshold. In the second method, a high pass filtered version of the original envelope is added to this envelope, resulting in a more natural response. However, it was difficult to achieve a correct adaptation response to low-frequency tone bursts. None of both methods were evaluated with cochlear-implant users.

3.2 Speech processing

In this section, a method to incorporate the rapid adaptation effect in the speech processing for cochlear implants is presented[4]. It includes a new algorithm to calculate the amplitude of the speech envelope, and it will be compared with the default CIS strategy. The high rate of stimulation of CIS processors allows the enhancement of peaks in the envelope with a duration of 10-20 ms. The two speech processing algorithms in this study consist of the following steps: prefiltering, band-pass filtering, envelope extraction, instantaneous compression, current mapping and pulse train modulation. The algorithms differ in the way the speech envelope is extracted, the other steps are identical. The processing for one frequency band is shown in figure 3.4: the left, bold trace in the flow diagram corresponds to the CIS algorithm, while the whole diagram corresponds to the EECIS algorithm.

3.2.1 Prefiltering and band-pass filtering

The input of the speech processing algorithm is digitized speech at a sample rate of 10 kHz and a resolution of 16 bit. Prefiltering is applied in order to compensate for the typical slope of -6 dB in the long term average

[4] Short-term adaptation is not incorporated. Due to its lower amplitude, it is not believed to have a substantial effect on speech perception.

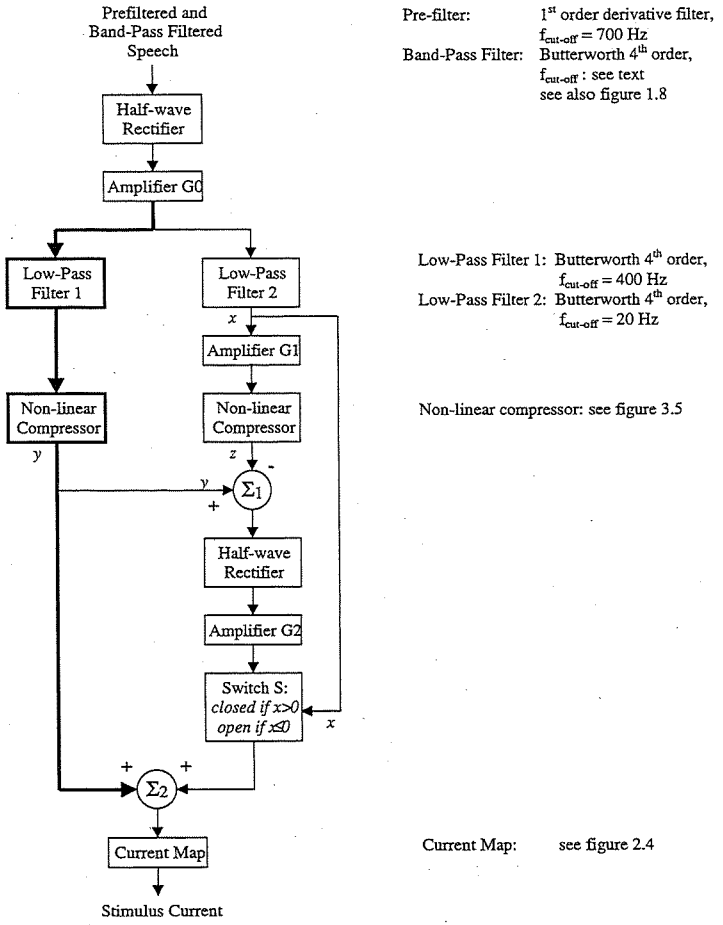


Fig. 3.4 Processing scheme for the standard envelope (left, bold trace) and the enhanced envelope extraction (both traces).

speech spectrum, starting from about 500 Hz (Byrne *et al.*, 1994). The pre-emphasis filter is a first order derivative filter, with the 3 dB cut-off frequency at 700 Hz, and a gain at DC of 0 dB. Next, the signal is filtered by different band-pass filters covering the speech frequency range (see figure 1.8 in chapter 1). The number corresponds to the number of active electrode channels in the cochlea of the implanted subject. Stimulation on non-active electrode channels do not elicit a hearing sensation due to atechnical failure or poor neural survival. The cut-off frequencies are linearly spread below 1 kHz and logarithmically above 1 kHz. For 8 channels, these are 100, 325, 550, 775, 1125, 1634, 2372, 3444 and 5000 Hz, and for 7 channels, these are 100, 357, 614, 934, 1421, 2161, 3287 and 5000 Hz.

3.2.2 CIS: standard envelope extraction

The remaining processing is identical for each channel and is shown for one channel in figure 3.4. The standard envelope is extracted in the left, bold trace. First, the signal is half-wave rectified[5], in contrast with the full wave rectification used originally in the CIS strategy (Wilson *et al.*, 1991). This is done because in the latter case the frequency, and thus the rate pitch, is doubled in the low frequency channels. This is not the case for the high frequency channels, since in that case the pitch appears as an envelope modulation and not as a single harmonic (see also chapter 4). Second, the speech envelope is smoothed by low-pass filtering the rectified signal (left, bold trace, low-pass filter 1: Butterworth 4^{th} order, cut-off frequency at 400 Hz). The resulting signal is amplified with a channel specific gain G0 to compensate for the greater attenuation caused by rectification in high-frequency channels. Third, the smoothed envelope is instantaneously compressed using the non-linear function shown in figure 3.5. This compression function is identical to the one used in the subjects' clinical speech processors and consists of a piecewise linear approximation of a square root curve. Sample values above 2^{14} are clipped. However, if the speech signal at the input has a resolution of 16 bits (values between -2^{15} and 2^{15}), only a very small number of samples exceed the value 2^{14} after band-pass filtering and smoothing.

[5] A half-wave rectifier preserves positive values and sets negative values to 0. A full-wave rectifier preserves positive values and inverts negative values.

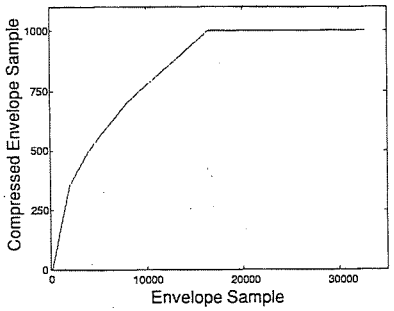


Fig. 3.5 Piecewise linear compression function. The input of the function (horizontal axis) is the 16 bit sample value of the envelope. The output is an integer value between 0 and 1000.

3.2.3 EECIS: enhanced envelope extraction

The enhanced envelope is the sum of the standard envelope extracted in the left, bold trace of figure 3.4 and a peak signal, extracted in the right trace whenever there is a sudden increase in the intensity of the input signal. Since there is no objective measure available that allows to settle the amplitude of the peak, a subjective evaluation by the subject under test is used (see discussion in section 3.5). For this implementation, only one extra parameter is needed to "fit" the algorithm to the cochlear implant user, namely the gain G2 of the peak signal. Another advantage of this implementation is that the only difference between both strategies is the presence of the peak signal. Hence, differences in performance are completely due to this peak signal. In short, rapid adaptation is simulated by increasing the amplitude of the speech envelope at sudden intensity increases.

The first step in the right trace of figure 3.4 is the derivation of another envelope from the rectified signal, using a low-pass filter with a much lower cut-off frequency (signal x, right trace, low-pass filter 2: Butterworth 4^{th} order, cut-off frequency at 20 Hz). This envelope is amplified (G1) and compressed, using the same compression function as the one in the left, bold trace. An example of the two envelopes is shown in figure 3.6. The solid curve shows the compressed envelope y extracted in the left, bold trace. The dashed curve shows the envelope z extracted in the right trace. The speech

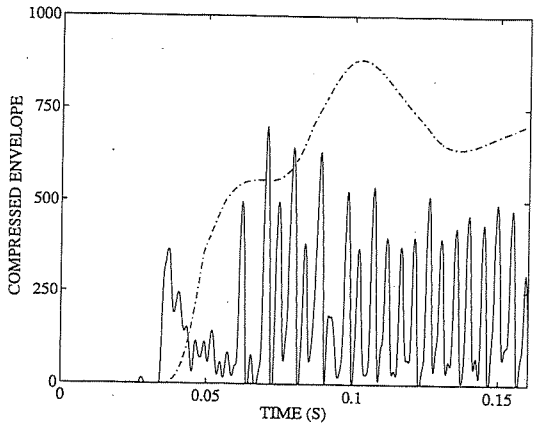


Fig. 3.6 Compressed envelopes of syllable /ka/ in channel 8 (frequency band between 3444 Hz and 5000 Hz) as a function of time. The solid line is the envelope of the fast filter (cut-off frequency at 400 Hz) and corresponds to signal *y* in figure 3.4. The dashed line is the amplified output of the slow filter (cut-off frequency at 20 Hz) and corresponds to signal *z* in figure 3.4.

signal is the syllable /ka/ (male speaker) in channel 8, so the band-pass filtering has cut-off frequencies of 3444 Hz and 5000 Hz. The burst of the /k/ occurs between 0.03 s and 0.05 s, the /a/ starts at 0.06 s.

The lower cut-off frequency of low-pass filter 2 serves two goals. First, more high frequency components are rejected, leading to a smaller ripple. It is clear from the example that no modulation according to the fundamental frequency (see chapter 4) is present anymore in the dashed envelope. Second, the filter has a bigger delay, so the output of this filter lags the output of the other one. The amplification G1 is chosen such that the 'slow' envelope *z* is lying above the 'fast' envelope *y* for the stationary part of the signal (the vowel). A suitable value is an amplification factor of 6, for all channels. Using this value, the criterion was met for all speech tokens in a large set consisting of consonants in three different vowel contexts /a/, /i/ and /u/ and of vowels in /pVt/ context, recorded from two male and two female speakers. However, at the onset of the speech utterance, the fast envelope *y* will always be above the slow one (*z*), due to the bigger delay of

low-pass filter 2. This part is the peak signal to be extracted. This is done by subtracting the slow envelope from the fast one (Σ1 in figure 3.4), thereby only retaining the positive values using half-wave rectification. The duration of the peak is in the order of magnitude of 10 ms, which is in correspondence with the time constant values of the rapid adaptation as measured by Delgutte (1980). The extracted peak signal is amplified by a factor G2 between 1 and 2, the exact value depending on a subjective evaluation by each subject (see section 3.3.2). The peak signal can be eliminated by setting G2 to 0.

The enhanced envelope is the summation of the standard speech envelope, extracted in the left trace, and the peak signal, extracted in the right trace (Σ2 in figure 3.4). It is important to notice that this addition is made after the compression of the signals. Otherwise, the peak signal would be reduced substantially. An example of the standard envelope (middle plot) and of the enhanced envelope (lower plot) in all channels for the syllable /ka/ is shown in figure 3.7. Notice the difference in amplitude for the envelope of the burst, corresponding to the /k/. In general, the enhanced envelope emphasizes the onset of the typical intensity increases in speech.

The switch S in figure 3.4 is used to prevent 'false peaks'. These can occur when there is a rapid decay in intensity, immediately after a sudden increase in intensity. In that case, the ringing caused by the low-pass filtering leads to a temporal negative output, even while the input remains positive. This effect is more prominent for the output *x* of the slow low-pass filter 2, since it has a larger time constant. This results in 'false peaks': the fast envelope *y* of the left, bold trace will be above the slow envelope *z* of the right trace, only because the latter is negative. This is illustrated in figure 3.8 for the syllable /to/ in channel 8: the false peak occurs at the arrow. By opening switch S when the output *x* of low-pass filter 2 is negative, these false peaks will not be added.

3.2.4 Current mapping and pulse modulation

The envelope samples are mapped to current values (last block at the bottom of figure 3.4), taking the dynamic range of the channel of the subject under test into account. A linear function is used, which maps an envelope sample value of 1 to the current threshold, and a value of 1000 to the most comfortable current level (see chapter 2, figure 2.4). The following parameters correspond to the standard strategy used in the Laura cochlear implant. The current envelopes in all channels are modulated with pulse

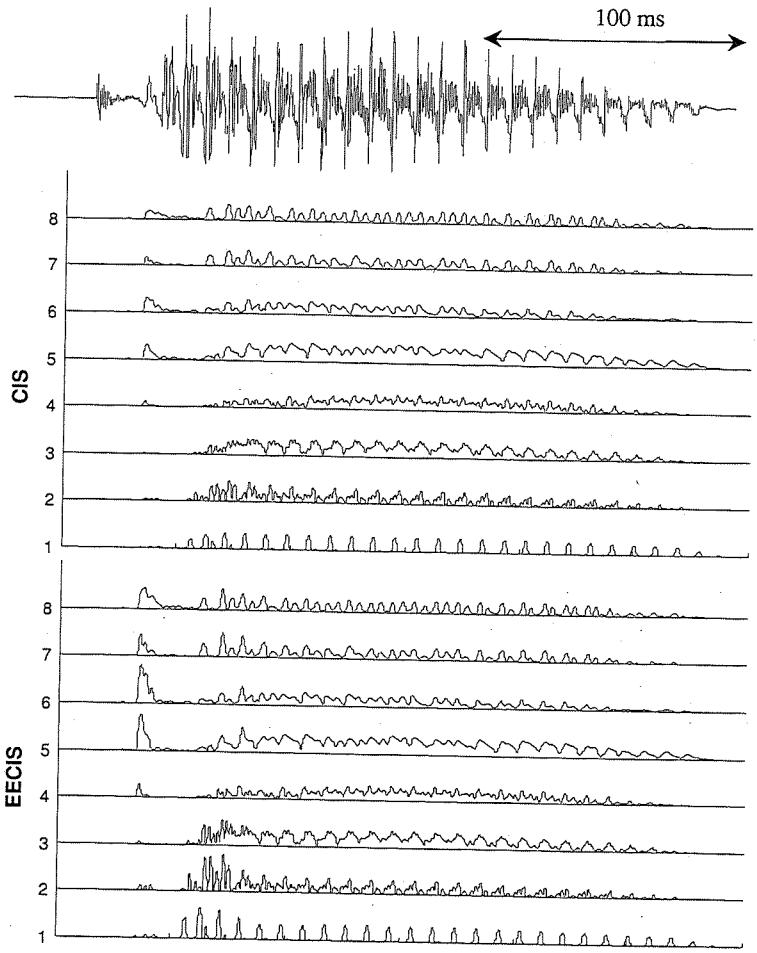


Fig. 3.7 The upper plot shows the sampled waveform of the syllable /ka/ as a function of time. The middle plot shows the corresponding envelopes in 8 channels for the CIS algorithm. The lower plot shows the envelopes for the EECIS algorithm. The value of G2 equals 1.5. The channel numbers are indicated before each envelope trace. Channel 1 corresponds to the most apical channel, channel 8 to the most basal channel.

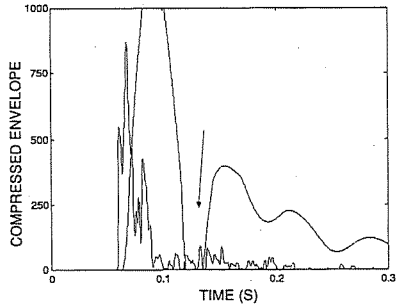


Fig. 3.8 Compressed envelopes of syllable /to/ in channel 8 as a function of time. The solid line is the envelope of the fast filter and corresponds to signal y in figure 3.4. The dashed line is the amplified output of the slow filter and corresponds to signal z in figure 3.4. The arrow indicates where the 'false peak' would occur.

trains with an overall rate of 10000 pps (pulses per second). If 8 channels are used, the pulse rate per channel is 1250 pps, while this rate equals 1429 pps in the 7 channel case. The width of each pulse is constant and equal to 40 µs/phase. According to the CIS strategy, the pulse trains of all channels are interleaved. The channel order of stimulation is staggered: 1, 4, 7, 2, 5, 8, 3 and 6, where channel 1 corresponds to the most apical channel and channel 8 to the most basal channel. Stimulation on the non-active electrode channel is skipped for subjects using only 7 channels.

3.3 Identification experiments

Several speech intelligibility tests were performed in order to compare both algorithms, the standard envelope extraction (CIS) and the enhanced envelope extraction (EECIS).

3.3.1 Subjects

Eight post-lingually deafened users of the Laura cochlear implant participated in this experiment. Table 3.1 lists some relevant information about each subject. All subjects use the Lauraflex speech processor, in which the CIS strategy is implemented. Columns 7 and 8 show the results

for vowel and consonant identification tasks obtained in a previous study (van Wieringen and Wouters, 1999a). All the tests were taken during two sessions of about 2.5 hours.

Table 3.1: Some details of the eight subjects who took part in this experiment. Column 6 shows the number of active electrode channels (#ch). Columns 7 and 8 show the results for an identification task of 10 vowels in /h/-vowel-/t/ context, and 16 consonants in /a/-consonant-/a/ context, respectively (van Wieringen and Wouters, 1999a). A suitable value for G2 (last column) could not be found for subjects HS and JH (see text).

Subject	Age (yrs)	Duration of profound deafness (yrs)	Aetiology	Implant experience (yrs)	#ch	Vowel score (%)	Consonant score (%)	Value of G2
AP	59	3	Cochl. otosclerose	1	8	49	29	1.5
DD	32	1	Trauma	3	8	61	35	1
GG	69	14	Meningitis	10	7	17	28	1.5
SV	29	6	Unknown	4	8	73	58	1.5
WM	48	5	Progressive	1	8	49	51	1.5
AE	42	3	Menière	2	7	79	51	2
HS	44	12	Congenital	2	8	40	35	*
JH	16	3	Meningitis	3	8	53	45	*

3.3.2 Fitting

Thresholds (THR) and most comfortable levels (MCL) were obtained at the beginning of the first session and were identical for both strategies. While for the CIS strategy the MCL value corresponds to the maximum possible pulse amplitude, this is not the case for the EECIS strategy: due to the addition of the peak signal, the envelope can temporarily exceed the MCL value. The processing schemes are designed such that the full dynamic range is used if the input speech samples are within the full 16 bit range. Therefore, the loudness of the words was always at a comfortable level for all subjects, and no correction of the overall stimulation level was needed.

The height of the peak signal is determined by the value of amplification factor G2. This gain is estimated for each subject by a subjective procedure as follows. Three voiceless stop consonants (p,t,k) in intervocalic /a/ context of a male speaker, processed by both strategies were presented to the subject. Voiceless stop consonants in /a/ context were used

since for these phonemes a large peak signal was prominent in all channels. If the value of G2 was too low, a difference could not be heard between both strategies for any of the words. If the value was too high, the word, processed by the EECIS strategy sounded distorted: the peak signal sounded as a short, loud, non-speechlike crack. For six subjects a suitable value could be found, varying between 1 and 2 (see Table 3.1). They reported that the words with the enhanced envelope sounded more clear, distinct and intelligible, as if they were better articulated. The other subjects, HS and JH, represent two extreme cases. Subject HS could not hear any difference, even when applying the technically highest possible currents in the peak signals. For subject JH, the words with the enhanced envelope always sounded distorted. The height of the peak signal had to be made so small that he could not hear a difference anymore. Both subjects were excluded from further testing. Possible reasons for this effect are not obvious and are discussed in section 3.5.

3.3.3 Stimuli and procedure

Two speech sets were used: stop consonants in vowel context and monosyllabic words of the form consonant-vowel-consonant (CVC). All the words within each set were balanced in root-mean-square level. The first set included nonsense words of the form VCV where V={/a/,/i/,/u/} and C={/p/,/t/,/k/,/b/,/d/,/g/}. Stop consonants were used, since the effect of the algorithm incorporating the rapid adaptation is largest for these phonemes compared to others. This is also the case in normal hearing, where large adaptation peaks are seen at, or shortly after stop consonants (Delgutte, 1997). All the words were pronounced by 2 male and 2 female speakers. Different contexts and different speakers were used to test the robustness of the new algorithm. Presentation was in blocks of 51 words: 6 stop consonants of 4 speakers were presented twice in each block, plus three randomly chosen words at the beginning of the block, which were not included in the analysis. Within each block, the vowel context and the processing strategy were the same. Twelve blocks were run for each subject in a so-called ABAB design, i.e. each strategy was used in every other block. At the start of each test session, the subjects could listen to the six stop consonants in intervocalic /a/ context of one male speaker, processed by both strategies. This way, they could get familiar with the words in each strategy. Since after testing no significant differences in identification scores were observed between these 'training' words and the others used in the test (other speakers and other contexts), all responses were included in the analysis. Response alternatives for the identification task were displayed on a graphics

tablet and the listeners were required to indicate the alternative with the tablet pen to register their response. Feedback was not given. Analysis consisted of calculation of percentage correct responses and construction of confusion matrices. Such matrices permit the calculation of relative feature information transmission (Miller and Nicely, 1955; see chapter 1). Two features were used: VOICING, with categories *unvoiced* {p,t,k} and *voiced* {b,d,g}, and PLACE OF ARTICULATION, with categories *front* {p,b}, *middle* {t,d} and *back* {k,g}.

In order to evaluate both strategies in a wider context, the subjects also completed an identification task with monosyllabic CVC words. These were part of the Flemish recording of the NVA list (Wouters *et al.*, 1994), which consists of equivalent sublists of 12 Dutch / Flemish words. These words contain little redundant information. Perceptual experiments revealed that the number of statistically independent phonemes varied between 2.4 and 2.7 for normal hearing and hearing impaired subjects (Bosman *et al.*, 1992). The words were recorded from one male speaker and six sublists were presented for each strategy. The response for the first word in each sublist was not included in the analysis. Percent correct phoneme scores were calculated for initial consonant, medial vowel and final consonant separately. For each of these groups, confusion matrices were constructed. The matrices of the 6 subjects were added together and feature analysis was performed (Miller and Nicely, 1955). For the consonants, the features were MANNER (6 categories), PLACE (3 categories) and VOICING (2 categories). For the vowels, the features were DURATION, first formant frequency (F1) and second formant frequency (F2), all having 3 categories.

The APEX set-up was used for the generation and presentation of the stimuli, as described in chapter 2. All the speech tokens used for the perceptual evaluation are processed off-line and presented electrically to the subject. The envelope value and the peak value before current mapping, and the channel number of each electrical pulse are stored in files on hard disc or CD-ROM. During the test, these files are read by the APEX program, and the values are sent to the DSP board connected to the computer. The DSP program requires subject dependent values for the following calculations: the amplification G2 of the peak signal, which will be added on the envelope signal, and the mapping of the envelope values to current values within the dynamic range. The DSP program generates the time code to drive the internal part of the cochlear implant device. The DSP board is connected to the external coil of the implant system of the subject.

3.3.4 Results

3.3.4.1. Stop consonants

Table 3.2 gives an overview of the results of the stop consonant identification task, for each vowel context separately. The results averaged across all subjects are also shown in figure 3.9. The biggest effect of the new algorithm was observed for stop consonants in intervocalic /a/ context, for which there was an absolute increase in identification score of 9 %. The effect was smaller for the /i/ context (+ 4 %) and there was a small negative effect for the /u/ context (- 1 %). For each context, the results for the two algorithms were compared in a paired samples t-test (using SPSS 7.5 for Windows): the difference in identification scores was significant for the /a/ context ($t[5]=3.15$, $p=0.026$), but not for the /i/ context ($t[5]=1.37$, $p=0.23$) and the /u/ context ($t[5]=-0.67$, $p=0.53$).

Table 3.2: Percentage correct responses of each subject for the stop consonant identification task. The results are given for each vowel context separately. The last row shows the results, averaged across all subjects.

	/aCa/		/iCi/		/uCu/	
	CIS	EECIS	CIS	EECIS	CIS	EECIS
AP	26	46	26	43	32	40
DD	36	51	34	30	43	39
GG	39	46	39	35	47	45
SV	46	46	64	67	40	32
WM	57	63	49	58	36	40
AE	73	80	67	73	56	47
Average	46	55	47	51	42	41

The analysis of relative feature information transmission revealed that the two features PLACE OF ARTICULATION and VOICING contribute to the better performance for the EECIS algorithm in the /a/ context. The relative transmission of the feature PLACE OF ARTICULATION, averaged across the data of all subjects, varied from 21 % for the CIS strategy to 29 % for the EECIS strategy. The relative transmission of the feature VOICING increased from 58 % to 61 %. Since this measure is more sensitive for small numbers, it can be concluded that the better performance is mainly due to the better transmission of the feature PLACE OF ARTICULATION. This is very encouraging, since previous studies have shown that this feature is very

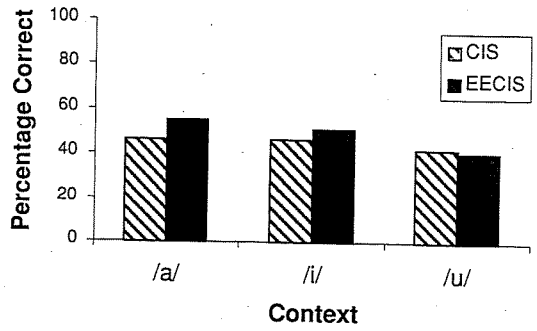


Fig. 3.9 Percentage correct stop consonant identification for the vowel contexts /a/, /i/ and /u/, and for the algorithms CIS and EECIS.

difficult to perceive for cochlear implantees (Tyler and Moore, 1992; Dorman, Soli, Dankowski *et al.*, 1990).

3.3.4.2. CVC words

A summary table of the identification scores for the CVC words test is presented as Table 3.3. The scores, averaged across all subjects and the percentages transmitted information of all features are shown in figure 3.10. Only small differences between the results of both processing schemes are observed for the consonants. The 3 % increase in percentage correct responses for the initial consonants is not significant (t[5]=1.12, p=0.31), neither is the 2 % decrease for the final consonants (t[5]=-0.62, p=0.56). The new strategy does not seem to have a notable effect on the information transmission of any feature. On the other hand, identification of the medial vowel is significantly improved (t[5]=3.45, p=0.018): average scores increased from 39 % correct for the CIS strategy to 46 % correct for the EECIS strategy. This was mainly due to the better transmission of the formant frequencies: from 10 % to 16 % for the first formant frequency and from 13 % to 21 % for the second formant frequency. This is a surprising result, since only the onsets of the vowels are enhanced in each of the 7 or 8 channels, and only if the preceding consonant has less energy than the vowel in that channel.

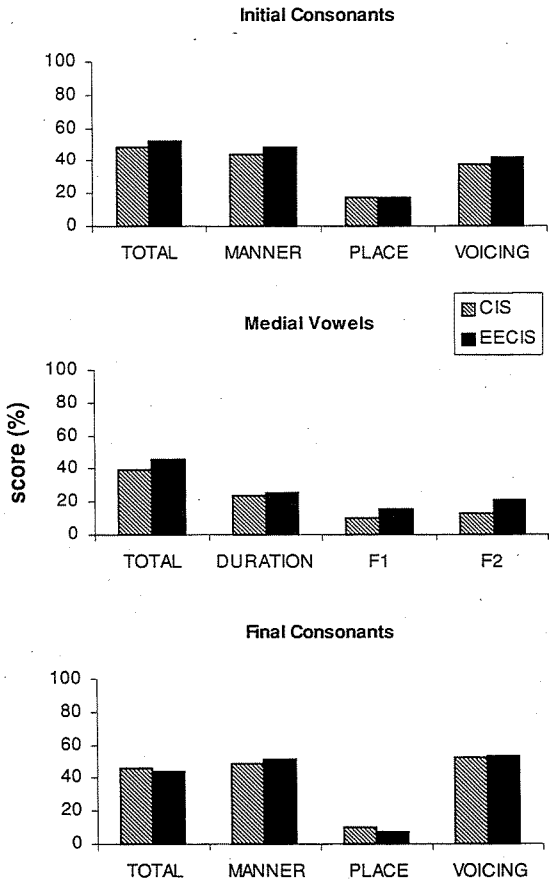


Fig. 3.10 Percentage correct phoneme identification (TOTAL) and relative transmitted information of the features for the initial consonants (upper plot), medial vowels (middle plot) and final consonants (bottom plot) of CVC words, for the algorithms CIS and EECIS.

Table 3.3: Percentage correct responses of each subject for the CVC identification task. The results are shown for the initial consonant, medial vowel and final consonant separately. The last row shows the results, averaged across all subjects.

	Initial Consonant		Medial Vowel		Final Consonant	
	CIS	EECIS	CIS	EECIS	CIS	EECIS
AP	50	50	48	61	44	47
DD	39	52	24	27	35	29
GG	36	42	15	17	33	33
SV	47	44	21	35	38	35
WM	61	58	50	58	56	61
AE	61	65	73	77	68	62
Average	49	52	39	46	46	44

3.4 Signal analysis

The peak signals were further investigated to find possible underlying reasons for the observed differences in improvement for the different vowel contexts. The peak signals were extracted separately by the EECIS algorithm, so the mean peak values at the onset of the second syllable (stop-vowel) could easily be calculated. The mean peak value is defined as the average value of all non-zero peak samples across all channels. Mean peak values were calculated for the eight-channel and the seven-channel algorithm. The correlation between the obtained measures for both cases was very high ($r=0.997$, $p<0.001$), so only the data for the eight-channel algorithm were retained for further analysis. One-way analysis of variance on the factor SPEAKER also revealed that there were no significant differences in the mean peak value of each syllable across the four speakers ($F[3,51]=0.23$, $p=0.88$). Moreover, all correlations between the mean peak values for all possible pairs of speakers were significant at the 0.01 level, and had values between 0.6 and 0.7. This indicates that the peak signal is not dependent on the speaker, which means that the EECIS algorithm is robust with respect to differences in pronunciation. For further analysis, the mean peak values were averaged across all speakers.

Figure 3.11 illustrates the relationship between the mean peak value (horizontal axis) averaged across all speakers, and the mean absolute difference in identification scores, averaged across all listeners (vertical axis)

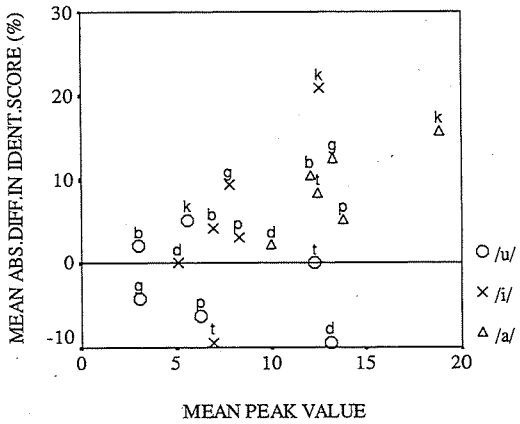


Fig. 3.11 Absolute difference in identification score between both algorithms, averaged across all listeners, in function of the mean peak value for the 6 consonants in the 3 vowel contexts, averaged across all speakers.

for each stop consonant in each vowel context. The ordering on the horizontal axis reveals that the mean peak values are highest for the /a/ context and lowest for the /u/ context, and those for the /i/ context lie in between. One explanation is that peak values are averaged across all channels. In the case of the /a/ context, peaks are prominent in all channels, leading to a large average value. On the other hand, only a limited number of channels contain peak signals when the vowel context was /i/ (lowest and highest channels) or /u/ (only lowest channels). The orderings of the mean peak values for the different stops are consistent with the data of Repp and Lin (1989) for English stops. They also found the highest onset amplitudes for velar stop consonants (/k/,/g/) preceding /a/ or /i/, while alveolar stop consonants (/t/,/d/) had the highest onset amplitudes if followed by /u/.

In figure 3.11, fourteen out of the eighteen tokens have non-negative values on the vertical axis, which means that for these tokens identification scores for the EECIS algorithm are at least the same, but mostly higher, compared to those of the CIS algorithm. There is a moderate, but significant positive correlation between the mean peak value for each token and the difference in identification score between both algorithms for the corresponding token ($r=0.51$, $p<0.05$). Within each context separately,

correlations were 0.78, 0.80 and –0.41 for the /a/, the /i/ and the /u/ context, respectively. The negative correlation for the /u/ context possibly reflects the confusion that arises, since alveolar stop consonants had the largest peak signals, as opposed to the /a/ and /i/ context.

3.5 Discussion

The proposed technique to enhance the speech envelope is based on the rapid adaptation seen in the neural discharge rate at the onset of a sound stimulus. However, it is very difficult to ensure that quantitatively the same effect in the neural response is obtained. One way would be to make single fiber recordings of the response of electrically stimulated auditory nerves in animals to pulse train stimuli with and without peak signals. Another way would be to measure evoked potentials in humans with the same stimuli. On the other hand, it can be assumed that, at least qualitatively, the desired effect is obtained, since increasing the current amplitude increases the discharge probability of the electrically stimulated nerve fibers. A side effect of this current increase is that there is also a corresponding recruitment of more fibers with an increase in the synchronization between them. This is unlike the events occurring in normally adapting fibers. An alternative technique would be to increase the stimulation rate instead of the amplitude of the pulses. However, the current technology of available cochlear implant systems does not allow yet the necessary flexible variations in pulse rate over a large interval.

It is difficult to determine the value of amplification factor G2 (height of the peak), in order to create a similar effect as in normal hearing. Since an objective, quantitative measure was not available, a subjective measure was used, i.e. the response of the subject. The advantage of the proposed technique is that the peak signal is extracted separately from the standard speech envelope, so a peak signal, 'fitted' to the subject, can be added to the standard speech envelope. For each subject, the interval of suitable values for G2 was rather small: the ratio between the smallest value, resulting in an effect and the largest value, resulting in a distorted sound, was not larger than 2. Across all subjects, the ratio between the smallest and the largest value used in the experiments was also 2. The value of G2 was kept the same across all channels, which makes the proposed technique suitable for clinical use. A possible improvement of the strategy is to make this value channel dependent.

The fitting of the peak signal before testing also revealed that not all subjects benefit from the new algorithm. One subject could not hear any difference when a peak signal was added, while for another subject words always sounded distorted in that case. A possible explanation for this is that the electrical-neural interface of these subjects differ from those of the other subjects, caused by e.g. differences in the number and distribution of surviving neurons, physiologic function of these neurons, and distance between stimulating electrodes and these neurons.

An objection might be that the improvements in identification of stop consonants are due to the amplification of the bursts. In other words, the same results could possibly be obtained if the words were simply amplified. There are, however, three reasons to refute this. First, the enhanced envelope extraction technique only amplifies the onsets and thus performs a selective amplification. In the case of the stop consonants, the bursts are amplified, relative to the following vowel. It is shown that increasing the consonant-vowel intensity ratio improves consonant recognition by hearing-impaired listeners (Kennedy *et al.*, 1998). The proposed envelope extraction technique can be seen as a method to obtain this effect. Second, adaptation can also be seen as selective amplification. Plots of neural activity show that the discharge rate for stop consonants is higher than what might be expected from their energy (Delgutte, 1997). Third, it is not feasible to increase the stimulation level for the CIS scheme to match that of the consonants in the EECIS scheme. A peak amplification factor G2 of 2 means that the onset is amplified with a factor 3. If all current pulses would be amplified by this factor, the loudness of the speech stimuli would be unbearable.

The obtained results for stop consonants are context dependent. The average improvement in percentage correct identification was largest for stops in /a/ context, smaller for stops in /i/ context, and slightly negative for stops in /u/ context. This effect is probably due to the fact that the average values of the peaks are largest for stops in /a/ context and smallest for stops in /u/ context, in part caused by the different number of channels containing prominent peak signals. This was also the reason why stops in /a/ context were used in the fitting procedure, since the purpose of the fitting was to determine the maximum possible value of the peak amplification factor G2. A different fitting method for the other contexts could compensate for the context sensitive outcomes, but it is not feasible to incorporate this in a real time speech processor, since the following vowel is not known a priori. The problem can partly be overcome by applying a compression on the peak signals, making the observed differences in mean peak value smaller.

However, compression will obviously have no effect in channels containing no peak signals (/i/ and /u/ context). Another approach is to increase the value of the peak amplification G2, even if this makes the speech signal sound unnatural. Care has to be taken though, since performance can drop if the consonant-vowel intensity ratio is increased too much (Kennedy *et al.*, 1998).

The results for the CVC words are more difficult to interpret. No significant differences were seen between the results for initial consonants, nor for final consonants. There are several possible explanations for this. First, the consonant set consisted of more than only stop consonants: 38% of the initial consonants and 30 % of the final consonants were stops. Second, the medial vowel was obviously not always /a/, while the best results were obtained for identification of stop consonants in /a/ context. Third, the subjects did not focus their attention entirely on the consonants alone when they identified the words. On the other hand, their performance in vowel identification increased, mainly due to the better transmission of the formant frequencies. Presumably, the peak signals enhanced the spectral contrast between the preceding consonant and the vowel, analogous to the adaptation effect (Delgutte, 1997).

In general, there were significant improvements, but they were rather small. Yet, the obtained results were immediate, requiring no training and no feedback. Greater improvements might be achieved if the subjects could use the new processing scheme for an extended time period. Real time implementation in a wearable speech processor is feasible, given the low complexity of the algorithm. At this moment, the EECIS algorithm is implemented in an extended, experimental version of the wearable Lauraflex processor.

3.6 Summary and conclusions

In summary, a new method to code the speech envelope in CIS processors for cochlear implants is proposed, which incorporates the rapid adaptation seen in the response of auditory nerves to sound stimuli. Analysis of the adaptation peaks revealed that the new strategy is robust with respect to differences in pronunciation. Significant improvements in speech intelligibility were observed in some conditions, while there was never a significant decrease in the transmission of any of the speech features. The biggest effect of the new algorithm was observed for stop consonants in intervocalic /a/ context, for which there was an absolute increase in

identification score of 9 %. The better performance was mainly due to the better transmission of the feature PLACE OF ARTICULATION, which is traditionally difficult to perceive for cochlear implantees. Also, the average identification score of the medial vowel of the CVC words was significantly improved by 7 %. This was mainly due to the better transmission of the formant frequencies. All these results demonstrate that some perceptual advantage is gained by including rapid adaptation in the speech processing for cochlear implants.

Chapter 4

Coding of the fundamental frequency in CIS speech processors

Abstract

The pitch of a periodic sound is strongly related to its fundamental frequency (F0), and provides several cues for the perception of speech and music. In section 4.1, it is described how this information is coded in the auditory nerve, and how it can be represented in a cochlear implant system. At the output of a CIS processor, the fundamental frequency appears as temporal fluctuations in the envelopes. To investigate wether an enhancement of these fluctuations results in a better perception of F0, a new speech processing algorithm is designed to control and increase the depth of these fluctuations (F0 CIS). The new algorithm is described in section 4.2, as well as the default CIS strategy and the CIS strategy in which the F0 fluctuations in the envelope are removed (FLAT CIS). Two experiments with four users of the Laura cochlear implant were carried out to examine the influence of the modulation depth of these envelope fluctuations on pitch discrimination. The first experiment is described in section 4.3. Prior to the evaluation of the three algorithms, the subjects were presented with two sinusoidally amplitude modulated pulse trains on a single electrode channel differing in modulation frequency (δf = 20%). They were asked to indicate the stimulus with the highest pitch. The results were dependent on the three parameters that were varied across the conditions: modulation depth, electrode channel and modulation frequency. In general, performance decreased for smaller values of the modulation depth. In the second experiment, described in section 4.4, an adaptive 2-down 1-up procedure was used to find the smallest difference in F0 that is still discriminable for the three algorithms. In contrast with experiment 1, all available channels (7 or 8) were stimulated. Rms-balanced synthetic vowels, /a/ or /i/, served as input for the algorithms. In general, performance was poorest for the FLAT CIS strategy, where changes in F0 are only apparent as changes of the average amplitude in the channel outputs. This indicates that the temporal coding of F0 in the speech envelope provides an additional and important cue for pitch perception. No significantly better results were obtained for the F0 CIS strategy compared to the default CIS strategy. The relation between the results of both experiments is discussed in section 4.5. The default CIS strategy results in envelope modulation depths for which sub-optimal scores were obtained in some cases of the first experiment. However, no improvement is seen for the FLAT CIS strategy. This indicates that less modulation is needed if all channels are stimulated with synchronous F0 fluctuations.

Geurts, L. and Wouters, J. (2000a). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," submitted to J.Acoust.Soc.Am.

4.1 Introduction

4.1.1 Fundamental frequency

All periodic signals can be decomposed into a series of discrete sinusoids, called the harmonics. The frequencies of these harmonics are an integer multiple of a common frequency, the fundamental frequency (F0). For speech, all voiced phonemes are periodic, since they are the result of a periodic vibration of the vocal chords, followed by a filtering in the vocal tract. The shape of the vocal tract settles the magnitude of the harmonics, and is related to the timbre, which is specific for the phoneme. The fundamental frequency corresponds to the vibration speed of the vocal cords, and the pitch of the speech signal is strongly related to it. The value is strongly dependent on the speaker and can vary from around 100 Hz for male voices to above 250 Hz for children voices. The fundamental frequency provides cues to intonation, question-statement contrasts, the identity of the speaker, and his emotional state. Also, it is essential for the recognition of a melody, sung or played on a musical instrument.

Yet, not all the harmonics – even not the first one – have to be present in the sound, in order to evoke a pitch corresponding to F0. As an example, a speech signal transferred through a telephone line still has the same pitch as the original, although all frequency components below 300 Hz are removed, thus including the first harmonic of an adult voice. This low pitch associated with a group of harmonics is often called "residue pitch" or "periodicity pitch". A good overview of pitch perception can be found in Moore's book (1989).

4.1.2 Temporal coding of F0 in the auditory nerve

The phenomenon of the residue pitch demonstrates that the fundamental frequency is not solely coded by the auditory nerves at the tonotopic place corresponding to F0, but also in the other regions of the cochlea. A complete overview article about periodicity coding in the auditory system is presented by Langner (1992). A distinction has to be made between harmonics that are resolved by the auditory system and the unresolved harmonics. In the former case the neural representation of the harmonic is insignificantly influenced by neighboring harmonics and its corresponding pitch can be heard out. In the latter case, the response of the auditory nerves results from the interference of several harmonics, and shows a periodicity

corresponding to F0. Whether a harmonic is resolved or not, depends on the bandwidth of the frequency channel in the cochlea containing the harmonic (see figure 1.4), and the frequency interval between the harmonics. The probability that the harmonic is unresolved increases for higher harmonic frequencies and lower frequency intervals.

Resolved harmonics

To illustrate the neural temporal coding of a pure tone (or a resolved harmonic), an interspike interval histogram is often used (Rose *et al.*, 1967). It is constructed by measuring the time intervals between successive spikes of the auditory nerve in response to a pure tone with constant intensity and frequency. The range between zero and the maximum interval is divided in many shorter intervals, called bins. For each bin, the number of times that the value of an interspike time interval was within the range of the bin is counted. A few histograms for one auditory nerve in response to several pure tones with different frequencies are shown in figure 4.1. The characteristic frequency of the nerve is 1.6 kHz, and the frequency of the stimulus and the mean response rate are indicated above each histogram. Notice that the nerve also responds to tones with frequencies that are very different from the characteristic frequency, as also illustrated in figure 1.5 in chapter 1.

The nerve discharges at most once during each cycle of the pure tone. The plots show that the most likely intervals are an integer multiple of the stimulus period. This also means that the time instant of the nerve discharge is related to the phase within the stimulus period. The histograms also indicate that the frequency of a pure tone is not only coded by the distribution of neural activity along the basilar membrane, according to the tonotopic organization, but also by the timing between successive neural discharges.

Unresolved harmonics

If harmonics fall into the same frequency channel of the cochlea, they produce an amplitude modulation with a frequency corresponding to their difference frequency. In that way the fundamental can be encoded in the neural response, even if it is not physically present in the stimulus. One way to investigate this response, is to present sinusoidally amplitude modulated (AM) tones and to construct period histograms for the measured spike intervals (Joris and Yin, 1992). The difference with interval histograms is

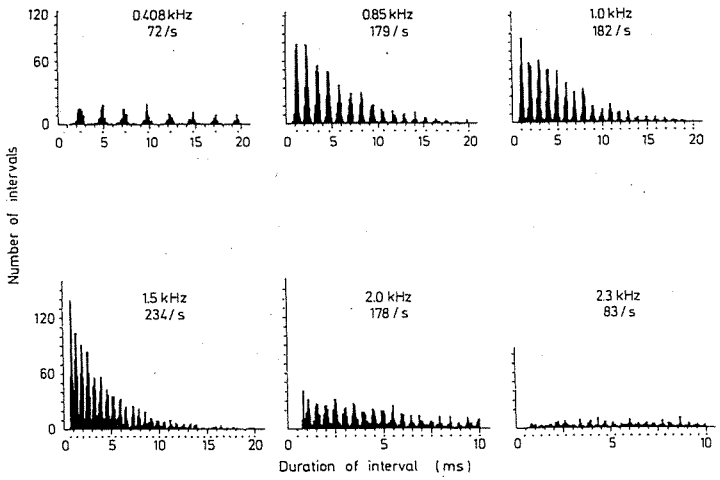


Fig. 4.1 Interspike interval histograms for a single auditory nerve with a characteristic frequency of 1.6 kHz. The frequency of the pure tone and the mean response rate are indicated above each histogram. From Moore (1989, Fig. 1.17)

that not the interval between successive spikes is measured, but the phase within each cycle of the modulating frequency at which the spike occurs. Some examples are shown in figure 4.2 for an AM tone with a modulation frequency of 100 Hz and a carrier frequency equal to the characteristic frequency (20.2 kHz, measurements taken from a cat). At the left of each plot, the histogram is shown for 2 cycles, and at the right the half-wave rectified AM stimulus for comparison. The modulation depth of the stimulus is defined as the difference between the largest and the smallest amplitude of the signal envelope divided by the sum of those amplitudes, and is indicated at the top of each plot.

It is clear that the temporal envelope of the input is still preserved in the response, moreover, the response appears to be more modulated than the stimulus. If the gain of the modulation depth is plotted as a function of the modulating frequency, one obtains a low-pass filter characteristic of which the cut-off frequency depends on the characteristic frequency of the nerve fiber. This effect is not due to a limitation in the temporal response of the

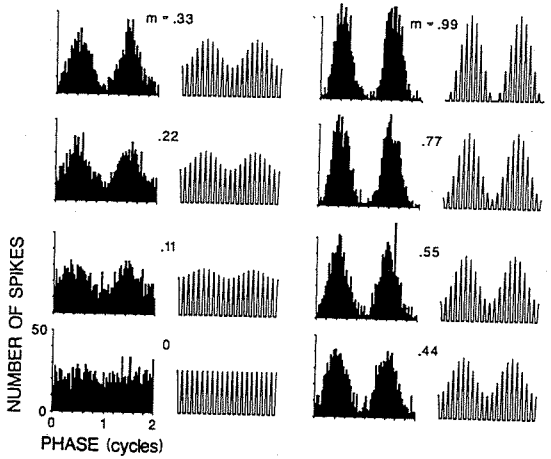


Fig. 4.2 Comparison of period histograms and half-wave rectified AM stimuli for several values of the modulation depth, indicated above each plot. From Joris and Yin (1992, Fig. 1)

nerve fiber, but to the fact that the side-band components of the AM stimulus fall outside the bandwidth of the cochlear filter (see figure 1.4). However, the cut-off frequencies for most fibers lie well above the range of typical fundamental frequencies of speech. In other words, most auditory nerve fibers allow the temporal representation of the fundamental frequency of a voiced speech signal. The important role of the temporal coding of the fundamental frequency was illustrated in a study of Cariani and Delgutte (1996). They predicted the pitch of several complex tones on the basis of interspike intervals between successive and nonsuccessive spikes, so called all-order intervals. The most frequent all-order interval present in the distribution pooled over many fibers corresponds to the pitch heard in psychophysical experiments.

Harmonic complexes

The response of a specific auditory nerve fiber to a vowel, or any harmonic complex, depends on its characteristic frequency and on the amplitudes of the harmonics in the vicinity of that frequency. In the case the nearest harmonic is resolved by the cochlear filter or that this harmonic is very dominant compared to the surrounding ones (e.g. the harmonic closest

to a formant frequency), the time representation of the neural response shows little or no modulation according to F0, but to the harmonic. On the other hand, nerve fibers with a characteristic frequency between two formants respond to several harmonics, and due to beating, they show a clear envelope modulation according to F0 (Miller and Sachs, 1984; Palmer, 1990).

4.1.3 F0 in speech processors for cochlear implants

Since only a limited number of electrode channels are available in a cochlear implant system, a relatively large frequency region is assigned to it. This results in a poor frequency selectivity, or in other words, harmonics are likely to be unresolved. In that case, the outputs of the band-pass filters will show a modulation according to F0. In most common speech processors, the envelope in each frequency band is typically estimated by low-pass filtering the rectified output of the band-pass filters (see chapter 1). If the cut-off frequency of the low-pass filter is chosen above the range of typical F0s of speech, then F0 appears as a temporal fluctuation in the speech envelope (for an illustration, see figure 4.5). In a CIS processor, brief pulses are presented to each electrode channel at a fixed high rate of about 1000 pulses per second. These pulse trains are modulated with the extracted speech envelopes. In the feature extraction algorithms that are or were used in the Nucleus system, the fundamental frequency itself is also estimated from the input signal and settles the repetition rate of the pulses applied on the electrodes (Blamey *et al.*, 1987; Skinner *et al.*, 1991).

The sensitivity of a cochlear implant subject to changes in pulse presentation rate or modulation frequency is described in several psychophysical studies. An unmodulated pulse train with a higher repetition rate will elicit a higher pitch. (Eddington *et al.*, 1978; Shannon, 1983a; Tong and Clark, 1985; Townshend *et al.*, 1987). This is often called "rate pitch" as opposed to the "place pitch" that corresponds to the tonotopic organization of the cochlea. The highest frequency limit for which a subject can still detect a change in pitch when the frequency changes, is strongly dependent on the subject, even if the same type of device and stimuli are used. This limit generally varies around 350 Hz (Eddington *et al.*, 1978; Simmons *et al.*, 1981; Blamey *et al.*, 1984), which implies that temporal coding of F0 is feasible. At the moment, there are reports on three subjects showing even a limit of around 1 kHz (two subjects in Townshend *et al.*, 1987; one (SR2) in Wilson *et al.*, 1994). McKay *et al.* (1994) investigated the pitch associated with sinusoidally amplitude modulated pulse trains and

reported that these stimuli could be ranked by modulation frequency in the same way that unmodulated stimuli could be ranked by pulse rate, provided that the carrier rate is high enough (at least 800 Hz for a modulation frequency of 200 Hz). Wilson (1997) investigated the neural response by measuring action potentials in the cochlea, evoked by sinusoidally amplitude modulated pulse trains. He concluded that the ratio between carrier rate and modulation rate should be at least four in order to obtain a good representation of the stimulus envelope. This ratio, obtained from physiological measurements, is in agreement with several psychophysical data (McKay *et al.*, 1994; Wilson, 1997). The Laura device can stimulate at an overall rate of 10000 pulses per second (pps), so implementing the CIS strategy on all eight channels results in a pulse rate of 1250 pps on each channel. This implies that modulation frequencies of up to about 300 Hz can be encoded correctly, which is well above the range of fundamental frequencies of normal speech. In fact, a change of the fundamental frequency has two effects at the output of the CIS strategy. First, the average amplitude in each channel will change, mainly due to a different attenuation of the harmonics in the sidebands of the band-pass filters (for an illustration, see figure 4.7). This cue will be referred to as the "average amplitude cue". Second, the period of the fluctuations in the envelope, which corresponds to the period of the fundamental frequency, will change too. This temporal cue will be referred to as the "envelope modulation cue".

A new speech processing algorithm is developed which enhances the "envelope modulation cue" by increasing the modulation depth of the envelope fluctuations. Prior to the evaluation of this algorithm, it is first investigated how the modulation depth of the envelope fluctuation influences the discrimination of two sinusoidally amplitude modulated (SAM) pulse trains on a single channel with different modulation frequencies (experiment 1). McKay *et al.* (1995) showed that if the modulation depth of a modulated pulse train is too small, a pitch corresponding to the carrier frequency is perceived, instead of the modulating frequency[6]. Hence, it is expected that performance degrades for smaller values of the modulation depth, since the carrier frequency is the same for both stimuli. In the second experiment of this study, the *just noticeable F0 difference* is measured for three CIS strategies, which stimulate all available channels, using synthetic vowels. In the first strategy (FLAT CIS), the temporal F0 fluctuations are removed, so

[6] In contrast with this study, McKay *et al.* (1995) did not use sinusoidally amplitude modulated pulse trains. Only two pulse levels were used, with the highest one occurring only once in each modulation cycle.

there is only an "average amplitude cue". The second strategy (CIS) is the default strategy (see chapter 1), in which also "envelope modulation cues" are present. In the new, third strategy (F0 CIS), the envelope fluctuations are enhanced compared to the default strategy, by increasing the modulation depth of these fluctuations. A comparison of the first two algorithms permits to determine to which extent the "envelope modulation cue" improves pitch perception corresponding to F0. Jones *et al.* (1995) made a related study based on the SMSP strategy, developed for the Nucleus device (McDermott *et al.*, 1992). They concluded that the inclusion of a temporal cue according to F0 did not increase performance in several tasks involving pitch perception. However, they did not accurately measure the pitch discrimination abilities of their subjects for the different strategies. By comparing the results of the second and the third CIS algorithm, it is possible to determine whether the enhancement of the "envelope modulation cue" results in a better frequency discrimination.

4.2 Speech processing

The three speech processing algorithms in this study consist of the following steps (see also description of CIS strategy in chapter 1): band-pass filtering, envelope extraction, logarithmic compression, current mapping and pulse train modulation. The processing schemes for one frequency band are shown in figure 4.3. The scheme at the top corresponds both to the FLAT CIS and the CIS algorithm. The only difference between both is the value of the cut-off frequency of the low-pass filter. The scheme at the bottom corresponds to the F0 CIS algorithm. In experiment 2, the input of the speech processing algorithms are synthetic vowels, downsampled from 20 kHz to 10 kHz and with a resolution of 16 bits (see section 4.4.1).

FLAT CIS and CIS

The signal is first filtered by several band-pass filters (Butterworth 4th order) covering the speech frequency range (see figure 1.8 in chapter 1). The number of band-pass filters corresponds to the number of active electrode channels in the cochlea of the implanted subject (see table 4.1). The cut-off frequencies are linearly spread below 1 kHz and logarithmically above 1 kHz (for the values, see section 3.2.1. in chapter 3). Next, the envelope is calculated by full-wave rectification and a smoothing filter. This filter is a Butterworth 4th order low-pass filter with a cut-off frequency of 50 Hz for the FLAT CIS algorithm and 400 Hz for the CIS algorithm. The former is well below the range of F0 values of adult speech, the latter is

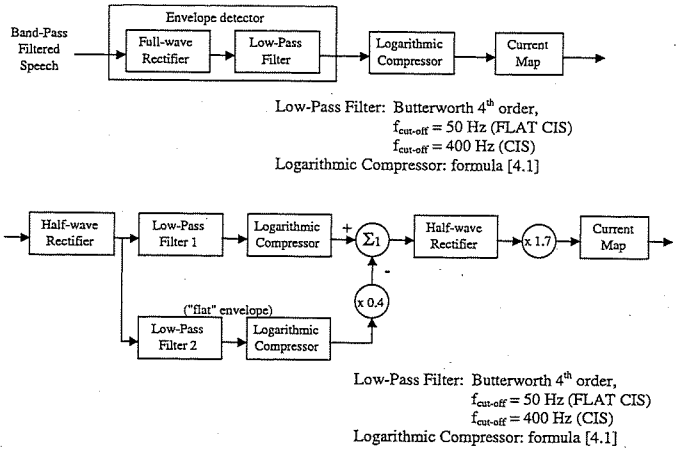


Fig. 4.3 Processing schemes for the three algorithms: FLAT CIS and CIS at the top, and F0 CIS at the bottom.

clearly above that range. As a result, the F0 fluctuations are clearly present in the envelope of the CIS algorithm, and are absent in the envelope of the FLAT CIS algorithm. This is illustrated in figure 4.4 where both envelopes for channel 5 are shown on the same plot. The input was the synthetic vowel /a/ with F0 equal to 250 Hz. Note that the low cut-off frequency of 50 Hz used in the FLAT CIS algorithm is unlikely to affect speech intelligibility. Two independent studies concluded that speech reception scores of normal hearing listeners are nearly unaffected after low-pass filtering the speech envelope at 50 Hz (Drullman *et al.*, 1994; Shannon *et al.*, 1995; see also chapter 3, section 3.1.3).

In the next step, the smoothed envelope is instantaneously compressed using the following formula, in which A and B are chosen such that the function does not show any discontinuity at the transition points (x=128 and x=16384).

$$\begin{aligned} y &= 0 && ,x \le 128 \\ &= A.\log(x) + B && ,128 < x < 16384 \\ &= 1000 && ,x \ge 16384 \end{aligned} \qquad [4.1]$$

F0 CIS

The scheme of the F0 CIS algorithm is somewhat more complicated. There are three major differences compared with the other two algorithms: the band-pass filtering, the rectification and the calculation of the envelope.

- Instead of using one 4th order band-pass filter for each channel, a 2nd order filter is used through which the signal passes twice: once in the normal direction and once in the reversed direction[7]. This way, the phase distortion is zero, i.e. all frequency components are delayed with the same amount of samples. If the normal procedure is followed, the maxima of the F0 fluctuations in adjacent channels do not coincide, as will be illustrated further. In the lower two channels, this time difference can equal several milliseconds. This way, the periodicity in the response of neurons that are sensitive to stimulation on two or more channels, might be blurred. The cut-off frequencies on the other hand are identical in all algorithms.

- The filtered signal is half-wave rectified instead of full-wave. The difference is apparent mainly in the lower frequency channels, where individual harmonics can be resolved for high values of F0. Half-wave rectification reflects better the phase-locked behavior of auditory nerve fibers, and results in a larger modulation depth in the output signal.

- For the calculation of the envelope, two 4th order Butterworth low-pass filters are first employed with cut-off frequencies of 400 Hz and 50 Hz. In the output of the former filter, the F0 fluctuations are present in the envelope, while these are absent in the output of the latter (see also figure 4.4). By subtracting an attenuated version of the "flat" envelope from the fluctuating envelope, the modulation depth is increased (summation in figure 4.3). The attenuation factor is chosen such that the "flat" envelope is always

[7] This implies that a real-time implementation of F0 CIS is not possible. As an alternative, band-pass filters with a finite impulse response could be used. These have the property that phase distortion is zero. This alternative approach was not followed here in order to restrict the differences with the two other algorithms.

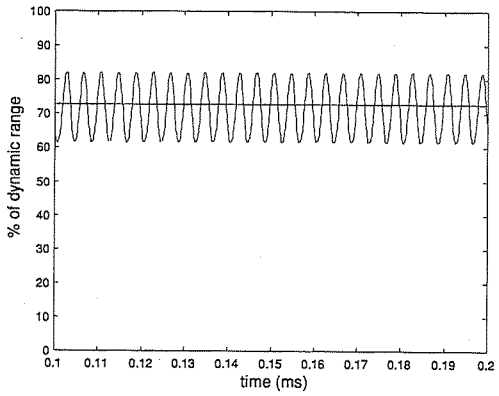


Fig. 4.4 Envelopes of synthetic vowel /a/ (F0 = 250 Hz) in channel 5 (frequency band between 1125 Hz and 1634 Hz) as a function of time. The solid line is the envelope of the low-pass filter of the CIS algorithm (cut-off frequency at 400 Hz). The dashed line corresponds to the envelope of the low-pass filter of the FLAT CIS algorithm (cut-off frequency at 50 Hz). In the latter case, the F0 fluctuations in the envelope are no longer present.

below the other in the high frequency channels[8]. From inspection of the envelopes derived from the stimuli used in experiment 2, the rather conservative value 0.4 was chosen. Negative values after summation have no meaning and are rejected (half-wave rectifier). To compensate for the resulting smaller amplitude, the signal in each channel is amplified with a factor 1.7.

In the three algorithms, the envelope samples are mapped to current values (last block of each scheme in figure 4.3), taking the dynamic range of the channel of the subject under test into account. A linear function is used, which maps an envelope sample value of 1 to the current threshold I_{THR}, and a value of 1000 to the most comfortable current level I_{MCL} (see also figure 2.4). The current envelopes in all channels are modulated with pulse trains

[8] In the low-frequency channels, it is not possible to achieve this. The modulation depth of the "fluctuating" envelope is already 100 %, since half-wave rectification is used.

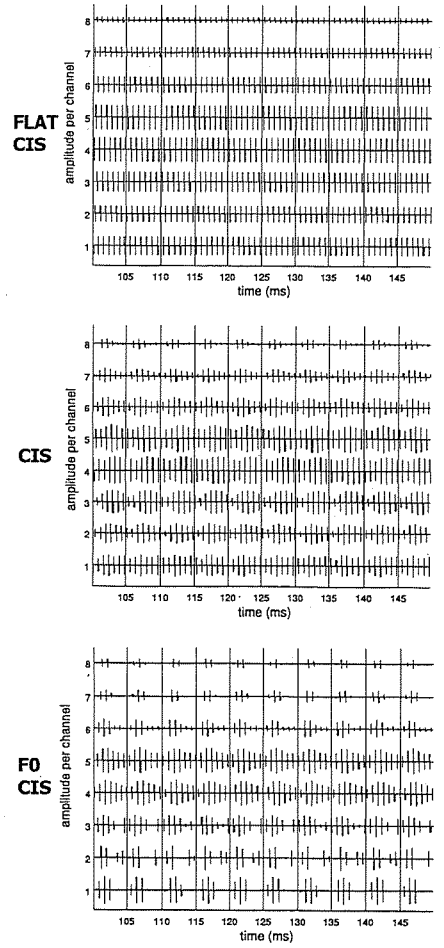


Fig. 4.5 The pulsatile output of the three CIS algorithms for the synthetic vowel /a/ (F0 = 200 Hz) as a function of time: FLAT CIS, default CIS, and F0 CIS. The channel numbers are indicated before each envelope trace.

with an overall rate of 10000 pps. The width of each pulse is constant and equal to 40 µs/phase. The channel order of stimulation is staggered: 1, 4, 7, 2, 5, 8, 3 and 6. Examples of the output of each algorithm for the synthetic vowel /a/ (F0 = 200 Hz) are shown in figure 4.5. The plot at the top is the output of the FLAT CIS strategy. Clearly, no F0 fluctuations are present in the envelopes. The plot in the middle is the output of the default CIS strategy, now showing a clear periodicity in the envelope corresponding to F0. Notice the frequency doubling of the F0 fluctuations in channel 1, due to the full-wave rectification (compare with channel 2). There also seems to be some asynchrony between channels 4,5 and 6, due to the phase distortion: the local maxima of the F0 fluctuations do not coincide in time. The plot at the bottom is the output of the F0 CIS algorithm. Clearly, the modulation depth of the F0 fluctuations is increased in all channels. Also, the synchronization between the channels is increased, and the periodicity of fluctuations in the output of channel 1 corresponds to F0.

4.3 Experiment 1

4.3.1 Subjects

Four post-lingually deafened users of the Laura cochlear implant participated in this study. All tests of experiment 1 and experiment 2, were taken during two sessions of about 2.5 hours. Some details about the subjects can be found in table 4.1. It was essential for this study that they were sensitive to changes of 20 % of the modulation frequency of a SAM pulse train, at least in the lower frequency range around 150 Hz and on one electrode pair. This sensitivity was investigated in a pilot experiment, which resulted in the rejection of four other subjects for further testing. No relation between this sensitivity with any objective parameter, such as those given in table 4.1, could be found. There even was one subject, who was an experienced musician before he became deaf, who could not consistently identify the stimulus with the highest pitch at the lowest modulation frequencies and the highest modulation depth.

Two versions of the electrode array of the Laura implant exist, and both consist of eight bipolar electrode pairs, called channels (see figure 1.11). These are numbered from apex to base. Subjects SV and JH use type 5.6: the distance between the two electrodes of one channel is 1.3 mm, and the distance between the midpoints of adjacent channels is 2.05 mm. For subjects AE and WM, who use type 5.8, these values are respectively 2 mm

and 3 mm, which means that their channels physically overlap. Two channels were selected for the first experiment, a channel close to the apical end of the electrode array and a channel close to the basal end. There was no selection criterion needed for this experiment, so selection was rather random. Subjects SV, AE and JH participated also in a previous study, for which also two well separated channels were selected (Carlyon *et al.*, 1999; chapter 5). These channels were the most apical and the most basal channel for which a 100 pps pulse train with biphasic pulses of 40 µs/phase wide and pulse amplitude below the technical limit of 1.5 mA, could be perceived as comfortably loud. For subject WM, the same channels as for JH were chosen.

Table 4.1: Some details of the four subjects who participated in this experiment. Columns 6 and 7 indicate the channels that were selected for experiment 1. The last column indicates the number of active channels used in experiment 2. The same channels are daily used in the subjects' clinical processors.

Subject	Age (yrs)	Duration of profound deafness (yrs)	Aetiology	Implant experience (yrs)	Apical channel	Basal channel	Number of active electrode channels
SV	30	6	Unknown	5	2	6	8
WM	49	5	Progressive	2	2	7	8
AE	42	3	Menière	2	4	8	7
JH	17	3	Meningitis	4	2	7	8

4.3.2 Stimuli and Procedure

For the generation and the presentation of the stimuli, the definition of the experimental procedure and the collection of the responses, the APEX set-up was used. All stimuli were SAM pulse trains. First, an unmodulated pulse train with an amplitude of "100.0" (see chapter 2), a pulse rate of 1250 pps and a pulse width of 40 µs per phase was generated. Next, the pulses are modulated with the following discrete function, which is evaluated each 100 µs:

$$f(t) = 1 - \frac{m'}{2}.(1 + \sin(2.\pi.f_m.t)) \qquad [4.2]$$

f_m is the modulation frequency and the relative modulation depth m' can vary between 0 and 1. The output of the modulating function varies between 1-m'

and 1, so the amplitude of largest pulse of the SAM pulse is not greater than "100.0". During the experiment, the pulses are mapped within the dynamic range of the subject's channel (see figure 2.4).

In each trial, two stimuli with different modulation frequencies were presented and the subjects had to indicate the stimulus with the highest pitch. An answer was correct if the stimulus with the highest modulation frequency was selected. Correct response feedback was always given on the computer screen. The length of each stimulus was 400 ms, and the silent interval between the stimuli was 500 ms. The two frequencies in each pair were either 150 Hz and 180 Hz, or 250 Hz and 300 Hz. The order of presentation in each trial was randomized. The level of the largest pulse of each SAM pulse train always corresponded to the upper limit of the dynamic range. Several values of the modulation depth were used, but were kept constant within each pair. The exact values were chosen from the set {5%, 10%, 20%, 40%, 99%} and were determined in a pilot experiment. The highest value was always 99 % and for the lowest value, the subject should score around chance level (p = 50 %). In this experiment, the number of presentations was 80 in all conditions (2 channels, 2 frequency pairs, and 1 to 4 modulation depths). Each next trial consisted of a stimulus pair from any condition, randomly chosen.

The relative modulation depth m' can be calculated using equation 4.3. I_{max} and I_{min} are the minimum and the maximum pulse amplitude in the SAM pulse train, respectively. The upper and the lower limits of the dynamic range are denoted by I_{MCL} and I_{THR}, respectively:

$$m' = \frac{I_{\max} - I_{\min}}{I_{MCL} - I_{THR}} \qquad [4.3]$$

The threshold, or the minimum level of the dynamic range, was the current level for which a 500 ms unmodulated pulse train of 1250 pps was just audible. The maximum level of the dynamic range equaled the amplitude of the largest pulse of a 500 ms long SAM pulse train (1250 pps carrier rate, 100 Hz modulation frequency) at comfortable loudness level. The level of the smallest pulse of this SAM pulse train was always kept at threshold. Table 4.2 represents the limits of the dynamic range for each subject and channel. Also, the absolute modulation depths (m) are given, following the definition in equation 4.4.

$$m = \frac{I_{\max} - I_{\min}}{I_{\max}} \qquad [4.4]$$

Table 4.2: Lower (I_{THR}) and upper (I_{MCL}) limits of the dynamic range for each subject and channel (3rd and 4th column). The last 5 columns indicate the absolute modulation depth that corresponds to the relative modulation depth indicated at the top of each column. If the condition was not tested, no figure is given (-).

Subject	Channel	I_{THR} (μA)	I_{MCL} (μA)	absolute modulation depths m at...				
				m'=5%	m'=10%	m'=20%	m'=40%	m'=99%
SV	2	80	500	-	8.4	16.8	33.6	83.2
	6	260	750	-	6.5	13.1	26.1	64.7
WM	2	150	570	-	7.4	14.7	29.5	72.9
	7	140	570	-	-	15.1	30.2	74.7
AE	4	350	1100	3.4	6.8	13.6	-	67.5
	8	450	1150	3.0	6.1	12.2	-	60.3
JH	2	100	470	-	7.9	15.7	31.5	77.9
	7	110	620	-	8.2	16.5	32.9	81.4

Stimuli were not balanced in loudness for the following reasons. For both stimuli in a trial, length, carrier rate, minimum and maximum current level - and thus modulation depth – were identical, so also mean current level, root mean square current level and total injected electric charge is the same. Therefore, it is not likely that both stimuli have a different loudness. This was also confirmed by all the subjects. In the case that loudness would be slightly different, it would be totally due to the difference in modulation frequency. The subjects are allowed to use any cue that emerges from a change of this parameter.

4.3.3 Results

Each plot in figure 4.6 shows the percentage correct answers for one subject. The conditions are indicated at the bottom of the figure. In general, it is clear that if the modulation depth is too small, the subjects performed at chance level. The minimum value of the modulation depth needed for a reliable discrimination (e.g. a score greater than 75 %) depends on the condition and the subject. This value varies from 10 % (subject AE, 150 Hz on both channels; 250 Hz on basal channel) to more than 40 % (subject SV, 250 Hz on apical channel). Notice that for each subject, there are one or more conditions for which the criterion of 75 % correct responses is not met for any modulation depth (e.g. subjects JH and WM at 250 Hz). Also, there

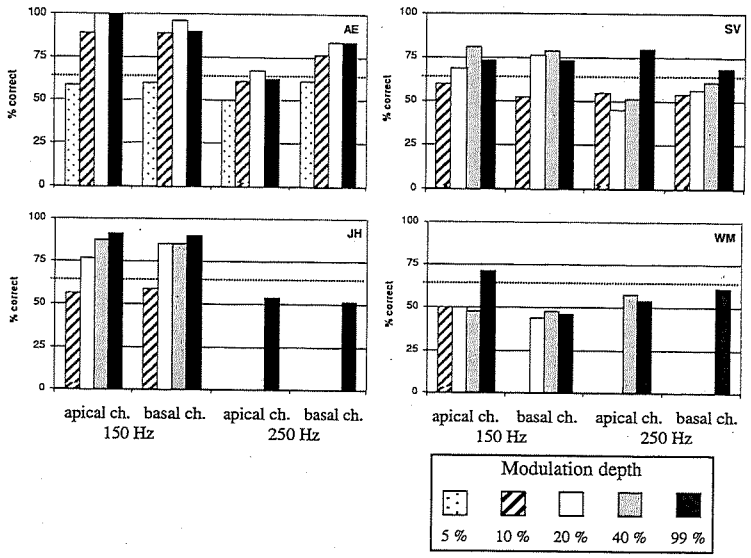


Fig. 4.6 Percentage correct scores for experiment 1. Each plot shows the results of another subject. The channel and lowest modulation frequency of the stimulus pair is indicated at the bottom. Scores above the dashed line are significantly higher than the 50 % chance level ($p<0.01$).

seems to be a saturation effect in some conditions: enlarging the modulation depth beyond a certain value does not notably increase the score (e.g. subject AE, 250 Hz; subject SV, 150 Hz). Subject AE was the best performer: her scores were the highest on average, and she could hear pitch differences in both frequency regions (150 Hz and 250 Hz). Subject SV's results were similar, but his scores were somewhat lower. Subject JH was clearly sensitive to modulation frequency changes in the low frequency region (150 Hz), but not in the high frequency region (250 Hz). His scores were still at chance level at the largest modulation depths. Subject WM was the poorest performer. He never reached the 75 % criterion, and only scored significantly above chance ($p<0.01$) in one condition: 150 Hz, apical channel, 99 % modulation depth.

4.3.4 Discussion

Two subjects were not sensitive to modulation frequency changes from 250 Hz, while the two others performed still above chance at 250 Hz, but worse compared to the 150 Hz condition: either lower scores were obtained at the same modulation depth (subject AE) or optimal performance was reached at higher modulation depths (subject SV). There are two possible explanations for this effect. First, in order to hear a pitch difference, the relative change in frequency or pulse rate must be larger at higher frequencies (Tong and Clark, 1985; Townshend *et al.*, 1987; McKay *et al.*, 1994) . In the presented experiments, the relative change was always 20%, so this value might be below the detection limit at 250 Hz for subjects JH and WM. A second possible explanation is that the pitch of a modulated pulse train is less salient at higher frequencies for the same modulation depth. This is supported by the fact that temporal modulations are more easily detected at lower frequencies (Shannon, 1992; Busby *et al.*, 1993). So, this might imply that also higher modulation depths are needed at higher frequencies in order to perform equally well as for lower frequencies in a pitch discrimination task.

At first sight, subjects with a larger dynamic range seem to obtain better overall scores. Indeed, the correlation between the absolute dynamic range (I_{MCL}-I_{THR}) for each channel of each subject and the average of the scores at 10%, 20% and 40% modulation depth[9] is 0.723, which is significant ($p=0.043$). This indicates that subjects with a larger dynamic range need less modulation depth to judge on the pitch of a temporally modulated train. Obviously, the variance in the data can not be explained only in terms of the dynamic range, but also the subject's sensitivity to rate or frequency variations has to be taken into account.

McKay *et al.* (1995) concluded that the modulation depth of a modulated pulse train should be greater than 100 µA in order to elicit a pitch that equals the pitch of an unmodulated pulse train with the rate equal to the modulation frequency. This value corresponded to at least a quarter of the dynamic range up to almost the whole dynamic range, dependent on the subject. Although one should always be cautious when comparing inter-

[9] For subject AE, the scores at m'=40% were not measured. The missing values were estimated by using the average scores at 20% and 99%. The error of the estimate will be small since the scores at both modulation depths do not differ substantially.

subject results, their findings can be related to the results of experiment 1. The optimal performance was obtained at modulation depths ranging from 20 % of the dynamic range (subject AE) via 40 % (subject SV, 150 Hz) to 99 % (subject SV, 250 Hz; subjects JH and WM). This suggests that the highest discrimination ability is achieved when the pitch of the modulated pulse train is the lowest, i.e. close to the pitch of an unmodulated pulse train with a pulse rate equal to the modulation frequency.

From physiological measurements of the neural response to modulated electric pulse trains, Litvak *et al.* (1999) concluded that the modulation gain[10] was much higher than the one obtained in acoustical stimulation, and that the synchronization to the modulation already saturated at modulation depths of 10 %. More stochasticity, i.e. less synchronization, in the response would result in a better approximation of natural acoustical hearing. However, the best performances in experiment 1 were obtained for modulation depths much larger than the aforementioned 10 %. In other words, more stochasticity might improve the mimicking of normal hearing, but will not necessarily result in a better performance of the electric listener.

4.4 Experiment 2

In the second experiment, the smallest noticeable difference in F0 was measured for the three algorithms using synthetic vowels. Unlike experiment 1, the stimuli thus consisted of modulated pulse trains on all available active channels.

4.4.1 Stimuli and procedure

All stimuli were generated from synthetic vowels, which were created using the Klatt synthesizer available in the CSRE software package[11]. The length of the stimuli was 500 ms, and the highest sample rate (20 kHz) was chosen. The formant frequencies and bandwidths corresponded either to /a/ or /i/, and were representative for the Flemish language (Govaerts, 1978). The values were kept constant over all fundamental frequencies. The exact values are given in table 4.3. When using the Klatt synthesizer, the value of the fundamental period has to equal an integer multiple of the sampling

[10] The modulation gain is the ratio between the modulation depth in the response and the modulation depth in the stimulus.

[11] Computerised Speech Research Environment, Version 4.5, AVAAZ Innovations, Inc.

period. The fundamental period of the synthesized vowels ranged from 2.7 ms (F0 = 370 Hz) to 6.7 ms (F0 = 149 Hz).

Table 4.3: Formant frequencies (Fi) and bandwidths (Bi) in Hz of the synthetic vowels used in experiment 2.

Vowel	F1	B1	F2	B2	F3	B3	F4	B4	F5	B5
/a/	890	110	1360	120	2475	160	3530	275	4830	325
/i/	225	40	2280	90	2820	120	3200	120	3500	150

The stimuli were further processed using Matlab routines. The first 100 ms were truncated, since there was a large transition effect showing a DC shift at the output of the Klatt synthesizer. Second, the vowels were gated at the initial 50 ms and at the final 50 ms, using a cosine ramp. Finally, all vowels were balanced in root mean square level. However, this does not guarantee that loudness can not be used as a cue when two vowels with different F0 must be discriminated. An alternative experiment would include loudness roving, but this approach was not adopted for two reasons. First, it could make the task more difficult, even if loudness would not be the cue that is used to discriminate. Second, the experiments attempt to reveal the importance of the temporal coding of F0. So, if the performance for the FLAT CIS algorithm is due to the loudness cue, and if the performance for the other algorithms is significantly better, another cue has to be used.

Within each channel, the mean amplitude of the pulses changes when F0 is varied. This is mainly due to the fact that the attenuation of the harmonics in the side bands of the band-pass filters depends on the frequency of the harmonic, and the latter changes along with the fundamental frequency. This change of average pulse amplitude in each channel can alter the perceived loudness and even the perceived timbre. This cue corresponds to the aforementioned "average amplitude cue". Figure 4.7 illustrates this "average amplitude cue" by showing the average pulse amplitude in each channel for two synthetic vowels (/i/) with different fundamental frequencies (250 Hz and 300 Hz). The vowel with the lower F0 seems to have more energy in the most apical channel, and less energy in the other channels.

To illustrate the effect of the different algorithms, the outputs of all vowels were further analyzed. For each synthesized vowel, all F0 fluctuations within the middle 200 ms were extracted, and both the minimum and the maximum pulse level within each fluctuation was determined. The average over all extracted fluctuations was calculated for both the minimum

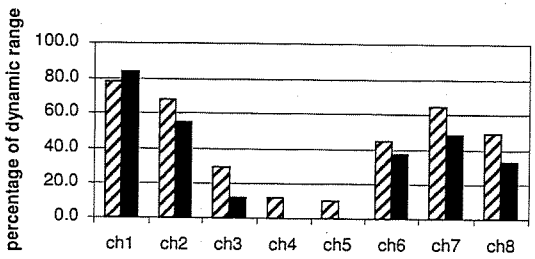


Fig. 4.7 Average pulse amplitude in each channel for two vowels. The dashed bars correspond to an /i/ with F0 = 300 Hz. The solid bars correspond to an /i/ with F0 = 250 Hz.

and the maximum pulse level. Figure 4.8 (/a/) and figure 4.9 (/i/) show both values for the three algorithms as a function of the fundamental frequency. Each plot is valid for another channel. The vertical axis represents the relative dynamic range: "0" corresponds to the threshold level, "100" to the most comfortable level. The dashed lines correspond to the FLAT CIS strategy, the thin solid lines to the default CIS strategy, and the thick solid lines to the F0 CIS strategy. For the FLAT CIS strategy, the mean minimum and mean maximum pulse levels almost completely overlap, illustrating that there is hardly any modulation corresponding to F0 left in the signal. In the case that there is still a non-zero modulation depth, it is far below the detection limits measured in experiment 1. The curves for the FLAT CIS strategy also illustrate the "average amplitude cue": the mean amplitude varies as a function of F0 in all channels. However, many of these curves show a clear non-monotonic behavior, which can make this cue very ambiguous. For the default CIS strategy (thin solid lines), there is a clear gap between both curves, which corresponds to the relative modulation depth. In case of the vowel /a/, there is a large depth at the lowest F0s (43.9 % at F0 = 149 Hz, averaged across all channels), but it decreases rapidly for higher F0s (17.5 % at 250 Hz, 14.0 % at 370 Hz). The same pattern is seen in the plots for the vowel /i/, but in this case the depth at the lowest F0s is already smaller (31.5 % at 150 Hz, 17.5 % at 250 Hz, 16.2 % at 370 Hz). This pattern is opposite to the pattern needed for good F0 discrimination, e.g. the results of experiment 1 for subject SV clearly indicate that a higher modulation depth is needed for higher F0s in order to achieve the same performance as for low F0s. He performed at chance for a modulation depth of 20 %, which is above the average depth obtained for F0 equal to 250 Hz

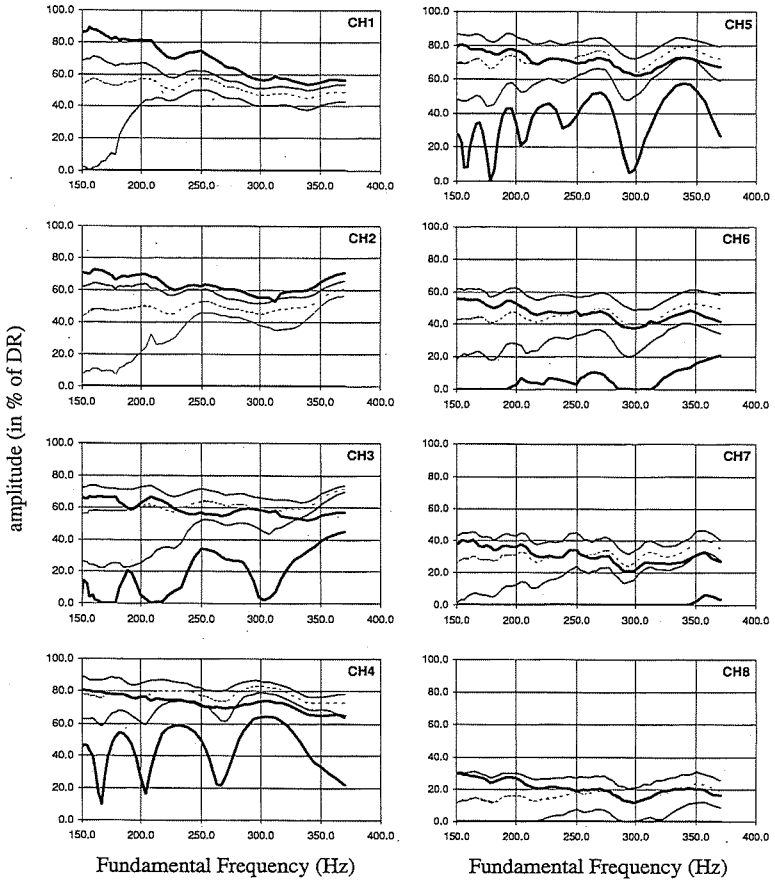


Fig. 4.8 Minimum and maximum pulse level within each fundamental period for the vowel /a/ for the three algorithms FLAT CIS (dashed lines), default CIS (solid thin lines) and F0 CIS (solid thick lines). The horizontal axis corresponds to the fundamental frequency (in Hz), the vertical axis corresponds to the relative dynamic range. Each plot shows the curves for the channel indicated at the top right. Note that the two curves for the FLAT CIS algorithm coincide almost completely.

(17.5 % for both vowels). So, he might benefit from a new algorithm that increases the modulation depth.

The F0 CIS is successful in accomplishing this (thick solid lines): the mean modulation depth varies from 53.8 % (F0 = 149 Hz) to 38.6 % (250 Hz) and 35.4 % (370 Hz) for the vowel /a/. For the vowel /i/ these values are 48.9 %, 38.4 % and 40.3 % respectively. Notice also that in many channels, the average minimum pulse is zero, i.e. at threshold level. In some cases, the minimum level shows a strong fluctuation when plotted as a function of F0. This effect is also visible for the default CIS algorithm, but to a lesser extent. It is inherent in the way the speech signal is analyzed in the frequency domain, namely by using band-pass filters with a flat response in a broad pass band, and steep slopes beyond the cut-off frequencies. Phase and attenuation of individual harmonics can differ substantially if e.g. the filter contains only one harmonic in the pass band close to the center frequency as opposed to the case where the filter contains two harmonics far from the center frequency. Another difference in the output between the F0 CIS and the default CIS algorithm, is the frequency weighting. The maximum pulse levels in the two lowest frequency channels are higher for the F0 CIS strategy, and are higher in the remaining channels for the default strategy. To compensate this, one could apply a channel specific gain in the F0 CIS algorithm, instead of the factor 1.7 (see description of F0 CIS in section 4.2).

The same four subjects as for experiment 1 participated in this experiment. Again, the APEX set-up was used for the presentation of the stimuli and the collection of the responses. To estimate the smallest difference in F0 that could be discriminated, a 2-down 1-up adaptive procedure was used (Levitt, 1971). Two vowels with different F0, further called the *standard* and the *signal*, were presented and the subject had to indicate the stimulus with the highest pitch. Correct response feedback was always given. The fundamental frequency of the standard stimulus was fixed to either 149 Hz or 250 Hz, so there were twelve conditions in total: two phonemes (/a/ and /i/), two standard F0s and three algorithms. Initially, the signal had a F0 of 250 Hz or 370 Hz respectively. After two successive correct answers, the F0 of the signal was lowered, while after one incorrect response, it was increased again. Initially, the stepsize of the fundamental period was 0.4 ms (standard F0 = 149 Hz) or 0.2 ms (standard F0 = 250 Hz). After four reversals[12], the stepsize was reduced to 0.1 ms or 0.05 ms

[12] A reversal occurs if a series of correct answers (at least two in this case) is interrupted by an incorrect answer, or vice versa (see also chapter 2).

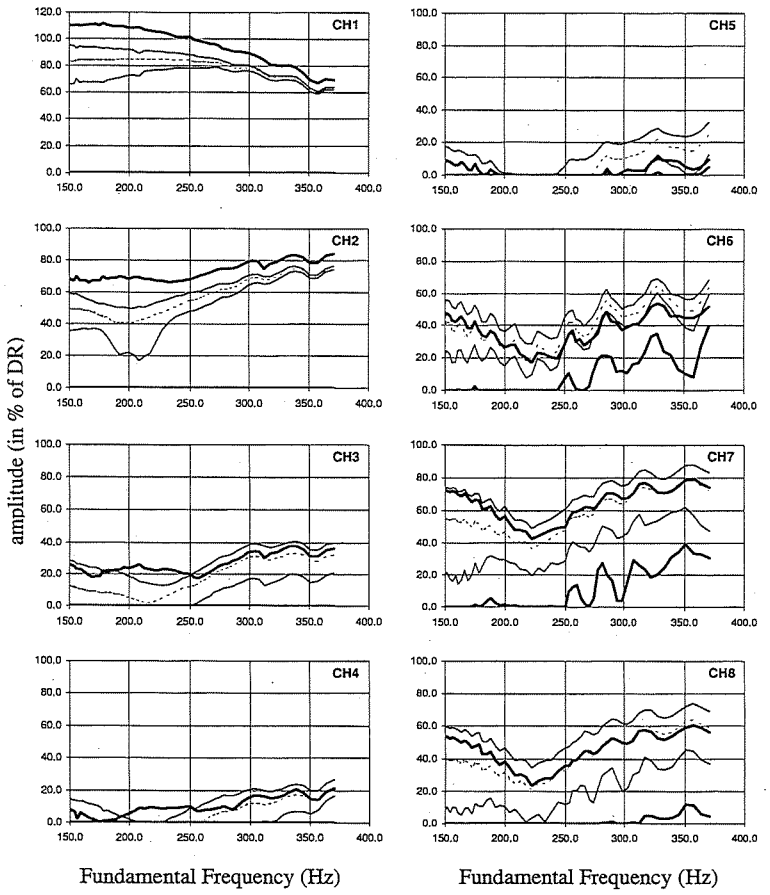


Fig. 4. 9 Same as figure 4.8, but for the vowel /i/.

respectively. The procedure stopped after twelve reversals, and the mean value of the fundamental period at the last eight reversals was calculated. This value corresponds to the fundamental period of a signal that could be discriminated in approximately 71 % of the cases (Levitt, 1971). For each condition, the whole procedure was repeated four times, and the mean and the standard deviation of the results were calculated. In some conditions, the subjects could not discriminate both stimuli, even at the largest difference in F0. In those cases, the percentage correct responses was calculated at this difference. Obviously, it is expected that these scores are lower than the aforementioned 71%.

4.4.2 Results

The results are shown in figure 4.10 for each subject separately. The horizontal axis corresponds to the conditions, the vertical indicates the F0 difference, which is the inverse of the fundamental period. In general, the largest F0 differences were obtained in the FLAT CIS algorithm. Still, in many cases the adaptive procedure converged to a value below the starting value. This indicates that the average amplitude cue can be used to distinguish between two vowels with different F0.

However, the fact that smaller F0 differences were obtained for the two other algorithms indicates that the inclusion of the temporal cue increases the ability to make a judgement about the pitch. All subjects obtained significantly lower F0 differences for these two algorithms at the lower frequency (150 Hz). The F0 difference that could still be heard varies between 6 Hz and 20 Hz across the subjects. Subjects SV and AE, who were sensitive to pitch changes above 250 Hz in experiment 1, also benefit from the envelope modulation cue for the vowel /a/ at higher frequencies. Their results vary between 12 Hz and 19 Hz. The other two subjects do not obtain better results, since they do not have the necessary sensitivity. For all subjects, very good performance was obtained for the FLAT CIS algorithm in the </i/ - 250 Hz> condition, so that there was no room for improvement. Underlying reasons for these unexpected results will be discussed in the next section. No significant differences are observed between the results of the default CIS and the F0 CIS algorithms. This indicates that there is a saturation effect, as also observed in experiment 1: further increase of the modulation depth does not improve the pitch discrimination ability.

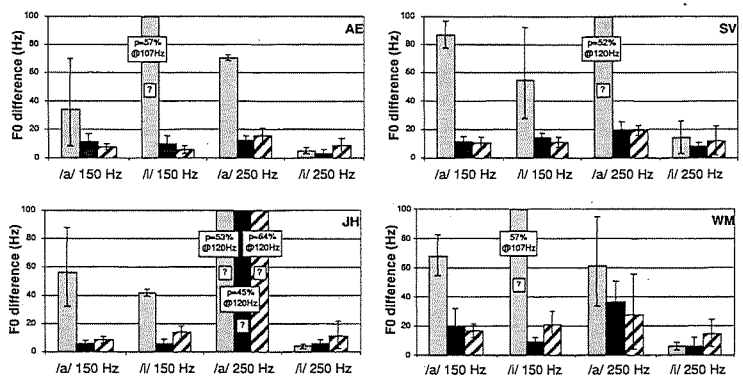


Fig. 4.10 Smallest discriminable fundamental frequency differences, as measured in experiment 2. Each plot shows the results of another subject. Vowel and fundamental frequency of the standard stimulus are indicated at the bottom. The three algorithms are: FLAT CIS (gray bars), default CIS (black) and F0 CIS (hashed). If the subject's results did not converge to some value, a large bar is displayed, and the percentage correct responses at the indicated F0 difference is given.

4.4.3 Discussion

The very good performance in the </i/ - 250 Hz> condition can be explained from figure 4.9. Above 250 Hz there is a clear quasi-monotonic relationship between the fundamental frequency of the vowel and the average maximum pulse level in each fundamental period. This is the case for all three algorithms. Moreover, the slopes of the curves are rather steep, meaning that a small change in F0, results in a substantial change in pulse level in each channel. While this level increases for smaller F0s in the lowest frequency channel, it is the other way around for the other channels. This can result in a pitch change according to tonotopic place coding, but also in a timbre or even loudness change. The latter was reported by subject JH, while the other three subjects still reported that they heard a change in pitch. For the lower frequencies (150 – 250 Hz), the slopes of the curves in figure 4.9 become very shallow or even change sign. This inconsistency

probably explains why performance is a lot poorer for the </i/ - 150 Hz> condition, when there are no envelope modulation cues (FLAT CIS). The same is true for the vowel /a/: the pulse level changes non-monotonically with fundamental frequency, which makes it hard to use this cue only to identify the stimulus with the highest F0.

4.5 General discussion

The results of both experiments are consistent in most cases. In experiment 1, subjects SV and AE showed a clear sensitivity to modulation frequency variations, both for the low and the high F0 frequencies. In experiment 2, they obtained better results when the envelope modulation cue was added (FLAT CIS versus other two), except for the condition </i/ - 250 Hz> for which there was no room for improvement. Subjects JH and WM could not detect modulation frequency variations from 250 Hz in experiment 1, so no smaller F0 differences were obtained in experiment 2, when the envelope modulation cue is added. Obviously, a cochlear implant user will not benefit from a speech processing algorithm that adds or enhances a cue for which s/he is not sensitive.

However, there appear to be some inconsistencies. From the results of experiment 1, one would predict that some subjects should benefit from the F0 CIS algorithm in some conditions. E.g., subject SV performed at chance for a modulation depth of 20 % at the modulation frequency of 250 Hz. In the default CIS strategy, the modulation depth varies between 12 % and 24 % for the synthetic vowel /a/ (F0 = 250 Hz) used in this study. Also subject WM performed rather poorly in experiment 1, and performed only significantly above chance in one condition, where modulation depth was 99 %. However, no improvement is seen in their results of experiment 2 when F0 CIS is compared with the default CIS strategy. There are two possible explanations for this finding. First, not all channels were tested in experiment 1, so it is still possible that there are other channels for which less modulation depth is needed in order to achieve maximum performance. However, for subject WM, several channels were tried in a pilot experiment, and no better results were obtained for those channels. Second, and more likely, the main difference between both experiments is that in experiment 1 only one channel is stimulated, while in experiment 2 all channels are stimulated. It is possible that less modulation depth is needed when the modulation according to F0 is presented on all channels. This can also explain the relatively good performance of subject WM in experiment 2. At the low frequencies (F0 around 150 Hz), he clearly benefits from the

inclusion of F0 fluctuations in the envelope, and can detect F0 differences of 10 to 20 Hz. It is also possible that an improvement in the results of the F0 CIS algorithm would be seen if other stimuli were used. E.g., if a low frequency sine wave is passed through the default CIS algorithm, the frequency of the envelope fluctuation is doubled due to the full-wave rectification. This is a disadvantage, since modulation frequency discrimination is more difficult at higher frequencies for cochlear implant users.

The results demonstrate the importance of the temporal coding of F0 for the perception of its corresponding pitch. In the SPEAK strategy, the bandwidths of the extracted envelopes are limited to 200 Hz, which can fall below F0. However, the steepness of the smoothing filter is far from infinite, which means that enough F0 modulation might remain in the envelope. A greater concern is the use of the relatively low pulse rate of 300 pps (see chapter 1). According to several physiological and psychophysical studies that recommended a carrier rate equal to 4 times the modulation rate (see section 4.1.3), only modulation frequencies up to 75 Hz can be encoded properly in the SPEAK strategy. However, in the aforementioned studies, single channel stimuli were used, in contrast with the multiple channel stimuli at the SPEAK processor output. In analogy with the fact that less modulation depth is required, it might also be true that a lower carrier rate suffices if all channels contain synchronous envelope modulations. This may explain why Jones *et al.* (1995) did not see any improvement in pitch related tasks if an explicit temporal code according to F0 was added to the SPEAK strategy.

Using the algorithms that included a temporal cue (default CIS and F0 CIS), the four subjects could hear F0 deviations from 150 Hz between 4% and 13%. Two subjects were not sensitive to pitch changes above 250 Hz in experiment 1, but the other two could hear F0 differences between 5% and 8% at 250 Hz in experiment 2. These data are comparable with measurements of the rate pitch difference limens (DL)[13] obtained in other studies. McKay *et al.* (1994) found DLs between 2-15 % at 150 Hz, and between 2-20 % around 250 Hz, while one out of the six subjects was not sensitive to rate changes from 250 Hz. Townshend *et al.* (1987) obtained rate pitch DLs between 10% and 30% at 150 Hz, and between 10% and 20% at 250 Hz, while one out of the three subjects was not sensitive to rate changes from 250 Hz. It might also be interesting to compare the observed

[13] This is the smallest difference in rate pitch that can still be perceived.

DLs for harmonic complexes with those obtained psycho-acoustically with normal hearing listeners. Houtsma and Smurzynski (1990) measured the DLs of complex tones (F0 equal to 200 Hz) from which the lower harmonics were removed. Averaged across 4 listeners, they found a DL of less then 1 Hz (0.5 %), even when the first six harmonics were removed. Cullen and Long (1986) observed a similar value for unfiltered acoustic pulse trains with a rate of 200 Hz. In other words, normal hearing listeners outperform implantees with an order of magnitude. The question of whether this is due to a better temporal or a better place coding (or the combination of both) still remains unresolved. However, the latter reason might be more plausible, given the facts that place coding is rather coarse in a cochlear implant and that the temporal representation of amplitude modulated stimuli is well preserved in the neural response for high carrier rates (Wilson, 1997).

In this study, synthetic vowels were used, in order to be able to control all the parameters of the vowel. Formant frequencies and bandwidth were kept constant, and only the fundamental frequency was varied. This is unlike real speech: speakers with higher F0s, such as women and children, typically have higher formant frequencies too, due to the smaller physical dimensions of the vocal tract. It is not unlikely that these cues would also be used if a comparable task as in experiment 2 was performed with real life speech.

4.6 Summary and conclusions

A new speech processing algorithm was designed, which increases the modulation depth of the F0 fluctuations at the output. To investigate the influence of this depth on pitch discrimination, two experiments were conducted. In the first, it was investigated how the modulation depth of a sinusoidally amplitude modulated pulse train on one electrode channel influences the ability to judge on the perceived pitch. Optimal performance in a pitch discrimination task depended on the subject, the channel (apical or basal) and the frequency range (150 Hz - 180 Hz or 250 Hz – 300 Hz). The relative modulation depths for which the highest scores were obtained, varied between 20 % and 99 %. In the second experiment, the smallest discriminable F0 difference was measured for three algorithms using synthetic vowels. In the first algorithm (FLAT CIS), the temporal variation in the envelope according to F0 was removed, so subjects had to make a judgement on the basis of relative amplitude changes. They all performed very well in one condition (</i/ - 250 Hz>), due to the unambiguous relation between F0 and the average amplitude in each channel. For the other

conditions (</i/ - 150 Hz>, </a/ - 150 Hz> and </a/ - 250 Hz>), performance was much poorer, and in some cases at chance level. If the temporal cue was added in the envelopes (default CIS and the new F0 CIS), significantly better results were obtained, provided that the subject was sensitive to modulation frequency changes in the corresponding frequency region. In the default CIS algorithm, modulation depths are obtained at which some subjects scored sub-optimal in experiment 1. So, it was expected that the F0 CIS would result in a better performance. However, this was not the case. A possible reason is that less modulation depth in the envelope variation according to F0 is needed, if all channels are stimulated.

Chapter 5

Detection of small across-channel timing differences by cochlear implantees

Abstract

In order to improve existing speech coding strategies for cochlear implants, more insight is needed in the perceptual effect of parameter variations of electrical stimulation of the auditory nerve. Changing the order of stimulation in the Continuous Interleaved Sampling (CIS) strategy results in changes of the time interval between the electrical pulses applied on different channels. Five post-lingually deafened users of the Laura cochlear implant participated in several experiments in which their sensitivity to changes of this time interval is investigated. In section 5.1, the relation of these experiments with speech processing algorithms is explained and previous comparable studies with cochlear implants and equivalent psycho-acoustical studies are summarized. In the first experiment (section 5.2), the subjects were presented with two trains of biphasic pulses applied concurrently to two widely separated channels. They could all discriminate between stimuli where pulses on the two channels were nearly synchronous (inter-channel delay = 0.1 ms) and those with a longer delay applied to one channel. All showed an asymmetry, being more sensitive when the longer delay was on either the more basal or, depending on the listener, the more apical channel. In the second experiment (section 5.3), a "masking" pulse train was presented on intermediate channels to prevent that the two target channels stimulate a common neural population. However, this did not substantially affect the results. In the next two sections, two supplementary experiments are described. In the third experiment, it is checked whether the (place) pitch of all the channels used in the previous experiments agrees with the tonopic organization or not. The task in the fourth experiment was to discriminate two dual-channel pulse trains with the same amount of delay, but with opposite directions (basal or apical channel first). Possible explanations and consequences of all the results are discussed in section 5.6. One of the main conclusions is that performance was not mediated by neurons responding to both channels, but that the findings demonstrate a genuine sensitivity to the relative timing of stimulation applied to discrete populations of auditory nerve fibers. The results also provide the first demonstration that one can exploit across-channel timing differences in a speech processor to change the perception of electrical stimulation by cochlear implantees.

Carlyon, R.P., Geurts, L. and Wouters, J. (2000). "Detection of small across-channel timing differences by cochlear implantees," Hearing Research 141,140-154

5.1 Introduction

A CIS processor stimulates all electrode channels in the cochlea sequentially in a predefined order. Various alternatives have been suggested, such as "from apex to base" (1-2-3-4-5-6-7-8), "from base to apex" (8-7-6-5-4-3-2-1) or a staggered order, e.g. as used in the 8 channel Laura implant (1-4-7-2-5-8-3-6). In the latter case, the time interval between stimulation on well separated frequency regions is shorter. It is reported that for one particular subject, speech reception scores were better for the apex-to-base order than for the staggered order (Lawson *et al.*, 1997; Dorman and Loizou, 1997). However, until now no extensive speech intelligibility tests are conducted in order to evaluate the suggested orders thoroughly. A better approach though to investigate the perceptual effect of changing this order or the impact of it for speech processors would be to employ more elementary stimuli on a restricted number of channels. One of the effects of using a different order is that the time interval between the pulses on any two channels changes. In the apex-to-base order, the delay of stimulation between adjacent channels is low, e.g. equal to 0.1 ms if the overall pulse rate is 10000 pulses per second (pps), and is large between widely separated channels, up to 0.4 ms if eight channels are stimulated[14]. The same holds for the base-to-apex order, except that the sign or the direction of the delay reverses. When a staggered order is used, the delay between adjacent channels is higher, and between more separated channels is lower. In this chapter, several experiments are described in which the perceptual effect of this time delay is investigated. Two pulse trains are applied to two widely separated channels, and the timing difference between both pulse trains is varied from 0.1 ms to 2 ms. The purpose is to examine to which extent a cochlear implantee can discriminate between stimuli having different values for this delay. Care will be taken that the two channels activate neurons from different regions in the cochlea, so that the subjects can not use within-channel cues, which emerge from nerve fibers that are sensitive to electric stimulation on both channels.

Two similar psychophysical studies have concluded that implantees have very poor sensitivity to across-channel timing differences. Tong and Clark (1986) presented two listeners with 100-pps pulse trains concurrently to two electrode channels, and asked them to discriminate between an inter-

[14] Since the sequence of the presentation is continuously repeated, the delay can never exceed the value of 0.4 ms if eight channels are stimulated at an overall rate of 10000 pps.

channel delay of 5 ms and a larger delay. Thresholds increased from less than 0.5 ms at small electrode separations to 1.5 ms and 2.5 ms (depending on the listener) at the widest separations tested (3.0 and 6.0 mm respectively). They concluded that the lower thresholds obtained at small channel separations were due to the two channels exciting a common population of auditory nerve fibers. McKay and McDermott (1996) presented periodic pulse trains to each of two electrode pairs, with each period of each train consisting of one "large" (high-current) pulse and four smaller ones. They asked listeners to discriminate between the case where the delay between the large pulse on one electrode channel and that on the other was 1 ms, and conditions where this delay was longer. They found that performance was at chance when the spatial separation between the electrode channels was greater than 4 mm. McKay and McDermott concluded that listeners could not compare the temporal pattern of activation between channels, which stimulated non-overlapping populations of auditory nerve fibers. One reason for the generally poor performance in the studies by Tong and Clark and by McKay and McDermott may be that the reference stimulus did not consist of simultaneous pulses on the two channels. Listeners in acoustic experiments have reported that the percept produced by an above-threshold asynchrony is that of the sound breaking into two separate sources (Carlyon and Shackleton, 1994). It may be that once this segregation has occurred, further across-frequency timing differences are not easily detected (c.f. van Noorden, 1975). If so, then implantees may be better at discriminating between a reference stimulus in which the pulses on two channels are simultaneous and a signal containing a small across-channel delay, than they are at discriminating between two large delays (c.f. Tong and Clark, 1986). A problem with testing this hypothesis is that, in order to avoid direct electrical interaction between channels, most modern devices prohibit truly simultaneous stimulation. Our approach was to use a reference stimulus which was as synchronous as possible on the two channels (delay = 0.1 ms), and to perform additional controls for the use of within-channel cues. These included, in addition to the selection of widely separated channels, the presentation of masker signals in electrode channels between the two to be compared.

Equivalent psycho-acoustical experiments investigated the sensitivity to timing differences between the envelopes of stimuli occupying widely-separated frequency regions (Strickland *et al.*, 1989; Yost and Sheft, 1989; Carlyon, 1994). Two studies have shown that normally hearing listeners can detect phase differences of about 30-60° between the envelopes of sinusoidally amplitude-modulated tones with different carrier frequencies

(Strickland *et al.*, 1989; Yost and Sheft, 1989). According to the tonotopic organization of the cochlea, the use of different carrier frequencies ensured that different neural populations are stimulated. In those experiments, care was taken to eliminate cues based on interactions in neural populations having center frequencies between the two carriers. Typical precautions included a narrowband noise presented in a frequency region intermediate to the two carriers, and the use of well-separated carrier frequencies. Carlyon (1994) performed some similar measurements using acoustic stimuli more akin to the pulse trains commonly used in cochlear implant research. He measured the detection of timing differences between pairs of multi-component complex tones that had been filtered into separate frequency regions. For these stimuli, which resembled bandpass filtered pulse trains, thresholds were approximately 2.5 ms at all fundamental frequencies. It was concluded that ringing on the basilar membrane prevented thresholds from dropping below this value. This ringing corresponds to the decaying oscillating displacement of the basilar membrane after the presentation of a very short click, as also seen in the impulse response of a mechanical damper-spring system. As a result, a single acoustic click can cause several neural discharges, which makes the neural representation of the stimulus more blurred. Another striking difference between acoustic and electric stimulation is the distribution of the neural response latencies[15]. At least at current levels well above the threshold of single neurons, auditory-nerve responses to electric pulses are very tightly time-locked, and can have latencies whose standard deviations are as low as 8 μs (van den Honert and Stypulkowski, 1987). All this implies that the sensitivity of implantees to across-channel timing differences might exceed that of normal listeners to across-frequency asynchronies. If this is true, it would show that acoustical stimulation underestimates the ability of the central auditory system to compare the pattern of activity between different populations of auditory nerve fibers. Also, it can have an important impact on the design of speech processors for cochlear implants

5.2 Experiment 1: Psychometric functions

5.2.1 Subjects

Details of the five post-lingually deafened subjects who participated in the experiments are shown in table 5.1. At least four sessions of about 2.5

[15] The latency is the time delay between the onset of the stimulus and the onset of the neural response.

hours were needed to complete all the experiments of this study. The last two columns of table 5.1 show the percentage of vowels (10-alternative forced choice) and consonants (16-alternative forced choice) correctly identified by each listener in the study by van Wieringen and Wouters (1999a). All five scored above the average (42%, sd=19) obtained by 25 listeners on the vowel test, and all except AP performed above the average (33%, sd=13) consonant identification. Four other listeners took part in a pilot experiment with adaptive procedure (see next section) but were not selected to participate in the present study. One of them was congenitally deaf and had difficulty in performing all tasks that she was tested on, including the adaptive task and the loudness balancing used here. Two others could do the loudness balancing but could not track a threshold in the adaptive across-channel asynchrony experiment. Both had performed below average on the vowel and consonant identification tests (van Wieringen and Wouters, 1999a). A fourth could do both tasks but was rejected because the only two channels which could give rise to sufficiently loud percepts for the present stimuli were channels 2 and 4, which were deemed too close to preclude significant peripheral interactions. Note that in the default CIS strategy and in the subjects' clinical Lauraflex processor, a much higher stimulation rate is used, resulting in much lower current levels to achieve normal loudness.

Table 5.1: Details of the five listeners who participated in the experiments. The last two columns show the percentage of vowels and consonants recognized correctly by each listener in a previous study (van Wieringen and Wouters, 1999a). Vowels: 10-alternative forced choice, /hVt/ context. Consonants: 16-alternative forced-choice, /aCa/ context.

Listener	Age (yrs)	Duration of profound deafness (yrs)	Aetiology	Implant experi-ence (yrs)	Vowel score (%)	Con-sonant score (%)
JH	16	3	Meningitis	3	53	45
AE	41	3	Menière	1	79	51
DD	31	1	Trauma	2	61	35
AP	59	3	Cochlear otoscler.	1	49	29
SV	29	6	Unknown	4	73	58

5.2.2 Stimuli and procedure

The APEX set-up was used for the generation and the presentation of the stimuli, the definition of the experimental procedure and the collection of the responses. The stimuli consisted of two concurrent pulse trains applied to different channels of the Laura cochlear implant. The pulses in each train were biphasic with a duration of 40 µs/phase and were presented at a rate of 100 pps. The pulse trains were 400-ms long and were gated on and off with 50-ms linear ramps, the purpose of which was to discourage the use of differences in onset time between the two channels, as in the psycho-acoustical study by Carlyon (1994). The two "target" channels were chosen for each subject to be as far apart as possible (as measured along the length of the electrode array) with the constraint that each channel should be capable of producing a sound of sufficient loudness (see below) for a 100-pps 40 µs/phase pulse train. Details of the target channels and the version of the electrode array are shown for each listener in table 5.2. X-rays of the prostheses implanted in two listeners were available and are shown in figure 5.1. Note that for these two listeners the target channels, indicated by solid lines, were separated by approximately one half-turn of the cochlea. (The significance of the electrodes connected by dashed lines will be explained in a later section).

Table 5.2: Columns two and three show the apical and basal target channels used for each listener. The figures in parentheses give the separation between the electrodes in each channel, in millimeters. The fourth column shows the separation between the two channels measured from points mid-way between the two electrodes (measured longitudinally, along the axes of the electrode arrays) in each channel. The version of the Laura electrode array implanted in each listener is given in the penultimate column. The current level of the 100-pps pulse trains for equal normal loudness are indicated in the last column.

Subject	Apical Pair (width, mm)	Basal Pair (width, mm)	Separation (mm)	Electrode array version	Current, µA (apical,basal)
JH	2 (3)	7 (1.3)	11.1	5.6	(600,723)
AE	4 (3)	8 (2)	7.5	5.7	(1000,1050)
DD	3 (1.3)	6 (1.3)	6.2	5.6	(550,940)
AP	2 (3)	7 (3)	10.0	5.8	(663,750)
SV	2 (3)	6 (1.3)	9.1	5.6	(675,800)

Before the main experiment, the currents on the two target channels were adjusted to be equally loud. First, the current on either the basal (listeners AP and SV) or apical (JH, AE, DD) channel was adjusted, with a

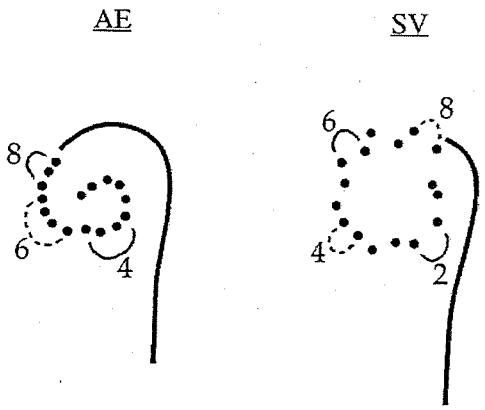


Fig. 5.1 Tracings of X-rays showing electrode placement for listeners AE and SV. The two electrodes making up each channel used in experiment 1 are joined by solid lines. Those joined by dashed lines are the maskers used in experiments 2 and 4. The different appearance of the X-rays for the two listeners arises from the fact that AE was implanted with version 5.7 of the array whereas SV was implanted with version 5.6. (Thanks to Dr. P. Govaerts of the Medical Institute Sint-Augustinus in Antwerp for providing the X-ray tracings.)

resolution of 10 µA, until it was judged to correspond to the middle point, labeled "normal", on a five-point scale ranging from "very quiet" to "very loud". The current level on the other channel was then adjusted so that it was judged to be equal in loudness to the first, following a procedure in which pulse trains on the "fixed" and "adjustable" channels were presented in the order "fixed-adjustable-fixed", with a 500-ms gap between them. At the start of half of the loudness matching runs the loudness of the adjustable channel corresponded to "loud" on the five-point scale; on the other half it corresponded to "quiet". Listeners AE, AP, and SV performed six loudness matches each; time constraints meant that listeners JH and DD each performed only two matches. All of the adjusted levels were averaged for each listener to obtain the current to be used for that channel in the main part of the experiment, and are shown in the right-most column of table 5.2. Note that any minor differences between the loudnesses of the two individual channels would not have provided a reliable cue in the main part of the

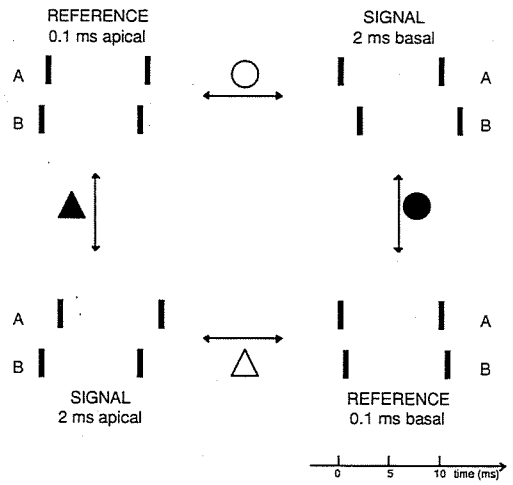


Fig. 5. 2 Schematic representation of the relative timing of the pulses applied to the basal (B) and apical (A) target channels in the experiments. Two subsequent pulses in each channel are shown. The top row shows a trial in which the signal comes second and contains a 2-ms basal delay; in the bottom row the signal comes first and contains a 2-ms apical delay. The open symbols indicate the comparisons made during each trial of the main part of experiment 1 (circle = basally delayed signal, triangle = apically delayed signal). Filled symbols correspond to the additional condition in which the same channel was delayed for both the reference and signal stimuli. The 0.1-ms asynchronies in the reference stimuli have been exaggerated for clarity.

experiment, in which both channels were always stimulated for every presentation.

In the main experiment listeners were required to discriminate between two stimuli in which both channels were stimulated concurrently, and which differed only in the relative timing between the two channels. Each stimulus pair consisted of a reference stimulus (the *standard*) , with the smallest possible delay, and a so called *signal*, mostly having a larger delay. Two conditions were used. In one condition, the reference stimulus had pulses in the more basal channel

delayed by 0.1 ms (onset to onset) relative to those on the more apical channel, and the signals had delays on the apical channel. In the other condition, the pulses in the reference stimulus were delayed by 0.1 ms on the apical channel, and the pulses in the signal were delayed on the basal channel. In each case, five different values of signal delay, ranging from 0.1 to 2.0 ms, were used. A schematic representation of two successive pulses in each pulse train is shown for both conditions in figure 5.2. The top row corresponds to a stimulus pair where the signal has a 2-ms basal delay, the bottom row contains an apically delayed signal. In this experiment, the subjects had to compare two stimuli as indicated by the open symbols on the figure. (The meaning of the filled symbols will be described later.)

The reference and signal were presented in random order with an inter-stimulus delay (offset-onset) of 500 ms. Listeners were required to say whether the signal was presented first or second. In each block of trials the channel which was delayed in the signal interval was always the same, and the reference stimulus always contained a 0.1-ms delay on the other channel. Each block consisted of ten trials at each inter-channel delay. The order of the blocks was randomized with the constraint that blocks of trials with basally delayed signals were interleaved with those containing apically delayed signals. Ten blocks were run for each signal delay, making a total of 100 trials per point.

As no feedback was given after each trial, the decision as to which of two stimuli to label the signal and which the reference is to some extent arbitrary. Note however that all listeners had participated in a pilot experiment using an adaptive procedure (see chapter 2) in which feedback was given, with the response being labeled as correct whenever the stimulus with the longer inter-channel delay was identified as the signal. Note also that the main experiment included some trials where the signal contained a 0.1-ms delay on one channel and the reference contained an identical delay on the other channel. Because the signal was defined as the apically-delayed stimulus in some blocks and as the basally-delayed stimulus in others, this meant that above-chance performance on one of these conditions would necessarily lead to below-chance performance on the other. Indeed, had feedback been provided in these trials, listeners would have received inconsistent information as to which stimulus to treat as the signal.

5.2.3 Results

The open symbols in each panel of figure 5.3 show the psychometric functions for one listener. Note that in this and in subsequent figures, a deviation from chance performance in either direction (above or below 50% correct) indicates that subjects could discriminate between the standard and signal. The most salient features of the results are that there are substantial inter-listener differences, but that all listeners performed well above chance in at least one condition. Listeners JH, AE, DD and AP could all distinguish between a stimulus having a 0.1-ms delay on the basal channel and one having a 0.1-ms apical delay. The first three of these listeners (left column) identified the basally-delayed stimulus as the "signal", and so their score for the basal signal is above 50%, and that for the apical signal below 50%. Listeners JH, AE and AP also did much better for basal than for apical signals over a range of larger delays. The opposite was true for listener AP. Listener SV differed from the others in that he could not discriminate between basal and apical delays of 0.1 ms. He did, however, also show an asymmetry in that he was generally better at detecting longer delays on the basal than on the apical channel. A further unique feature of SV is that he was the only one who explicitly described the signals as containing an "echo". The others described the signals using the Flemish word "zachter", which means "softer " both in the sense of "quieter" and "more muffled".

Before discussing the results in more detail, it is worth noting that, although there was substantial variability between listeners, the reliability of the measures for a single listener and condition was very high. This is demonstrated by a reliability coefficient of 0.972 (calculated using the Spearman-Brown formula (Ferguson and Takane, 1989, p466-478)) between the first and second halves of the responses making up each data point in figure 5.3.

At first sight, the results of listeners JH and AE (left column) suggest that they were exquisitely sensitive to delays on the basal channel, and less sensitive to apical delays. However, an alternative explanation stems from the fact that the reference stimulus for the basally delayed signals was different from that used for the apically delayed signals. It may be that this reference stimulus, in which the apical channel was delayed by 0.1 ms, sounded "special" in the sense of producing a percept that was qualitatively different from all other stimuli used in the experiment. Similarly, to account for the results of listener AP, one would have to propose that the reference containing a 0.1 ms basal delay was special. Support for this interpretation

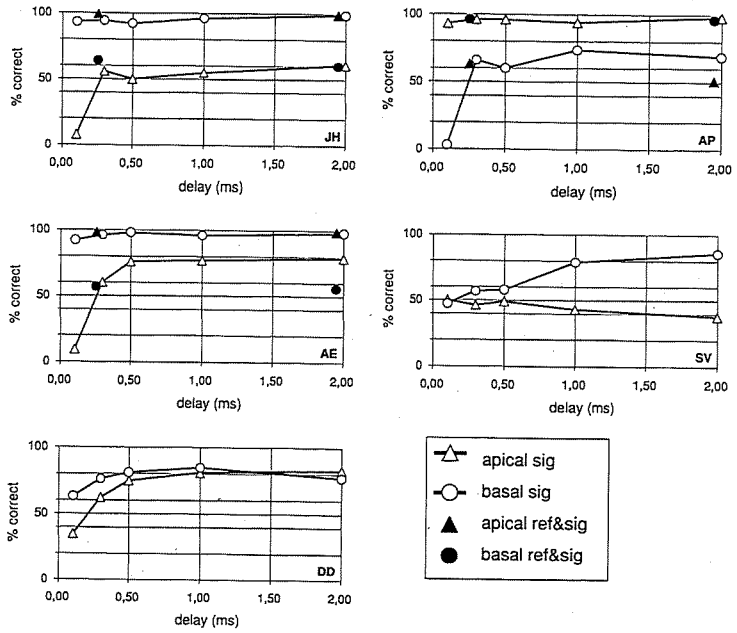


Fig. 5.3 Connected open symbols show sensitivity averaged across 100 trials/point for conditions where the signal had a delay on the basal (circles) or apical (triangles) channel. In these conditions the reference had a 0.1-ms delay on the other channel. Filled unconnected symbols show data from a condition where the reference and signal had delays on the same channel; they have been shifted horizontally by a small amount for clarity. Note that although the ordinate is labeled "% correct", any deviation from chance performance (50%) indicates that listeners could discriminate between the standard and signal.

comes from an additional experiment in which JH, AE, and AP had to discriminate between a basally or apically delayed signal and a reference having a (shorter) 0.1-ms delay in the *same* direction. This experiment can be understood with reference to figure 5.2 if one assumes that the two stimuli to be compared in each trial are shown in the same column (comparison indicated by the filled symbols), rather than in the same row.

The filled symbols in figure 5.3 show that the asymmetry in performance reversed, so that JH and AE were better at detecting the apical delay, and AP was better at detecting the basal delay. In other words, the level of performance was determined by the reference stimulus rather than by the signal. Note, however, that not all of the results can be interpreted in terms of one stimulus sounding "special", as listeners AE, DD, and AP performed above chance both when the reference contained a basal delay and the signal an apical delay, and *vice versa*. In addition, the explanation cannot account for the data of SV, who could not discriminate between basal and apical delays of 0.1 ms but who was sensitive to longer delays on the basal channel.

5.2.4 Discussion

There are two broad "classes of explanation" for the high level of performance observed in experiment 1, and, in particular, for the finding that four listeners were particularly sensitive to a 0.1-ms delay on either the apical (JH, AE, DD) or basal channel (AP). First, it may be that some neuron(s) central to the auditory nerve receive inputs from fibers innervating different regions of the cochlea, and respond in a way which depends on the relative timing of those inputs. For example, it may be that the response is stronger when the inputs are nearly synchronous than when they are asynchronous, leading to the asynchronous signals sounding "softer" than the reference stimuli. In order to explain the observed asymmetries in performance, this account would have to assume that the tolerance of such a mechanism to deviations from synchrony differed for basal and apical delays, and also differed across listeners. A second class of explanation is that, despite the wide electrode separation, the stimulation on the two channels interacted within a common population of auditory nerve fibers, presumably located between the sites of the two "target" channels. Following the distinction made by Shannon (1983b), this explanation can be sub-divided into two separate forms of interaction. The first occurs with simultaneous stimulation and results from a direct summation of the charge applied to the two electrode channels. This interaction is unlikely to occur for the following reasons:

- The pulses on both channels do not overlap in time. Since the pulsewidth is 40 µs/phase and the delay is 100 µs, there is even a gap of 20 µs between both pulses. However, the nerve membrane acts as a capacitive load, so it is not unlikely that the charge injected by the first current pulse interferes with the charge injected by the second pulse (Cartee *et al.*, 1997; Shepherd and Javel, 1999).

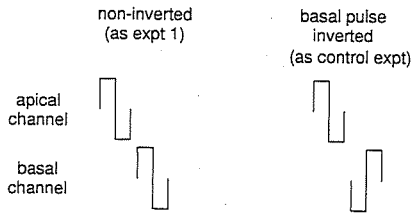


Fig. 5.4 Two nearby pulses on two channels as in the stimuli of experiment 1 (left), or as in the control experiment with the basal pulse inverted (right). A residual charge from the first pulse will interfere differently with the second pulse. If this charge is added in one case, it will be subtracted in the other case. This manipulation might alter the results obtained in experiment 1, if the charge summation hypothesis is true.

- The probability for this effect is substantially lowered by using biphasic pulses. Indeed, if the membrane capacitor is charged by the first phase, it is again discharged by the second phase by the same amount. In the design of the Laura current source, care is taken that the charges of both phases are balanced (Peeters, S., personal communication).

However, a control experiment was still conducted to test the charge summation hypothesis by reversing the polarity of the pulses applied to the basal channel (see figure 5.4). This reverses the sign of the summation of any residual charge from the first pulse with the charge of the second pulse. The three listeners (JH, AE, AP) which performed almost perfect in discriminating between a stimulus with a 0.1-ms delay on the basal channel and one with an identical delay on the apical channel, did this task again with the pulses on the basal channel reversed. The obtained scores were almost identical, the difference varying between 1% and 3%, which is obviously not a significant difference.

The second form of interaction occurs even with non-simultaneous stimulation and results from an electrical "spread of excitation", allowing a neural population to respond to a composite of the temporal patterns applied to the two channels. Experiment 2 controls for both suggested forms of interaction by presenting a train of masking pulses to channels intermediate to the two targets.

5.3 Experiment 2: Interspersed Maskers

5.3.1 Stimuli and procedure

Even though the previous control experiment ruled out explanations in terms of charge summation, it is possible that the stimulation of the two target channels excited a common population of auditory nerve fibers. This hypothesis was tested in experiment 2, by presenting maskers to channels that were intermediate to the two target channels. In order to reduce confusion effects arising from listeners not knowing which channel(s) to listen to, the maskers started 400 ms before the first target stimulus and ended 400 ms after the end of the second in each two-interval forced-choice (2IFC) trial. This would have caused them to form a separate auditory stream from the targets (e.g. Dannenbring and Bregman, 1978), whilst still exciting activity in neural populations intermediate to the two target sites. A further quality difference was provided by presenting the maskers at a rate of 1000 pps rather than the 100 pps used for the signals.

In experiment 2a, the masker was applied to a single channel with a position intermediate to the two targets as measured along the length of the electrode array[16]. This was channel 4 for all listeners except AE, for whom channel 6 was used. Prior to the main experiment, a 400-ms portion of the masker was loudness-balanced twice to each of the two target channels, once starting with a current level which was judged to be quieter than the target and once starting with a level judged to be louder. The masker level was then set to the mean of the loudness matches to the two targets. In the main part of experiment 2a the delay between the pulses of the masker and the target stimuli was not randomized but was such that listeners could not reliably use intervals between masker and target pulses to perform the task.[17]

[16] This way of masking starts from the assumption that the spread of excitation is also along the length of the electrode array.

[17] The shortest time between a masker pulse and a subsequent target pulse on the *non-delayed* channel was always 0.3 ms in the first interval of each trial and 0.4 ms in the second, regardless of whether the stimulus was a reference or a signal. Hence only the interval between the masker pulses and those on the *delayed* channel could possibly provide a cue. In experiment 2a, the delay between the two target channels in the signal interval was 1 ms, and, as this was equal to the masker period, the time between adjacent masker and target pulses was the same as in the reference stimulus.

Four listeners were tested at basal and apical delays of 1 ms. Listener AE was only tested with signals which were delayed on the basal channel by 1 ms. No feedback was provided.

In experiment 2b, the possible spread of excitation in the other direction around the cochlea was neutralized by applying a dual masker using the masker channel of experiment 2a plus a channel basal to the basalmost target. This was channel 8 for all listeners tested. The level of the additional masker was loudness-balanced to the targets in the same way as in experiment 2a, with the level of each masker being set to the mean of the loudness matches to each of the two targets. In this part of the experiment the more basal masker was always advanced by 0.1 ms *re* the apical masker, but the timing between the masker and target pulses was randomized from presentation to presentation. The two electrodes used for each masker channel are connected by dashed lines in the X-ray tracings of figure 5.1. Listener AE could not participate in experiment 2b (with a dual masker) because the use of the basalmost channel as one of the targets meant that we could not select a more basal masker. The alternative method - of using a masker more apical than the apical target (channel 4) - was precluded by the facts that channel 3 overlapped physically with channel 4 (see figure 1.11, bottom), that channel 2 could not be stimulated at a level high enough for its loudness to match that of the targets, and that channel 1 was not functioning at all. The other four listeners were tested with signal delays on the target channel that had given the best performance in experiment 1: this was the basal channel for JH, DD, and SV, and the apical channel for AP. The signal delay used was selected separately for each listener to correspond to one which had led to a high level of performance in experiment 1. This was 2 ms for SV and 1 ms for the other three listeners.

Unlike the case in experiments 1 and 2a, each trial in experiment 2b was followed by correct-answer feedback. This made what we expected to be a taxing task (attending to two channels whilst ignoring two others) more acceptable to the listeners, but meant that a comparison of experiments 1 and 2b would have confounded the effects of the masker and of feedback. Therefore experiment 2b alternated blocks of 25 trials in which the dual-channel masker was present with similar blocks in which it was absent; feedback was provided in both cases. In both parts of experiment 2 all measurements were based on a total of 100 trials for listeners JH, DD, AP and SV, and on 50 trials for listener AE.

5.3.2 Results

Figure 5.5 shows the results for a signal delay of 1 ms in experiment 2a, in which a single masker was presented (solid bars), together with the data from the corresponding condition of experiment 1 (hashed bars). It can be seen that the masker did not have a substantial effect on either the overall level of performance or the asymmetry between conditions with basally- and apically delayed signals. The results obtained with the dual maskers of experiment 2b are shown by the filled bars of figure 5.6, with the data obtained for those same delays in the absence of a masker shown by the open bars. Again, it can be seen that the maskers had only a modest effect overall. In addition, a comparison of the new measures obtained with no masker and those originally measured in experiment 1 (hashed bars) shows that the introduction of feedback had only a minor effect on performance.

The split-half reliability coefficients for the data in figures 5.5 and 5.6 were 0.974 and 0.808 respectively. The coefficient for figure 5.6 was lower than that in the other experiments, but the difference between the means of the first and second half of all responses was small (45.1 vs. 44.6) and not significant (paired samples t-test, $t[11] = 0.458$, $p = 0.656$).

The results of experiment 2 make it unlikely that, in experiment 1, listeners were detecting across-channel timing differences by attending to auditory nerve fibers with characteristic frequencies between those of the regions in the cochlea occupied by the target electrodes. This conclusion rests, however, on the assumption that the masker channels excited auditory nerve fibers with characteristic frequencies between those excited by the targets. This assumption was tested in experiment 3.

5.4 Experiment 3: Pitch judgements

5.4.1 Stimuli and procedure

In experiment 3 listeners compared the pitch of two 400-ms pulse trains presented sequentially to two different channels. The two channels were selected from the masker and target channels used in experiment 2, and all possible combinations were tested. The two channels were presented in random order, separated by a 500-ms pause. The listener was asked to identify the interval containing the higher pitch. No feedback was given. The results are based on 40 judgements per pair for listeners AP, DD, and

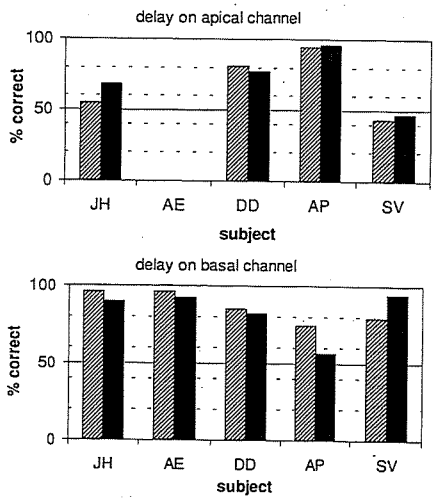


Fig. 5.5 Data from experiment 2a in the absence (hashed bars) or presence (filled bars) of a masker on a single channel. Pulses in the signal were always delayed by 1 ms. Note that although the ordinate is labeled "% correct" in both parts of the figure, any significant deviation from chance performance (50%) indicates that listeners could discriminate between the standard and the signal.

JH, on 20 judgements per pair for SV, and on 50 judgements per pair for AE. Because pulse rate affects pitch (e.g. Eddington *et al.*, 1978; Shannon, 1983a; Pijl and Schwarz, 1995) we presented all stimuli at the same rate. We chose the rate of 1000 pps that was used for the masker channels in experiment 2. The alternative approach, of using a rate of 100 pps for all channels, was not adopted because it was not always possible to generate a sufficiently loud percept on both masker channels at that rate. As the pulses on the target channels had been presented at a rate of 100 pps in previous experiments, the levels on those channels at a rate of 1000 pps had to be re-balanced against the stimuli on the masker channels. To do this, each target channel was adjusted twice to each masker channel: once starting from a quieter level and once from a louder level.

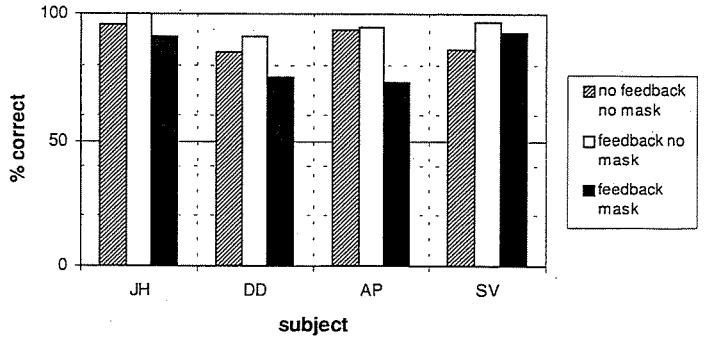


Fig. 5.6 Open and filled bars show performance obtained in the absence and presence of a dual masker in experiment 2b, in which feedback was provided. The pulses in the signal were delayed by 1 ms on the basal channel for listeners JH and DD, by 1 ms on the apical channel for AP, and by 2 ms on the basal channel for SV. The hashed bars show data from the corresponding conditions of experiment 1 in which there was no masker and no feedback.

5.4.2 Results

Table 5.3 shows the percentage of trials on which the more basal channel in each comparison was judged to have the higher pitch, with the data for each listener being shown in a separate matrix. It can be seen that this value exceeded 50% for all listeners and all comparisons, with only one exception: listener DD comparing channels 6 and 8. In all but four of the 27 comparisons listeners scored at least 85%, indicating that the electrodes were not only arranged tonotopically but were easily discriminable. Note also that two out of the four exceptions occurred for listener DD, who, overall, performed worse than the other four listeners at this pitch discrimination task. If this reflected greater channel interaction, and if such interaction were responsible for the ability of some listeners in experiment 1 to detect 0.1-ms timing differences between channels, then one might expect DD to have performed well at that task. Inspection of figure 5.3 shows that, in contrast, he was relatively poor at discriminating between 0.1 ms basal and apical delays.

Table 5.3: Results of experiment 3. Each part of the table shows a matrix indicating the percentage of trials in which one listener identified the more basal channel as having the higher pitch, for various pairs of channels. Channels, which were used as targets in experiments 1 and 2, are shown in bold type in the column and row labels; those used as maskers in experiment 2 are shown in italics.

JH	*4*	7	*8*
2	98	98	100
4		100	93
7			53

AP	*4*	7	*8*
2	85	98	100
4		95	100
7			65

AE	*6*	**8**
4	100	100
6		98

SV	*4*	**6**	*8*
2	95	100	100
4		85	100
6			100

DD	*4*	**6**	*8*
3	60	100	90
4		95	88
6			40

One possible limitation in generalizing from the results of experiment 3 to those of experiment 2 comes from the difference in masker pulse rate between the two experiments (1000 vs. 100 pps). It is possible that the spatial pattern of activation elicited by the masker channels differed between the two experiments, either as a result of the rate difference *per se*, or of the change in current resulting from the loudness-balancing procedure (this difference had a mean value of 2.5 dB, with a range of 0.9 to 5.4 dB across listeners and electrodes). However, it seems unlikely that such differences would have reversed the effective ordering of channels along the neural array, particularly given evidence that for loudness-balanced stimuli the perceptual dimensions corresponding to rate and place are orthogonal (Tong *et al.*, 1983).

5.5 Experiment 4: Discriminating between 2-ms basal and apical delays

5.5.1 Stimuli and procedure

In the introduction it was suggested that the detection of across-channel timing differences might require a reference stimulus in which the pulses on the two channels were either synchronous or very nearly so. However, the results of listener SV in experiment 1 suggest that this may not always be the case. As with the other listeners he showed an asymmetry, being better at discriminating basally delayed than apically delayed signals from the reference stimuli. Unlike that of the other listeners, however, his asymmetry could not be attributed to the different reference stimuli used in the two conditions, as these were indiscriminable (evidenced by his chance performance at the two 0.1-ms delays). Hence, it was expected that he would be able to discriminate between a basal delay of 2 ms (which he had reliably labeled as a "signal" in experiment 1) and an identical apical delay (which he had not). Experiment 4 measured the discrimination between 2-ms apical and basal delays for all listeners, both in the presence and absence of a masker on one (listener AE) or two (JH, DD, AP, SV) channels. These maskers were the same as in experiment 2, with the minor exception that the starting phase of listener AE's single-channel masker was randomized from presentation to presentation. Blocks of 25 trials with and without a masker were alternated until 100 trials of each had been collected for every listener. Feedback was given, with the "correct answer" stimulus being defined as that signal which had been consistently labeled as such in experiment 1 (basal delay for all except AP).

5.5.2 Results

Figure 5.7 shows that in the absence of a masker listeners AE and, as predicted, SV could perform the task reliably. Hence for these two listeners the detection of across-channel timing differences was possible even when the stimulation to the two channels always contained an asynchrony as large as 2 ms. However, this was not the case for the other three listeners. On average, the maskers had at most a small effect.

Experiment 4 is similar to one condition in the study by McKay and McDermott (1996), in that the signal and reference differed only in the order of stimulation on the two channels. They presented amplitude-modulated

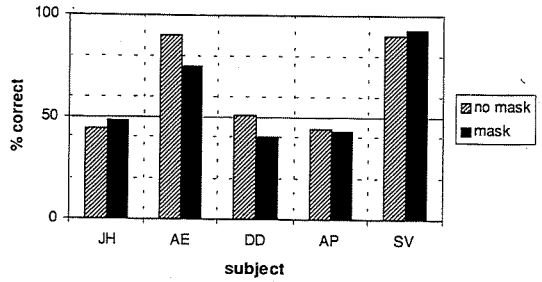


Fig. 5.7 Percentage of trials on which listeners correctly identified the stimulus with a 2-ms apical (listener AP) or basal (other listeners) delay as the signal in experiment 4. Solid bars indicate the condition with a masker present. Hashed bars are for the condition with no masker. Note that, although the ordinate is labeled "% correct", any significant deviation from chance performance (50%) indicates that listeners could discriminate between the standard and signal.

pulse trains to each of two channels, with each train consisting of one large and four small pulses repeated with a period of 10 ms. The "channel reversal" comparison occurred when the delay between the large pulses on the two channels was 9 ms in the signal interval and 1 ms in the reference interval. One of their listeners could perform this task at a channel separation of 3.75 mm when the current applied to the two channels was constant throughout a block of trials, but performance dropped to chance level when the current was randomized from presentation to presentation. As was the case with three of our five listeners, the other four of McKay and McDermott's participants could not do the task even in the absence of current randomization.

5.6 General discussion

5.6.1 Comparison with previous measures of across-channel timing sensitivity

All five listeners in the present study could reliably detect differences between the timing of pulses applied to widely separated channels of their cochlear implants. This differs from the results of previous studies (Tong

and Clark, 1986; McKay and McDermott, 1996), where, except at small (<3mm) channel separations, listeners performed at chance. One possible explanation for the discrepancy, described in the introduction, is that the present study was the only one to include inter-channel delays of less than 1 ms. Consistent with this explanation, the good performance of four listeners was largely due to one stimulus, having an inter-channel delay of 0.1 ms, sounding different from all other stimuli used in the experiment. Also, three out of five listeners could not discriminate between two stimuli with equal and opposite delays of 2.0 ms on the two target channels. However, this does not explain the whole pattern of results, as two listeners (SV and AE) could perform such a comparison, and three could detect differences between pairs of stimuli that did not include the one that sounded "special".

Four further differences between our study and those of Tong and Clark (1986) and of McKay and McDermott (1996) come from the method of stimulation. The most obvious of these is that our listeners were implanted with the Laura device, whereas those participating in the earlier studies were users of the Nucleus-22 implant. If our results were due to interactions at the level of the auditory nerve, then this might be attributed to poorer spatial selectivity in the Laura device. Differences between the devices are hard to evaluate, because extensive psychophysical measurements of the Laura implant's spatial selectivity have not been made, and, even if they had, the large inter-listener differences commonly observed within the same device (e.g. Shannon, 1983b; Chatterjee and Shannon, 1998) would hinder a comparison. However, the ineffectiveness of interspersed maskers in experiment 2, combined with the finding in experiment 3 that the channels were arranged tonotopically, argues against an explanation of our results in terms of poor spatial selectivity. Further arguments against this explanation come from the high discriminability of the two targets in experiment 3, and the fact that the listener showing the worst discrimination between channels in that experiment did not perform well at the across-frequency comparisons in experiment 1.

The second difference in stimulation is that we used a pulse duration of 40 μs/phase, compared to the 100 μs/phase used by Tong and Clark (1986) and for most of McKay and McDermott's (1996) listeners. There is psychophysical evidence that pulse duration affects the slope of the temporal integration function (Pfingst *et al.*, 1991), a finding which may reflect an effect on the temporal pattern of neural activity evoked. It is conceivable that this, in turn, might affect the ability of listeners to compare firing patterns between different populations of auditory nerve fibers. However,

the slope of the integration function is constant for durations between 50 and 500 µs/phase (Pfingst *et al.*, 1991), and so it is unlikely that such differences could account for the discrepancy between our results and those obtained previously with a pulse duration of 100 µs/phase. Alternatively, one could argue that the shorter pulse durations used here necessitated higher current levels, and that this resulted in a reduction of spatial selectivity. This seems unlikely, given not only the arguments against a "spatial selectivity" interpretation described above, but also McKay and McDermott's (1997) experiments on the effect of pulse duration on the detection of across-channel timing differences. They reported that the spatial separation beyond which performance deteriorated was smaller at a pulse duration of 50 or 100 µs than at 266 µs, suggesting that spatial selectivity is best at short pulse durations. This conclusion held even though the pulses of different durations were matched for loudness, so that the current level increased with decreasing pulse duration.

A third difference arises from the fact that, unlike the abrupt gating used in the other two studies, our stimuli were turned on and off with 50-ms ramps. We can think of no reason why this would have resulted in a marked improvement in performance.

A fourth, perhaps more plausible, reason for the difference between our results and those of McKay and McDermott (1996) is that they randomized the current applied to the two channels from presentation to presentation whereas we did not. Our decision not to randomize the current level was based on two considerations. First, if the two channels did excite a common population of auditory nerve fibers, then we would expect their relative timing to affect primarily a percept of "roughness" or pitch (Tong and Clark, 1986), rather than influencing loudness. We therefore chose to control for such cues using interspersed maskers rather than by roving the level. Second, as McKay and McDermott (1996) have pointed out, one would expect a randomization of current level to degrade performance even if the cue being used were not loudness. As it turned out, the label ("zachter") used by some of our subjects to describe the signals suggests that they may have sounded quieter than the reference stimuli, and so loudness randomization could well have caused a substantial degradation in performance. However, we should stress that, although listeners' perceptual descriptions are informative, the crucial issue is whether the different percepts used to perform the task arose from interaction at the level of the auditory nerve, or whether they reflect the operation of more central processes. The results of our experiments point to the latter alternative.

5.6.2 Comparison with sensitivity to timing differences between ears

The results presented here show that the timing of electric pulses is well-preserved in the auditory periphery, and we have argued that mechanisms central to the auditory nerve can process timing differences between channels. If this is so, then, given the extremely fine sensitivity of binaural mechanisms to timing differences between acoustic stimuli presented to opposite ears, we might expect a similar acuity to electric interaural timing differences (ITDs). Evidence against this prediction comes from a study by van Hoesel and Clark (1997), who showed that even when the electrodes in the two ears stimulated similar regions of the cochlea, the threshold ITD was 0.5 ms for one of their two listeners and 2.0 ms for the other. More recently, Wilson (1997) has briefly reported thresholds of approximately 150 µs for one binaurally implanted listener. This latter threshold, although greater than the 10µsec observed for analogous acoustic stimuli in normal listeners (Hafter and Dye, 1983), reveals a sensitivity of the same order as reported here. Unfortunately, the very small number of patients implanted binaurally, combined with the large inter-listener differences commonly observed in implant psychophysics, makes it impossible to decide on a "typical" result for the detection of electric ITDs. However, it is perhaps worth noting that, acoustically, thresholds for detecting a difference in ITD between two sounds is lowest when one of the sounds is diotic, and increases when both stimuli have non-zero ITDs (Yost, 1974). This means that if the latency of the auditory nerve response to electrical stimulation differs between the two ears, then the threshold ITD will be an overestimate of the minimum interaural asynchrony which the listener is capable of detecting. One possible source of such a latency difference would occur if the integrity of auditory nerve dendrites differed between the two ears, leading to differences in the site of activation and hence in latency (for a discussion, see Shepherd and Javel, 1997).

5.6.3 Importance for speech processors

In previous studies it is shown that cochlear implant subjects are sensitive to differences between the timing of pulse stimulation on proximate channels (Tong and Clark, 1986; McKay and McDermott, 1996). Our experiments reveal for the first time that cochlear implant subjects are also sensitive to changes in the delay of presentation on widely separated channels. This means that it is possible to alter the perception of electrical

stimulation by varying this delay. The question now arises how this effect can be used in speech processing schemes. As mentioned in the introduction, changing the order of stimulation in a CIS processor results in a change of the time interval between the pulses on any two channels. In the case of the Laura implant (eight channels, 10000 pps overall rate), this time interval can vary between -0.4 ms and 0.4 ms, which is in the range of detectable delays for all subjects but SV. Moreover, for subjects JH and AE, the stimulus with a 0.1-ms apical delay sounded markedly different from the others. The same is true for subject AP and the stimulus with the 0.1-ms basal delay. Therefore it is not unlikely that a CIS processor with a channel order such that the delay between separated channels is very short (e.g. 1-5-2-6-3-7-4-8) might produce a different percept to them than a CIS processor with a channel order in which these delays are long (e.g. 1-2-3-4-5-6-7-8). However, it can not be concluded from our experiments which channel order would convey the speech information better. Moreover, it is also not clear yet for which information in a speech signal such a fine timing sensitivity is needed. One option is to implement the suggested orders in the subjects' speech processors and to evaluate and to compare their speech reception scores.

5.7 Summary and conclusions

The main goal of this study was to examine whether implantees can perceive a change in the timing differences between two pulse trains presented on widely separated channels. For one out of five subjects, the detectable delay was 1 ms. Three other subjects could almost perfectly distinguish between a 0.1-ms delay on the apical channel and the same delay on the basal channel. This exceeds the sensitivity of normally hearing listeners on a similar task using analogous acoustic stimuli by an order of magnitude. Hence, the data show that the monaural auditory system is capable of a degree of timing sensitivity between auditory channels that can not be shown with acoustic stimulation. By using masking signals on channels intermediate to the target channels, within-channel cues were ruled out. This implies that the information carried by two distinct neural populations had to be compared somewhere more central to the cochlea. The experiments also revealed that across-channel timing differences can be used to alter the perception of electric stimulation. More specifically for speech processing, using a different order of stimulation in a CIS processor might influence the perception of the processed speech signal. However, it can not be concluded from the results which order would be optimal.

Chapter 6

Conclusions and future directions

In this chapter the main conclusions of all studies will be summarized and suggestions for future experiments will be given.

6.1 APEX

For the evaluation of speech processing stategies or the control of any psycho-physical experiment with a cochlear implant user, a tool was needed to create stimuli, to present them to the subject, to collect his/her responses and to determine the sequence of stimuli, depending on the method of the experiment. Therefore, APEX was developed to enable any researcher to program and conduct such an experiment. Figure 2.1 shows the concept of the whole set-up. Stimuli, free of subject dependent parameters, are created off-line in Matlab and stored in files on hard disk or CD ROM. The stimuli and the procedure of the experiment are defined in a simple text file. At present, nine commonly used methods of psychophysics and speech intelligibility tests are implemented in APEX. The PC program controls the experiment, and sends the stimuli to a DSP board. On this board, some subject parameters are included in the stimuli, and the code for the implanted device is generated. The response of the subject is collected through a graphics tablet and a pen, or through the computer mouse. If necessary, the PC program performs some statistics on the responses and writes the results to the input file. The APEX platform has several advantages:

- No limitations are imposed on the implanted device, i.e. all degrees of freedom of the implant can be employed. In the case of the Laura device, a biphasic (40 µs/phase) current pulse with any amplitude and polarity can be generated on any channel every 100 µs.

- The use of Matlab allows for easy stimulus editing. In the case of psychophysical experiments, all commonly used types of stimuli can be created using a limited number of commands and custom made routines. Also, new speech algorithms are easily implemented in Matlab by using routines of Matlab's Signal Processing Toolbox. The alternative of implementing all the stimulus generation in real time on PC or DSP board is very time consuming, more complex and needs a lot of specific programming effort.

- No computer programming or engineering skills are needed for the stimulus editing and the programming of a new experiment. A little experience in Matlab is sufficient.

- The APEX platform is made for experiments with users of the Laura cochlear implant system, but its concept is portable to any other device. Also, acoustical presentation of stimuli is possible.

Until now, many experiments have been conducted with the APEX platform. The concept and the implementation was suitable for all the experiments that were designed for the studies in this thesis (Geurts and Wouters, 1999; Geurts and Wouters, 2000a; Carlyon, Geurts and Wouters, 2000) and for several completed or ongoing studies by other researchers (van Wieringen and Wouters, 1999a; van Wieringen and Wouters, 1999b).

6.2 Mimicking neural adaptation in speech processors for cochlear implants

The neural response to an electric stimulus differs in several aspects from the response to acoustic sounds. One of them is the adaptation effect, which corresponds to the high neural activity at the onset of an acoustic stimulus, and the decay in the response during the stationary part of that stimulus. According to previously described studies, this effect seems to be less prominent in the electric case. A new speech processing algorithm, EECIS, is developed in which the rapid adaptation is mimicked by a substantial increase of the current level each time the intensity at the input suddenly increases. Though this temporary increase only lasts for about one hundredth of a second, considerable changes in the implantee's speech perception are observed. The processed words sounded more intelligible, as if they were better articulated, compared to the default CIS strategy. Significant increases are also seen in speech reception scores. The percentage correctly identified stop consonants in intervocalic /a/-context increased on the average by 9%, the individual results ranging from a 0% to 20% increase. Also more vowels from the monosyllabic word list were correctly identified: the average increase was 7%, with individual increases between 2% and 14%.

A very relevant question that arises from the used method is: how can one be convinced that the effect is correctly mimicked? Unfortunately, there are no means yet to compare the neural response to an acoustic stimulus with the response to the same stimulus processed by the EECIS algorithm. Therefore, in the proposed algorithm, the duration of the added peak is determined by a few parameters, which can be modified in order to match the durations found in preceding physiologic studies. The amplitude of the added peak is determined by the subjective response of the patient, and it is possible that an exaggerated adaptation effect will occur in the resulting neural response, since the highest gain of the peak was chosen for which the speech signal did not sound distorted. Nevertheless, as discussed in chapter 3, one can be assured that at least qualitatively a similar effect is obtained. Figure 6.1 illustrates a method to approach the optimal, quantitatively equal response.

The bottom path represents a model that predicts the neural response of a number of nerve fibers with different characteristic frequencies from a given acoustic signal at the input. Several scientists have developed such auditory pathway models on the basis of physiological facts or measurements (Meddis *et al.*, 1990; Patterson *et al.*, 1995; Delgutte *et al.*, 1997). The purpose of a speech algorithm is to convert the acoustic signal into a series of current waveforms or pulses on the available electrode channels such that the neural response resembles the one obtained from the auditory pathway model. A second model (top right in figure 6.1) is needed that predicts the neural response from an electric signal applied on the electrode channels. Such a model can be based on measurements of the so called electrically evoked whole-nerve action potential (EAP) (Brown *et al.*, 1990; 1996). This potential is generated each time one or more nerve fibers are excited by an electric stimulus, and its amplitude is a measure of the total number of excited nerve fibers. The electrode-nerve fiber interface models are developed recently by several scientists (e.g. Wilson *et al.*, 1994; Van Immerseel *et al.*, 1999). One way to obtain a new speech algorithm is to make a cascade of an auditory pathway model and the inverse model of the electrode-nerve fiber interface. The latter would have to predict the necessary electric stimulus for a given neural response.

The suggested method was not adopted for the following reasons. First, at the time this study commenced, there were no reliable electrode-nerve fiber interface models available. Second, the purpose of this study was to evaluate the perceptual effect of adding adaptation only. Indeed, if an algorithm evolving from the suggested method would be compared with the

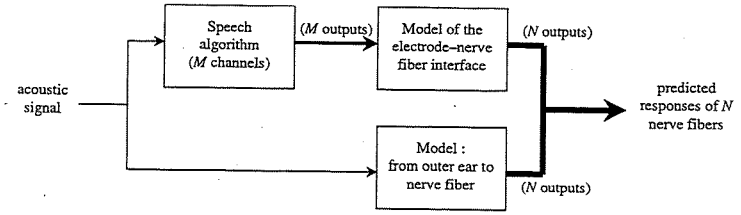


Fig. 6.1 This scheme is a guideline to develop a speech processing algorithm that results in the same neural response as obtained with the acoustical presentation of the speech stimulus.

CIS algorithm, more differences in the electric stimulus patterns would be obtained besides the adaptation effect. Any resulting differences seen in the speech reception scores would then not be due only to the adaptation effect. Third, and more important, the new algorithm had to be simple, requiring only a limited amount of computing power and memory. This way, the algorithm could be implemented in real time on a wearable processor, such as the Lauraflex+. This would allow the implantee to use the new strategy continuously for longer periods (several weeks). In the present study, the data were collected without the subject having the opportunity to get familiar with the EECIS strategy. It is suggested that a new algorithm is significantly disadvantaged when it is compared with an algorithm for which the subject is experienced (Wilson *et al.*, 1993; Tyler *et al.*, 1986). In other words, the obtained increase in scores may be larger after a training period of at least several weeks.

6.3 Temporal coding of the fundamental frequency in cochlear implants

In a CIS processor, the fundamental frequency of a voiced phoneme is coded as a temporal fluctuation in the speech envelope that modulates the pulse trains. The period of this fluctuation corresponds to the inverse of the fundamental frequency. In the presented study, it is investigated how the modulation depth of these fluctuations influences the perception of pitch. In the first experiment, two sinusoidally amplitude modulated pulse trains with different values of the modulation frequency are presented on one channel. The discrimination ability initially grows from chance to a certain level as the modulation depth increases, and in some cases saturates beyond a certain

depth. For some subjects and conditions, the depth at which optimal performance is achieved is larger than the depths obtained with vowels processed by the default CIS algorithm. Therefore, a new algorithm, F0 CIS, is developed in which the modulation depth can be controlled by varying a few parameters. The F0 CIS algorithm is compared with the default CIS algorithm and a third algorithm, FLAT CIS, in which the F0 fluctuations are removed. In the latter case, the subjects have to judge on the basis of relative amplitude differences in each channel. In this second experiment, the smallest detectable difference in fundamental frequency is measured using synthetic vowels. For the FLAT CIS algorithm, the performance varies from chance level, even at the largest F0 differences, to excellent in one condition (vowel /i/, F0 = 250 Hz). Except for the latter condition, performance was always significantly better for the two other algorithms, indicating that the temporal cue is important, if not necessary, for pitch perception. Despite the prediction based on the results of the first experiment, no significant improvements are obtained with the F0 CIS algorithm. This is probably due to the fact that less modulation depth is needed in order to obtain the same performance as for a single channel, if all channels are stimulated with synchronous fluctuations.

Another approach to obtain a suitable neural representation of the fundamental frequency is again the method illustrated in figure 6.1. In other words, an alternative experiment would compare the default CIS algorithm with an algorithm that mimics normal physiology as closely as possible. It is possible though that one of the outcomes of such study would be that for the default CIS algorithm, the nerve response is too tightly phase-locked[18] to the stimulus, and that the modulation depths at the output of the new algorithm are smaller. Indeed, Litvak *et al.* (1999) showed that for modulation depths of 20% - which is within the range of values obtained with the default CIS algorithm - the timing of the neural response is restricted to a very narrow range of phase angles of the modulator of a SAM pulse train. On the other hand, a modulation depth of 2% resulted in a response more like the response observed with normal acoustic stimulation. However, the current experiments illustrate that no decrease in the performance is seen for modulation depths for which a too high degree of phase-locking would be obtained. In other words, more stochasticity in the neural response, as

[18] A response is phase-locked to some stimulus if the period histogram (see chapter 4) shows a clear, single local maximum. "Too tightly phase-locked" means that this local maximum is very narrow and large relative to the rest of the histogram.

observed with normal acoustic stimulation, does not guarantee that performance will increase.

The presented study showed that the temporal coding of the fundamental frequency is beneficial for the perception of the pitch. Likewise, it might be worthy to include a temporal representation of other frequencies than the fundamental in a speech coding algorithm, such as the formant frequencies. Both formants are very important for the recognition of vowels and also the transitions of the formant frequencies that occur between a vowel and an adjacent consonant conveys information about the consonant (van Wieringen and Pols, 1995). In current speech coding algorithms the formants are coded according to the tonotopic organization of the cochlea (see chapter 1). However, physiologic measurements revealed that auditory nerve fibers having a characteristic frequency close to a formant frequency tend to phase-lock to this frequency (Young and Sachs, 1979; Miller and Sachs, 1984). Unfortunately, before formant frequencies can be coded temporally, there are several obstacles to overcome yet. The most important is that many cochlear implant subjects are not sensitive to frequency or rate changes above 300 Hz. This is well below the range of second formant frequencies, and also below many common first formant frequencies. As long as no technique has been found to increase this limit, a temporal coding of this frequency will not improve speech intelligibility. It is not unlikely though that this boundary reflects a limitation in the processing more central to the cochlea. A comparable psycho-acoustical task to pulse rate discrimination is to distinguish between two amplitude modulated noises, differing in modulation frequency, but occupying the same cochlear region. Such experiments revealed an upper frequency limit between 300 Hz and 500 Hz (Burns and Viemeister, 1976; 1981), which is comparable to the one obtained with implantees . A second obstacle is that the selectivity of current electrode channels is not high enough to prevent that neurons respond to two adjacent channels, which may lock on different frequencies.

6.4 Sensitivity to temporal delays between two channels of a cochlear implant

Many of the parameters of the CIS strategy are chosen rather arbitrarily, and little is known about the perceptual effect of changing these. One example is the order of stimulation, which determines the relative timings of stimulation between any two channels. In the last study of this thesis, it was

investigated to which extent a cochlear implant user is sensitive to the relative timing of two pulse trains on two widely separated channels. It was found that the sensitivity of some subjects exceeded the technical limits of the device. In other words, they could still discriminate two stimuli with the smallest possible difference in delays, being 200 µs. This exceeds the sensitivity of normally hearing listeners on a similar task using analogous acoustic stimuli by an order of magnitude. Another interesting finding is the observed asymmetry in the psychometric functions: some subjects performed better with delays on the basal channels, while one subject performed better with apical delays. For some subjects, this was due to the use of different reference stimuli, but again, the only difference between both reference stimuli was the direction - basal or apical - of the delay. One other subject could not discriminate both references, so the asymmetry was also manifested at larger delays.

The experiments demonstrated that one can exploit across-channel timing differences to change the perception of electrical stimulation. However, at the moment, there is no immediate obvious application for this finding for the following reasons. First, there seems to be no information in a speech signal that needs such a fine sensitivity. Second, changing the order of stimulation does not only affect the relative timing of stimulation on two channels. Other effects will occur too, such as the masking of a pulse on a given channel by a pulse on an adjacent channel, which will increase with decreasing time interval between both pulses. Third, more experiments are required in order to provide a better understanding of the observed phenomenon.

The experiments also demonstrated an aspect of electrically stimulated nerve fibers that was already observed with physiological measurements, namely that the response is very tightly time-locked to the stimulus. This means that the observed latency is short and does not show much variability, compared to acoustical stimulation. Due to this difference, the performance of the cochlear implant subjects exceeded that of normal hearing listeners by an order of magnitude. Again, the neural response to electric stimuli is different from the response to acoustic stimuli, but that does not mean that this is a disadvantage, as also shown in the previous study on the coding of the fundamental frequency. On the contrary, due to this time-locking, the time instant of the neural discharging can be controlled very accurately.

6.5 Final conclusions

The presented work consisted of one technical creation and three studies with cochlear implant users. A computer based set-up was developed to conduct psychophysical experiments and to design and evaluate new speech processing algorithm, thereby exploiting all degrees of freedom of the Laura cochlear implant. In the first study, it was investigated how one phenomenon of the physiology of normal hearing, namely rapid neural adaptation, can be included in a CIS-like speech processor and if it provides a higher speech intelligibility. Some phonemes were identified better, and although the increases in the scores were small, they were significant and obtained without training or feedback. The topic of the second study was the coding of the fundamental frequency of voiced phonemes. It was shown that the temporal cue, manifested as fluctuations in the speech envelope, is very important for pitch perception, and that optimal performance is already obtained for modulation depths of around 20 %, provided that many channels are stimulated. In the last study, it was demonstrated for the first time that implantees are sensitive to small timing differences between two pulse trains on widely separated channels. All studies provide a better understanding of cochlear implant physics and hopefully, they bring the world one or more steps closer to the ultimate speech processor for cochlear implants.

References

Blamey, P.J., Dowell, R.C., Tong, Y.C. and Clark, G.M. (1984). "An acoustic model of a multiple-channel cochlear implant, " J.Acoust.Soc.Am. 76, 97-103

Blamey, P.J., Dowell, R.C., Clark, G.M. and Seligman, P.M. (1987). "Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant," J.Acoust.Soc.Am. 82, 38-47

Bosman, A., Smoorenburg, G.F. and Bronkhorst, A.W. (1992). "Relations between phoneme scores and syllable scores for normal-hearing and hearing-impaired subjects," in *The Auditory Processing of Speech, From Sounds to Words*, edited by M.E.H.Schouten (Mouton de Gruyter, Berlin, New York, 1992), 357-362

Brown, C.J., Abbas, P.J. and Gantz, B. (1990). "Electrically evoked whole-nerve action potentials: Data from human cochlear implant users," J.Acoust.Soc.Am. 88, 1385-2210

Brown, C.J., Abbas, P.J., Borland, J. and Bertschy, M.R. (1996). "Electrically evoked whole-nerve action potentials in Ineraid cochlear implant users: Responses to different stimulating electrode configurations and comparsion to psychophysical responses," J. of Speech and Hearing Research 39, 453-467

Burns, E.M. and Viemeister, N.F. (1976). "Nonspectral pitch," J.Acoust.Soc.Am. 60, 863-869

Burns, E.M. and Viemeister, N.F. (1981). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J.Acoust.Soc.Am. 70, 1655-1660

Busby, P.A., Tong, Y.C. and Clark, G.M. (1993). "The perception of temporal modulations by cochlear implant patients," J.Acoust.Soc.Am. 94, 124-131

Byrne, D., Dillon, H., Tran, K, *et al.* (1994). "An international comparison of long-term average speech spectra," J.Acoust.Soc.Am. 96, 2108-2120

Carlyon, R.P. (1994). "Detecting pitch-pulse asynchronies and differences in fundamental frequency," J.Acoust.Soc.Am. 95, 968-979

Carlyon, R.P. and Shackleton, T.M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?," J.Acoust.Soc.Am. 95, 3541-3554

Carlyon, R.P., Geurts, L. and Wouters, J. (2000). "Detection of small across-channel timing differences by cochlear implantees," Hearing Research 141,140-154

Cariani, P.A., and Delgutte B. (1996). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J.Neurophysiol. 76, 1698-1716

Cartee, L.A., van den Honert, C. and Finley, C. (1997). "Refraction and summation properties of cochlear neurons in response to intrameatal electrical stimulation with pulse pairs," in *Abstracts of the 1997 Conference on Implantable Auditory Prostheses, Asilomar, U.S.A.*

Chatterjee, M. and Shannon, R.V. (1998). "Forward masked excitation patterns in multielectrode electrical stimulation," J.Acoust.Soc.Am. 103, 2565-2572

Clark, G.M., Black, R., Forster, I.C., Patrick, J.F., and Tong, Y.C. (1978). "Design criteria of a multiple-electrode cochlear implant hearing prostheses," J.Acoust.Soc.Am. 63, 631-633

Cullen, J.K.Jr. and Long G.R. (1986). "Rate discrimination of high-pass-filtered pulse trains," J.Acoust.Soc.Am. 79, 114-119

Dannenbring, G.L. and Bregman, A.S. (1978). "Streaming vs. fusion of sinusoidal components of complex tones," Percept. Psychophys. 24, 369-376

Delgutte, B. (1980). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J.Acoust.Soc.Am. 68, 843-857

Delgutte, B. (1997). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W.J.Hardcastle and J.Laver (Oxford, Blackwell, 1997), 507-538

Delgutte, B., Hammond, B.M. and Cariani, P.M. (1997). "Neural encoding of the temporal envelope of speech: Relation to modulation transfer functions," in *Psychophysical and Physiological Advances in Hearing: Proceedings of the 11th International Symposium on Hearing, Grantham, U.K.*, edited by A.R.Palmer, A.Rees, A.Q.Summerfield, and R.Meddis (Wurr Publishers, London, 1997), 595-603

Dorman, M., Soli, S., Dankowski, K., Smith, L., McCandless G. and Parkin, J. (1990). "Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant," J.Acoust.Soc.Am. 88, 2074-2079

Dorman, M.F., Loizou, P.C., Fitzke, J., Tu, Z. (1998). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J.Acoust.Soc.Am. 104, 3583-3585

Drullman, R., Festen, J.M. and Plomp, R. (1994). "Effect of temporal envelope smearing on speech perception," J.Acoust.Soc.Am. 95, 1053-1064

Dynes, S.B.C and Delgutte, B. (1992). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea," Hearing Research 58, 79-90

Eddington, D.K., Dobelle, W.H., Brackman, D.E., Mladejovsky, M.G., and Parkin, J.L. (1978). "Auditory prostheses research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol. 87, Suppl. 53, 1-59

Eggermont, J.J. (1985). "Peripheral auditory adaptation and fatigue: A model oriented review," Hearing Research 18, 57-71

Erber, N.P. and Alencewicz, C.M. (1976). "Audiological evaluation of deaf children," J.Speech and Hearing Disorders 41, 256-267

Ferguson, G.A. and Takane, Y. (1989). "Statistical Analysis in Psychology and Education," McGraw-Hill.

Fu, Q.-J. and Shannon, R.V. (1998). "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," J.Acoust.Soc.Am. 104, 2570-2577

Fu, Q.-J., Shannon, R.V. and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," J.Acoust.Soc.Am. 104, 3586-3596

Geurts, L., Wouters, J., Govaerts, G., Peeters, S., Snellings, J. and Van Compernolle, D. (1996). "An experimental set-up for evaluating new speech processing strategies for cochlear implants," Acta Acustica 82, Proceedings of the Forum Acusticum 1996 Antwerpen, S221

Geurts, L. and Wouters, J. (1999). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J.Acoust.Soc.Am. 105, 2476-2484

Geurts, L. and Wouters, J. (2000a). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," submitted to J.Acoust.Soc.Am.

Geurts, L. and Wouters, J. (2000b). "A concept for a research tool for experiments with cochlear implant users", submitted to J.Acoust.Soc.Am.

Govaerts, G. (1978). "Perceptuele structuren van synthetische en natuurlijke klinkers," Psychologica Belgica 18-1, 27-67

Greenwood, D.D. (1990). "A cochlear frequency-position function for several species--29 years later," J.Acoust.Soc.Am. 87, 2592-2605

Hafter, E.F. and Dye, R.H. (1983). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," J.Acoust.Soc.Am. 73, 644-651

Handel, S. (1989). "Listening, an introduction to the perception of auditory events," (The Massachusetts Institute of Technology Press, 1989)

Hartmann, R., Topp, G. and Klinke, R. (1984). "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea," Hearing Research 13, 47-62

Hodgkin, A.L. and Huxley, A.F. (1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve," J.Physiol. 117, 500-544

Holdsworth, J.W. and Patterson, R.D. (1993). "Analysis of waveforms," UK Patent No. GB 2-234-078-B, UK Patent Office, London

House, W.F., and Urban, J. (1973). "Long term results of electrode implantation and electronic stimulation of the cochlea in man," Ann. Otol. 82, 504-517

Houtsma, A.J.M. and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," J.Acoust.Soc.Am. 87, 304-310

Javel, E., Tong, Y.C., Shepherd, R.K. and Clark, G.M. (1987). "Responses of cat auditory-nerve fibers to biphasic electrical current pulses," Ann. Otol. Rhinol. Laryngol. 96, Suppl. 128, 26-30

Jones, P.A., McDermott, H.J., Seligman, P.M. and Millar, J.B. (1995). "Coding of voice-source information in the Nucleus cochlear implant system," Annals of Otology, Rhinology and Laryngology 104, suppl. 166, 363-365

Joris, P.X. and Yin, T.C.T. (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J.Acoust.Soc.Am. 91, 215-232

Kennedy, E., Levitt, H., Neuman, A.C. and Weiss, M. (1998). "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners," J.Acoust.Soc.Am. 103, 1098-1114

Killian, M.J.P., Klis, S.F.L. and Smoorenburg G.F. (1994). "Adaptation in the compound action potential response of the guinea pig VIIIth nerve to electric stimulation," Hearing Research 81, 66-82

Langner, G. (1992). "Periodicity coding in the auditory system," Hearing Research 60, 115-142

Lawson, D.T., Wilson, B.S., Finley C.C., Zerbi, M., Cartee, L.A., Roush, P.A., Farmer, J.C., Tucci, D.L. (1997). "Cochlear implant studies at Research Triangle Institute and Duke University Medical Center," Scandinavian Audiology 26, Suppl. 46, 50-64

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. 49, 467-477.

Litvak, L., Delgutte, B., Eddington, D.E. and Cariani, P.A. (1999). "Auditory-nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains," in *Abstracts of the 1999 Conference on Implantable Auditory Prostheses, Asilomar, U.S.A.*, 62

Loizou, P.C. (1998). "Mimicking the human ear", IEEE Signal Processing Magazine 15, 101-130

Lopez-Poveda, E.A., O'Mard, L.P. and Meddis, R. (1997). "A revised computational inner hair cell model," in *Psychophysical and Physiological Advances in Hearing: Proceedings of the 11th International Symposium on Hearing, Grantham, U.K.*, edited by A.R.Palmer, A.Rees, A.Q.Summerfield, and R.Meddis (Wurr Publishers, London, 1997), 112-121

McDermott, H.J., McKay, C.M. and Vandali, A.E. (1992). "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," J.Acoust.Soc.Am. 91, 3367-3371

McKay, C.M., McDermott, H.J. and Clark, G.M. (1994). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J.Acoust.Soc.Am. 96, 2664-2673

McKay, C.M., McDermott, H.J. and Clark, G.M. (1995). "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth," J.Acoust.Soc.Am. 97, 1777-1784

McKay, C.M. and McDermott, H.J. (1996). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J.Acoust.Soc.Am. 100, 1081-1092

McKay, C.M. and McDermott, H.J. (1997). "Perceptual effects of interpulse interval and pulse duration at suprathreshold levels," 1997 Conference on Implantable Auditory Prostheses, Asilomar, CA.

Meddis, R. (1986). "Simulation of mechanical to neural transduction in the auditory receptor," J.Acoust.Soc.Am. 79, 702-711

Meddis, R. (1988). "Simulation of auditory-neural transduction: Further Studies," J.Acoust.Soc.Am. 83, 1056-1063

Meddis, R., Hewitt, M.J. and Shackleton, T.M. (1990). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," J.Acoust.Soc.Am. 87, 1813-1816

Miller, G.A. and Nicely, P.E. (1955). "An analysis of perceptual confusions among some English consonants," J.Acoust.Soc.Am. 27, 338-352

Miller, M.I and Sachs, M.B. (1984). "Representation of voice pitch in discharge patterns of auditory-nerve fibers," Hearing Research 14, 257-279

Moore, B.C.J. (1989). "An introduction to the psychology of hearing," 3rd edition (Academic Press, 1989).

Palmer, A.R. (1990). "The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers," J.Acoust.Soc.Am. 88, 1412-1426

Parkins, C.W. (1989). "Temporal response patterns of auditory nerve fibers to electrical stimulation in deafened squirrel monkeys," Hearing Research 41, 137-168

Patterson, R.D., Allerhand, M.H. and Giguère, C. (1995). "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," J.Acoust.Soc.Am. 98, 1890-1894

Peeters, S., Offeciers, F.E., Kinsbergen, J. , Van Durme, M., Van Enis, P., Dykmans, Ph. and Bouchataoui, I. (1993). "A digital speech processor and various encoding strategies for cochlear implants," Progress in Brain Research 97, 283-291

Pfingst, B.E., DeHaan, D.R. and Holloway, L.A. (1991). "Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. I:Phase duration and stimulus duration," J.Acoust.Soc.Am. 90, 1857-1866

Pickles, J.O. (1988). "An introduction to the physiology of hearing," 2nd edition (Academic Press, 1988).

Pijl, S. and Schwarz, D.W.F. (1995). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J.Acoust.Soc.Am. 98, 886-895

Plomp, R. and Mimpen, A.M. (1979). "Improving the reliability of testing the speech reception threshold for sentences, " Audiology 18(1), 43-52.

Ranck, J.B. (1981). "Extracellular stimulation," in *Electrical Stimulation Research Techniques*, edited by M.Patterson and R.P.Kesner (Academic Press, 1981), 1-36

Relkin, E.M. and Doucet, J.R. (1997). "Is loudness simply proportional to the auditory nerve spike count?" J.Acoust.Soc.Am. 101, 2735-2740

Repp, B.H. and Lin, H. (1989). "Acoustic properties and perception of stop consonant release transients," J.Acoust.Soc.Am. 85, 379-396

Rose, J.E., Brugge, J.F., Anderson D.J. and Hind, J.E. (1967). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J.Neurophysiol. 30, 769-793

Sellick, P.M., Patuzzi, R. and Johnstone, B.M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," J.Acoust.Soc.Am. 72, 131-141

Shannon, R. V. (1983a). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hearing Research 11, 157-189.

Shannon, R. V. (1983b). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hearing Research 12, 1-16.

Shannon, R.V. (1992). "Temporal modulation transfer function in patients with cochlear implants," J.Acoust.Soc.Am. 91, 2156-2164

Shannon, R.V., Zeng, F., Kamath, V., Wygonski, J. and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 270, 303-304

Shepherd, R.K. and Javel, E. (1997). "Electric stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status," Hearing Research 108, 112-144

Shepherd, R.K. and Javel, E. (1999). "Electrical stimulation of the auditory nerve: II. Effect of stimulus waveshape on single fibre response properties," Hearing Research 130, 171-188

Simmons, F., White, R., Walker, M. and Mathews, R. (1981). "Pitch correlates of direct auditory nerve stimulations," Ann.Otol.Rhinol.Laryngol. 90, 15-18

Skinner, M.W., Holden, L.K., Holden, T.A., Dowell, R.C., Seligman, P.M., Brimacombe, J.A. and Beiter, A.L. (1991). "Performance of postlingually deaf adults with the Wearable Speech Processor (WSP III) and mini speech processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear and hearing 12, 3-22

Skinner, M.W., Holden, L.K., Holden, T.A. and Demorest, M.E. (1995). "Comparison of procedures for obtaining thresholds and maximum acceptable loudness levels with the Nucleus cochlear implant system, " J. of Speech and Hearing Research 38, 677-689

Smith, R.L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," J.Acoust.Soc.Am. 65, 166-178

Stevens, S.S. (1951). "Mathematics, measurement and psychophysics," in *Handbook of Experimental Psychology*, edited by S.S. Stevens (Wiley, New York), pp. 1-49

Strickland, E., Viemeister, N.F., Fantini, D.A. and Garrison, M.A. (1989). "Within-versus cross-channel mechanisms in detection of envelope phase disparity," J.Acoust.Soc.Am. 86, 2160-2166

Tong, Y.C., Blamey, P.J., Dowell, R.C. and Clark, G.M. (1983). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant," J.Acoust.Soc.Am. 74, 73-80

Tong, Y.C. and Clark, G.M. (1985). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," J.Acoust.Soc.Am. 77, 1881-1888

Tong, Y.C. and Clark, G.M. (1986). "Loudness summation, masking, and temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea," J.Acoust.Soc.Am. 79, 1958-1966

Townshend, B., Cotter, N., Van Compernolle, D. and White, R.L. (1987). "Pitch perception by cochlear implant subjects," J.Acoust.Soc.Am. 82, 106-115

Tyler, R.S., Preece, J.P., Lansing, C.R., Otto, S.R. and Gantz, B.J. (1986). "Previous experience as a confounding factor in comparing cochlear-implant schemes," J. of Speech and Hearing Research 29, 282-287

Tyler, R.S. and Moore B.C.J. (1992). "Consonant recognition by some of the better cochlear implant patients," J.Acoust.Soc.Am. 92, 3068-3077

van den Honert, C. and Stypulkowski, P.H. (1987). "Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli," Hearing Research 29, 207-222

Van der Borght, G. (1994). "Ontwikkeling van een spraakprocessor ter aansturing van een cochleair implant," Master's thesis, Katholieke Industriële Hogeschool Antwerpen, 1994.

van Dijk, J.E., van Olphen, A.F., Langereis, M.C., Mens, L.H., Brokx, J.P. and Smoorenburg, G.F. (1999). "Predictors of cochlear implnat performance," Audiology 38, 109-116

van Hoesel, R.J.M. and Clark, G.M. (1997). "Psychophysical studies with two binaural cochlear implant subjects," J.Acoust.Soc.Am. 102, 495-507

Van Immerseel, L., Houben, V. and Peeters, S. (1999). "Models of auditory nerve response to electrical stimulation in cochlear implants as an aid to optimise information transfer to the neural system," in *Abstracts of the 1999 Conference on Implantable Auditory Prostheses, Asilomar, U.S.A.*, 74

van Noorden, L.P.A.S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology.

van Wieringen, A. and Pols, L.C.W. (1995). "Discrimination of single and complex consonant-vowel- and vowel-consonant-like formant transitions," J.Acoust.Soc.Am. 98, 1304-1312

van Wieringen, A. and Wouters, J. (1999a). "Natural vowel and consonant recognition by Laura cochlear implantees," Ear and Hearing 20, 89-103

van Wieringen, A. and Wouters, J. (1999b). "Gap detection in single- and multiple-channel stimuli by Laura cochlear implantees," J.Acoust.Soc.Am. 106, 1925-1939

Volta, A. (1800). "On the electricity excited by the mere contact of conducting substances of different kinds. Phil.Trans.Roy.Soc.London 90, 403-431

von Békésy, G. (1960). "Experiments in hearing," McGraw-Hill, New York

Westerman, L.A. and Smith, R.L. (1984). "Rapid and short-term adaptation in auditory nerve responses," Hearing Research 15, 249-260

Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K and Rabinowitz W.M. (1991). "Better speech recognition with cochlear implants," Nature 352, 236-238

Wilson, B.S., Lawson, D.T., Finley, C.C. and Wolford, R.D. (1993). "Importance of patient and processor variables in determining outcomes with cochlear implants," J. of Speech and Hearing Research 36, 373-379

Wilson, B.S., Finley C.C., Zerbi, M. and Lawson, D.T. (1994). "Speech processors for auditory prostheses," Seventh Quarterly Progress Report, NIH project N01-DC-2-2401. Neural Prosthesis Program, National Institutes of Health, Bethesda, MD.

Wilson, B.S., Lawson, D.T., Zerbi, M. and Finley C.C. (1995). "Speech processors for auditory prostheses," Final Report, NIH project N01-DC-2-2401. Neural Prosthesis Program, National Institutes of Health, Bethesda, MD.

Wilson, B.S. (1997). "The future of cochlear implants," British Journal of Audiology 31, 205-225

Wouters, J., Damman, W. and Bosman, A.J. (1994). "Vlaamse opname van woordenlijsten voor spraakaudiometrie", Logopedie 7(6), 28-33.

Yost, W.A. (1974). "Discrimination of interaural phase differences," J.Acoust.Soc.Am. 55, 1299-1303

Yost, W.A. and Sheft, S. (1989). "Across-critical-band processing of amplitude-modulated tones," J.Acoust.Soc.Am. 85, 848-857

Young, E.D. and Sachs, M.B. (1979). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of audiotory-nerve fibers," J.Acoust.Soc.Am. 66, 1381-1403

Appendix

Example of stimulus editing with the custom-made Matlab functions

The first step in the synthesis of a new stimulus is to create an empty matrix with twice as many columns as there are channels, and as many rows as ten times the length of the stimulus in milliseconds. The following command generates the matrix for a 100 ms long stimulus, containing pulses on two channels.

```
x=zeros(100*10,2*2);
```

The second step is to generate a pulse train on one or more channels. Amplitude, channel, mode, inter-pulse interval (in ms) and time instance of the first pulse (in ms) can be specified with the function *train.m*. The first command below creates a pulse train on channel 1, inter-pulse interval is 0.8 ms, amplitude is 100, mode is 1 (40 µs/phase, apical electrode cathodic first) and the first pulse starts at 0 ms. The second command creates an identical pulse train on channel 2, starting with a delay of 0.4 ms.

```
x=train(x,1,0.8,100,1,0);
x=train(x,2,0.8,100,1,0.4);
```

The pulse trains are amplitude modulated with the following commands. First a sinusoidal amplitude modulation is applied with a modulation frequency of 100 Hz (period = 10 ms) and a modulation depth of 50 %. The sine wave starts at 0 degrees at time instant 0 ms on channel 1, and at time instant 5 ms on channel 2. The third command applies a window with linear ramps having a rise/fall time of 20 ms.

```
x=samtrain(x,1,10,0.5,0);
x=samtrain(x,2,10,0.5,5);
x=gate(x,20);
```

The function *show.m* displays the stimulus on a time scale for each channel. Figure 2.3. in chapter 2 displays the signal x created with the previous commands.

Finally, the following command compresses stimulus x, with channel 1 having the highest priority. There is only one mapping, which has to start at time instant 0 ms. The stimulus will be saved in the file "example.stm":

```
col=compress(x,[1 2],[0],'example.stm')
```

This function will give the following output, meaning that there are 125 pulses in each channel (length is 100 ms and inter-pulse interval is 0.8 ms), and that no pulses overlap in time.

```
col=
   125     0
     0   125
```

List of publications

Publications in international journals

Geurts, L. and Wouters, J. (1999). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J.Acoust.Soc.Am. 105, 2476-2484

Carlyon, R.P., Geurts, L. and Wouters, J. (2000). "Detection of small across-channel timing differences by cochlear implantees," Hearing Research 141,140-154

Geurts, L. and Wouters, J. (2000a). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," submitted to J.Acoust.Soc.Am.

Geurts, L. and Wouters, J. (2000b). "A concept for a research tool for experiments with cochlear implant users", submitted to J.Acoust.Soc.Am.

Publications in national journals

Wouters, J., Geurts, L., Peeters, S., Vanden Berghe, J., and van Wieringen, A. (1998). "Developments in speech processing for cochlear implants," Acta Oto-Rhino-Laryngologica Belg. 52, 129-132.

Articles in proceedings of international meetings

Geurts, L., Govaerts, G., Peeters, S., Van Immerseel, L. van Wieringen, A. and Wouters, J. (1996). "Speech processing for cochlear implants: the implementation and evaluation of phase-locking in high rate coding strategies," Proceedings of the Workshop on the Auditory Basis of Speech Perception 1996 Keele University (UK), 265-268

Contributions to conferences

Geurts, L., Wouters, J., Govaerts, G., Peeters, S., Snellings, J. and Van Compernolle, D. (1996). "An experimental set-up for evaluating new speech processing strategies for cochlear implants," Acta Acustica 82, Proceedings of the Forum Acusticum 1996 Antwerpen, S221

Geurts, L., Wouters, J. and Carlyon, R. (1998). "Speech envelope modification for CIS: speech perception results with the LAURA cochlear implant," presented at the Annual Meeting of the Belgian Society of Audiology 1998, Brussels

Geurts, L. and Wouters, J. (1998). "Speech envelope modification for CIS: speech perception results with the LAURA cochlear implant," presented at the 4th European Symposium on Paediatric Cochlear Implantation 1998 's-Hertogenbosch (NL), 67

Geurts, L. and Wouters, J. (1998). "Speech envelope modification for CIS: speech perception results with the LAURA cochlear implant," presented at the Short Papers Meeting of the British Society of Audiology 1998, London (UK)

Geurts, L., Wouters, J. and Carlyon, R. (1999). "Detection of small across-channel timing differences by cochlear implantees," presented at the Annual Meeting of the Belgian Society of Audiology 1999, Affligem

Geurts, L. and Wouters, J. (1999). "F0 coding in CIS processors: effect of modulation depth of F0 beats on pitch discrimination," in Abstracts of the 1999 Conference on Implantable Auditory Prostheses, Asilomar, U.S.A.

Patents

Geurts, L., Wouters, J. and Peeters, S. (1998). "Enhanced Envelope C.I.S.", Patent No. PHN 16.966 EP-P, *filed* 1/6/1998.

Curriculum vitae

Luc Geurts was born in Genk (Belgium) on the 23rd of March 1970. In 1988 he graduated from the Heilig Hart College in Lanaken. Between 1988 and 1993 he studied Applied Sciences at the University of Leuven (KULeuven), where he received his degree of Masters in Electrical Engineering. In September 1993, he started teaching electronics at the Catholic Institute for Industrial Sciences and Technology of Limburg. From 1 July 1994 on, he combined this job during 2.5 years with a research position as a Ph.D. student at the Laboratory for Experimental ORL of the KULeuven. Since January 1997, he worked full-time on his project "Sound processing for cochlear implants: coding of temporal information".