



Horen

EEN MYSTERIE

Reinier Plomp

Horen

EEN MYSTERIE

[binnenkant omslag]

Reinier Plomp
2020

Inhoudsopgave

Voorwoord	5	4 De perceptie van spraak	56
1 Inleiding	8	<i>Op zoek naar spraakeenheden</i>	
Micro tegenover macro	10	De droom van zichtbare spraak	57
Perifere en centrale processen	13	Fonemen als bouwstenen?	59
'Schone' tegenover 'vervulde' geluiden	15	Coarticulatie als natuurlijk proces	61
Horen als mysterie	16	Conclusies	63
2 De waarneming van een enkele toon	20	5 De perceptie van spraak	66
Het oor als een frequentie-analysator	20	<i>De verstaanbaarheid van lopende spraak</i>	
De raadselachtige toonhoogte	22	De variaties van het spraaksignaal	66
De multidimensionele klankkleur	26	Spectrale en temporele versmering:	
De stabiele luidheid	28	de rol van coarticulatie	69
Nagalm	29	Van fonemen naar woorden: de rol van de context	71
Conclusies	31	Van woorden naar zinnen: de rol van redundantie	75
3 De waarneming van gelijktijdige tonen	34	Invloed van storende geluiden bij normaal- en	
Het continu horen van een door ruis onderbroken toon	34	slechthorenden	78
Het scheiden van gelijktijdige geluiden	40	Conclusies	81
Consonantie en dissonantie	43	6 Twee tijdsbeelden	84
Horen met twee oren	50	Delft in 1946	84
Conclusies	53	Amerika en Europa	86
		Noten	97

Horen, een mysterie

is een uitgave van

NVA, Nederlandse Vereniging voor Audiologie

Auteur

Reinier Plomp

Redactie

Els de Jong

Jeroen Taalman

Vormgeving

Paul Roos, Delft

Druk

Wilco, Amersfoort

©2021 NVA

Voorwoord

Na de Tweede Wereldoorlog is het wetenschappelijk onderzoek van gehoor en spraak pas goed op gang gekomen, vooral in de Verenigde Staten. Ik acht mij gelukkig deze onstuimige ontwikkeling van klein begin tot volwassen wetenschap te hebben meegemaakt. In dit geschrift heb ik in het kort een indruk willen geven van wat ik in die jaren als de essentie van het hoorproces ben gaan beschouwen, kortom: ‘de wetten van het horen’. Dus géén algemeen overzicht, maar een persoonlijke visie. Daarbij heb ik mijn aandacht vooral gericht op de mysterieuze aspecten van wat de ‘black box’ presteert vóór geluid tot een gewaarwording komt, wordt geïdentificeerd en herkend.

Aan dit verhaal heb ik twee tijdsbeelden toegevoegd die de betreffende periode typeren. Het eerste schildert Delft met zijn Technische Hogeschool zoals ik deze in 1946 aantrof in een toestand die herinnerde aan de dertiger jaren. Het tweede handelt over het gevarieerde gezelschap van voornamelijk Amerikaanse gehoor- en spraakonderzoekers waarin ik terecht kwam en zo heb leren waarderen.

Reinier Plomp,

2020

1

Inleiding

1 Inleiding

In ons dagelijks leven worden we vrijwel permanent omringd door geluiden. Deze zijn van een enorme verscheidenheid, variërend van de menselijke stem tot de vele geluiden die ons informatie verschaffen over onze omgeving, in huis, op straat, in auto, bus of trein.

We stuiten hier al meteen op de vraag wat we bedoelen met het woord ‘geluid’. Met deze term kunnen de trillingen worden aangeduid die via de lucht het oor treffen. Als de frequenties van de trillingen buiten het hoorbare gebied vallen, spreekt men van infrageluid bij zeer lage frequenties, van supra- of ultrasoon geluid bij zeer hoge frequenties. Geluid is hier dus een meetbaar fysisch verschijnsel.

Maar hiermee is niet alles gezegd. Men kan ook iets anders bedoelen: niet wat het oor ingaat, maar wat we gewaarworden. We spreken dan van de *perceptie* van geluiden, van hoe ze worden gehoord. Tegenover het *objectieve* van het binnenkomende geluid staat het *subjectieve* van de gewaarwording. De klinkers van de spraak (a, o, ie, enz.) onderscheiden zich *fysisch* in hun spectrale samenstelling, terwijl ze *perceptief* als verschillend worden gehoord. Terwijl we het fysische geluid kunnen vastleggen in meetbare grootheden als frequentie en amplitude, missen we vergelijkbare termen bij geluiden zoals ze worden waargenomen. We moeten het dan doen met omschrijvingen als hoog en laag, luid en zacht, schril en dof, zonder over exacte cijfers te beschikken.

Dit dubbelzinnige karakter van het begrip ‘geluid’ hoeft geen probleem te zijn als we het onderscheid maar zorgvuldig in de gaten houden. Hier zal het in beide betekenissen worden gebruikt. In feite zal het juist gáán om de relatie tussen het

binnenkomende fysische geluid en het in het bewustzijn ervaren geluid. Het hoorproces zal als een ‘black box’ worden behandeld, waarvan we alle aandacht geven aan het verband tussen ‘ingang’ en ‘uitgang’ maar ons minder bekommeren om wat er fysisch, (neuro)fysiologisch en cognitief allemaal in het zwarte kastje gebeurt.

Deze opzet wijkt bewust af van wat inleidende leerboeken plegen te bieden. Zij geven juist alle aandacht aan wat er *in* het kastje gebeurt, hoe fysische en fysiologische processen in het perifere gehoororgaan de trillingen omzetten in elektrische zenuwpulsen die via de gehoorbaan naar de hersenen worden geleid. Dankzij de moderne digitale technieken is er in het laboratorium meer belangstelling gekomen voor hoe het brein met de ontvangen informatie omgaat. Experimenten hebben aangetoond dat deze centrale verwerking veel méér inhoudt dan een ‘passieve’ omzetting van electrofysiologische signalen in gewaarwordingen. Die verwerking is niet alleen passief maar blijkt zich ook ‘actief’ te gedragen, waarbij niet altijd precies dat wordt gehoord wat aan het oor werd aangeboden. Dat maakt het hele proces een stuk ingewikkelder. We zullen hiervan vele frappante voorbeelden tegenkomen.

In het voorgaande heb ik mij beperkt tot het door een enkele bron geproduceerde geluid. In onze lawaaiige samenleving komen we die situatie echter nog maar zelden tegen. Gewoonlijk hebben we niet met een enkele bron te maken, maar horen wij allerlei geluiden gelijktijdig en dwars door elkaar heen. Wolfgang Metzger publiceerde in 1953 een boeiend boek over visuele waarnemingen onder de titel: *Gesetze des Sehens*, dat voor mij een grote inspiratiebron geweest is. Zijn betoog over het zien even onderbrekend, schrijft hij (verkort):

“Wat het oor presteert, is eenvoudig fabelachtig. Terwijl ik hier schrijf, slaat mijn oudste zoon met een pook tegen de kachel, de jongste maakt tevreden geluidjes in zijn kinderwagen, de kerkklok slaat, voor het huis stopt een auto, een van de meisjes speelt in een andere kamer piano, haar moeder spreekt aan de huisdeur met een bode, en daarbij kan ik nog het krassen van mijn potlood op het papier horen. In de luchttrillingen die mijn oor treffen, is dat alles tot een enkel, hoogst complex geheel gesuperponeerd. Ongetwijfeld presteert het oor hier meer dan het oog. Waarom beperken de psychologen zich zo eigenzinnig tot onderzoek van het zien?”

Metzger had gelijk. Zoals we nader zullen ontdekken, was er in zijn tijd nog nauwelijks enige aandacht voor de wijze waarop gelijktijdige geluiden als gescheiden kunnen worden gehoord; in feite werd het probleem in het geheel niet beseft. Juist hieraan zal ruime aandacht worden gegeven.

Teneinde het betoog in de volgende hoofdstukjes niet door nadere uiteenzettingen te onderbreken, zal hier nog op een tweetal algemene aspecten worden ingegaan.

Micro tegenover macro

Wetenschappelijk onderzoek van het horen en de spraak is in de tweede helft van de negentiende eeuw serieus op gang gekomen. Eén naam moet met ere worden genoemd: Hermann von Helmholtz. Geen ander heeft zich wetenschappelijk zo veelzijdig ontplooid. In 1842 gepromoveerd op een fysiologisch proefschrift, schreef hij vijf jaar later een baanbrekende verhandeling over het fundamentele vraagstuk van het behoud van energie. Daarnaast besteedde hij jaren aan de bestudering van het zien en

het horen. Het hier voor ons relevante resultaat was zijn in 1863 verschenen magistrale verhandeling *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Al spoedig in het Engels vertaald, kan dit boek als de basis van het verdere onderzoek worden beschouwd. Toen ik in 1953 in het onderzoek van spraak en gehoor dook, vormde dit bijna een eeuw oude werk de basis, aangevuld met het in 1938 verschenen *Hearing - Psychology and Physiology* van de psycholoog Stanly Smith Stevens (1906-1973) en de fysioloog Hallowell Davis (1896-1992).

Terwijl reeds lang bekend was dat in het door trillende snaren geproduceerde geluid deeltrillingen kunnen worden onderscheiden, werd voor het eerst door Jean-Baptiste Joseph Fourier (1768-1830) wiskundig bewezen dat elke willekeurige periodieke trillingsvorm beschreven kan worden als de som van een reeks van sinussen waarvan de frequenties veelvoudig zijn van de grondfrequentie van de trilling. Dit is de beroemde fourieranalyse die zo'n grote betekenis in de natuurkunde heeft gekregen. Het is niet vreemd dat Helmholtz de sinusvormige trilling als bouwsteen van het geluid beschouwde en zijn beschouwingen daarop baseerde. Voor hem bestond elke gehoorde toon uit een som van sinusvormige deeltonen met frequentieverhoudingen 1:2:3:4 enz., de harmonischen, waarvan de laagste de toonhoogte bepaalt en de relatieve sterkten de klankkleur (timbre). Terwijl Helmholtz in zijn boek nog ruime aandacht aan deze meervoud van deeltonen besteedde, beperkten Stevens en Davis zich vrijwel geheel tot het juist hoorbare verschil in toonhoogte en luidheid en verdere basale eigenschappen van de sinusvormige trilling. Voor hen vormde de sinus de bouwsteen van het horen en kon het horen van meer complexe geluiden hierop worden teruggebracht. In deze overtuiging werden zij door vele onder-

zoekers na hen gevolgd. Hiermede sloot men perfect aan bij het wetenschappelijke ideaal elk probleem te ontleden in zijn elementen en deze afzonderlijk te bestuderen. Dit ideaal drukte ik uit met het woord ‘micro’.

Het is echter niet moeilijk met een eenvoudig voorbeeld de eenzijdigheid van deze aanpak aan te tonen. De bekende Duitse fysicus Max Planck (1858-1947) lichtte in 1909 in lezingen aan de Columbia University te New York het fundamentele verschil toe tussen de Eerste en Tweede Wet van de thermodynamica. Stel je twee waarnemers voor, waarbij de eerste als microscopische waarnemer de bewegingen van een enkele molecuul in een vat bestudeert. Hij concludeert dat deze bewegingen de wet van behoud van energie gehoorzamen, bekend als de Eerste Wet van de thermodynamica. De tweede waarnemer, de macroscopische, gaat daarentegen na hoe de verzameling moleculen zich als gehéél gedraagt. Hij ontdekt dat de verzameling naar maximale entropie (complete wanorde) streeft, de Tweede Wet van de thermodynamica.

Het punt van Planck was dat de microscopische waarnemer die Tweede Wet nooit zal ontdekken omdat het een statistische wet is, geldend voor een verzameling, maar niet relevant voor het individuele molecuul. Het voorbeeld illustreert dat, zelfs binnen de enkele discipline van de fysica een op een subsysteem gericht onderzoek tekortschiet in het verklaren van het systeem als geheel. Vandaar het woord ‘macro’. Tekenend is Plancks conclusie: “Het lijkt me dat zich een vergelijkbare moeilijkheid aanbiedt in de meeste problemen van het intellectuele leven.”

Inderdaad, als zowel microscopische als macroscopische waarnemers nodig zijn om een natuurkundig verschijnsel adequaat te beschrijven, dan kunnen we zeker zijn dat dit nog veel sterker

geldt voor de hoogst complexe processen van de perceptie. De psycholoog Howard Gardner maakt in zijn boek *The mind's new science - A history of the cognitive revolution* (1985) onderscheid tussen “moleculaire of kleinschalige” en “molaire of grootschalige” analyse-eenheden, hierbij met Planck vergelijkbare termen gebruikend. Hij wijst er op dat Helmholtz en de eveneens beroemde psycholoog Wilhelm Wundt (1832-1920) al bewust waren van het belang van beide benaderingen. De in de titel van zijn boek genoemde “revolutie” heeft zijns inziens te eenzijdig de traditionele moleculaire benadering gevolgd en daarbij de nieuwe ontwikkeling van de molaire Gestaltpsychologie veronachtzaamd.

Gardner's kritiek is terecht. Ook bij de bestudering van de zintuigen hebben microscopische onderzoekers de hoofdrol gespeeld. Zij waren (en zijn) ervan overtuigd het huis van de perceptie te kunnen bouwen door het op elkaar metselen van stenen, één voor één. De op sinusvormige trillingen gebaseerde tonen zijn hiervan een voorbeeld. Maar het gaat nog veel verder. Nog lang had de waarneming van een reeks op elkaar volgende tonen of de samenklank van gelijktijdige tonen nauwelijks aandacht. En ook nu nog wordt het verstaan van spraak door velen gezien als de optelsom van de perceptie van individuele, elkaar opvolgende klanken: de fonemen, vergelijkbaar met de letters van het alfabet.

Perifere en centrale processen

Leerboeken van en overzichten over de audiologie, het vakgebied van het gehoor, geven de indruk dat het horen geheel of vrijwel geheel bepaald wordt door processen van het perifere gehoororgaan. Maar dat is te beperkt. Dat we achtereenvolgende tonen van verschillende frequenties in toonhoogte onderschei-

den, kan men inderdaad als een prestatie van de cochlea ('slakkenhuis') beschouwen, maar als men in die reeks een melodie herkent, is er veel méér aan de hand. Dat geldt in nog veel sterkere mate als een stroom van spraakklanken gehoord worden als een betekenisvolle zin. Na de *auditie*, als we ter onderscheiding het proces van het orgaan zo mogen aanduiden, volgt de *cognitie* als de mentale activiteit die de informatie verwerkt en interpreteert op basis van in het geheugen opgeslagen vroegere waarnemingen.

Vanaf zijn eerste levensdag benut de mens de voortdurende stroom van binnenkomende akoestische ervaringen om zijn of haar kennis van de omringende wereld uit te breiden. Wat eerst als een willekeurig geluid wordt gepercipieerd, blijkt na een poosje in combinatie met het zien vertrouwd te raken als het spreken van een persoon of het klappen van een deur. Zoals we vooral bij de spraak zullen zien, is de betrouwbaarheid van die interpretatie van vele factoren afhankelijk. Om een extreem voorbeeld te noemen: als korte toonstoten afgewisseld worden met stootjes ruis, zijn we vast overtuigd een continue toon te horen, los van de ruisstoten. Of als twee personen tegelijk spreken, merken we niet dat de stemmen elkaar behoorlijk verminken. Kortom, het geheugen is niet passief alsof een naslagboekje wordt geraadpleegd, maar is actief bezig het gehoorde te duiden, al of niet correct. Wat we horen hangt mede af van wat we verwachten te horen. We zullen nog heel wat frappante voorbeelden tegenkomen.

Het is begrijpelijk dat het *horen* van geluiden en het *duiden* van geluiden als behorend tot verschillende disciplines elkaar wetenschappelijk moeilijk vonden. Stevens besteedde in het samen met Davis geschreven boek vooral veel aandacht aan de luidheid zonder zich over de rol van de herkenning te bekommeren. De

perceptie van spraak komt in het gehele boek niet voor. Sinds 1950 is hieraan wel enige verbetering gekomen, maar van werkelijke integratie is nog geen sprake.

In dit geschrift heb ik alle aandacht aan de cognitie naast de auditie willen besteden. Als men het horen als een 'black box' met geluiden *in* en geluiden *uit* beschouwd, is dit geboden, wil men een goed begrip van het horen krijgen. Het zal blijken dat de voorliefde van onderzoek op het 'micro' niveau veel resultaten heeft opgeleverd waarvan de betekenis voor het alledaags horen op zijn minst als dubieus kan worden geacht.

'Schone' tegenover 'vervulde' geluiden

Met deze laatste opmerking ben ik op een terrein geland dat mij na aan het hart ligt. Onderzoek in het laboratorium, hoe belangrijk en onmisbaar ook, heeft de permanente neiging de werkelijkheid buiten het gebouw te veronachtzamen. De keuze voor 'schone' geluiden lijkt een vanzelfsprekende noodzaak te zijn, maar is ook hachelijk als men het onderzoek daartoe beperkt. Dat geldt ook als men denkt dit gevaar te kunnen bezweren door een gelijktijdig tweede 'schoon' geluid in het onderzoek te betrekken. De selecte keuze van condities beperkt de resultaten.

Met een aan de literatuur ontleend voorbeeld zal ik het onverwachte effect van een tweede stimulus toelichten. Dit is onderzocht door van een gesproken woord eerst alleen het eerste foneem uit te snijden en aan proefpersonen aan te bieden. Vervolgens herhaalt men het experiment door het tweede foneem toe te voegen, daarna het derde, tot het woord wordt herkend. Deze aanpak lijkt zindelijk, maar er gaat een riskante suggestie van uit. Het versterkt de veronderstelling dat het verstaan kan worden vergeleken met het opzoeken van een woord in een woordenboek. Dit door Marslen-Wilson geïntroduceerde 'co-

hortmodel' is blijkbaar door wetenschappers zo zinnig gevonden dat het Wikipedia heeft gehaald. Maar heeft het enige waarde? Een woord als 'buitenhuis', met de eerste paar fonemen door ruis gemaskeerd, wordt zonder problemen onmiddellijk herkend. Het brein heeft geen op fonemen als letters beschouwde bouwstenen nodig om een woord te herkennen; een fragment als routine is al genoeg om eerdere of latere 'beschadigingen' te overbruggen. Lopende spraak zou eenvoudig onverstaanbaar zijn als we zó sterk van de individuele fonemen afhankelijk zouden zijn.

Een ander voorbeeld. Met name de verschillen in de uitspraak van woorden als *pa*, *ta* en *ka* hebben veel aandacht gehad. Als men zich echter realiseert dat deze subtiele verschillen in de praktijk door onnauwkeurigheid van articuleren en maskering door omgevingsgeluiden niet of nauwelijks worden waargenomen, kan men zich terecht afvragen wat de zin van dergelijk onderzoek is. Maar genoeg hierover; zie deel 4.

Horen als mysterie

In de titel heb ik bewust het woord 'mysterie' gebruikt, een woord dat in de wetenschap niet graag wordt gehoord. Het gaat er toch immers om steeds verder door te dringen in het geheim van deze wereld en door gedegen onderzoek het deksel op te tillen dat deze geheimen verbergt. Helmholtz – fysioloog én fysicus – veronderstelde dat de basilaire membraan die de cochlea van het binnenoer in twee met vloeistof gevulde delen scheidt, uit parallelle dwarse snaren bestaat, afgestemd op de frequenties van het geluid. Het was de geniale Georg von Békésy (1899-1972) die het lukte de trillingen als lopende golf langs de basilaire membraan waar te nemen, met de plaats van maximale amplitude afhankelijk van de trillingsfrequentie; hiervoor ontving hij in

1961 de Nobelprijs. Hiermee was wel enige klaarheid verkregen in de rol van het perifere gehoororgaan, maar wat gebeurt er met de in de haarcellen geïnitieerde neurale activiteiten als deze het brein bereiken?

In 1665 hield de Deense anatoom Niels Stensen (1638-1686), bekend geworden als Nicolaus Steno, te Issy bij Parijs een lezing waarin hij argumenteerde dat, willen wij begrijpen wat het brein doet, wij het als een machine uit elkaar moeten nemen om te weten te komen hoe het werkt. We ontmoeten hier weer typisch de gedachte dat de som der delen het geheel bepaalt. Met deze woorden begint Matthew Cobb zijn boek *The idea of the brain* (2020). De auteur beschrijft de talloze uiteenlopende opvattingen die in de loop der geschiedenis zijn verkondigd, zonder tot een enigermate overtuigend beeld te komen. Zoals te verwachten was, speelt bij de nieuwste overwegingen de vergelijking met de digitale computer een grote rol. Maar ook hier struikelen we over voetangels en klemmen. De computeraanpak wordt gekenmerkt door de vaak lange reeks achtereenvolgende logische beslissingen. Alleen zó kan het besluit vallen of het om een huis, auto of stoel gaat. Maar ook dat resultaat staat nog ver af van de werkelijkheid. Stel dat ingewikkelde, langdurige en gebrekkige proces tegenover de vrijwel onmiddellijke wijze waarop wij gedetailleerd een voorwerp herkennen en kunnen beschrijven. Hier staan twee werelden tegenover elkaar.

De langdurige en vergeefse pogingen het brein te begrijpen tonen naar mijn overtuiging aan dat we hier met een mysterie te maken hebben, dat niet door de mens kan worden onthuld. Zoals eerder gezegd, daar houdt de wetenschap niet van, maar daar moet zij zich toch bij neerleggen.

Voor ons probleem houdt dit in dat de 'black box' van het horen niet als een 'voorlopige' beschrijving van onze kennis moet worden beschouwd, maar dat het echt om een zwarte doos gaat, ontoegankelijk voor ons verstand. Uiteraard, en dat hangt met het karakter van een mysterie samen, kan hier geen bewijs worden gegeven. De in het vervolg aan de orde komende raadselachtige verschijnselen moeten gezien worden als toelichtingen van dat hier van een mysterie sprake is.

2

De waarneming van een enkele toon

2 De waarneming van een enkele toon

Zoals we visuele objecten van elkaar kunnen onderscheiden, zijn we in staat twee of meer geluiden tegelijkertijd te horen. Dat geldt voor een concert zowel als voor het geroezemoes van stemmen. Die geluiden hebben alle hun eigen fysische spectra die in de lucht bijna onontwarbaar bij elkaar worden opgeteld. Toch blijkt het gehoororgaan in hoge mate in staat te zijn die geluiden weer te ‘ontmengen’, een verwonderlijk proces waarop in het volgende deel nader wordt ingegaan. Hier beperk ik mij, om het niet al te ingewikkeld te maken, tot een enkele klank, een enkele toon, zoals deze door een muziekinstrument en het menselijke stemorgaan wordt voortgebracht.

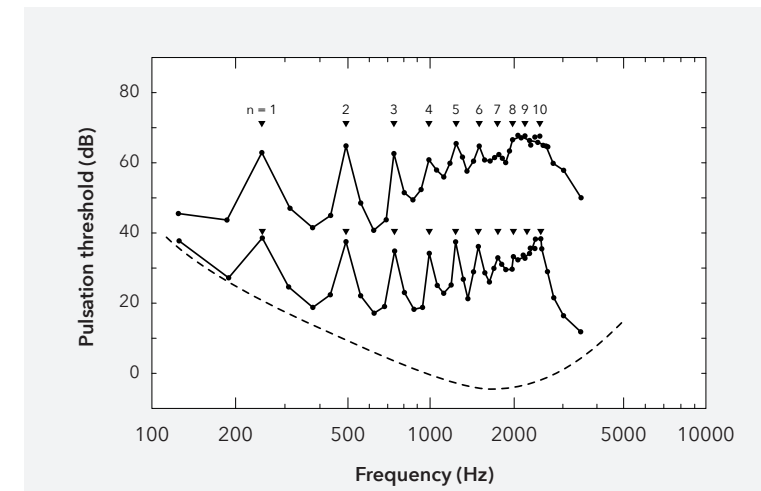
Dat het voor individuele tonen al zo ingewikkeld is, blijkt reeds uit het feit dat zij bij gelijke hoogte en luidheid heel verschillend kunnen klinken. Dat heeft te maken met die spectrale samenstelling waarover we het eerst zullen hebben.

Het oor als een frequentie-analysator

Zoals eerder is genoemd, kunnen geluidstrillingen sterk in frequentie, het aantal trillingen per seconde, uiteenlopen. Dat we dit kunnen horen, wijst er op dat het orgaan het geluid analyseert en in frequentie weet uiteen te leggen. Het gaat om een fysisch proces dat op ingenieuze wijze in de cochlea, het ‘slakkenhuis’ van het gehoororgaan wordt uitgevoerd. In dit proces worden de sinusvormige componenten (denk aan wat gezegd is over de vondst van Fourier) van een willekeurige trillingsvorm van elkaar gescheiden in lage en hoge tonen over een bereik van ca. duizend (16 Hz - 16.000 Hz). Van de karakteristieke golfvorm van een muziektoon en een spraakklank blijft in feite weinig over. Deze analyse heeft echter zijn grenzen. Nauwkeurige experimen-

tenten hebben aangetoond dat de scheiding beperkt blijft tot de laagste 5 à 6 harmonischen; de hogere blijven meer en meer samengevoegd, zie Figuur 2.1. Eberhard Zwicker (1924-1990) was de eerste die nauwkeurig de filtereigenschap van het gehoor in kaart bracht en de term ‘kritieke bandbreedte’ introduceerde, ruwweg samenvallend met eenderde octaaf (26 %).¹

Dit is echter maar de helft van het verhaal. Terwijl de sinusvormige componenten min of meer gescheiden via de gehoorzenuw naar het centrum in de hersenen worden geleid, vindt daar een soort reconstructie plaats. Over het hoe en wat is geen zinnig woord te zeggen, maar het resultaat is dat we de trilling keurig als één geheel plegen te horen. Met enige moeite is het mogelijk de laagste harmonischen van een lang aangehouden toon individueel waar te nemen, maar in de dagelijkse praktijk is van het



Figuur 2.1 Pulsatiedrempel van twee tonen van 250 Hz bestaande uit 10 harmonischen.¹

gehele proces van scheiding en vereniging niets te merken. Een fysische toon is wéér een toon, of deze nu dof of helder, als een *aa* of als een *oo* klinkt. Zowel de luidheid, de toonhoogte en de klankkleur is behouden gebleven.

De raadselachtige toonhoogte

Uiteraard zijn de wetenschappers nieuwsgierig geweest hoe het totale proces omgaat met de twee voor iedere toon zo karakteristieke fysische eigenschappen: de periodiciteit en de golfvorm. Helmholtz kan hier met ere worden genoemd. Met de theorie van Fourier als uitgangspunt ging hij in *Die Lehre von den Tönempfindungen* uitvoerig op het proces in. Met slimme proeven demonstreerde hij dat een willekeurige toon inderdaad door het oor in sinusvormige componenten wordt opgesplitst. Zijn voorkeur voor de sinusvormige toon als basis van het horen bracht hem tot de conclusie dat de grondtoon de gehoorde toonhoogte bepaalt terwijl de amplituden van de harmonischen in hun onderlinge verhoudingen de klankkleur opleveren. We zullen ons hier eerst richten op de toonhoogte waarover een interessante controverse is ontstaan die wel een eeuw het denken heeft beheerst. Daarna zal de klankkleur aan de orde komen.

Zoals gezegd, was Helmholtz stellig overtuigd van de unieke rol van de grondtoon. Hij beseftte wel dat deze in veel gevallen qua sterkte overtroffen wordt door de hogere harmonischen, maar daar had hij wel een oplossing voor: door niet-lineaire vervorming van het signaal in de cochlea wordt een eventueel afwezige of zwakke grondtoon opnieuw gecreëerd.

Er bestond echter een eeuwenoude traditie volgens welke de toonhoogte op de *periodiciteit* van de trilling is gebaseerd. In 1841, ruim vóór Helmholtz, demonstreerde August Seebeck (1805-1849) met behulp van een sirene hoe bepalend de periodi-

citeit voor de toonhoogte is. Onmiddellijk kwam hij in conflict met Georg Ohm (1789-1849), de grondlegger van de bekende wet van Ohm, een vurige aanhanger van de theorie van Fourier voor het horen. Toen Helmholtz in 1863 zijn meesterwerk schreef, waren Seebeck en Ohm reeds lang overleden. Zonder veel moeite kon hij zijn lezers van de unieke rol van de grondtoon overtuigen.

De impact van Helmholtz's boek is zo groot geweest dat het een eeuw geduurd heeft voordat de controverse van de grondtoon en de periodiciteit opnieuw serieus ter discussie kwam. Toen ik mij in 1953 in de problematiek van het horen ging verdiepen, was het werk van Helmholtz de gezaghebbende bron. Ook Stevens en Davis volgden in 1938 min of meer vanzelfsprekend de opvattingen van Helmholtz. Seebeck wordt nergens genoemd, de eventuele rol van de periodiciteit bleef onbesproken.

De eer van een eerste kritiek op de overheersende rol van de grondtoon komt toe aan de fysicus Jan Schouten (1910-1980). Na zijn promotie in 1937 verschoof hij in het natuurkundig laboratorium van Philips te Eindhoven (internationaal bekend als het 'Natlab') zijn aandacht van het zien naar het horen. In navolging van Seebeck gebruikte hij een (optische) sirene voor het produceren van periodieke geluidsignalen. Zoals te verwachten was, leidde deze aanpak tot de conclusie dat bij het horen van de toonhoogte aan de periodiciteit van niet door het oor te scheiden hogere harmonischen een belangrijke rol moet worden toegekend. Op grond van zijn proeven concludeerde Schouten dat de aanwezigheid van de grondtoon niet essentieel is (1940)². De controverse Helmholtz-Seebeck over de oorsprong van de toonhoogte was herboren. De periodiciteit van de door het gehoororgaan niet te scheiden hogere harmonischen, door hem als 'residu' aangeduid,

kan de toonhoogte verklaren. Hiermede was het probleem van de ‘missing fundamental’ geschapen. Omdat toonhoogte zo fundamenteel is voor de theorie van het horen, zal ik de verdere discussie hierover in het kort weergeven, ook omdat hierbij de bijdragen van Nederlandse onderzoekers groot is geweest.

In de 50er jaren liep het vraagstuk van de toonhoogte als een hernieuwde controverse echt uit de hand. In 1953 promoveerde te Utrecht de medicus G.A. Hoogland (KNO)³. Zijn dissertatie, met de pakkende titel *The missing fundamental*, was tot stand gekomen onder leiding van de fysicus Jaap Groen (1912-1973). De titel geeft al onverbloemd het conflict aan. Als aanhanger van Helmholtz was Groen er van overtuigd dat de experimenten van Schouten niet juist konden zijn. Hoogland stelde in zijn proefschrift dat in het sirene-experiment de grondtoon door vervorming van het signaal toch een rol moet hebben gespeeld. Hier was een waar wetenschappelijk conflict geschapen! Het verhaal ging dat het verzoek van Schouten om bij de promotie vanuit het publiek te mogen opponen niet werd toegestaan.

Blijkbaar werd het probleem van de ‘missing fundamental’ snel wereldkundig, want tijdens de voorjaarsbijeenkomst 1954 van de Acoustical Society of America kon J.C.R. Licklider (1915-1990) een beslissende demonstratie geven. Als de grondtoon bepalend is voor de toonhoogte zou een bandje maskerende ruis rond de grondtoon de veronderstelde rol van deze toon teniet moeten doen. Licklider liet horen dat dit inderdaad het geval is: de hogere harmonischen leveren inderdaad de toonhoogte van de afwezige grondtoon op. Hiermede was het probleem ten gunste van Schouten beslecht; na bijna een eeuw bleek Helmholtz toch ongelijk te hebben. In september 1955 bezocht Licklider Nederland en herhaalde zijn demonstratie in het laboratorium van Groen, waarbij ook Hoogland, de Boer en schrijver dezes aanwezig waren.

In dezelfde jaren benaderde Egbert de Boer in het kader van zijn proefschrift (1956)⁴ het probleem in een ingenieus experiment van een heel andere kant. Hij startte met een toon bestaande uit vijf sinusvormige componenten:

$$800 + 1000 + 1200 + 1400 + 1600 \text{ Hz}$$

en verschoof dit complex over 50 Hz tot

$$850 + 1050 + 1250 + 1450 + 1650 \text{ Hz.}$$

Luisteraars namen deze verschuiving waar alsof de toonhoogte versprong van 200 Hz naar ca. 210 Hz. Twee verschillende interpretaties werden voorgesteld. Als we aannemen dat de vijf tonen door het gehoororgaan van elkaar worden onderscheiden, kan er van patroonherkenning sprake zijn resulterend in een toonhoogte die zo goed mogelijk recht doet aan de verschoven frequenties. De beste keus is:

$$833,3 + 1041,7 + 1250 + 1458,3 + 1667,7 \text{ Hz,}$$

bestaande uit veelvouden van 208,3 Hz, dicht bij de waargenomen verschuiving naar 210 Hz.

Een alternatieve benadering is dat we de vijf tonen niet gescheiden maar als één periodieke golfvorm moeten zien. De Boer toonde aan dat ook de verschuiving van die golfvorm de sprong in toonhoogte kan verklaren. Intensief verder onderzoek heeft tot de opvatting geleid dat we niet te strak in termen van òf-òf moeten denken. Als algemene conclusie wordt aanvaard dat de door de frequentie-analyse gescheiden derde tot vijfde harmonischen het meest bijdragen aan de toonhoogte. Opgemerkt

kan nog worden dat de lage toonhoogte ook bij minder dan vijf harmonischen kan worden waargenomen. De toonhoogte is zeer bestand tegen storingen door andere geluiden.

Na deze eerste ‘moderne’ experimenten over de perceptie van de toonhoogte nam de belangstelling sterk toe, vooral in Nederland. Talrijke aspecten werden bestudeerd die hier verder niet ter sprake kunnen komen. Ik volsta met enige namen te noemen. In het programma van het in 1957 onder directie van Jan Schouten te Eindhoven opgerichte Instituut voor Perceptie Onderzoek (IPO) werd vanaf de opening veel aandacht aan het toonhoogtevraagstuk gegeven. Genoemd kunnen worden Roel Ritsma en Ben Lopez Cardozo, gevolgd door Julius Goldstein en Adriaan Houtsma. In het Instituut voor Zintuigfysiologie TNO te Soesterberg leverden Guido Smoorenburg, Tammo Houtgast en schrijver dezes hun bijdragen. Verder kunnen Diek Duifhuis en Frans Bilsen worden genoemd. Het waren niet de minsten; vrijwel allen hebben hun carrière later als hoogleraar afgesloten.

De waarneming van de toonhoogte is een bijzondere prestatie van het hoorproces. Eerst wordt in het perifere orgaan de periodieke trilling in harmonischen ontleed, centraal worden ze weer verenigd tot een enkele klank. Die bijzonderheid zal nog sterker blijken als in het volgende hoofdstuk de samenklank van tonen aan de orde komt; iedere complexe toon treft zijn eigen analyse en uit de warreling van al die sinusvormige componenten weet het gehoor feilloos die te verenigen die in het aangeboden geluid bij elkaar hoorden.

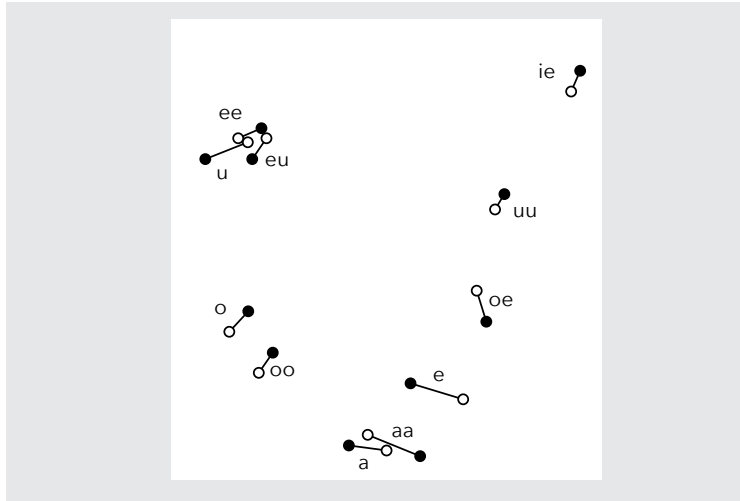
De multidimensionele klankkleur

Met de klankkleur hebben we met een heel ander aspect van de toon te maken. Terwijl de frequentie langs een eendimensione-

le schaal van laag naar hoog kan variëren, wordt de klankkleur bepaald door de relatieve verdeling in intensiteit van de harmonischen. Tonen met sterke lage frequenties klinken dof, met sterke hoge frequenties scherp. Fysisch gaat het hier dus om een meerdimensionele grootte, die zich in het horen eveneens meerdimensioneel aandient. Elk muziekinstrument geeft zijn kenmerkende toon. De term ‘klankkleur’ drukt dit passend uit: net zomin als de verschillen tussen kleuren kunnen die tussen klanken van dezelfde toonhoogte langs een enkele schaal geordend worden.

Pas na ca. 1960 kwamen beschrijvingstechnieken beschikbaar waarmee de perceptieve verschillen tussen klanken multidimensioneel kunnen worden weergegeven. De wiskundige Joseph Kruskal (1928-2010) publiceerde in 1964 de volgende methode⁶. Stel dat we de klankverschillen tussen elf klinkers willen kwantificeren. We bieden dan aan proefpersonen achtereenvolgens alle mogelijke drietallen klinkers aan; bij elk drietal moet beslist worden welke twee het minst op elkaar lijken, dat levert 2 punten op, en welke twee het meest, 0 punten; het overgebleven paar krijgt 1 punt. Door voor alle tweetallen de behaalde punten bij elkaar op te tellen, krijgen we een matrix, vergelijkbaar met een afstandenmatrix van steden. Het algoritme van Kruskal berekent de beste multidimensionele configuratie van elf punten die recht doet aan de afstandenmatrix. In figuur 2.2 (pagina 28) geven de open punten de twee belangrijkste dimensies van de op deze wijze verkregen puntenwolk weer.

Op een vergelijkbare wijze kunnen we ook de spectrale verschillen tussen de elf klinkers weergeven. Het frequentiespectrum wordt dan voor elke klinker opgedeeld in 18 banden ter breedte van eenderde octaaf, ruwweg vergelijkbaar met de kritieke



Figuur 2.2 Resultaat van de matching van de belangrijkste spectrale verschillen (dichte punten) met de perceptieve verschillen (open punten) van elf klinkers.⁵

bandbreedte van het gehoororgaan. Zo krijgen we een 11x18 matrix van getallen. Een principale-componentenanalyse levert dan een puntenconfiguratie op die recht doet aan de spectrale verschillen. De twee belangrijkste dimensies van deze puntenwolk, in de figuur door dichte punten weergegeven, is gematcht met de perceptieve configuratie. We constateren een uitstekende overeenstemming, hetgeen betekent dat het signaal aan de uitgang van de black box getrouw het ingangssignaal weergeeft. Een vergelijkbaar resultaat wordt gevonden als we voor de fysica uitgaan van de frequenties van de twee belangrijkste klinkerformanten (de resonantiefrequenties van de mond-keelholte).

De stabiele luidheid

De derde eigenschap van tonen is de luidheid. Verandering van de toonhoogte en de klankkleur beleven we als verandering van

de toon zelf. Voor de luidheid ligt dit anders. We spreken van een *gelijkblijvende* toon die luider of zachter klinkt, in afhankelijkheid van het geluidsniveau.

Dat wil niet zeggen dat samengestelde tonen van dezelfde frequentie en geluidsniveau even luid zijn. De amplitudeverdeling van de harmonischen speelt ook een rol. De luidheid van een puur sinusvormige trilling is duidelijk minder dan van een toon van dezelfde fysieke intensiteit maar met een breed frequentiespectrum. Zo'n toon activeert een veel breder gebied van de cochlea en dat laat zich merken. Organisten hebben al lang geleden ontdekt dat de luide toon wordt gecreëerd als men orgelregisters combineert waarvan de grondfrequenties zich verhouden als 1:2:4:8 (pijplengten van resp. 8, 4, 2 en 1 voet), bekend als het *plenum*; dan krijgt men een toon met een breed spectrum.

Een hoogst opmerkelijke eigenschap van de luidheid is dat zij nauwelijks of niet wordt aangetast door andere gelijktijdige geluiden. We zouden verwachten dat de luidheid van een muziekinstrument of van de menselijk stem er ernstig onder lijdt als andere geluiden van vergelijkbare sterkte mee gaan klinken. De dagelijkse ervaring leert dat van enige hinder nauwelijks sprake is. Gelijktijdig klinkende stemmen kunnen elkaar door maskering onhoorbaar maken, maar storen elkaar niet in hun luidheid. Als dat wèl het geval zou zijn, zou elke stem voortdurend in luidheid moeten fluctueren, hoogst hinderlijk en vermoeiend. Van meerstemmige muziek zou weinig terecht komen. Maar niets daarvan.

Nagalm

Tot nog toe is er als vanzelfsprekend vanuit gegaan dat het door

de bron geproduceerde geluid rechtstreeks het oor bereikt. Maar die situatie komt eigenlijk alleen maar in de open lucht voor. In de meeste gevallen bevinden zich spreker en hoorder beiden in besloten ruimten. We hebben dan niet alleen met direct ontvangen geluid te maken maar ook met geluid dat via weerkaatsingen aan wanden enz. het oor bereikt. Naarmate het aantal achter-eenvolgende reflecties groter is, is de weg langer en komt de geluidstraal dus later en zwakker aan. Talloze reflecties uit verschillende richtingen worden bij elkaar opgeteld. Gunstig is dat de reflecties het geluid versterken, maar deze winst gaat teloor bij grote ruimten en/of akoestisch 'harde' wanden. Dan wordt het weerkaatste geluid vaak als storende nagalm ervaren.

Deze hinderlijkheid hangt samen met de mate waarin het geluid van moment tot moment in sterkte varieert. Ik beperk me tot de spraak omdat deze tal van snelle spectrale afwisselingen bevat die voor het verstaan belangrijk zijn. Naarmate de nagalm sterker is, zullen zwakkere componenten van het geluid meer en meer door de nagalm temporeel versmeerd worden, niet meer door het gehoor worden gedetecteerd ten nadele van de verstaanbaarheid.

Een markant criterium is de afstand tot de bron waar het directe en het indirecte geluid even sterk zijn, de *galmstraal*. Om een indruk te geven: in klaslokalen bedraagt deze 1 à 2 m, in concertzalen ruwweg het dubbele. Deze waarden tonen aan dat in de praktijk de meeste luisteraars zich in het indirecte veld bevinden en het geluid dus voornamelijk via de wanden ontvangen.

Tegenwoordig bestaan er methoden die de verstaanbaarheid met behulp van fysische metingen kunnen voorspellen. Grondslag is de modulatie-overdrachtsfunctie. Een band ruis van bijv. 1000-2000 Hz wordt sinusvormig met verschillende snelheden

in sterkte gevarieerd en gemeten wordt de mate waarin deze variaties bestand zijn tegen de versmering van de nagalm. Door deze handeling voor andere frequentiebanden te herhalen wordt een beeld verkregen hoe de verschillende frequenties van het spraakspectrum in de ruimte worden overgebracht. Tammo Houtgast en Herman Steeneken hebben deze aanpak bedacht en in 1971 gepubliceerd⁷. Vervolgens ontwikkelden zij op deze basis een maat, bekend als de STI (Speech Transmission Index) waarvan een latere versie, de RASTI (Room Acoustics STI), in 1988 internationaal als standaardmethode werd aangenomen. Sinds 2000 wordt een uitbreiding onder de naam STIPA (STI for Public Address systems) toegepast. Verschillende akoestische bedrijven hebben de methode omgezet in handzame apparatuur.

Conclusies

Zoals elke periodieke golfvorm mathematisch kan worden beschreven als de som van sinussen met frequentieverhoudingen 1:2:3:4:5 enz., kan een geluidstrilling fysisch worden opgevat als de som van sinusvormige trillingen. Het perifere gehoororgaan voert een hiermee vergelijkbare frequentie-analyse uit; niettemin wordt de trilling als één geheel gehoord.

Tonen onderscheiden zich van elkaar door hun toonhoogte, klankkleur en luidheid; de toonhoogte wordt voornamelijk bepaald door de frequenties van de lagere harmonischen, de klankkleur door de relatieve sterkte van de harmonischen onderling, de luidheid door hun totale sterkte.

Opmerkelijk is hoe alle drie eigenschappen resistent zijn tegen storingen van gelijktijdige geluiden.

3

De waarneming van gelijktijdige tonen

3 De waarneming van gelijktijdige tonen

Deel 2 handelde over een enkele constante toon als basisgeluid van het horen: de door een strijkinstrument, een blaasinstrument of de stembanden geproduceerde periodieke trilling. Dat was al knap ingewikkeld, maar bleef toch nog tamelijk abstract. Het wordt pas echt complex als er met die trillingen iets gaat gebeuren, als tonen elkaar opvolgen en vooral als tonen tegelijk klinken. Hierover gaat het huidige deel.

Begonnen wordt met de waarneming van elkaar in de tijd volgende geluiden. Vervolgens wordt ingegaan op de vraag hoe de harmonischen van *gelijktijdige* geluiden van elkaar worden onderscheiden om later weer te worden verenigd; een uiterst ingewikkeld proces. De interactie van tonen, zich uitend in *consonantie* tegenover *dissonantie*, komt daarna ter sprake. Tenslotte komt de bijdrage van het horen met twee oren aan de orde.

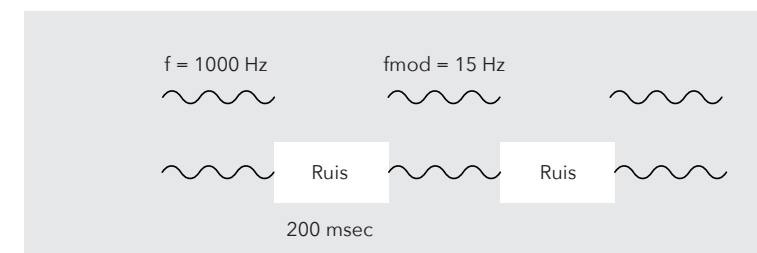
Het continu horen van een door ruis onderbroken toon

De psycholoog George Miller (1920-2012) van het Massachusetts Institute of Technology (MIT), later professor aan Harvard University, publiceerde in 1950 samen met J.C.R. Licklider (1915-1990) over korte toonstoten die als een continue toon gaan klinken als de onderbrekingen worden gevuld met ruis.⁸ We kunnen deze waarneming beschouwen als de auditieve variant van het in 1912 door Max Wertheimer (1880-1943) ontdekte verschijnsel dat twee op korte afstand van elkaar afwisselend oplichtende lichtbronnen gezien worden als een enkel verspringend licht. Hiermede was dus aangetoond dat het *continuïteitseffect* van de Gestalttheorie ook bij geluiden optreedt.

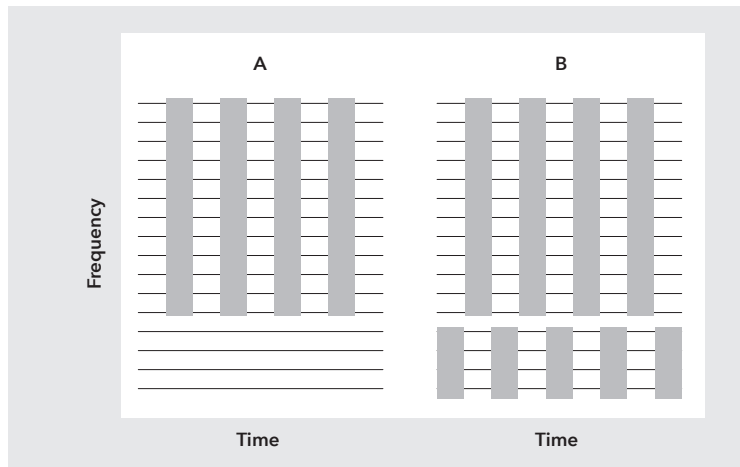
Een variant is de toonhoogte van de stoten te variëren. Figuur 3.1

illustreert het geval dat een toon van 1000 Hz wordt gemoduleerd met een frequentie van 15 Hz. Afgewisseld met stilte (bovenste reeks) is de modulatie niet erg duidelijk. Voegen we de ruisstoten toe (onderste reeks), dan is dat wél het geval en horen we een continu gemoduleerde toon dwars door de ruis heen. Dat is ook het geval als de toon in amplitude wordt gevarieerd. Het is een hoogst opmerkelijk verschijnsel.

Het zal duidelijk zijn dat het perifere gehoororgaan dit verschijnsel niet kan verklaren. Hier hebben we voor het eerst een treffend bewijs dat het hoorproces twee opvolgende fasen kent. In de eerste, het *auditieve* proces van het perifere gehoororgaan, vindt frequentie-analyse plaats. Daarna wordt het signaal in het *centrale* proces van het brein nader bewerkt. De frequentie-analyse van de trillingen in de cochlea kan fysisch-fysiologisch worden begrepen als een *passief* proces. Het centrale proces in de hersenen blijkt daarentegen van een geheel andere aard te zijn: de toonstoten worden *niet* gehoord zoals ze aan het oor werden aangeboden maar, zonder dat we ons dit bewust zijn of er iets aan kunnen doen, rijgen zij zich aaneen tot een continue toon. Het komt erop neer dat we het na elkaar horen van kortdurende geluiden niet helemaal kunnen vertrouwen: afwezig geluid kan



Figuur 3.1 Met 15 Hz gemoduleerde toonstoten van 1000 Hz worden afgewisseld met stilte of ruisstoten.



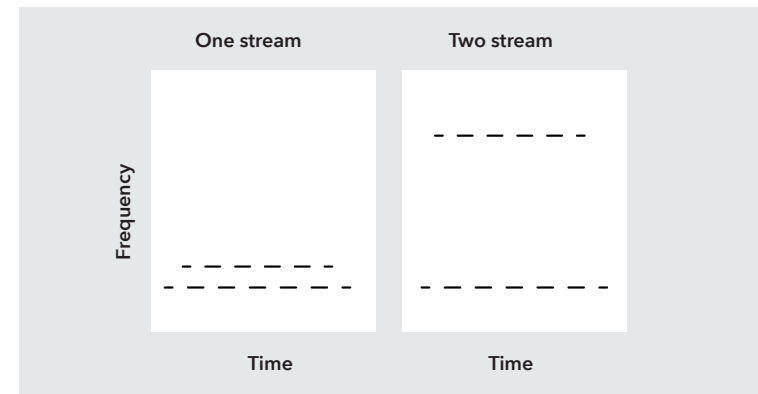
Figuur 3.2 De lijnen representeren de harmonischen van een kortdurende toon afgewisseld met een ruisband. In A blijven de laagste vier harmonischen continu klinken, in B worden ook zij met ruis afgewisseld, zij het in tegengestelde fase.

tóch worden gehoord. Het brein speelt een *actieve* rol bij het hoorproces. Hiervan zullen we in het vervolg nog vele bewijzen aantreffen. Hier beperken we ons tot nog een opvallende demonstratie.

Een variant is de sinusvormige trilling te vervangen door een trilling bestaande uit een groot aantal harmonischen. Figuur 3.2 toont links een uit 16 harmonischen bestaande toon waarvan de hogere 12 vijf keer per seconde door ruisstoten worden afgewisseld. De complexe toon klinkt ongestoord dwars door de ruis heen. Rechts wordt het nog wat geraffineerder: nu worden ook de 4 laagste harmonischen door ruis afgewisseld, maar zó dat op geen enkel moment alle 16 harmonischen tegelijk aanwezig zijn. Ook in dit geval laat het gehoor zich misleiden en klinkt de complexe toon continu. Albert Bregman heeft zich jaren met het

continuïteitseffect beziggehouden en geeft in zijn boek *Auditory scene analysis - The perceptual organization of sound* (1990) vele voorbeelden.

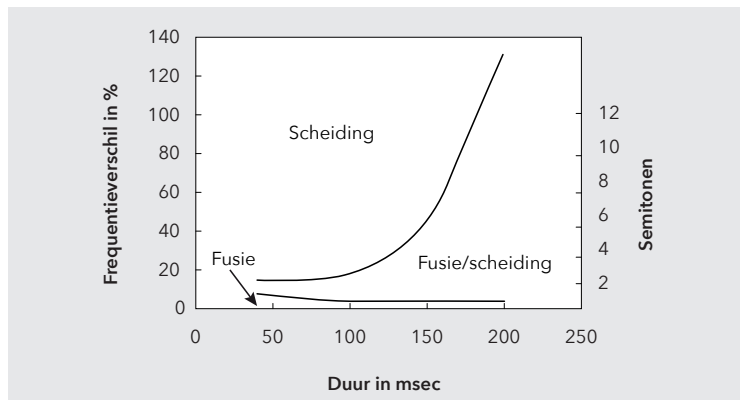
Alvorens op verwante zaken over te gaan, moet op een essentiële grens van het effect worden gewezen, namelijk de duur van de afwisselingen. Houdt men de toonstoten aan tot bijv. 1 seconde, dan is er geen sprake meer van continuïteit. Het verschijnsel treedt het meest sprekend op voor onderbrekingen korter dan 200 milliseconde. Deze waarde komt overeen met de gemiddelde duur van lettergrepen in lopende spraak. Zoals we in deel 5 zullen zien, is dit niet toevallig. Het continuïteitseffect heeft alles te maken met de resistentie van spraak tegen kortdurende verstoringen. Óók muziek en spraak worden als doorlopend gehoord als zij kort worden gemaskeerd door andere geluiden. Het brein schijnt te beslissen welke geluiden bij elkaar horen en herstelt een ‘beschadigd’ geluid tot wat het meest waarschijnlijke signaal is geweest.



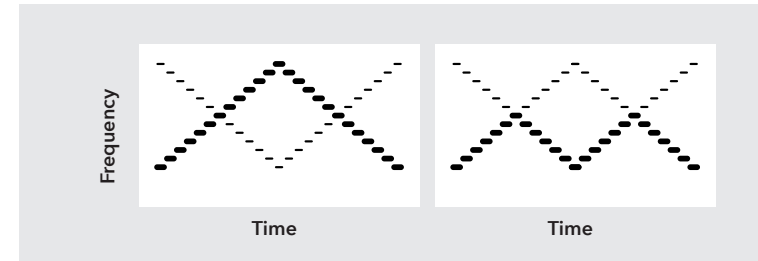
Figuur 3.3 Achtereenvolgende toonstoten met een klein frequentieverschil worden als een enkele stroom gehoord, bij een groot verschil als twee onafhankelijke stromen.

Miller en Heise experimenteerden in 1950 met toonstoten die afgewisseld werden met toonstoten van een andere frequentie.⁹ Bij een klein frequentieverschil horen we één stroom waarvan de toonhoogte op en neer springt zoals van een melodie. Bij toename van het frequentieverschil wordt dit steeds moeilijker en splitst de stroom zich in twee stromen, één van de lagere, de andere van de hogere toon, zie figuur 3.3. Deze overgang treedt geleidelijk bij toenemend frequentieverschil op. Er blijft een gebied waarbinnen men met enige inspanning één óf twee stromen kan horen, zoals door Leon van Noorden onderzocht, zie figuur 3.4.¹⁰ Deze splitsing speelt een grote rol in de muziek. De werken van Johann Sebastian Bach zijn er vol van. Hij wist zelfs snel elkaar afwisselende viooltonen de indruk van tweestemmigheid te geven.

Interessant is het horen van twee elkaar in toonhoogte kruisende stromen van afwisselende toonstoten. Men zou verwachten een stroom van laag naar hoog te horen gelijk met een stroom



Figuur 3.4 De scheiding van twee elkaar afwisselende toonreeksen als functie van de toonduur.



Figuur 3.5 Twee series van elkaar afwisselende toonstoten die elkaar kruisen worden alleen als kruisend gehoord als zij in klankkleur (sterkte harmonischen) verschillen (links), anders als een lagere samen met een hogere stroom.

in tegengestelde richting, elkaar kruisend als links in figuur 3.5. Dat blijkt niet het geval te zijn. Hardnekkig blijven we een toon horen die hoog varieert en een tweede die laag varieert, beide terugspringend bij het kruispunt, zoals rechts in de figuur. Dit gedrag verandert als men de twee tonen in klankkleur laat verschillen. Dan wordt het kruispunt keurig gepasseerd als links getekend. Het centrale hoorproces ‘veronderstelt’ blijkbaar dat fysisch identieke stromen elkaar niet plegen te passeren, terwijl dat bij in klank verschillende stromen wél wordt aanvaard. Hier gedraagt het oor zich wezenlijk anders dan het oog; het scheidt de gelijktijdige geluidstromen als van verschillende bronnen afkomstig. De perceptie probeert zich af te stemmen op de productie. Een duidelijke demonstratie van de nauw veronderstelde relatie tussen productie en perceptie.

De besproken experimenten betroffen stilzwijgend afwisselende toonstoten van vergelijkbare luidheid. Als we de sterkte van een van de stromen meer en meer reduceren, komt er een moment dat deze zwakke toonstoten zich verenigen tot een continue toon, net zoals dat in figuur 3.1 bij storende ruis het geval was.

Deze *pulsatiedrempel* is in figuur 2.1 al getoond, waar zij gebruikt werd als methode voor het bepalen van het frequentie-analyserend vermogen van het gehoororgaan.

Als twee stromen van geluidstoten elkaar afwisselen treedt nog een ander verschijnsel op: het blijkt moeilijker het verschil in de tijdsvolgorde aan te geven naarmate de twee stromen meer in frequentie verschillen. Dick Warren¹¹ (1970) heeft dit voor een extreem geval bestudeerd. Een toon met sterke harmonischen, een lage toon, een hoge toon en een ruisband, elk 200 milliseconde lang, wisselden elkaar eindeloos af. Proefpersonen bleken de volgorde van die sterk verschillende geluiden niet aan te kunnen geven. Dat begon pas te lukken als het tempo werd gehalveerd. Daarentegen kon de volgorde van 200 milliseconde durende spraakeenheden zonder problemen worden waargenomen. We constateren hier duidelijk het verschil tussen de perceptie van vertrouwde tegenover onbekende geluiden. Herkenning en ervaring spelen ook hier een belangrijke rol.

Het scheiden van gelijktijdige geluiden

In het dagelijkse leven zijn gelijktijdige geluiden meer regel dan uitzondering. Voortdurend worden we ermee overspoeld. Al zijn de trillingen van de verschillende bronnen in de lucht gesuperponeerd, toch worden ze verbluffend goed van elkaar onderscheiden. Vaak zal men van storingen of zelfs van lawaai spreken, maar in andere gevallen gaat het juist om het gelijktijdig klinken. Muziek is hiervan het beste voorbeeld.

Een in frequentie variërende toon wordt gemakkelijk los gehoord van een toon met een vaste toonhoogte. Heel anders ligt het bij het samenklinken van twee constante tonen, door een muziekinstrument of de mens voortgebracht, beide tonen bestaande uit een reeks harmonischen.

Stel het geval van een toon met een grondfrequentie van 250 Hz gelijktijdig met een toon met een grondfrequentie van 340 Hz, ieder bestaande uit zes harmonischen.

Het resultaat is:

$250 + 340 + 500 + 680 + 750 + 1000 + 1020 + 1250 + 1360 + 1500 + 1700 + 2040 \text{ Hz}$.

Van deze verzameling harmonische componenten zullen met de beperkte frequentie-analyse van de cochlea, de laagste 5 à 6 van elkaar gescheiden worden, de hogere zullen meer of minder vermengd blijven. Toch worden twee tonen gehoord, corresponderend met grondfrequenties van 250 resp. 340 Hz; in het centrale hoorproces van het brein worden de bij elkaar behorende harmonischen weer keurig tot een geheel verenigd. Hier vindt een proces plaats waarover in feite niets zinnigs kan worden gezegd. Zoals in de inleiding genoemd, hebben we met een 'black box' te maken waarvan we het fysische ingangssignaal kennen en ook horen wat er aan het eind uitkomt, maar daarmee is alles gezegd. Hoe het brein in staat is 'buitenwereld' en 'binnenwereld' zo perfect op elkaar af te stemmen, is een mysterie. Het enige wat we doen kunnen, is factoren opsporen die kennelijk dit proces bevorderen.

Het is hoogst opmerkelijk dat Helmholtz deze problematiek al heeft onderkend. Hij wijst er in zijn boek op dat als aan een reeds bestaande klank een tweede wordt toegevoegd, we dankzij de temporele verschillen in begin en eind de twee klanken kunnen onderscheiden. Bovendien zijn er bij synchrone stemmen verschillen in de wijzen waarop de tonen in- en uitklinken (*On the sensations of tones*, Dover Publications, blz. 59). Ik heb dit

inzicht in het scheiden van gelijktijdige geluiden bij geen enkel latere onderzoeker teruggevonden.

Het reconstructieproces wordt in hoge mate bepaald door de geluiden die aan de betreffende samenklank voorafgaan en volgen. Zoals we eerder zagen, voegt het gehoor volgens het continuïteitseffect op elkaar volgende tonen samen als zij qua toonhoogte en klankkleur bij elkaar passen. Daarnaast kunnen talrijke andere factoren worden genoemd. De muziek kent er vele. Een in frequentie of sterkte vibrerende toon wordt gemakkelijk onderscheiden. Vooral karakteristieke verschillen in het aan- en uitklinken zijn belangrijk. Strijk- en blaasinstrumenten worden van elkaar onderscheiden door het meer plotseling dan wel geleidelijk begin van de toon. Temporele zowel als spectrale verschillen dragen bij aan het verenigen van de bij elke toon behorende harmonischen.

Een interessante uitzondering is het orgel. De luisteraar heeft er geen besef van dat de toon niet door een enkele bron wordt voorgebracht, maar het product is van een aantal pijpen of tongen samen, al naar de registers die zijn ingeschakeld. Die verschillende bronnen beginnen bij het indrukken van een toets op vrijwel gelijke wijze te klinken en versmelten tot één geheel. Dit bewijst hoe essentieel karakteristieke verschillen in de productie van de toon voor het gescheiden horen zijn.

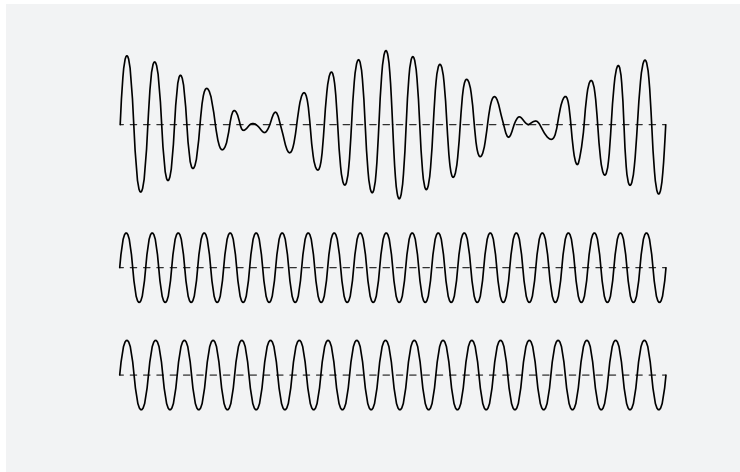
Bij het bespelen van instrumenten zowel als het beluisteren speelt de beperkte geluidsnelheid een rol. Voor goed samenspel moeten musici het inzetten van tonen op elkaar afstemmen. Voor de luisteraar is het van belang dat het samenspel zo getrouw mogelijk wordt gehoord. Tijdverschillen zijn in beide gevallen van belang.

De geluidsnelheid in lucht bedraagt 340 m per seconde, overeenkomend met 3 milliseconden per meter. Muziekpodia plegen maximaal 200 m² groot te zijn, waardoor de grootste afstand tussen de musici van een orkest minder dan ca. 15 m zal zijn, wat een maximaal tijdsverschil van 45 milliseconden inhoudt. Zaalakoestici houden aan dat de eerste reflecties van het geluid (via de zijwanden) binnen 30 milliseconden na het directe geluid bij de luisteraar moeten aankomen, willen zij goed met elkaar versmelten. Als we deze waarde ook voor het samenspel als criterium stellen, mogen we concluderen dat zij voor de genoemde podiumgrootte voor vrijwel alle leden van het orkest voldoet. De grote zaal van het Concertgebouw te Amsterdam heeft een podium van 160 m² en een tijdsverschil met de eerste reflectie van gemiddeld 21 milliseconden zodat deze meer dan 130 jaar oude zaal geheel aan het moderne criterium voldoet.

Consonantie en dissonantie

In het voorgaande is besproken hoe gelijktijdige geluiden, in de lucht gesuperponeerd, toch als schijnbaar los van elkaar kunnen worden waargenomen. Het ging om de vraag in hoeverre tonen ‘onbeschadigd’ het ingewikkelde en raadselachtige hoorproces overleven. Nu zullen we onder ogen zien hoe ze elkaar door interferentie kunnen storen, hoe de samenklank aangenaam, maar ook schrill kan klinken, afhankelijk van het frequentieverschil.

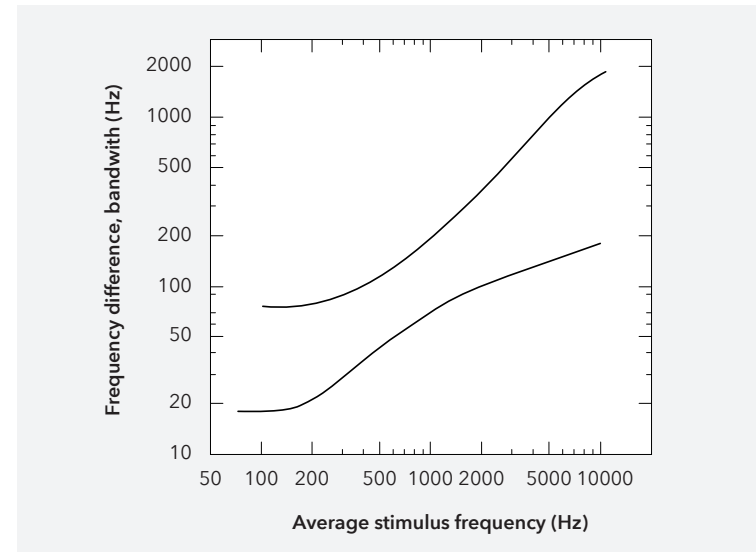
Het eenvoudigste geval is het gelijktijdig klinken van twee sinusvormige trillingen. Eerder hebben we vastgesteld dat het beperkte frequentie-analyserend vermogen van het perifere gehoororgaan sinusvormige trillingen met een frequentieverschil kleiner dan de kritieke bandbreedte niet van elkaar kan scheiden.



Figuur 3.6 Amplitudevariatie van twee sinustonen met een klein verschil in frequentie.

Figuur 3.6 toont hoe twee sinussen met een klein verschil in frequentie in amplitude varieert; afwisselend zijn de trillingen in fase en worden opgeteld, dan wel in tegenfase en heffen elkaar op. Voor de waarneming levert dit de bekende zwevingen als muziekinstrumenten niet goed op elkaar zijn afgestemd.

Langzame zwevingen zijn goed te horen en klinken niet onaangenaam. Vergroot men het frequentieverschil, dan volgen de zwevingen elkaar sneller en sneller op en gaat de samenklank schriller klinken om uiteindelijk over te gaan in het los van elkaar komen van de tonen. De onderste kromme in figuur 3.7 geeft aan hoe het frequentieverschil van maximale 'schrilheid' toeneemt als functie van de frequentie. Vraagt men proefpersonen in plaats van schrilheid of ruwheid naar de *dissonantie* van de samenklank, dan vindt men vergelijkbare waarden. Schrilheid en dissonantie kunnen dus als synoniemen worden beschouwd. De dissonantie verdwijnt vrijwel geheel als het frequentiever-

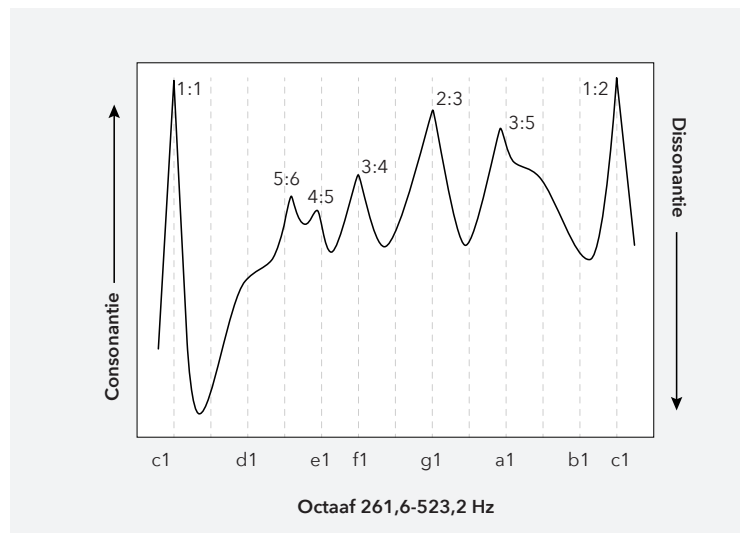


Figuur 3.7 Twee sinustonen met een klein frequentieverschil geven zwevingen die bij toenemend frequentieverschil meer en meer schrill gaan klinken met een maximum aangegeven door de onderste kromme. Daarna neemt de dissonantie af en verdwijnt bij frequentieverschillen groter dan de kritieke bandbreedte (bovenste kromme).

schil de kritieke bandbreedte overschrijdt, de bovenste kromme. Ruwweg is de dissonantie maximaal bij een kwart van de kritieke bandbreedte.

Gaat men van sinustonen over op complexe tonen, dan wordt de situatie zeer veel ingewikkelder. De harmonischen van de lagere toon met frequenties f_1 , $2f_1$, $3f_1$, enz. kunnen dissoneren met harmonischen van de hogere toon met frequenties f_2 , $2f_2$, $3f_2$, enz. Als de lagere toon varieert kan de tweede harmonische met frequentie $2f_1$ in de buurt komen van de grondfrequentie f_2 van de hogere. De dissonantie zal eerst toenemen, vervolgens afne-

men om bij $2f_1=f_2$ (het octaaf) geheel te verdwijnen. We hebben hier dus een eerste frequentieverhouding met een consonante samenklank. Voegen we een derde harmonische toe, dan zal de harmonische $3f_1$ met $2f_2$ tot dissonantie leiden, verdwijnend bij $3f_1=2f_2$ (2:3). Naarmate het aantal harmonischen in de samenklank toeneemt, neemt ook het aantal dissonante en consonante frequentieverhoudingen toe. Beperken we ons tot het octaaf, dan zijn de belangrijkste consonante verhoudingen 2:3 (kwint), 3:4 (kwart), 4:5 (grote tert), 5:6 (kleine tert), 5:8 (kleine sext) en 3:5 (grote sext). Opgemerkt moet worden dat de ‘consonantiegraad’ volgens deze reeks minder wordt. Helmholtz is de eerste geweest die het verloop van de consonantiegraad op grond van subjectieve indrukken grafisch heeft weergegeven. Een op experimentele basis meer recente poging is figuur 3.8.



Figuur 3.8 Mate van consonantie en dissonantie van complexe tonen als functie van de frequentie-afstand over een octaaf.

Aan Pythagoras (zesde eeuw voor Chr.) wordt de ontdekking van het verschijnsel van consonantie toegeschreven. Hij constateerde dat een trillende snaar die in het midden wordt vastgehouden een toon geeft die harmonieus samenklinkt met de toon van de snaar als geheel: de octaaf 1:2. Zo ontdekte hij ook het singuliere karakter van andere lengteverhoudingen, de kwint en de kwart. We hebben hier te maken met de vroegste natuurkundige wet. Uiteraard wist men toen niet dat de trillingsfrequentie omgekeerd evenredig is met de snaarlengte; dat werd pas in de zeventiende eeuw duidelijk.

De legendarische vondst van Pythagoras was de grondslag van de ontwikkeling van de Westerse muziek. In de Middeleeuwen heeft men geleidelijk ook de tertsen en sexten als consonanten aanvaard. Al de genoemde frequentieverhoudingen werden de ankerpunten van de twaalfdelige toonschaal die wij kennen.

Een probleem was dat deze met verdere halve tonen aangevulde toonschaal het vrij transponeren verhinderde. Op allerlei vernuftige manieren trachtte men hieraan tegemoet te komen tot in de zeventiende eeuw de knoop werd doorgehakt met de zogenoemde gelijkzwevende stemming, waarbij het octaaf mathematisch in tonen van gelijke intervallen van $2^{1/12}$, is verdeeld. Hoe groot de concessie was, laat de volgende tabel zien:

Kleine terts	5:6	1,200	1,189
Grote terts	4:5	1,250	1,260
Kwart	3:4	1,333	1,335
Kwint	2:3	1,500	1,498
Kleine sext	5:8	1,600	1,587
Grote sext	3:5	1,667	1,682
Octaaf	1:2	2,000	2,000

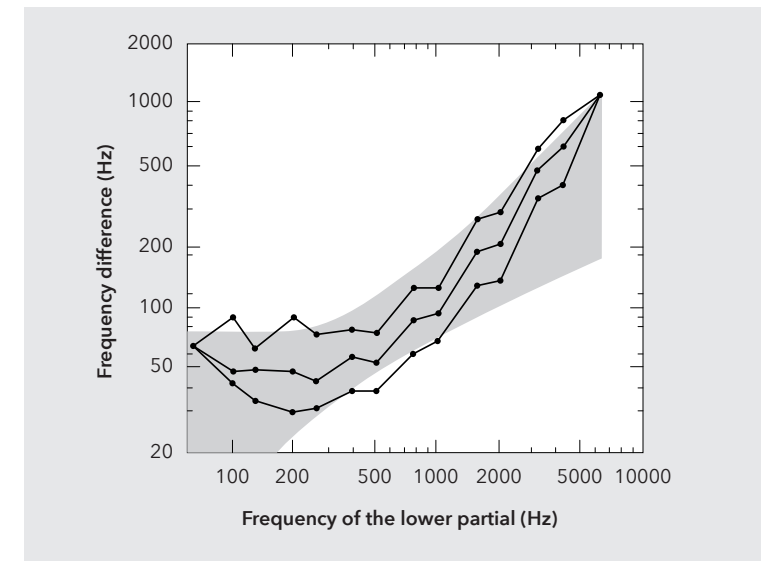
Vergelijking van de laatste twee kolommen toont dat de kwint als gevoeligste consonant uitstekend benaderd wordt, alsook de kwart. De tertsen komen er bekaaid af, zie ook figuur 3.8, hetgeen zwevingen oplevert die duidelijker hoorbaar worden naarmate het akkoord langer wordt aangehouden.

Het is in feite verbluffend dat de verdeling in 12 intervallen zo goed klopt. Verdeelt men het octaaf in 10 of 11 tonen, dan verdubbelt zich de grootste afwijking. Men moet de verdeling opvoeren tot 19, nog beter tot 31 tonen om zuiverder samenklanken te bereiken. De beroemde fysicus Christiaan Huygens (1629-1695) heeft zich met deze problematiek uitvoerig beziggehouden en pleitte voor de verdeling in 31 tonen. Adriaan Fokker (1887-1972), curator van het Fysisch Kabinet van het Teylermuseum te Haarlem, heeft in 1950 een instrument met deze verdeling laten bouwen, dat aldaar nog steeds aanwezig is en kan worden bespeeld.

In de muziek wisselen meer consonante en meer dissonante akkoorden elkaar af. In het eerste geval liggen de gelijktijdig aanwezige deeltonen verder van elkaar af dan in het tweede. We mogen verwachten dat bij consonante akkoorden de afstanden tussen gelijktijdige harmonischen meer in de buurt liggen van de bovenste kromme van figuur 3.7, en bij dissonante akkoorden in de buurt van de onderste kromme. De eerstgenoemde kromme geeft immers aan waar consonantie begint, de tweede waar dissonantie maximaal is.

Met behulp van de volgende procedure is dit voor twee composities statistisch nader onderzocht¹². Om de aanpak duidelijk te maken begin ik met de frequentie 523,2 Hz (c_2). Van een compositie worden alle akkoorden uitgeselecteerd die deze frequentie bevatten. Dat kan zijn als grondtoon, maar ook als tweede

harmonische van een toon van 261,6 Hz (c_1), derde harmonische van 174,6 Hz (f), vierde harmonische van 130,8 Hz (c) enz. Vervolgens wordt nagegaan hoe lang in deze gesommeerde akkoorden de 523,2 Hz toon gepaard gaat met een hogere toon op 1 semitoon afstand, waarbij deze toon ook weer een harmonische kan zijn geweest. Dit wordt herhaald voor tonen op afstanden van 2, 3, 4 of meer semitonen. Op deze wijze verkrijgt men een distributie van toonsafstanden (frequentieverschillen), gerekend vanaf 523,2 Hz, waarvan bepaald kan worden welk percentage van de tijd het frequentieverschil 25%, 50% of 75% was. Deze



Figuur 3.9 De resultaten van een statistische analyse van de frequentieafstanden in akkoorden van een muzikale compositie (Romanze van A. Dvořák's Strijkkwartet Op. 51 in E_b majeur) bij aanwezigheid van 6 harmonischen. De middelste kromme betreft 50% van de afstanden, de andere 25% en 75%. Het gearceerde gebied wordt begrensd door maximale schrilheid en kritieke bandbreedte volgens figuur 3.7.

exercitie kan worden uitgevoerd voor elk aantal aanwezig geachte harmonischen. Voor het geval van 6 meegerekende harmonischen is in figuur 3.9 (pagina 49) het resultaat bij 523,2 Hz zowel als bij een aantal andere frequenties weergegeven.

We kunnen constateren dat de spreiding van meer consonante versus meer dissonante akkoorden opvallend goed overeen komt met het experimenteel bepaalde gebied begrensd door de kritieke bandbreedte en de maximaal dissonante frequentie-afstand. Uiteraard is de keuze van 6 harmonischen willekeurig. Het feit dat bij meer of minder in de berekening betrokken harmonischen vergelijkbare resultaten worden gevonden, bevestigt de zinnigheid van de statistische benadering. Zoals verwacht kon worden, verschuiven de krommen enigszins naar de dissonante kant naarmate meer harmonischen worden meegerekend. Voor een tweede compositie werd het algemene beeld bevestigd. Kortom, de intervallen van muzikale akkoorden zijn goed afgestemd op het frequentie-analiserend vermogen van het gehoororgaan.

Horen met twee oren

In het vorige deel is buiten beschouwing gebleven dat we twee oren hebben. Dit ‘dubbele’ gehoor is niet strikt essentieel voor het horen, maar draagt er wel toe bij. Dankzij de samenwerking van de twee oren zijn we in staat de richting van geluiden vast te stellen, waardoor de waarneming een reconstructie van het geluidsbeeld wordt waarin de bronnen hun plaats krijgen. Bovendien worden gelijktijdige geluiden beter van elkaar onderscheiden. En, last but not least, leiden de veelvuldige reflecties van het geluid in besloten ruimten tot de unieke sensatie van het *ruimtelijk* horen dat het luisteren naar muziek een extra dimensie geeft. Neem het geval van een enkele bron recht van voren. Een verschuiving over een paar graden, overeenkomend met een

tijdverschil van slechts 20 microseconden tussen de oren, blijkt al waarneembaar te zijn. Daarnaast speelt een verschil in geluidniveau ten gevolge van de hoofdschaduw een rol. Verdraait men een storende ruisbron van recht van voren over 90 graden naar rechts of links, dan levert dit voor het verstaan van spraak dat van recht van voren komt een winst van meer dan 15 dB in verstaanbaarheidsdrempel op¹³. Deze aanzienlijke waarde demonstreert hoe belangrijk het binaurale horen in de dagelijkse werkelijkheid is. Gevonden is dat tijdverschil en hoofdschaduw ongeveer gelijke bijdragen leveren.

De genoemde winst betreft de uitzonderlijke conditie van afwezige geluidreflecties en geen extra stoorbronnen. De praktijk is meestal anders. De meest ongunstige situatie is een besloten ruimte met een groot aantal dicht op elkaar staande personen; in de vakliteratuur spreekt men over het *cocktailparty-effect*. Het geroezemoes van het geheel gaat dan samen met de aanwezigheid van meerdere stemmen op korte afstand. Als ruwe vuistregel kan gelden dat van de 15 dB slechts 3 dB overblijft. Hierbij moet nog wel worden opgemerkt dat naast het horen ook ‘liplezen’ (spraakafzien) een rol speelt, dat een verbetering van ca. 4 dB op kan leveren.

In de laatste decennia is er veel aandacht gekomen voor de betekenis van het binaurale horen voor de waardering van de akoestische kwaliteit van concertzalen. Na de Tweede Wereldoorlog was er een grote behoefte aan nieuwe zalen, vooral grotere voor meer dan 2000 personen. Het bleek daarbij dat, ondanks zorgvuldig akoestisch onderzoek, er zalen werden gerealiseerd die niet geheel aan de verwachtingen voldeden. Vergeleken met oude zalen, was het beleven van de ruimtelijkheid duidelijk minder. Uitgebreid onderzoek toonde aan dat de tijd tussen het directe geluid en de eerste weerkaatsing via de zijwanden voor

het ruimtelijk beleven van het geluid een kritische rol speelt. Het optreden van grotere tijdverschillen dan ca. 30 milliseconden, gevolg van bredere zalen dan de vroegere, bleek een negatieve factor te zijn. Men dacht dit te kunnen verhelpen met het ophangen van panelen aan het plafond die met reflecties onder 30 milliseconden de rol van de zijwanden zouden kunnen overnemen. Maar ook dat bleek niet de oplossing te zijn. Waarom voldeden deze klankkaatsers niet? Uitvoerige experimenten leidden tot een verrassende conclusie. Reflecties van boven leveren vrijwel gelijke signalen bij de oren op, resulterend in geringe ruimtelijkheid. De correlatie is veel kleiner als de geluiden van links en rechts komen, dus meer ruimtelijkheid geven. Het was bijna onmogelijk de zalen zo te corrigeren dat de ruimtelijkheid goed tot zijn recht kon komen.

Hoe moeilijk het is een grote concertzaal van uitstekende kwaliteit te ontwerpen kan worden toegelicht met de geschiedenis van de Philharmonic Hall in het Lincoln Center van New York City. De architecten schakelden het gerenommeerde adviesbureau *Bolt, Beranek and Newman* (BBN) in met als resultaat een tamelijk smalle rechthoekige zaal met 2400 zitplaatsen. Tegen het advies in werd het plan vergroot tot ca. 2600 plaatsen. Toen de zaal op 23 september 1962 met veel tam-tam in gebruik werd genomen, bleek deze niet aan de hoog gespannen verwachtingen te voldoen. Bij een bezoek, enkele jaren na de opening, viel mij de ongewoon grote diepte van de zaal op, hetgeen impliceert dat de plaatsen achterin de zaal moeilijk met de gewenste vroege reflecties via de zijwanden te bereiken zijn. Nadat kleine correcties niet voldoende bleken, kon, dankzij de financiële steun van de rijke muziekliefhebber Avery Fisher, een ingrijpende renovatie worden gerealiseerd, waarna de zaal in 1976 opnieuw werd geopend als Avery Fisher Hall. Maar ook daarmee was de kritiek

niet verstomd. Experts concludeerden dat de zaal nu te breed was voor de gewenste eerste reflecties; weerkaatsende panelen bleken niet in staat dit te compenseren en opnieuw moest aan de zaal worden gesleuteld. Het zal duidelijk zijn dat het ontwerpen van een goede concertzaal niet alleen een kunde, maar ook een kunst is. We mogen ons gelukkig prijzen dat de grote zaal van het Concertgebouw te Amsterdam, met de Grosser Musikvereinssaal te Wenen en de Symphony Hall te Boston, tot de beste zalen ter wereld wordt gerekend.

Conclusies

Tonen en spraak die korter dan ca. 200 milliseconden door storend geluid worden gemaskeerd, worden dwars door de storingen heen continu gehoord.

De mengeling van geluidstrillingen worden in de 'black box' van het gehoorproces gescheiden voor zover zij van verschillende bronnen afkomstig zijn, tegelijkertijd samengevoegd voor zover zij van dezelfde bron afkomstig zijn.

De grootste 'dichtheid' van harmonischen in muzikale akkoorden blijkt te liggen tussen de kritieke bandbreedte en ruwweg een kwart daarvan, overeenkomend met consonante resp. dissonante samenklanken.

Bij een spreker recht vóór de luisteraar levert binauraal horen een aanzienlijke winst op als het storende geluid van opzij komt; in geroezemoes van veel stemmen is de winst slechts gering.

4

De perceptie van spraak

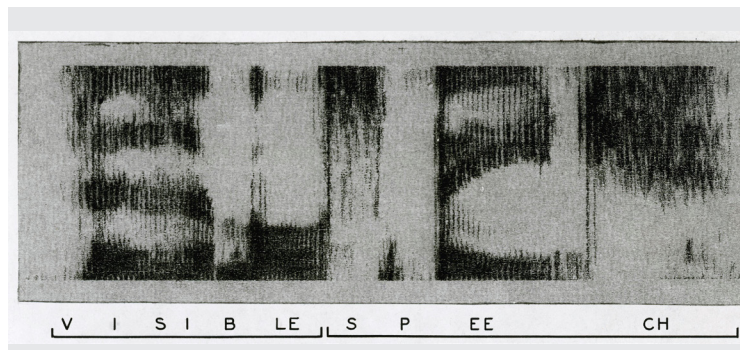
Op zoek naar spraakeenheden

4 De perceptie van spraak

Op zoek naar spraakeenheden

Zonder twijfel is de menselijke stem het belangrijkste geluid dat wij percipiëren. Ruw gesteld, gebruiken we onze ogen voor ruimtelijke oriëntatie en onze oren voor sociale communicatie. Algemeen wordt gedacht dat de blinde ernstiger gehandicapt is dan de dove, maar dat is slechts een oppervlakkige observatie. De geestelijke ontwikkeling van een doof kind wordt zeer vertraagd door het gehoorverlies dat het intermenselijk contact bemoeilijkt. We denken misschien dat lezen de taalkundige achterstand kan compenseren, maar dat is niet het geval.

De betekenis van de spraakcommunicatie is zo groot en kent zoveel aspecten dat ik de behandeling heb gesplitst in twee delen. Er bestaat namelijk een merkwaardig onderscheid tussen het laboratorium en de praktijk: de meeste wetenschappers, met name in de Verenigde Staten, gaan er van uit dat *lopende spraak* uit een opeenvolging van klanken als concrete bouwstenen bestaat en



Figuur 4.1 Spectrogram van de woorden "Visible speech".¹⁴

zijn daarom geneigd de aandacht vrijwel geheel op die eenheden te richten en hun minutieuze bestudering als de sleutel tot het verstaan van spraak te beschouwen. Daarom eerst dit deel over deze basiseenheden en hoe wij deze waarnemen, om in het volgende de aandacht op lopende spraak te richten.

De droom van zichtbare spraak

Op 9 november 1945 kondigde het gezaghebbende tijdschrift *Science* een opzienbarende uitvinding aan die het onderzoek van spraak in één klap op een hoger niveau zou brengen: de *geluidspectrograaf*. Het apparaat was in de kort tevoren geëindigde Tweede Wereldoorlog ontwikkeld door de beroemde Bell Telephone Laboratories. Voor het eerst was het mogelijk het dynamische verloop van spraak visueel te bestuderen. Een serie filters deelt het spraaksignaal op in bandjes van 300 Hz en het uitgangssignaal van elk bandje moduleert de helderheid van een lampje, aangebracht voor een met fosfor bedekte lopende band. Als resultaat geeft de oplichtende fosfor het verloop van het spraaksignaal weer, zie figuur 4.1.

Het is interessant de achtergrond van de uitvinding te kennen. Men hoopte in de oorlog dat het visueel kunnen 'lezen' van spraak zou kunnen bijdragen tot het ontcijferen van vertrouwelijke Duitse radiob berichten. Men zal toen niet geweten hebben dat in de USSR vergelijkbare proeven werden verricht. Alexander Solzhenitsyns autobiografische roman *In de eerste cirkel* onthult dat de auteur na het einde van de oorlog als gevangene in een geheim spraaklaboratorium de taak kreeg telefoongesprekken te analyseren om op deze wijze een Stalin onwelgevallige regeringsfunctionaris op te sporen. Solzhenitsyn noemt de Amerikaanse uitvinding en was met de werking bekend.

Wetenschappelijk gezien was de ontwikkeling van de geluidspectrograaf gebaseerd op de overtuiging dat de klinkers en medeklinkers kunnen worden gezien als de letters van het geschreven alfabet. Daniel Jones begon zijn boek *The phoneme: Its nature and use*, verschenen in 1950, als volgt:

“Bij het spreken denken wij een opeenvolging van geluiden te produceren waarvan de meeste een zekere tijd worden aangehouden; en als we naar spraak luisteren, denken we een overeenkomstige opeenvolging te horen. Het effect is zo stellig voor ons dat we in de regel geen bijzondere moeilijkheden hebben te zeggen wat de geluiden in woorden zijn, of in het toekennen van letters van het alfabetisch schrift.”

Deze één-op-éénkoppeling van spraakklanken en lettertekens was voor Jones vanzelfsprekend. Het toekennen van letters was voor hem hetzelfde als het herkennen van de klanken. Maar is dat terecht? Om een extreem voorbeeld te geven: kan iemand die niet vertrouwd is met het woord ‘angstschreeuw’ dit woord met de zeven op elkaar volgende medeklinkers foutloos opschrijven? Zelf moet ik het woord eerst in gedachten in letters omzetten alvorens het uit te schrijven. Dankzij onze kennis van het schrift kunnen we klanken benoemen. Die verwarring van gesproken en geschreven taal zullen we voortdurend tegenkomen.

Het is niet verrassend dat de geluidspectrograaf met gejuich werd begroet. De oorlog was wel voorbij, maar andere toepassingen leken voor de hand te liggen. De verwachting was hoog dat het apparaat zou kunnen worden toegepast als een spraakomzetter voor doven. In 1947 publiceerde Ralph Potter, de auteur van de boven genoemde aankondiging in *Science*, een dik boek

over de geluidspectrograaf onder de sprekende titel *Visible speech*. Meer dan de helft van het boek was gevuld met lessen hoe geluidspectrogrammen te lezen. Zoals de president van Bell Telephone Laboratories in het voorwoord schreef, zonder de militaire achtergrond te noemen: “Het was de hoop de telefoon geschikt te maken voor de totaal doven dat de ontwikkeling van een instrument voor het zichtbaar maken van spraakgeluiden werd ondernomen.” De slotlinea van het boek onthult het ongeremde optimisme van de ontwerpers, hier verkort weergegeven:

“Als het lukt jonge mensen verstaanbaar te leren spreken en verbeterde versies van zichtbare spraak te lezen als een geschreven taal met hetzelfde gemak als zij hun ABC leren, zal de methode een wijdverbreide toepassing kunnen vinden. De doven zullen de weg wijzen!”

Toen ik in 1966 voor het eerst de betreffende afdeling van Bell Telephone Laboratories in Murray Hill (N.J.) bezocht, was van dit enthousiasme niets meer te merken. Een latere versie van Kay Instruments heeft zijn nut bij spraaktraining bewezen, maar verder is het niet gekomen.

Waarom is het onmogelijk een spectrogram met hetzelfde gemak als geschreven tekst te lezen? Dat is de vraag die we eerst onder ogen moeten zien alvorens enige grip te krijgen op het verstaan van spraak in het algemeen.

Fonemen als bouwstenen?

Zoals eerder gezegd, is de traditie dat gesproken klinkers en medeklinkers, samen aangeduid als *fonemen*, vaste concrete klanken vertegenwoordigen corresponderend met hun alfabetische

symbolen. Laten we ervan uitgaan dat een foneem een specifiek stukje spraak is, een kleinste eenheid. Neem als voorbeeld het woord /bag/ (de streepjes geven aan dat het hier om fonetische symbolen gaat). Het is evident dat we niet na elkaar /b/, /a/ en /g/ als drie van elkaar onderscheiden fonemen uitspreken. Bij de /b/ houden we er rekening mee, dat deze door /a/ gevolgd wordt, met het gevolg dat de klank al kenmerken van de klinker bevat. Evenzo geldt dat de /a/ doorklinkt in de /g/. De mond-keelholte maakt een vloeiende beweging en de drie fonemen zijn in een hoge mate met elkaar versmolten. In vakjargon: we hebben te maken met *coarticulatie*. Aan de luisteraar de taak, ondanks die complicerende coarticulatie, de drie uitgesproken fonemen te identificeren.

Hockett heeft in 1955 een bloemrijk beeld van die taak gegeven; vrij vertaald:

*“Stel u een rij ongekookte eieren voor van verschillende grootte en kleur die op een transportband worden voortbewogen. Op een bepaald punt wordt deze band tussen de rollen van een wringer gevoerd die de eieren plet en met elkaar vermengt. De eieren vóór de wringer representeren de stimuli die het spraakmechanisme aansturen, de troep na de wringer de output van het spraakmechanisme. De inspecteur heeft de taak die troep te onderzoeken en op basis van de in min of meer ernstige mate verminkte dooiers, de verspreide eiwit en de verschillend gekleurde brokken eierschaal te beslissen welke eieren naar de wringer werden geleid. Let wel: hij hoeft de eieren niet te reconstrueren, maar ze enkel te identificeren.”*¹⁵

Hockett ziet de taak van de luisteraar als vergelijkbaar met die van de inspecteur.

In 1987, meer dan 30 jaar later, erkende Michael Studdert-Kennedy (1927-2017), hoofd research van Haskins Laboratories te New Haven (Connecticut) “nog geen vast begrip van coarticulatie te hebben”, kortom het als een raadselachtig proces te beschouwen.¹⁶ De grote aandacht van Hockett zowel als Studdert-Kennedy voor de coarticulatie toont dat beiden de fonemen als de bouwstenen van de spraak zagen en hun identificatie door de luisteraar als een schier onmogelijke opgave. Het is dan kwelend te beseffen dat de mens dit bij lopende spraak continu en zonder moeite doet.

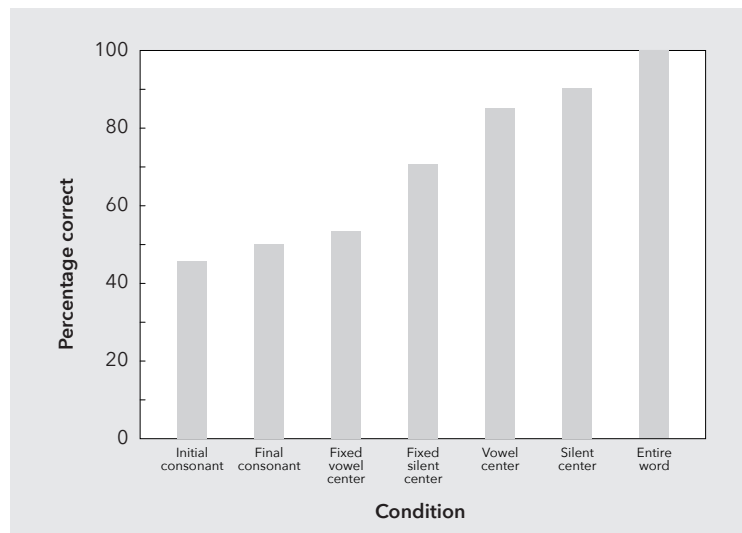
Sommige onderzoekers probeerden het probleem op te lossen door te veronderstellen dat de *perceptie* van fonemen in nauwe relatie met de *productie* moet worden gezien. De spreker heeft de intentie een /a/ uit te spreken en de luisteraar weet uit wat hij hoort die intentie zonder moeite op te pikken. De psycholoog Alvin Liberman (1917-2000), eveneens van Haskins Laboratories, heeft in de 50er jaren die veronderstelling nader uitgewerkt in een theorie die als de “motor theory of speech perception” in de literatuur bekend is geworden. Subtiele verschillen in de uitspraak van fonemen als /p/, /t/ en /k/ werden als evidentie gepubliceerd, maar waren veelal niet erg overtuigend (verdere voorbeelden heb ik elders gegeven).¹⁷ Hoewel vele onderzoekers terugschrokken van deze schimmige theorie, werd de foneem niettemin vrijwel algemeen (althans in de VS) geaccepteerd als de bouwsteen van de spraak, zowel van de productie als van de perceptie.

Coarticulatie als natuurlijk proces

Het voorgaande heeft duidelijk gemaakt dat door de genoemde onderzoekers en vele anderen coarticulatie als een voor de theorie complicerend verschijnsel werd en wordt gezien. Of dit werkelijk zo is, is op verschillende manieren onderzocht.

Cyril Harris (1917-2011), bekend als ontwerper van uitstekende concert- en operazalen (o.a. Lincoln Center, New York), was in 1953 bij Bell Telephone Laboratories betrokken bij onderzoek naar de merites van de geluidspectrograaf.¹⁸ Hij vond bijv. dat als /d/ van /dik/ verbonden werd met /k/ van /dak/ het resultaat niet alleen onnatuurlijk klonk maar feitelijk onverstaanbaar was. Zijn conclusie was dat omringende fonemen wel degelijk medebepalend zijn, dat coarticulatie veel méér is dan ‘beschadiging’. Dit bevestigde de ervaring dat spectrogrammen zo moeilijk leesbaar zijn.

Een andere aanpak betrof de vraag of geïsoleerde fonemen (‘in reïncultuur’) beter geïdentificeerd worden dan als zij door voorafgaande en volgende fonemen zijn ingesloten. Winifred Stran-



Figuur 4.2 Percentage correct geïdentificeerde klinkers waarvan alleen het aangegeven fragment van het woord /d(klinker)t/ aan de luisteraars werd aangeboden. De percentages zijn gemiddelden over 14 klinkers.

ge e.a. (1976)¹⁹ vonden dat 69% geïsoleerde klinkers correct werden herkend tegenover 91% klinkers in lettergrepen. Verder onderzoek toonde dat medeklinkers ‘sec’ al significante bijdragen leveren aan de herkenning van voorafgaande of volgende klinkers, zie figuur 4.2.²⁰ Deze en dergelijke bevindingen bevestigen de veronderstelling dat coarticulatie geen storend artefact is maar een essentieel aspect van de spraak.

Het is overigens opmerkelijk dat men in het verleden zo vanzelfsprekend uitging van het spectrogram als zinnig hulpmiddel voor het bestuderen van fonemen. De stilzwijgende veronderstelling was dat zij gekenmerkt worden door specifieke spectrotemporele eigenschappen. Maar dat is een vergissing. De fonemen onderscheiden zich door de verschillende wijzen waarop zij door het productiemechanisme van de mond-keelholte en de lippen worden voortgebracht. Onderzoek van de individuele fonemen zonder context is een hachelijke onderneming. We spreken plofklanken als de /k/, /p/ of /t/ altijd uit in combinatie met een klinker, juist omdat ze in feite geen eigen bestaan kennen. Daar komt nog bij dat zij voor of na een klinker behoorlijk verschillen. De twee /p/’s van /pip/ zijn identiek in hun productie, maar niet in hun fysische eigenschappen. Er is alle reden om deze fonemen te vergelijken met het inklinken van de toon van een snaar- of blaasinstrument. Dat inklinken kenmerkt het instrument.

Conclusies

Spraak bestaat niet uit een reeks van opeenvolgende losse eenheden, maar uit een continue stroom van klanken, door coarticulatie verbonden. Voorafgaande en volgende medeklinkers zijn doordrenkt van de tussenliggende klinker en dragen bij aan diens identificatie.

5

De perceptie van spraak

De verstaanbaarheid
van lopende spraak

5 De perceptie van spraak

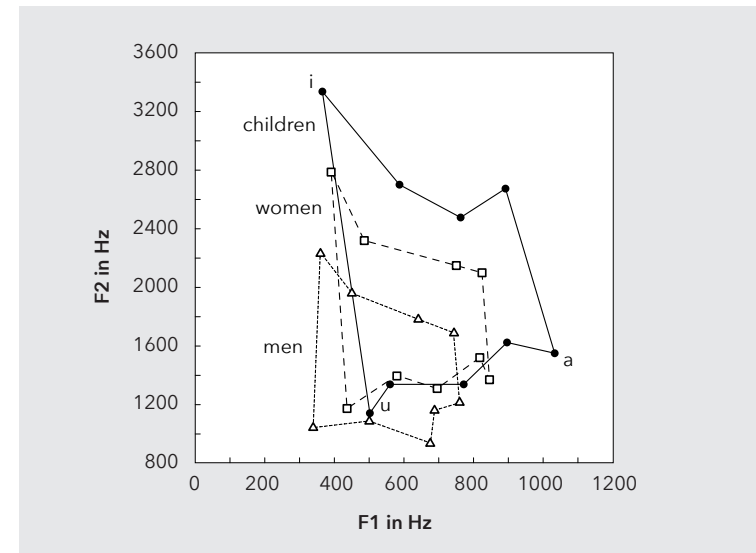
De verstaanbaarheid van lopende spraak

In het vorige deel concludeerden we dat het verstaan van spraak niet kan worden verklaard als het stuk voor stuk identificeren van fonemen als zelfstandige kleine eenheden. De fonemen zijn zó innig met elkaar verweven, werken zó sterk op elkaar in dat we naar grotere eenheden moeten zoeken. We komen al een stuk verder als we de lettergrepen of woorden als basiseenheden zouden mogen zien. Maar ook deze kunnen elkaar sterk beïnvloeden. In tegenstelling tot de geschreven taal, kent de gesproken taal geen ‘spaties’ tussen de woorden. Lopende spraak is inderdaad lopende spraak; vloeiend gaan de klanken in elkaar over. Pas tussen zinnen zijn er pauzes; zinnen bepalen de verstaanbaarheid.

De variaties van het spraaksignaal

De eerste hobbel die we moeten nemen bij het maken van de sprong van fonemen naar zinnen is de kwestie van de grote vrijheid in uitspraak. De bespreking van fonemen kenmerkte zich door de traditionele eigenschappen van het laboratorium: nauwkeurig uitgevoerde en reproduceerbare experimenten met correct sprekende personen. Daar is niets op tegen, integendeel. De vraag is echter wel of we op deze manier resultaten verkrijgen die representatief zijn voor de praktijk. Subtiële spectrale en temporele verschillen tussen in ‘reincultuur’ uitgesproken medeklinkers in /pa/, /ta/ en /ka/ kunnen hoogst interessant zijn voor de onderzoeker, maar zijn zij relevant in de alledaagse conversatie?

We kunnen dit probleem onderzoeken door in de eerste plaats



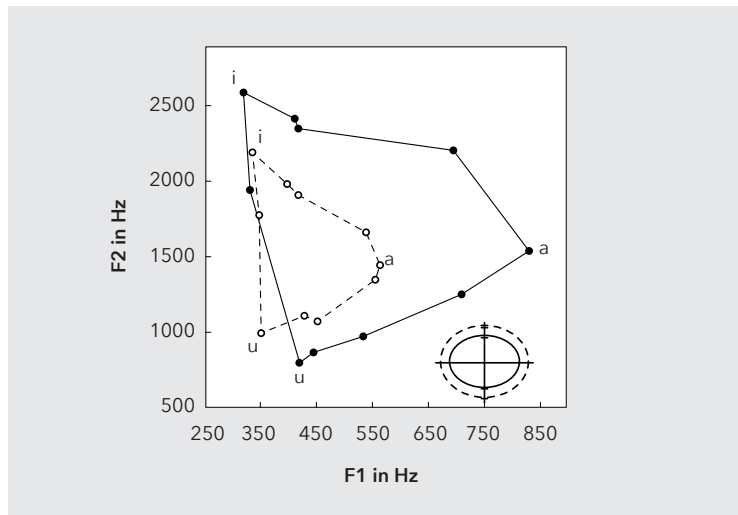
Figuur 5.1 Gemiddelde formantfrequenties F2 versus F1 van klinkers gesproken door mannen, vrouwen en kinderen, gebaseerd op data van Strange e.a. (1976).²¹

eens na te gaan in hoeverre klinkers en medeklinkers, keurig in het laboratorium uitgesproken door verschillende personen, fysisch met elkaar overeenkomen. Helaas is dergelijk onderzoek vrijwel uitsluitend bij klinkers uitgevoerd en hier zal ik mij toe moeten beperken. Figuur 5.1 geeft de gemiddelde frequenties van de belangrijkste twee formanten van klinkers uitgesproken door mannen, vrouwen en kinderen. De klinkers van de drie categorieën onderscheiden zich aanzienlijk in hun spectra. Voor correcte identificatie zijn verdere karakteristieke verschillen tussen mannen-, vrouwen- en kinderstemmen nodig.

De identificatie van klinkers neemt nog aanzienlijk in onzekerheid toe als we de losse klinkers van het laboratorium inruilen voor klinkers in zinnen uitgesproken. Florian Koopmans-van

Beinum (1980) vergeleek o.a. de formantfrequenties van geïsoleerde klinkers met die van klinkers ‘geknipt’ uit conversatiespraak. Figuur 5.2²² geeft het gemiddelde resultaat van twee mannelijke en twee vrouwelijke sprekers. De formantdriehoek in lopende spraak blijkt aanzienlijk kleiner te zijn dan van geïsoleerd uitgesproken klinkers, terwijl de spreiding duidelijk groter is. Ook het percentage correct herkende klinkers liep sterk uiteen van 90% van de geïsoleerde klinkers tegenover 33% van die uit de lopende spraak.

Shocky en Reddy (1975)²³ lieten proefpersonen luisteren naar zinnen in een voor hen vreemde taal. Zij vonden dat vier fonetici slechts 56% van de fonetische symbolen correct konden aan-



Figuur 5.2 Gemiddelde formantfrequenties F1 versus F2 van geïsoleerde klinkers (dichte punten) en klinkers uitgesneden uit lopende spraak (open punten). De ellipsen geven de gemiddelde standaarddeviatie van herhalingen door dezelfde spreker.

geven, bovendien dat slechts 50% van de transcripties overeen kwamen.

Deze drie onderzoeken tonen aan dat de frequentiespectra van klinkers en medeklinkers enorm uiteen kunnen lopen. Opnieuw een bewijs dat het verstaan van spraak niet op de identificatie van fonemen is gebaseerd. We moeten naar grotere eenheden zoeken.

Spectrale en temporele versmering: de rol van coarticulatie

We kunnen een stapje verder gaan door niet alleen naar de enkele foneem te kijken maar ook de direct voorafgaande en volgende bij de identificate te betrekken. In het vorige deel constateerden we dat de geïsoleerde fonemen niet als bouwstenen keurig op een rij liggen, maar in hoge mate verstrengeld zijn. De onderzoekers kwamen er niet onderuit deze menging als coarticulatie te aanvaarden. Kennelijk waren zij met deze complicerende factor niet blij.

In dit deel kijken we er heel anders tegen aan, niet negatief maar positief. Eigenlijk werd dat al bij figuur 4.2 duidelijk: de klinker werd het best geïdentificeerd als de voorafgaande /d/ en volgende /t/ mee werden aangeboden. Als de klinker werd geëlimineerd, gaven de twee medeklinkers samen met een stilte ertussen zelfs het op een na beste resultaat.

Voor het verkrijgen van een dieper inzicht in het proces van woordherkenning zijn reactietijden nuttig gebleken. Savin en Bever (1970)²⁴ vergeleken de tijd nodig voor het identificeren van de beginconsonant van een lettergreep met de tijd voor de gehele lettergreep. Tegen wat men zou verwachten, waren de reactietijden significant korter voor de lettergrepen dan voor de begin-

consonanten sec. Opnieuw een bewijs dat het benoemen van losse consonanten, althans de plofklanken, eigenlijk onmogelijk is, alleen zinvol in combinatie met klinkers.

De positieve rol van coarticulatie komt nog beter tot uiting als we ons niet beperken tot woordjes die met een enkele medeklinker beginnen, maar met twee of meer direct op elkaar volgende, zoals in ‘fris’, ‘sterk’, e.d. We spreken dan van clusters van medeklinkers. Het is opvallend, dat de pleitbezorgers van de fonemen als bouwstenen zich over deze complicatie nooit zorgen maakten.

Het is leerzaam even stil te staan bij hoe jonge kinderen leren spreken. De eerste uitingen bestaan uit een klinker voorafgegaan door een enkele medeklinker (/ba/, /ma/, e.d). Clusters van twee of meer medeklinkers als /bla/ blijken veel moeilijker en worden pas later geproduceerd. Speciale training is soms nodig om moeilijke combinaties correct te leren spreken. Dit is heel evident als het gaat om een vreemde taal. Deze ervaringen wijzen uit dat zowel sprekers als luisteraars weinig bewust zijn van de individualiteit van fonemen, maar de cluster min of meer als één geheel behandelen.

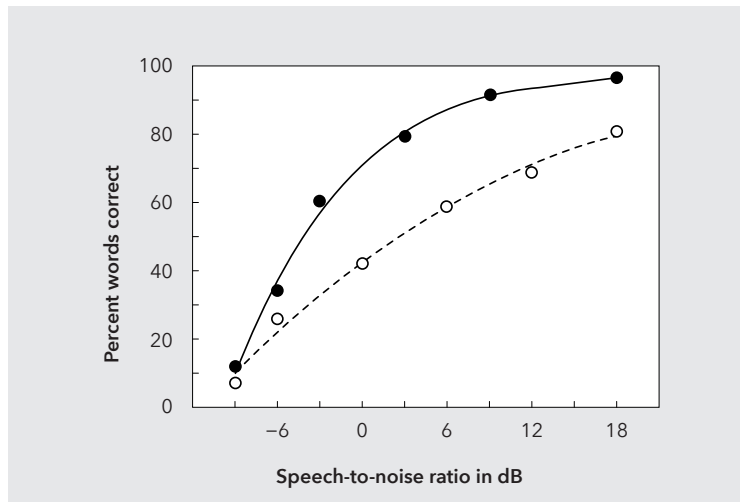
Hier komt het begrip *routine* aan de orde. Lopen, vangen van een bal, besturen van een auto, enz., al deze prestaties zijn routines bestaande uit een groot aantal spierbewegingen die we hebben leren combineren tot een ‘automatisme’ dat als routine wordt aangeduid. Het leren lezen is eveneens een goed voorbeeld. Naarmate we meer ‘geroutineerd’ raken, verschuift de herkenning zich van letters naar woorden als combinaties van letters. Het feit dat men in het verre verleden spaties tussen de woorden is gaan invoeren, illustreert deze overgang.

Zoals gezegd, hebben clusters net als alle routines tijd nodig om in te slijpen. Sommige zijn zo moeilijk dat ze werden en worden gebruikt als een herkenningscode. De theoretisch fysicus M.J.G. Veltman vertelde in een interview nadat hem in 1999 de Nobelprijs natuurkunde was toegekend, dat hij zijn computerprogramma de naam ‘schoonschip’ had gegeven omdat buitenlanders niet in staat zijn de routine ‘sch’ correct uit te spreken.

Van fonemen naar woorden: de rol van de context

Coarticulatie betreft de perceptie van één of enkele medeklinkers samen met een klinker, waarbij het verstaan nog buiten beschouwing blijft. Dit wordt anders als we grotere eenheden onder de loep nemen en ons af gaan vragen wat er gezegd is, waarbij het om herkenning gaat. Daarvoor wordt een beroep op ons geheugen gedaan. Woorden die onduidelijk worden uitgesproken of ons niet zo vertrouwd zijn, hebben de steun van de omringende tekst nodig. De context, de betekenisvolle samenhang van woorden in een zin, is voor het verstaan zeer wezenlijk.

De vroegste experimenten betreffende de significantie van de context werden door William Bagley (1874-1946) uitgevoerd. Met behulp van de door Edison uitgevonden fonograaf werden door hem rond 1900 interessante resultaten bereikt.²⁵ Acht medewerkers van de afdeling psychologie van Cornell University (New York) waren de luisteraars. Zij verstonden een verminkt woord aanzienlijk beter als het als het eerste woord van een zin werd gebruikt; midden in een zin was nog iets gunstiger. Bagley’s onderzoek was geïnspireerd door eerder werk van Cattell met geschreven tekst, hieronder nog te bespreken. Na 1950 begon onderzoek van de context op de verstaanbaarheid pas echt goed op gang te komen, waarbij niet van verminkte woorden wordt uitgegaan, maar de woorden tegen de achter-



Figuur 5.3. Percentage correct verstaane woorden geïsoleerd (open punten) of in zinnen (dichte punten) uitgesproken, als functie van de signaal-ruisverhouding.

grond van ruis worden aangeboden. George Miller en medewerkers (1951)²⁶ constateerden dat woorden in zinnen bij een aanzienlijk lagere signaal-ruisverhouding worden verstaan dan geïsoleerd uitgesproken, zie figuur 5.3.

Picket en Pollack (1963)²⁷ sneden een reeks woorden uit zinnen en vonden dat de woorden beter werden herkend naarmate de reeks langer was.

Het is opmerkelijk dat verschillende experimenten uitwezen dat niet alleen voorafgaande maar ook navolgende woorden aan de correcte herkenning bijdragen. Deze constatering schakelt theorieën uit die de identificatie van fonemen en woorden vergelijken met het raadplegen van een woordenboek waarbij de volgorde van de letters beslissend is.

Een ander interessant gegeven is dat talrijke woorden twee of meer geheel verschillende betekenissen kennen. Horen we zo'n woord in een zin dan bepaalt de context vrijwel automatisch welke betekenis de juiste is, of we op een bank zitten, de bank bezoeken of een ondiepte in rivier of zee bedoelen.

Het is verhelderend hier ook enige aandacht te geven aan de rol van de context bij geschreven taal. James Cattell (1860-1944) was hier een voorloper. Net afgestudeerd, verliet hij Amerika om bij de beroemde psycholoog Wilhelm Wundt (1832-1920) te Leipzig zich verder te oriënteren. Hier publiceerde hij zijn eerste artikel over de minimale belichtingstijden vereist voor het lezen van letters, woorden en zinnen. Cattell constateerde dat het benoemen van letters meer tijd kost dan van gehele woorden; zelfs zinnen werden gemakkelijker gevat dan woorden en letters. In zijn samenvatting van deze bevindingen concludeerde Edmund Huey (1870-1913) al in 1908:

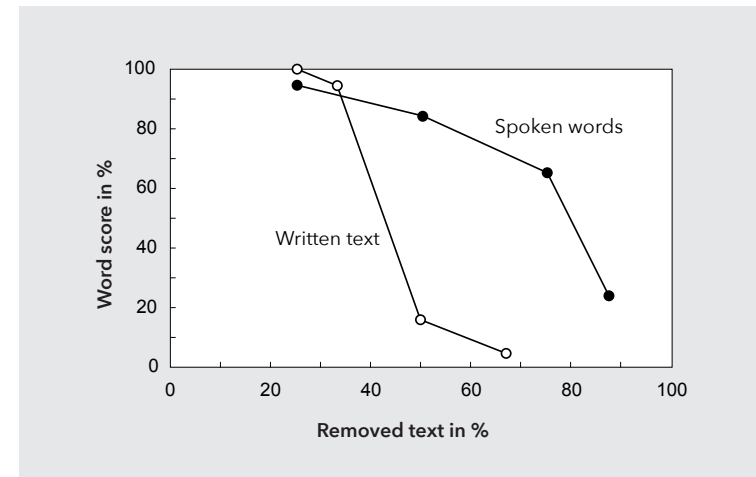
*“Waarom zou een bekend geschreven woord bij het zien niet even gemakkelijk herkend en benoemd worden als een huis of een muur? In de laatstgenoemde gevallen nemen we geen notitie van de verdiepingen en bakstenen als nadere delen, noch van de takken en bladeren bij het herkennen van een bepaalde struik of eikenboom. De compositie, de totale vorm, is het bepalende, zowel in de herkenning van letters, cijfers, woorden of voorwerpen van welke aard dan ook.”*²⁸

Dit commentaar klinkt alsof het gisteren geschreven is. Meer dan een eeuw geleden waren wetenschappers in staat met een primitieve tachistoscoop aan te tonen dat ook voor het *geschreven* woord geldt dat het geheel meer is dan de som der delen. Deze experimenten demonstreren dat er een onweerstaanbare neiging is bekende woorden als gehelen te zien, in plaats van als

een reeks van letters. We moeten de letters kennen om te weten dat MCMXCVI in Romeinse lettertekens hetzelfde is als 1996 in Arabische karakters. Naarmate we meer met het woordbeeld vertrouwd raken, gaan we de tekens als een eenheid zien. Dat is de meest efficiënte wijze van lezen. Kolers (1972)²⁹ heeft erop gewezen dat als een lezer iedere letter een voor een zou moeten identificeren, de snelheid van lezen beperkt zou blijven tot ongeveer 35 woorden per minuut, terwijl ervaren lezers 300 woorden per minuut halen.

Zoals we eerder noemden, is de signaal-ruisverhouding in de loop der jaren veel als criterium voor de verstaanbaarheid gebruikt. In 1963 onderzochten Miller en Isard³⁰ de rol van de context met zinnen die weliswaar uit dezelfde woorden bestonden, maar dan in verschillende volgorde. Grammaticaal correcte zinnen verdroegen een lagere signaal-ruisverhouding naarmate zij als 'zinniger' konden worden opgevat, terwijl niet-grammaticale zinnen de hoogste verhouding nodig hadden.

Het verschil tussen horen en lezen kan worden geïllustreerd met hun kwetsbaarheid voor storingen. Miller en Friedman (1957) lieten proefpersonen teksten van 300 karakters lezen waarvan een aantal was vervangen door onderstreepte spaties (onderstreept om ze niet te verwarren met echte). Figuur 5.4 toont het percentage woorden waarvan de ontbrekende karakters correct werden ingevuld. Ter vergelijking geeft de figuur dito het percentage woorden dat correct werd gehoord als vergelijkbare korte storingen werden ingevoerd. Ondanks het grote onderscheid in de twee situaties, illustreert de figuur dat onderbrekingen veel funester voor geschreven dan voor gesproken taal zijn. De hechtheid van grote verbanden als woorden en zinnen is bij horen veel groter dan bij lezen. Dat blijkt ook uit het feit dat,



Figuur 5.4 De open punten geven het percentage correct gereconstrueerde letters als functie van het percentage periodiek verwijderde letters (gebaseerd op data van Miller & Friedman, 1957)³¹. De dichte punten geven het percentage correct verstaane woorden die periodiek onderbroken werden door stille intervallen ter lengte van fonemen (gebaseerd op data van Miller & Licklider, 1950).³²

terwijl de spraak geen 'spaties' tussen woorden kent, deze voor het vlotte lezen onontbeerlijk zijn.

Van woorden naar zinnen: de rol van redundantie

Net als de samengesmolten klanken meer zijn dan de individuele fonemen, zijn zinnen als gehelen meer dan een reeks woorden. Het is onmogelijk vloeiend hardop te lezen als men de tekst niet kan overzien maar de woorden een voor een na elkaar zichtbaar worden gemaakt. Men moet voortdurend een flink deel van de volgende woorden kennen om de accenten goed te leggen. We hebben een continu verschuivend 'visueel venster' nodig om vlot te lezen. Evenzo kent het horen een 'auditief venster' om

het gesprokene als een soort automatisme te kunnen volgen. De routines komen vol tot hun recht bij zinnen als grotere spraak-eenheden dan lettergrepen en woorden.

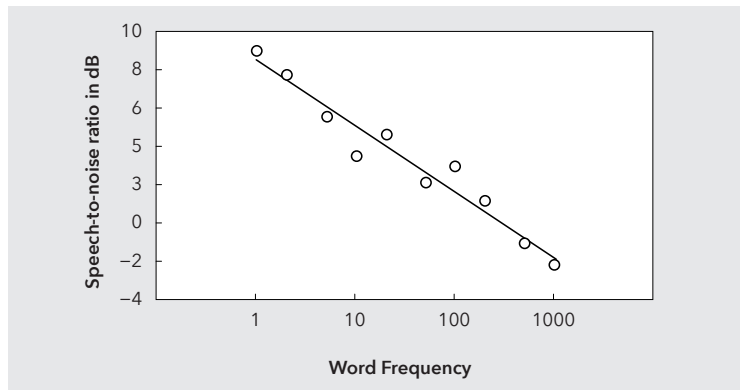
Spraak in de ‘moedertaal’, niet gestoord door andere geluiden, is gemakkelijk te verstaan. Zó gemakkelijk zelfs, dat we lange gesprekken kunnen voeren zonder vermoeid te geraken. Men pleegt in dit verband van *redundantie* te spreken: de lopende spraak bevat onder normale omstandigheden veel meer informatie dan we voor het verstaan nodig hebben. De redundantie impliceert dat de spraak aanzienlijk in informatie kan worden gereduceerd, bijv. door storende geluiden of andere stemmen, voor de ‘drempel’ wordt bereikt. Anders gezegd: de aanwezigheid van een ‘surplus’ kan beschouwd worden alsof de verstaanbaarheid van ongestoorde spraak meer dan 100% is. Het surplus suggereert dat we het spraaksignaal in zekere mate temporeel of spectraal kunnen versmeren waarbij verschillen meer en meer worden uitgeschakeld, zonder de verstaanbaarheid al te zeer aan te tasten.

We beginnen met temporele variaties. Hiervoor splitsen we het spraaksignaal in banden vergelijkbaar met de kritieke bandbreedte en dempen we de fluctuaties in intensiteit voor elke band met een laagdoorlaatfilter dat de omhullende versmeert. De vraag is dan tot welke waarde de filterfrequentie kan worden gereduceerd om zinnen nog net te verstaan. Dit is voor het eerst uitvoerig door Rob Drullman³³ onderzocht. Hij vond dat pas bij een beperking van de frequentie van de omhullende tot ca. 4 Hz de grens wordt bereikt. De verstaanbaarheidsdrempel in ruis ligt bij deze waarde ca. 6 dB hoger dan voor onbehandelde spraak. Deze waarde is groter dan de drempelverhoging van slechthorenden bij aanwezigheid van een of meer storende stemmen

(figuur 5.6), hetgeen betekent dat bij een dusdanige versmering het verstaan van gelijktijdige stemmen praktisch onmogelijk is.

Gaan we over van het temporele naar het spectrale, dan blijkt eveneens een aanzienlijke versmering toelaatbaar: zelfs bij bandbreedten van meer dan een octaaf zijn zinnen in stilte nog correct te verstaan (ter Keurs e.a., 1993).³⁴ Kijkt men op foneemniveau naar de effecten, dan blijkt spectrale versmering vooral de klinkers aan te tasten, terwijl de medeklinkers meer te lijden hebben van temporele versmering. Dit onderscheid is te verwachten: medeklinkers zijn korter en dynamischer dan klinkers.

De redundantie van spraak wordt in hoge mate bepaald door onze vertrouwdheid met de gesproken taal. Onze woordenschat blijkt al op jonge leeftijd enorm toe te nemen. Gevonden is dat een kind van 6 jaar ruwweg 14.000 woorden kent; voor volwassenen loopt de schatting, afhankelijk van de taalvaardigheid, uiteen van 27.000 tot 52.000. De grote vraag is hoe het brein in staat is vrijwel onmiddellijk op grond van de ontvangen informatie het juiste woord te treffen, en dat niet voor een enkel woord, maar bij lopende spraak voor de dichte opvolging van geheel verschillende woorden. De ‘black box’ van het geheugen heeft dit geheim niet prijsgegeven. Onderzoek heeft niet veel meer opgeleverd dan indicaties over de rol van factoren die bij het verstaan een rol spelen. Naarmate woorden vaker in geschreven tekst voorkomen, worden ze aanzienlijk gemakkelijker geïdentificeerd, zie figuur 5.5 (pagina 78). Echter niet alleen de vertrouwdheid speelt een rol maar ook de kans op verwarring met fonetisch vergelijkbare woorden.



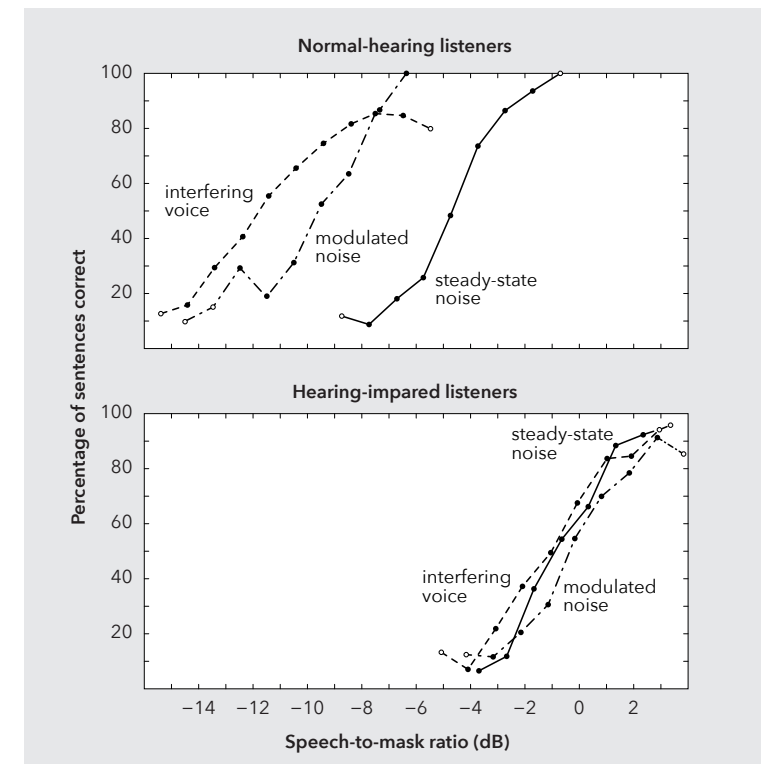
Figuur 5.5 Signaal-ruisverhouding voor het verstaan van 50% woorden als functie van de frequentie van woorden in geschreven tekst (gebaseerd op data van Howes, 1957).³⁵

Invloed van storende geluiden bij normaal- en slechthorenden

De meest voorkomende situatie is dat twee of meer stemmen gelijktijdig worden gehoord. Dan is de ene stem in feite het stoorgeluid van de andere. Zo wordt dit echter niet beleefd. Bij een beperkt aantal kunnen we elke spreker zonder moeite volgen alsof zijn of haar stem niet is aangetast. Toch storen stemmen elkaar zo sterk dat van moment tot moment delen onhoorbaar zijn. Het actieve proces is permanent bezig dit te herstellen.

Ook al lijken die wederzijdse storingen de verstaanbaarheid veelal niet aan te tasten, dan is dit een misleidende constatering. Zoals we zagen, kent de spraak een grote redundantie; dankzij deze overvloedigheid hebben we geen moeite met een in sterkte vergelijkbare storing, zoals een tweede stem. Maar er zijn grenzen. De beste manier om dit na te gaan is het verstaan van zinnen te meten als functie van de signaal-ruisverhouding, waarbij de te onderzoeken stem het signaal is, de storende stem

de ruis. Een uitvoerige studie hierover werd door Joost Festen uitgevoerd, in 1990 gepubliceerd.³⁶ Dat dit artikel na dertig jaar nog steeds tientallen keren per jaar wordt geciteerd (inmiddels in totaal ca. 600) toont dat de publicatie een schot in de roos was en dat het probleem nog onverminderd veel aandacht krijgt, in het bijzonder met betrekking tot de slechthoerendheid. De twee diagrammen van figuur 5.6 vatten de resultaten samen. De curven geven het percentage correct verstaane zinnen als func-



Figuur 5.6 Percentage correct verstaane zinnen als functie van de signaal-ruisverhouding, voor normaalhorende en slechthorende proefpersonen. (Festen en Plomp, 1990)

tie van de signaal-ruisverhouding. Het eerste diagram betreft gemiddelden van 20 jonge normaalhorenden, het tweede van 20 slechthorenden met gehoorverliezen uiteenlopend van 35 tot 58 dB. In beide groepen was het aantal mannen en vrouwen gelijk. We zien dat bij normaalhorenden het type stoorgeluid een groot verschil maakt. In ruis representatief voor het geroezemoes van veel stemmen zijn bij een geluidniveau gelijk aan die van de zinnen deze nog zonder problemen te verstaan. De 'drempel' van 50% verstande zinnen ligt bij ca. -5 dB signaal-ruisverhouding. De verstaanbaarheid neemt zelfs aanzienlijk toe als de ruis in twee banden in sterkte fluctueert, representatief voor de variaties in de spraak. In dit geval ligt de drempel bij een verhouding van ca. -9 dB. Vervanging van de fluctuerende ruis door storende zinnen levert nog ca. 2 dB verbetering op.

De samenvattende eindconclusie is dat de storing van een enkele stem ca. 7 dB gunstiger is dan het geroezemoes van veel stemmen. Als twee personen tegelijk spreken, kost het geen enkele moeite beiden te verstaan; de 100% bij -7 dB illustreert de redundantie. In deze conditie kan van een 'reserve' worden gesproken die kinderen en anderstaligen meer kans geeft. Heel anders ligt het bij storend geroezemoes. In die situatie zal voor alle drie genoemde categorieën gelden dat de spreker moeilijk of niet wordt verstaan; deze conditie is berucht als het 'cocktailparty-effect'.

Voor slechthorenden is storend geluid veel funester. Het type storing maakt vrijwel niets uit; de drie krommen vallen vrijwel samen. Dat de 50%-drempel bij ca. 0 dB signaal-ruisverhouding ligt kan veel verklaren. Het betekent dat een enkele storende stem het verstaan al ernstig aantast. Gelukkig kan dan spraakafzien (liplezen), overeenkomend met een verbetering van ruw-

weg 4 dB in drempelwaarde, nog enig soelaas bieden.

De hoge signaal-ruisverhoudingen die slechthorenden nodig hebben om spraak in aanwezigheid van storende geluiden te verstaan, wordt, zoals nader onderzoek heeft uitgewezen, door auditieve zowel als centrale factoren bepaald. Als auditieve factor kan het verminderd analysevermogen van het gehoororgaan genoemd worden; bij centrale factoren moet gedacht worden aan algemene vermindering van het vermogen gelijktijdige informatie te verwerken.

De krommen in de figuur geven gemiddelden weer. Het maakt echter veel uit of de stemmen al of niet beide van mannen of vrouwen zijn. Voor normaalhorenden geldt dat stemmen van vrouwen beter gescheiden worden dan stemmen van mannen. De beste resultaten werden gevonden bij sprekers van verschillend geslacht. De allerslechtste situatie is zinnen van dezelfde spreker als storing. Deze gegevens wijzen opnieuw uit dat naarmate stemmen meer in hun akoestische eigenschappen (toonhoogte, timbre enz.) van elkaar verschillen, zij beter worden gescheiden en verstaan.

Conclusies

De aanzienlijke spectrale en temporele verschillen in de uitspraak van klinkers en medeklinkers door mannen, vrouwen en kinderen zowel als binnen elk van deze groepen zijn zo aanzienlijk dat de subtiele kenmerken waarin zij zich in het laboratorium onderscheiden in de praktijk van het spreken en verstaan geen rol spelen.

Zowel het spreken als het verstaan worden bepaald door de mate van vertrouwdheid met de taal waarbij de als routines aangeleerde clusters van fonemen essentieel zijn.

De redundantie van lopende spraak is zo groot dat in alledaagse omstandigheden storende geluiden of een tweede gelijktijdige spreker geen problemen opleveren.

Slechthorendheid tast in sterke mate de redundantie van spraak aan, resulterend in een toenemend onvermogen gelijktijdige geluiden te scheiden.



6 Twee tijdsbeelden

Delft in 1946

Als 17-jarige liet ik mij op 6 september 1946 op Oude Delft 95, het deftige oude hoofdgebouw van de Technische Hogeschool, onder nummer 188 inschrijven als student in de technische natuurkunde. Delft was nog een geheel vooroorlogse, misschien wel negentiende-eeuwse stad. Een automatisch telefoonnet was er niet; in de telefooncentrale werden de verbindingen met de hand aangebracht, wat een groot aantal meisjes werk zal hebben verschaft. Het verhaal ging dat Delft als eerste op de lijst had gestaan om na de bevrijding te worden aangesloten, maar vrijwillig die plaats afstond aan Arnhem waar door de oorlogsomstandigheden de centrale was vernield. Elke avond fietsten mannen met lange stokken langs de grachten om de gaslantaarns aan te steken; 's morgens kwamen ze terug om ze weer uit te schakelen.

Maar het ergste komt nog: de binnenstad kende geen riole-ring. Niet dat alles zo in de grachten liep, maar de toiletten in de huizen waren nog ouderwetse plé's met beervaten. Twee keer per week kwamen platte schuiten door de grachten die de gevulde vaten voor schone omwisselden; het was duidelijk beter om dan door de Oostpoort de binnenstad te ontvluchten en de collegezalen op te zoeken. Ik weet dat nog zo goed omdat ik een voorkamertje op het adres Rietveld 131 betrokken had en het ritueel dus van dichtbij meemaakte. Mijn bescheiden verblijf keek uit op het Raam, een straat in een vrij onopvallend en schaars bebouwd gebied met als uitzondering het imposante toenmalige TH-gebouw van de civiele techniek. Later heb ik pas ontdekt dat die leegte een eeuwenoud gevolg was van de ontploffing van het kruithuis in 1654.

Maar nu de natuurkunde zoals ik deze aantrof. Ook hiervan moet gezegd worden dat zij weliswaar niet negentiende-eeuws was, maar toch wel heel sterk aan de dertiger jaren herinnerde. Het moest duidelijk zijn dat het in Delft niet zomaar om de natuurkunde ging, maar om de *technische* variant hiervan. Dat uitte zich in de volgende, voor een latere tijd onvoorstelbare zaken. Het curriculum van het eerste jaar omvatte niet alleen natuur- en wiskundige, maar ook enkele werktuigbouwkundige colleges. Professor E.C. von Pritzelwitz van der Horst leerde ons in het college "stoom" hoe de stoommachine werkte, met inbegrip van de ingenieuze schaar van Stephenson. Een ambachtschoolboekje was gelukkig voldoende om het geheim van de verbrandingsmotor te doorgronden. Maar ook het handwerk moest ons duidelijk maken waar het 'technische' voor stond. Gedurende enige maanden spoedde ik mij 's morgens al vóór 7 uur naar werktuigbouwkunde om een spijker te smeden en een ijzeren cylinder door hakken en vijlen om te vormen tot een kubus (wat uiteraard niet lukte). Dit aspect van de studie werd afgerond met het maken van een werktekening van een kussenblok. Inmiddels was het zo gaan vriezen, dat in februari de kolen op waren en we naar huis werden gestuurd.

Maar deze eigenaardigheden werden twee later ruim vergoed door de colleges theoretische natuurkunde van professor R. Kronig (1904-1995). In zijn jonge jaren had hij zich in de kring van de eerste atoomtheoretici onderscheiden; in 1939 werd hij te Delft benoemd waar hij tot zijn pensionering zou blijven. Zijn Duitse afkomst kenmerkte hem door de wijze waarop hij $1\frac{1}{2}$ kT op onvergetelijke wijze als 'driehalve kathee' uitsprak.



Audiohal TNO Soesterberg

Amerika en Europa

Met de opdracht onderzoek te gaan doen op een mij onbekend multidisciplinair gebied begon ik in 1953 bij de Rijksverdedigingsorganisatie TNO te Soesterberg. Bij de afdeling KNO te Utrecht kreeg ik tijdelijk een kamer. Maarten Bouman was mijn baas, professor C.W. Kosten van de TH had hem mij aanbevolen. Naast de problematiek van spraakoverdracht *in* en gehoorbescherming *tegen* lawaai, kreeg ik ruim baan mij in meer diepgaande vragen betreffende spraak en horen te verdiepen. Tot dit fundamentele onderzoek zal ik mij hier verder beperken.

Behalve de boeken van Helmholtz en Stevens & Davis, was er de *Journal of the Acoustical Society of America* (JASA) om mij op gang te helpen. Het doorbladeren van de Journal toonde dat de

VS het terrein beheerste. Hoewel ASA een onderdeel is van het American Institute of Physics, is de inhoud van het tijdschrift veel breder dan de fysica. Zowel de psychofysica als de (electro) fysiologie van het hoorproces werden in de artikelen bestreken. Voor mij de ideale bron om wegwijs te raken.

Twee zaken vielen mij vooral op. In de eerste plaats dat veel onderzoek stelde op de krachtsinspanning van de kort tevoren geëindigde oorlog. De aandacht in die jaren voor het opsporen van gestoorde of gecodeerde berichten had tot de *signaaldetectietheorie* geleid en dit werkte na de oorlog in de meetmethoden door. In 1954 hadden Wilson Tanner en John Swets de aanpak in de psychofysica geïntroduceerd. Veel researchprogramma's, met name bij Bell Telephone Labs en Massachusetts Institute of Technology (MIT), waren op de signaaldetectietheorie gebaseerd. Men onderzocht bij voorkeur de detectie van geluiden met behulp van de nu algemeen bekende *meerkeuzemethoden* (in het onderwijs door velen verfoeid!). In mijn eigen onderzoek heb ik mij hierbij aangesloten. Dat bracht wel een merkbare verarming van vraagstellingen met zich mee. Door alle aandacht voor de *drempel* was men nauwelijks geïnteresseerd in *bovendrempelige* verschijnselen als toonhoogte en klankkleur, onderwerpen die mij juist sterk interesseerden. Dat verschil in vraagstelling merkte ik pas goed toen ik in 1966 de laboratoria bezocht. Uiteraard waren dit zaken die ik in het begin nog niet geheel onderkende en daarom had ik de neiging mij aan te sluiten bij de Amerikaanse aanpak. Mijn eerste onderzoek getuigt hiervan: de drempel van toonstoten tegen een achtergrond van ruis, met de duur van de toonstoot als variabele, in 1959 gepubliceerd. Zoals verwacht mocht worden koos ik JASA, waar ik immers belangstelling voor dit onderzoek kon verwachten. Het een eeuw eerder verschenen boek van Helmholtz, in het



Manfred Schroeder

Engels vertaald, werd in de VS meer als historisch interessant beschouwd dan als inspiratiebron gebruikt. De eminente psycholoog Edward Boring, die het werk van Helmholtz zo goed kende en waardeerde, sloot zijn “Perspective” als inleiding op het boek van Stevens & Davis af met de woorden: “Zeker zijn we nu klaar voor een nieuwe *Lehre von den Tonempfindungen*”, blijkbaar zonder te beseffen dat alle aandacht in het nieuwe boek beperkt bleef tot de zuiver sinsuivormige toon, een abstractie vergeleken met Helmholtz’ behandeling.

Het leek mij interessant Helmholtz zijn rol als inspirator van het gehooronderzoek terug te geven door zijn *subjectieve* behandeling van tonen te confronteren met de in de VS zo gangbare, op de signaaldetectietheorie gebaseerde, *objectieve* aanpak. Zo begon ik mijn serie experimenten over het frequentie-analyserend vermogen, combinatietonen, consonantie, zwevingen en toonhoogte, vijf onderwerpen die als artikelen in 1964-67 in JASA werden gepubliceerd. Omdat het negentiende-eeuwse Duitse onderzoek in Amerika vrijwel onbekend was, voorzag ik die artikelen van uitvoerige historische inleidingen. Samen vormden zij in 1966 het experimentele deel van mijn dissertatie.

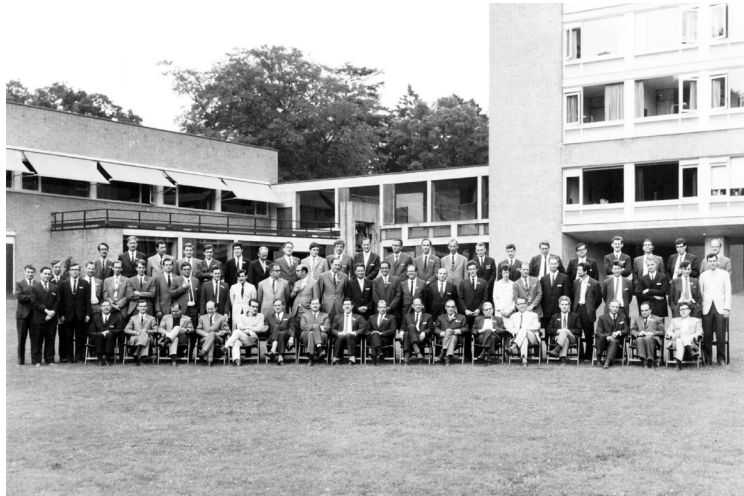
Dat jaar was voor mij de grote omslag. Eerder had ik vrijwel geen buitenlandse vakgenoten ontmoet. De om de drie jaar gehouden akoestische congressen (ICA) waren de enige gelegenheid en daar had ik slechts vluchtig met enkele vakgenoten uit de VS gesproken. Ik besepte dat om grondig op de hoogte te komen van wat daar speelde, ik de vakgenoten in hun laboratoria moest opzoeken. Aan die wens werd royaal tegemoetgekomen. Men gaf mij de gelegenheid na de promotie een ‘grand tour’ van drie maanden door de VS te maken, waarop ik mij voorbereidde door op ruime schaal exemplaren van mijn proefschrift rond te sturen en afspraken te regelen.

Begin september 1966 staken Rita en ik de oceaan over. Na New York trokken we naar het noorden waar ik drie weken besteedde aan de talrijke centra in Bosten en Cambridge. Vandaar ging het via de ‘midwest’ naar San Francisco en LA om via het zuiden en zuid-oosten begin december terug te keren in New York. Een onvergetelijke reis waarbij ik een veertigtal laboratoria bezocht en met een veelvoud aan beroemde en minder beroemde onderzoekers, zowel op het gebied van het gehoor als de spraak, kennismaakte. Overal werd ik bijzonder hartelijk ontvangen en kreeg ik alle gelegenheid de aan de orde zijnde vraagstellingen, de wijze van experimenteren en de beschikbare technische middelen te leren kennen. Ter illustratie pik ik er een aantal centra uit die de verschillende aspecten representeren.

Het spreekt vanzelf om als eerste het bezoek aan de beroemde Bell Telephone Laboratories te Murray Hill (N.J.) te noemen. Jim Flanagan had een heel programma samengesteld die mij gedurende twee dagen langs een groot aantal medewerkers deed rennen. De research van spraak en akoestiek stond onder leiding van Manfred Schroeder die, na zijn promotie in 1954 bij het Drittes Physikalisches Institut te Göttingen, bij Bell Labs was

terechtgekomen (in 1951 had ik zes weken te Göttingen in het genoemde instituut gewerkt en moet hem daar hebben ontmoet). Hij bood Rita en mij een diner aan waar wij voor het eerst kreeft aten, waar hij zo verzot op was.

Heel andere ervaringen waren de bezoeken te Boston en Cambridge. Het eerste betrof Stanley Smith Stevens ('Smitty'), beroemd om zijn 'power law', aan wie wij de kennis danken dat een toename van 10 dB verdubbeling van de luidheid oplevert. Toen ik hem voorzichtig vroeg wat hij vond van nieuwe en meer exacte psychofysische methoden zoals triadische vergelijkingen, bleek hij deze ontwikkeling geheel overbodig te achten. Hij onderwierp mij aan een simpel experiment waarin ik aan de ruwheid van drie stukjes papier een getal moest toekennen zonder deze te zien. Het was stellig een standaardproef voor ongelovige bezoekers; er kwam volgens hem precies de machtsfunctie uit die hij vroeger had vastgesteld. Een hoogst demagogische verto-



Symposium Driebergen

ning die mijn indruk versterkte dat deze wijze van experimenteren haar beste tijd had gehad. Het is interessant dat Stevens na zijn overlijden in 1973 werd opgevolgd door Dave Green, die met anderen juist de toepassing van de signaaldetectietheorie in de psychofysica had geïntroduceerd.

Het tweede bezoek bij George Miller was van geheel andere aard. Miller was een van de eersten op het gebied van de cognitie die ruim oog had voor het holistische karakter van de taal. Ook hij was blijkbaar gewend zijn bezoekers als proefpersoon te gebruiken. Deze twee hoogleraren vormden de top van de Amerikaanse psychofysica, met tal van leerlingen door het land verspreid.

Anderen kenmerkten zich door consequent de signaaldetectietheorie toe te passen. Dat konden lieden zijn met een psycholo-



Symposium Groningen

gische dan wel een biofysisch-electronische achtergrond. Tot de eersten behoorde Wilson Tanner die ik in Ann Arbor aantrof. Hij was een zo fanatieke aanhanger, dat als ik het woord 'drempel' liet vallen, de meisjes die hem bij zijn onderzoek omringden begonnen te giechelen: zij mochten alléén maar de vakterm 'd-prime' gebruiken! Onderzoekers van de tweede groep trof ik in het Research Laboratory of Electronics van MIT te Cambridge bij de medewerkers van de biofysicus Walter Rosenblith aan. Zij werkten toen nog in de vervallen houten barakken waar een kwart eeuw eerder de radar was ontwikkeld.

Hiermee ben ik aangekomen bij de electrofysiologie van de cochlea, waarvan als de bekendste onderzoekers op mijn programma stonden: Jozef Zwislocki (Syracuse), Nelson Kiang (Eaton-Peabody Laboratory, MIT), Peter Dallos (Evanston). De Nobelprijswinnaar Georg von Békésy kon ik helaas niet bezoeken want hij was juist naar Hawaii vertrokken. Wel had ik, na hem mijn proefschrift te hebben gestuurd, een lange brief van hem gekregen.

Voor gehooronderzoek ten behoeve van slechthorenden en doven waren de twee voornaamste centra: het beroemde Central Institute for the Deaf (CID) te St Louis en het Gallaudet College te Washington DC. De research van CID werd door Hallowell Davis geleid vanuit zijn kleine werkkamer waarvan de wanden geheel bedekt waren met getuigschriften en oorkonden, zeer indrukwekkend. Het door Thomas Gallaudet in 1864 opgerichte College, inmiddels Gallaudet University, is de enige universiteit waar het onderwijs uitsluitend in gebaren wordt gegeven. James Pickett, hoofd van de research, leidde mij rond; het viel mij op dat de collegezaal, ondanks het spreekverbod, als *Auditorium* stond aangeduid.



Ira Hirsh

Ik sluit deze rondgang langs het Amerikaanse spraak- en gehooronderzoek af met een tweetal kennismakingen van een meer historisch karakter. De eerste betreft de fysicus Harvey Fletcher die o.a. op het gebied van het verstaan van spraak bij Bell Telephone ruim zijn sporen had verdiend en als trouwe mormoon, teruggekeerd naar zijn geboortegrond, zijn laatste jaren besteedde aan de Brigham Young University te Provo (Utah) met onderzoek van orgeltonen. De tweede 'historische' figuur was Ernst Glen Wever, pionier van de omzetting van mechanische in elektrische stimuli in het oor, mede bekend door zijn boeken, met wie ik in zijn instituut in Princeton kennismaakte.

Het zal duidelijk zijn dat de grand tour voor mij onvergetelijk is geweest. Ik leerde hoe in het 'mekka' van gehoor- en spraakonderzoek werd gewerkt en wat de meeste aandacht had, anders gezegd, wat in de 'mode' was. Er was echter wel enig verschil met mijn voorliefden. Op toengebied hadden naast detectie-drempels ook bovendrempelige verschijnselen mijn belangstel-

ling, op spraakgebied gold dat voor verstaanbaarheid tegenover de subtiele verschillen tussen fonemen, zoals die in Haskins Laboratories van Franklin Cooper te New Haven werden bestudeerd.

Deze eerste persoonlijke ontmoetingen bepaalde mijn verdere nauwe contacten met de VS. Dit moge blijken uit de voorkeur voor JASA als publicatiemedium (50 artikelen als auteur of co-auteur), de frequente deelname met medewerkers en promovendi aan meetings van de ASA en de vele incidentele symposia en researchbesprekingen waaraan ik vanaf 1970 deelnam.

Maar nu eerst het belangrijkste en geheel onverwachte gevolg van mijn reis. In 1967/68 werd mij gevraagd een internationaal symposium over gehooronderzoek te organiseren. Het was een verbazingwekkend aanbod: pas gepromoveerd, niet aan enige universiteit verbonden, en dan gevraagd te worden gekwalificeerde onderzoekers, waaronder topfiguren, uit te nodigen voor een conferentie onder mijn leiding! Het symposium zou worden gesponsord door NATO Science Committee. Na enige aarzeling nam ik het aanbod aan onder voorwaarde dat ik vrij zou zijn in de keuze van deelnemers. Dit werd mij toegezegd. Ik zag het als een ideale gelegenheid om het bescheiden aantal tamelijk geïsoleerd werkende Europese onderzoekers met elkaar en vooral ook met hun Amerikaanse collega's in contact te brengen.

Zo kon met de organisatie worden begonnen. Egbert de Boer (UvA), Roel Ritsma en Ben Lopes Cardozo (beiden IPO, Eindhoven) nodigde ik uit met mij te brainstormen over de invulling van het programma. We stelden een lijst van uit te nodigen personen samen die resulteerde in 15 deelnemers uit VS, 9 uit UK, 10 uit Duitsland en 7 uit andere buitenlanden, met de psy-

chofysica en de fysiologie gelijkelijk vertegenwoordigd. Over het thema kwamen we tot een ideale conclusie: de frequentie-analyse en de detectie van periodiciteit, een problematiek die de psychofysica zowel met de mechanica/hydrodynamica (de cochlea) als met de electrofysiologie (de actiepotentialen) verbond. Daar kon ruime belangstelling voor worden verwacht en zo kwam de titel van het symposium tot stand: "Frequency Analysis and Periodicity Detection in Hearing".

Het symposium vond van 23 tot 27 juni 1969 in het conferentiecentrum Hydepark te Driebergen plaats. Vrijwel alle genodigde sprekers en verdere deelnemers hadden toegezegd. Van de 34 sprekers noem ik de fysiologen J.J. Zwislocki (VS), P.J. Dallos (VS), J.E. Rose (VS), I. Whitfield (UK), W.E. Reichardt (Duitsland), H. Spoendlin (Zwitserland) en de psychofysici E.C. Carterette (VS), I. Pollack (VS), B. Scharf (VS), E. Zwicker (Duitsland), J.F. Schouten (Nederland) en schrijver dezes. Levendige discussies vonden plaats. Andere hoogtepunten waren de demonstratie door André Lehr van het mobiele carillon van Eijsbouts over de raadselachtige toonhoogte van klokken en de tocht met koetsjes langs de Vecht met het einddiner in het Muiderslot, waar Rita zong en Jan Schouten een tafelspeech hield waarin hij de namen van de deelnemers op ludieke wijze met elkaar tot één verhaal verbond. Het langdurige geklap bij de officiële sluiting getuigde van het grote succes van de ontmoeting. Met Guido Smoorenburg verzorgde ik het schriftelijke verslag dat een jaar later werd gepubliceerd.

Het symposium had een onvoorzien, maar verstrekkend vervolg. Besloten werd het niet bij dit eerste contact van Europese onderzoekers te houden, maar te laten volgen door periodieke bijeenkomsten, ook weer het gehele terrein van het gehooron-

derzoek omvattend. Dat dit vervolg een permanent karakter kreeg, bewijst het feit dat in 2018 het achttiende “International Symposium on Hearing” plaats vond, met de negentiende in 2021 te Lyon gepland, meer dan een halve eeuw na de eerste!

Het intensieve contact met de VS is ook voor de klinische audiometrie van grote betekenis geweest. De aandacht voor de perceptie van tonen en spraak leidde mij bij mijn aanstelling aan de VU tot de problemen van slechthorenden in geroezemoes. Bezoeken in de VS met promovendi aan CID, waar Ira Hirsh Hallowell Davis had opgevolgd, Gallaudet University, waar Sally Revoile inmiddels de research leidde, en City University of New York (CUNY) met Harry Levitt en zijn eminente medewerkers waren het gevolg. Samen met Ton Mimpen werd de ‘speech reception threshold’ (SRT) testmethode ontwikkeld, met 50% correcte herkenning van zinnen tegen een achtergrond van ruis als maat. Publicatie van de methode leidde tot overname in verschillende landen. In de VS breidde Sick Soli de test uit tot de ‘hearing in noise test’ (HINT). Onder deze naam vond de testmethode brede toepassing. Sick Soli zelf verzorgde dit in het House Ear Institute te LA voor een grote groep landen zodat de test in meer dan 20 talen in gebruik is gekomen. In 2008 redigeerde Sick een speciaal nummer van *International Journal of Audiology* dat aan mij werd opgedragen en waarvan hij me in Amsterdam het eerste exemplaar aanbood. Dit was een waardig slot van mijn frequente contacten met Amerikaanse collega’s gedurende vele jaren en tevens van mijn bijdrage aan de audiologie.

Noten

Hoofdstuk 2

- 1 Houtgast, T. (1974). *Lateral suppression in hearing*. Dissertatie, VU Amsterdam.
- 2 Schouten, J.F. (1940). *The residue and the mechanism of hearing*. Proc. Kon. Ned. Akad. Wetensch. 43, 991-999.
- 3 Hoogland, G.A. (1953). *The missing fundamental*. Dissertatie, Rijksuniversiteit Utrecht.
- 4 de Boer, E. (1956). *On the ‘residue’ in hearing*. Dissertatie, Universiteit van Amsterdam.
- 5 Pols, L.C.W., van der Kamp, L.J.T., & Plomp, R. (1969). *Perceptual and physical space of vowel sounds*. Journal of the Acoustical Society of America 46, 458-467.
- 6 Kruskal, J.B. (1964). *Nonmetric multidimensional scaling: A numerical method*. Psychometrika 29, 115-129.
- 7 Houtgast, T. & Steeneken, H.J.M. (1971). *Evaluation of speech transmission channels by using artificial signals*. Acustica 25, 355-367.

Hoofdstuk 3

- 8 Miller, G.A. & Licklider, J.C.R. (1950). *The intelligibility of interrupted speech*. Journal of the Acoustical Society of America 22, 167-173.
- 9 Miller, G.A. & Heise, G.A. (1950). *The trill threshold*. Journal of the Acoustical Society of America 22, 637-638.
- 10 Van Noorden, L.P.A.S. (1975). *Temporal coherence in*

the perception of tonen sequences. Dissertatie, Technische Universiteit Eindhoven.

- 11 Warren, R.M. & Warren, R.P. (1970). *Auditory illusions and confusions*. *Scientific American* 223, 30-36.
- 12 Plomp, R. & Levelt, W.J.M. (1965). *Tonal consonance and critical bandwidth*. *Journal of the Acoustical Society of America* 38, 548-560.
- 13 Bronkhorst, A.W. & Plomp, R. (1988). *The effect of head-induced interaural time and level differences on speech intelligibility in noise*. *Journal of the Acoustical Society of America* 83, 1508-1516.

Hoofdstuk 4

- 14 Potter, R.K., Kopp, G.A. & Green, H.C. (1947). *Visible speech*.
- 15 Hockett, C.F. (1955). *A manual of phonology*.
- 16 Studdert-Kennedy, M. (1987). *The phoneme as a perceptuomotor structure*. In A. Alport, e.a. (Eds), *Language perception and production* (pp.667-84).
- 17 Plomp, R. (2002). *The intelligent ear*.
- 18 Harris, C.M. (1953). *A study of the building blocks in speech*. *Journal of the Acoustical Society of America* 25, 962-969.
- 19 Strange, W., Verbrugge, R.R., Shankweiler, D.P. & Edman, T.R. (1976). *Consonant environment specifies vowel identity*. *Journal of the Acoustical Society of America* 60, 213-224.
- 20 Strange, W. & Bohn, O.S. (1998). *Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies*. *Journal of the Acoustical Society of America* 104, 488-504.

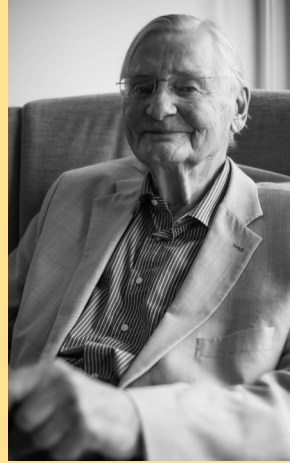
Hoofdstuk 5

- 21 Strange, W., Verbrugge, R.R., Shankweiler, D.P. & Edman, T.R. (1976). *Consonant environment specifies vowel identity*. *Journal of the Acoustical Society of America* 60, 213-224.
- 22 Koopmans-van Beinum, F.J. (1980). *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. Dissertatie Universiteit van Amsterdam.
- 23 Shockey, L. & Reddy, R. (1975). *Quantitative analysis of speech perception. Results from transcription of connected speech from unfamiliar languages*. In G. Fant (Ed.), *Proceedings of the speech communication Seminar*, Stockholm.
- 24 Savin, H.B. & Bever, T.G. (19670). *The nonperceptual reality of the phoneme*. *Journal of Verbal Learning and Verbal Behavior* 9, 295-302.
- 25 Bagley, W.C. (1900-01). *The apperception of the spoken sentence: A study in the psychology of language*. *American Journal of Psychology* 12, 80-130.
- 26 Miller, G.A., Heise, G.A. & Lichten, W. (1951). *The intelligibility of speech as a function of the context of the test materials*. *Journal of Experimental Psychology* 41, 329-335.
- 27 Pollach, I. & Pickett, J.M. (1963). *The intelligibility of excerpts from conversation*. *Language and Speech* 6, 165-171.
- 28 Huey, E.B. (1908). *The psychology and pedagogy of reading*.
- 29 Kolars, P.A. (1972). *Experiments in reading*. *Scientific American* 227, 84-91.
- 30 Miller, G.A. & Isard, S. (1963). *Some perceptual consequences of linguistic rules*. *Journal of Verbal Learning and Verbal Behavior* 2, 217-228.

- 31 Miller, G.A. & Friedman, E.A. (1957). *The reconstruction of mutilated English texts*. Information and Control 1, 38-55.
- 32 Miller, G.A.. & Licklider, J.C.R. (1950). *The intelligibility of interrupted speech*. Journal of the Acoustical Society of America 22, 167-173.
- 33 Drullman, R., Festen, J.M. & Plomp, R. (1994). *Effect of temporal envelope smearing on speech reception*. Journal of the Acoustical Society of America 95, 1053-1064.
- 34 ter Keurs, M., Festen, J.M, & Plomp, R. (1993). *Effect of spectral envelope smearing on speech reception. II*. Journal of the Acoustical Society of America 93, 1547-1552.
- 35 Howes, D. (1957). *On the relation between the intelligibility and frequency of occurrence of English words*. Journal of the Acoustical Society of America 29, 296-305.
- 36 Festen, J.M. & Plomp, R. (1990). *Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing*. Journal of the Acoustical Society of America 88, 1725-1736.

[binnenkant omslag]

Reinier Plomp (1929) heeft met zijn werk grote invloed gehad op de psycho-akoestiek. Zijn indrukwekkende loopbaan van meer dan 40 jaar leverde talrijke publicaties op die nog altijd geciteerd worden. Hij heeft daarmee een stevig fundament gelegd onder, zoals hij dat zelf noemt, de audiologie met een psychofysische oriëntatie. **Horen, een mysterie** is een geactualiseerde uitgave van het in de VS verschenen en nog altijd veelgelezen *The Intelligent Ear: On the Nature of Sound Perception* (2002).



Plomp begon zijn carrière na de Tweede Wereldoorlog bij TNO in Soesterberg. Visuele en auditieve waarneming stonden in de belangstelling, mede ingegeven door de Koude Oorlog. Hij ontwikkelde interesse in spraakverstaan in achtergrondlawaai en de rol van het brein in het horen. Al snel zag hij dat de frequentie-analyse van het binnenoor, hoewel zeer belangrijk voor de hoorfunctie, bij lange na niet kon verklaren hoe horen werkt.

Baanbrekend onderzoek kwam voort uit een meer holistische benadering van horen en perceptie, met inmiddels wijdverbreid geaccepteerde inzichten. Als hoogleraar aan de Vrije Universiteit te Amsterdam begeleidde hij vele promovendi, waarvan enkele ook het hoogleraarschap bereikten en zo het gedachtegoed verder tot ontwikkeling brachten.

Horen, een mysterie is geschreven voor diegenen die in de klinische audiologie werkzaam zijn en belangstelling hebben voor de beginselen van het horen.