

Stembronmodellering: een overzicht

Bert Cranen

Vakgroep Taal en Spraak, Katholieke Universiteit Nijmegen

Inleiding

Stem is een voor de communicatie onmisbaar instrument. Tevens is het zo alledaags dat vrijwel niemand zich bewust is van de complexe processen die zich afspelen bij stemgeving. Voor de gelukkige bezitter van een goed functionerende stem is dat nauwelijks een bezwaar te noemen. Echter, voor degenen die geacht worden hulp te bieden aan patiënten met een minder goed functionerende stem, is een gedegen fundamentele kennis over het stemapparaat een begerenswaardig maar moeilijk te verkrijgen goed.

Dat die kennis moeilijk te verkrijgen is, heeft direct te maken met de relatieve ontoegankelijkheid van het stemapparaat. Hierdoor wordt het verzamelen van meetgegevens over het functioneren onder natuurlijke omstandigheden bemoeilijkt. Verder blijkt voor een fundamenteel begrip van stemvorming kennis nodig uit verschillende domeinen (akoestiek, aero-dynamica, weefselmechanica, anatomie, etc) die maar ten dele in isolatie bestudeerd kunnen worden.

Kennis over stem blijkt geconcentreerd bij twee beroepsgroepen. De eerste bestaat uit hulpverleners in de klinische praktijk; de tweede beroepsgroep wordt gevormd door mensen uit de automatiseringswereld die zich er op toelagen kunstmatige spraak te genereren. Op hen rust de taak een computer uit te rusten met een stem die aangenaam en weinig mechanisch klinkt (zoals bijvoorbeeld ten behoeve van hulpmiddelen voor visueel gehandicapten of telefonische informatiediensten die 24 uur per dag bereikbaar moeten zijn).

De kennis die over stemgeving is opgebouwd is vaak vastgelegd in de vorm van *modellen*, meestal een stel wiskundige relaties die de *belangrijkste aspecten van het stemproductieproces* op een kwantitatieve manier met elkaar in verband brengen. Er blijkt een enorme variëteit aan stembronmodellen te zijn. Daar zijn op zijn minst twee redenen voor aan te geven. Ten eerste hangt het van de doelstelling van de onderzoeker af wat als "de belangrijkste aspecten" ervaren wordt. Ten tweede is het van belang te onderkennen dat voor de ontwikkeling van een kwantitatief model meetgegevens nodig zijn waarmee het model te ontwikkelen is en waarmee de adequaatheid van het model vast te stellen is. In de

praktijk blijkt echter dat van veel eigenschappen die men idealiter zou willen modelleren geen betrouwbare metingen te verkrijgen zijn. Dit is met name waar als het gaat om in-vivo metingen van de mechanische eigenschappen van het stemplooiweefsel en om gedetailleerde, drie-dimensionale afmetingen van de stemplooi-geometrie als functie van de tijd.

De rekenmodellen die ontwikkeld zijn om de verschillende aspecten van stemgeluidsproductie te kunnen beschrijven en onderzoeken, kunnen in dit artikel onmogelijk allemaal met enige mate van diepgang besproken worden. Daarom zullen we eerst een globale inventarisatie maken van de meetmethoden die toepasbaar zijn om de processen die zich afspelen bij fonatie in kaart te brengen. De achterliggende gedachte hierbij is dat dat bevindingen met de verschillende soorten modellen die in de literatuur gerapporteerd worden alleen op waarde geschat kunnen worden indien er inzicht bestaat in de mogelijkheden en onmogelijkheden om een gegeven model te toetsen. Daarna zal een korte, globale schets gegeven worden van het soort modellen en beschrijvingstechnieken die ontwikkeld zijn om verschillende relevant geachte aspecten van het stemgevingsproces te beschrijven. Tenslotte zullen we als illustratie nader ingaan op een specifiek model waarmee gepoogd wordt de relatie tussen de bewegingen van onvolledig sluitende stemplooien en resulterend stemgeluid nader te onderzoeken. Het zal blijken dat dit model voospelt dat niet de aan- of afwezigheid van een lekopening bepalend is voor de kwaliteit van het stemgeluid, maar eerder het al of niet raken van de processu vocalis.

Fysische aspecten van stemgeluid

Het spraakgeluid zoals we dat kunnen registreren met een microfoon, wordt gevormd door de *luchtdruk* variaties bij de mondopening. Die drukvariaties ontstaan wanneer de *luchtstroom* vanuit de longen door de trillende stemplooien periodiek wordt onderbroken. *De primaire bron van het spraakgeluid is de (pulserende) luchtstroom bij de stemplooien.* De spraakklank die uiteindelijk wordt waargenomen hangt af van de stand van de articulatoren die de vorm van de mond/keel-holte (*aanzetstuk*) bepalen.

Het blijkt mogelijk om het spraakgeluid te ontleden in twee componenten: (a) een component die veroorzaakt is door het aanzetstuk, en (b) een component die louter het gevolg is van de trillende stemplooien. Die laatste component wordt aangeduid met *stemgeluid*.

Voor veel mensen omvat stem veel meer dan alleen een akoestisch fenomeen. In dit artikel zullen we echter de volgende werkdefinitie hanteren: *Stem is het akoestische verschijnsel dat optreedt wanneer de stemplooiweefsels interageren met de luchtstroom vanuit de longen.*

Het zal duidelijk zijn dat wanneer men de generatie van het stemgeluid met een model in detail wil nabootsen, men inzicht zal moeten hebben in tal van fysische processen die veel verder gaan dan akoestiek alleen. Vragen die meteen

als relevant in het oog springen zijn: Wat is de de samenhang tussen luchtdruk en luchtstroming? Waardoor raken stemplooien in trilling en waar halen die stemplooien hun energie vandaan om te trillen? Hoe beïnvloedt de stemplooitruilling het stemgeluid? Heeft de vorm van het aanzetstuk ook invloed op de luchtstroom tussen de stemplooien en wellicht zelfs op de stemplooitruilling?

Het begin van het antwoord op deze vragen ligt uiteraard bij metingen. Zodra men op basis van de eerste metingen een model heeft opgebouwd kan men door vergelijking van modeluitkomsten met werkelijk waargenomen verschijnselen proberen de gebouwde modellen te verfijnen. Dat is de standaard cyclus die bij modellering meestal doorlopen wordt. Door meer of minder eisen te stellen aan het aantal verschijnselen dat door een model adequaat beschreven c.q. verklaard dient te worden kan men het model meer of minder ingewikkeld maken. In de navolgende secties zullen we verschillende fysische sub-domeinen nader uiteenrafelen en zo een beeld opbouwen van de processen die voor meting en modellering in aanmerking komen.

Stromende lucht en geluid

Het meest eenvoudige spraakproductiemodel dat men zich kan voorstellen is een *stembron* die een *pulserende luchtstroom* genereert gevolgd door een *filter* dat het aanzetstuk modelleert. We zullen dit model dat ook ten grondslag ligt aan de meeste spraaksynthesatoren nader leren kennen onder de naam *lineair bron/filter-model*.

Luchtdruk en luchtstroom zijn ten nauwste met elkaar verbonden. Toch dient men een scherp onderscheid te maken tussen die twee. Luchtstroom (aangeduid met het symbool U) is het aantal cm^3 lucht dat er per seconde op een bepaalde plaats voorbij komt en is direct gerelateerd aan de hoeveelheid *kinetische energie* die in de lucht is opgeslagen. Luchtdruk (aangeduid met het symbool p) wordt uitgedrukt in het aantal N/m^2 of vaak ook nog in cm water druk en is gerelateerd aan de hoeveelheid *potentiële energie* die in de lucht zit opgeslagen. Omdat de totale hoeveelheid energie (kinetische + potentiële) niet spontaan kan vermeerderen, betekent dit in de praktijk dat een luchtdruk-variatie altijd gepaard moet gaan met een luchtstroomvariatie en omgekeerd.

Hoewel de meesten bij luchtstroom zullen denken aan het netto transport van luchtdeeltjes ten behoeve van de zuurstofopname door het lichaam, is het belangrijk om te onderkennen dat deze grootheid behalve de ademstroom ook een akoestische component bevat. De luchtstroom bevat naast een langzaam fluctuerende component dus ook snelle variaties die op elke plaats in het spraakproductiesysteem dezelfde informatie bevat als het druksignaal. Daarmee zij overigens niet beweerd dat de golfvormen van de twee grootheden altijd in een simpele relatie tot elkaar staan. Op één plaats in het spraakproductiesysteem is er echter wel een hele simpele relatie tussen akoestische druk en luchtstroom en daar wordt bij spraakproductiemodellering dankbaar gebruik van gemaakt. Bij de lippen is de druk nl. bij benadering gelijk aan de eerste afgeleide van de luchtstroom:

$$P_{mond} = \frac{dU_{mond}}{dt} \quad (1)$$

Deze bijzondere omstandigheid, gecombineerd met de observatie dat er ook een relatief eenvoudige relatie bestaat tussen de luchtstroom bij de stemplooien en de luchtstroom bij de lippen, maakt dat het stemgeluid bestudeerd kan worden door te kijken hoe de luchtstroom bij de stemplooien verloopt (of, als de relatie tussen glottale luchtstroom en luchtstroom aan de lippen bekend is, door de druk c.q. luchtstroom aan de mond te meten).

Het is een misvatting te denken dat *stromingsenergie* (dat is, de kinetische energie die zit opgeslagen in de bewegende luchtdeeltjes) altijd vrij komt in de vorm van geluid. Bedenk bijv. dat vrijwel alle mensen uitstekend in staat zijn om nagenoeg geluidloos in en uit te ademen. De kinetische energie van de luchtdeeltjes (of althans een deel daarvan) komt pas vrij in de vorm van hoorbare *akoestische energie* wanneer er een *voldoende snelle variatie in die luchtstroom* wordt bewerkstelligd. Wanneer er fluctuaties in de luchtstroom veroorzaakt worden, moeten deze variaties in kinetische energie volgens de wetten van de fysica gepaard gaan met variaties in potentiële energie (lees drukfluctuaties). Het zijn die drukvariaties die de toehoorder waarneemt als geluid mits de snelheid van de fluctuaties zich in de juiste frequentieband (ca. 20–20000 Hz) bevinden.

Luchtstroomvariaties kunnen regelmatig of juist onregelmatig van aard zijn. Stochastische (random) luchtstroomvariaties en bijbehorende luchtdrukvariaties treden op wanneer de luchtstroom een obstructie of een sterke vernauwing op zijn weg tegenkomt waardoor er turbulenties ontstaan. Bij spraakproductie wordt van dit principe op systematische wijze gebruik gemaakt om stemloze spraakklanken zoals /f/, /s/, /χ/, /p/, /t/, /k/ te genereren. Soms ook is het een ongewild verschijnsel zoals bij zwaarademende sprekers.

Regelmatige luchtstroomvariaties treft men tijdens spraakproductie aan wanneer de stemplooien trillen. De luchtstroom wordt door de trillende stemplooien met een min of meer vaste frequentie periodiek onderbroken. Hierdoor ontstaat het geluid dat in de wandeling stem genoemd wordt en dit geluid vormt de basis onder de zogenaamde stemhebbende spraakklanken, zoals klinkers (/a/, /o/, /u/, /y/ etc.), de sonoranten /w/, /l/, /j/, en de nasalen /m/ en /n/ etc. Uiteraard kunnen ook spraakklanken geproduceerd worden waar zowel een periodieke component als een stochastische component in zit. Voorbeelden van dat soort klanken zijn /v/, /z/, /b/, /d/.

Als het de open en dicht gaande stemspleet (*glottis*) is die kan worden aangemerkt als de primaire oorzaak van het stemgeluid, is het een essentiële vervolgvraag op wat voor manier de veranderingen in *stemspleet-geometrie* corresponderen met bepaalde karakteristieken van het stemgeluid. Uiteraard zou men ook graag begrijpen hoe het komt dat de stemplooien in trilling kunnen geraken en blijven.

Merk op dat de oorspronkelijke akoestische vragen langzaam zijn omgevormd in vragen over het mechanische gedrag van de stemplooiweefsels.

Trillende stemplooien

Het gedrag van de stemplooien is het beste te begrijpen wanneer men ze beschouwt als een *mechanisch resonantie-systeem*, analoog aan een schommel, een veer met een gewicht eraan, of de tuien van de Rotterdamse Erasmusbrug. Al dit soort systemen hebben de eigenschap dat als je er op goed gekozen tijdstippen energie aan toevoert maar ze voor de rest van de tijd de vrije loop laat, de systemen in een trilling raken waarvan de frequentie van de trilling bepaald is door de mechanische eigenschappen van dat systeem. Zo ook bij stemgeving: wanneer men de stemplooien opvat als een elastisch materiaal waar op gezette tijden een duwtje tegen gegeven wordt, dan raakt het elastische materiaal in trilling (een zogenaamde eigentrilling) waarbij de trilling bepaald is door de vorm en de mechanische eigenschappen van het weefsel. Omdat de stemplooien drie-dimensionale structuren zijn, kunnen de eigentrillingen ook componenten in drie dimensies hebben waardoor de bewegingen er zeer complex uit kunnen zien. Door bepaalde spieren aan te spannen of juist te ontspannen, veranderen de vorm en/of mechanische eigenschappen van het elastische materiaal en daarmee het trillingsgedrag. Voor een meer gedetailleerde beschrijving waarop verschillend trillingsgedrag kan worden geanalyseerd in termen van eigentrillingen wordt verwezen naar Titze (1994).

Zoals gezegd, de trillingsfrequentie en de trillingswijze hangen primair af van de weefseleigenschappen. De precieze manier waarop tijdens elke trillingscyclus energie aan het stemplooiweefsel wordt toegevoerd doet er dan ook niet zozeer toe, als het maar gebeurt. Immers, elke trilling is bedoeld om uit te sterven tenzij er van tijd tot tijd voldoende energie wordt toegevoerd om de onvermijdelijke verliezen te dekken. Om tijdens een trillingscyclus energie aan de stemplooien toe te kunnen voeren is het niet alleen noodzakelijk dat kracht wordt uitgeoefend op de stemplooien, maar zoals Titze (1994) duidelijk maakt, zeker zo belangrijk is dat deze kracht netto in dezelfde richting is als beweging van de stemplooien zelf (dus naar buiten tijdens de uitwaartse beweging, en naar binnen tijdens de inwaartse beweging). Uiteraard mag gedurende delen van de trillingscyclus het krachtenverloop wel tegengesteld zijn aan de bewegingsrichting, zolang er maar een *netto energieoverdracht naar de stem plooieweefsels* toe is.

De enige kracht die op de stemplooien kan worden uitgeoefend, wordt geleverd door de druk die er heerst in de langstromende lucht. Als de druk groter is dan de tegenkrachten van het weefsel (de resultante van elastische veerkracht en de massastraagheidskrachten ten gevolge van de stemplooien die nog in beweging zijn), dan worden de stemplooien uiteen gedreven. In het andere geval zullen de stemplooien naar elkaar toe bewegen. De wet van energiebehoud (voor stationair stromende gassen ook wel de wet van Bernoulli genoemd) zegt dat elke luchtstroomverandering vergezeld moet gaan van een drukverandering. Als een gas sneller stroomt (een hogere kinetische energie heeft) moet dat ten koste gaan van de druk. Als het gas gedwongen wordt langzamer te stromen (bijvoorbeeld door de zich sluitende stemplooien), levert dit een hogere druk op. Men kan zich

voorstellen dat onder gunstige timing-condities het drukverloop in de langsstromende lucht de stemplooien in trilling kan brengen en dat die trillingen door de drukvariëaties die erdoor opgewekt worden zichzelf in stand kunnen houden: de stemplooien vormen een mechanische oscillator.

De condities waaronder de stemplooien in trilling kunnen komen en blijven, zullen beïnvloed worden door tal van factoren. We noemen er twee. De eerste factor is de mate waarin de onder- en bovenrand van de stemplooien uit fase kunnen bewegen, een eigenschap die ten nauwste gerelateerd is aan de stijfheid van het weefsel. In een stemspleet die tijdens een glottale cyclus van convergerend naar divergerend verandert is de drukverdeling anders dan in een stemspleet die een uniforme doorsnede heeft.

De tweede factor is de vorm van het aanzetstuk. De drukken die opgewekt worden in de holten onder en boven de stemplooien bepalen uiteraard ook in sterke mate de drukken tussen de stemplooien. Omdat de massatraagheid van de lucht die in het aanzetstuk zit is, bepalend is voor het gemak waarmee de luchtstroom kan worden geblokkeerd en weer op gang gebracht, moet verwacht worden dat de akoestische belasting van het stemapparaat ook invloed heeft op de condities waaronder het in trilling komt.

Verschillende categorieën stembronmodellen

Uit het bovenstaande wordt duidelijk dat voor een werkelijk compleet stembronmodel een simulatie vereist is van het luchtstromingsgedrag, het trillingsgedrag van de stemplooiweefsels, de krachten die deze trillingen opwekken, alsmede van de akoestische fenomenen die door de combinatie van al deze factoren veroorzaakt worden. Modellen waarbij dat wordt nagestreefd worden meestal aangeduid met *zelf-oscillerende stembronmodellen* en behoren over het algemeen tot de meest complexe in hun soort. Voorbeelden van dit type model zijn het twee-massa model (Ishizaka & Flanagan, 1972) en het body-cover model (Titze, 1994).

Het vergt geen erg diepgaand fysisch inzicht om vast te stellen dat bij dit soort modellen, zelfs al maakt men nog zulke simplistische aannamen, het aantal onbekenden in het mathematische vergelijkingstelsel dat moet worden opgelost snel oploopt. Daarnaast moet vanwege de omvang van het te modelleren domein heel veel bekend zijn over parameters waarmee de eigenschappen van de stemplooiweefsels beschreven kunnen worden, zoals veerkracht, massa, verliezen ten gevolge van inwendige wrijving (visceuze demping), afmetingen, gelaagdheid in meerdere weefseltypes en de invloed daarvan op het mechanische gedrag etc. Aangezien de meeste van deze gegevens niet of slechts bij heel grove benadering in-vivo te verkrijgen zijn, kunnen dit type modellen voorlopig dan ook slechts gebruikt worden voor het onderzoeken van tamelijk globale kenmerken van het stemgevingsproces. Extrapolaties van het gedrag van deze modellen naar specifieke situaties in de werkelijkheid moeten met grote voorzichtigheid gedaan worden.

Een tweede categorie stembronmodellen wordt gevormd door de klasse die men zou kunnen aanduiden met *parametrische akoestische stembronmodellen*.

Het betreft hier zuiver descriptieve modellen waarbij de parameters van het model de luchtstroom aan de glottis beschrijven. Het doel van deze categorie modellen is het akoestische stemgeluid, los van articulatie, zo treffend en zuinig mogelijk te karakteriseren. De parameters zijn eerder perceptief gemotiveerd dan dat geprobeerd wordt de fysica van het productieproces te beschrijven. Voorbeelden hiervan zijn bijvoorbeeld het veelgebruikte Liljencrants/Fant-model (Fant, Liljencrants & Lin, 1985) en het Klatt-model (Klatt & Klatt, 1990).

Tenslotte is er nog een categorie die een soort middenpositie inneemt tussen de reeds genoemde. Hierbij wordt de weefselmechanica buiten beschouwing gelaten en wordt alleen de relatie tussen stemspleetgeometrie en resulterend stemgeluid in ogenschouw genomen. Er wordt aangenomen dat men een zinnige schatting kan maken van het verloop van het stemspleetoppervlak en men probeert geen antwoord te geven op de vraag door welk krachtenspel die stemspleetgeometrie tot stand is gekomen (daarbij dus ook meteen negerend hoe die op zijn beurt weer beïnvloed zou kunnen worden door de opgewekte drukvariaties). Een goed voorbeeld van een dergelijk model is beschreven door Titze (1984).

Wanneer weefselbewegingen onderwerp van onderzoek zijn, is de klasse van zelf-oscillerende stemplooi modellen het aangewezen instrument. Voor meer akoestisch getinte onderzoeken bevatten dit soort modellen naar mijn mening echter te veel niet controleerbare vrijheidsgraden en zou men zich moeten beperken tot een wat minder ruim domein. In de rest van dit artikel zullen we daarom eerst heel kort ingaan op de vraag wat voor meetgegevens te verkrijgen zijn met niet al te invasieve methoden om daarna via het veelgebruikte lineaire bron-filtermodel te komen tot fysisch meer realistische modellen.

Directe en indirecte metingen aan het stemapparaat

Directe observatie van stemplooi bewegingen en stemspleetgeometrie tijdens fonatie is mogelijk met behulp van een endoscoop (star of flexibel). Met behulp van stroboscopische belichtingstechnieken, of in sommige gevallen beter nog, door de endoscoop te koppelen aan ultra-snelle CCD-camera's kan men het trillingsgedrag van het stemplooiweefsel binnen een glottale periode bestuderen. Een fenomeen dat direct opvalt wanneer men naar vertraagd afgespeelde opnamen van fonatie in borstregister kijkt, is het feit dat de onderkant van de stemplooiën in fase voorlopen op de bovenkant (veelal aangeduid met de term *verticaal faseverschil*). Dit fenomeen lijkt een cruciale rol te spelen in de verklaring waarom stemplooiën in trilling komen en blijven onder normale fonatie-condities en heeft daarom veel aandacht gekregen in allerlei zelf-oscillerende stembronmodellen. Helaas is dit verschijnsel erg moeilijk op quantitative wijze in kaart te brengen. Ten eerste omdat bij directe visuele observatie van bovenaf tijdens delen van de glottale cyclus geen weefselbewegingen van de onderste delen van de stemplooiën zichtbaar zijn en ten tweede omdat het erg moeilijk is om gecalibreerde oppervlaktemetingen te doen via een endoscoop.

In die gevallen waar men voldoende heeft aan informatie over het verloop van de oppervlakte van de stemspleet (die is immers bepalend voor de lucht-

stroom die zich kan ontwikkelen), is de complete beeldinformatie van een endoscoop overbodig en kan men zijn toevlucht nemen tot fotoglottografie. Men dient zich wel te bedenken dat wat voor gewone beelden geldt, ook voor een fotoglottogram geldt: het is onduidelijk welk deel van de lichtmodulatie heeft plaats gevonden door de onderkant en welk deel door de bovenkant van de stemplooien en een calibratie van het fotoglottogramsignaal in termen van cm^2 glottisoppervlak behoort al helemaal niet tot de mogelijkheden.

Door gebruik te maken van andere afbeeldingstechnieken zoals ultrageluid, röntgen of MRI zou men kunnen hopen de stemplooien van de zijkant te observeren en zo verschillen in weefselbewegingen tussen onderkant en bovenkant van de stemplooien wel in absolute termen zichtbaar te maken. Echter door tal van tegenwerkende factoren, waar we in het kader van dit artikel niet nader op in kunnen gaan, zijn dit tot nu toe geen methodes gebleken waarmee voldoende quantitative gegevens onder min of meer natuurlijke omstandigheden vergaard kunnen worden. Een andere techniek waarmee men vanaf de zijkant naar de stembanden 'kijkt' is de electroglottogram (EGG) meting. Omdat een EGG het contactoppervlak tussen de twee stemplooien representeert en geen onderscheid kan maken tussen contact dat veroorzaakt is door horizontale dan wel verticale bewegingscomponenten van het weefsel, is ook het meten van een EGG een veel te indirecte techniek om gedetailleerde gegevens over weefselbewegingen te verkrijgen.

Tenslotte, voor het meten van luchtdrukken tijdens spraakproductie zijn er minatuur druktransducers die kunnen worden ingebracht zonder de spraakproductie al te zeer te verstoren en waarmee drukvariaties met frequenties tot 5000 Hz zonder problemen gevolgd kunnen worden (Cranen & Boves, 1985). Voor de meting van luchtstroom (dat is de ademhalingscomponent plus de akoestische component) zijn helaas geen transducers beschikbaar die rechtstreeks het gewenste signaal opleveren en die bovendien voldoende klein en robuust zijn om (via de neus) in het spraakproductieapparaat te worden ingebracht. Voor luchtstroommetingen neemt men daarom meestal zijn toevlucht tot het meten van de orale luchtstroom waaruit men via inverse filtering de luchtstroom door de glottis schat (zie volgende paragraaf).

Modellering van de akoestiek

Het feit dat het onmogelijk lijkt om met voldoende tijdsresolutie alle details van de stemplooibewegingen te observeren, vormt samen met de vaststelling dat het toch de variërende luchtstroom door de stemspleet (in het vervolg aangeduid met *glottale luchtstroom*) is welke de primaire bron van het stemgeluid vormt, extra aanleiding om voorlopig te focussen op de details van die luchtstroom in samenhang met de luchtdruk ter plaatse.

Directe observatie van de luchtstroom ter hoogte van de stemplooien is niet mogelijk zonder de normale spraakproductie te verstoren. Daarom wordt meestal

een indirecte meettechniek gebruikt: *inverse filtering*. Het principe van deze methode is gebaseerd op het lineaire bron/filter-model dat hierna beschreven wordt. De basisgedachte achter de methode is dat men vrij nauwkeurig weet welke fysische principes ervoor zorgen dat de golfvorm van de luchtstroom bij de stemplooiën verandert op zijn weg naar buiten door het aanzetstuk. Bijgevolg kan men dan ook de omgekeerde weg kan bewandelen en een goede schatting krijgen van de luchtstroom bij de stemplooiën door de luchtstroom aan de mond te meten. Hiervoor zijn namelijk wel meetinstrumenten beschikbaar in de vorm van gezichtsmaskers waarin een pneumotachograaf is opgenomen (bijvoorbeeld Rothenberg, 1977). We zullen het lineaire bron/filter-model in een later stadium tevens gebruiken als het vertrekpunt voor meer verfijnde akoestische modellering.

Het lineaire bron/filter model

Om spraakproductie te kunnen nabootsen op een computer is naast een stembron uiteraard ook een model van de keel-, neus- en mondholte (*aanzetstuk*) nodig. De meest gebruikte modellering van het spraakproductie-systeem is er een waarbij het aanzetstuk wordt opgevat als een lineair filter dat het stemgeluid omvormt in de uiteindelijke spraakklank. Er is geen wederzijdse beïnvloeding tussen bron en filter.

De term lineair slaat op het feit dat het spraaksignaal aan de lippen op elk moment wordt berekend als een lineaire combinatie van luchtstroomwaarden op eerdere momenten aan het uiteinde van het aanzetstuk (bij de lippen) mits de luchtstroom aan het begin van het spraakkanaal (bij de glottis) op dat moment zelf maar bekend is. Voor het frequentie-domein heeft dit tot gevolg dat het amplitudespectrum van het spraaksignaal (mits uitgedrukt in dB) kan worden opgevat als de som van een bron-spectrum en de overdrachtskarakteristiek van het aanzetstuk (zie figuur 1). Deze mathematisch uiterst aantrekkelijke ontleding in additieve componenten is in overeenstemming de intuïtieve notie dat we in spraak een fonatiecomponent en een articulatie-component kunnen onderscheiden die onafhankelijk van elkaar te beregelen zijn.

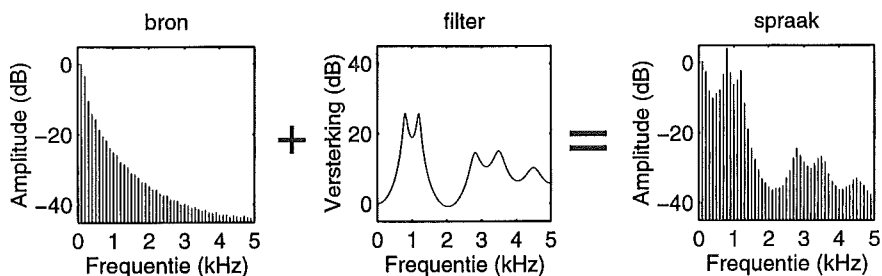


Fig. 1. Bij het lineaire bron/filter-model wordt het amplitudespectrum van het signaal aan de lippen (uitgedrukt in dB) opgevat als de som van een bronspectrum en een filter-overdrachtskarakteristiek die het aanzetstuk representeert.

Het is de ontleding in additieve spectrale componenten die ten grondslag ligt aan het inverse filter procedé. Het spectrum van een binnenkomend spraaksignaal (of luchtstroom-signaal bij de mond) wordt ontleed in twee componenten: één component waarmee het trendmatige verloop in de energie als functie van de frequentie verdisconteerd wordt (dit deel wordt gedefiniëerd als het bronspectrum) en één component die de spectrale variaties rond de trend (formantstructuur) verdisconteert. Merk op dat de scheiding in een bron- en een aanzetstuk-component alleen eenduidig kan wanneer het gerechtvaardigd is om te veronderstellen dat alle spectrale structuur van het aanzetstuk afkomstig is. Zodra er redenen zijn om te veronderstellen dat het bronspectrum naast het trendmatige verloop ook een zekere eigen spectrale structuur heeft, dient men externe criteria te hebben om uit te maken welk deel van de spectrale structuur in het signaal aan de lippen moet worden toegerekend aan de bron en welk deel aan het aanzetstuk.

Laten we het lineaire bron/filter-model gebruiken om te illustreren hoe bepaalde golfvormkenmerken van de luchtstroom bij de stemplooien van belang zijn voor het stemgeluid. In figuur 2 en 3 zijn door de computer berekende, kunstmatige voorbeelden te zien van een stel luchtstroompulsen als functie van de tijd (gestyleerd volgens het Liljencrants/Fant-model). Deze luchtstroompulsen zouden model kunnen staan voor pulsen zoals die ontstaan bij klinkerproductie door een man. Beide figuren tonen aan de linker kant (van onder naar boven) (a) de luchtstroom bij de stemplooien; (b) de eerste afgeleide van dat signaal (dat is, de snelheid waarmee de luchtstroom bij de stemplooien verandert); en (c) het spectrum van (b). Aan de rechter kant van de figuren ziet men (d) de luchtstroom aan de lippen; (e) de eerste afgeleide van dat signaal wat bij benadering gelijk is aan de druk bij de lippen, d.i. het spraaksignaal zoals men dat met een microfoon aan de mond zou meten; en (f) het spectrum van (e). Duidelijk waarneembaar is hoe zich iedere 10 ms een zelfde soort gebeurtenis afspeelt. Het luchtstroomsignaal is een *periodiek signaal* waar in dit geval ca. 5 ms de luchtstroom vrijwel nul is (de stemspleet is dicht), terwijl de volgende 5 ms de luchtdeeltjes in beweging zijn (door de zich openende en sluitende stemspleet op weg van de longen naar de buitenlucht). Het verschil tussen figuur 2 en 3 zit vooral in de manier waarop de onderbreking van de luchtstroom t.g.v. de stemspleetsluiting plaats vindt: zeer abrupt bij figuur 2 en veel geleidelijker bij figuur 3.

De specifieke "klankkleur" die bij deze luchtstroomverschijnselen hoort, is het beste te visualiseren met behulp van een spectrum (zie afbeeldingen (e) en (f) in de betreffende figuren). Een spectrum laat zien op welke manier de geluidsenergie verdeeld is over lagere en hogere frequenties. Een eigenschap van een periodiek signaal is dat het een spectrum heeft waarbij alleen frequentiecomponenten aanwezig zijn die een frequentie hebben welke een veelvoud is van de trillingsfrequentie van de stemplooien. In de figuren 2 en 3 is goed te zien dat de luchtstroomsignalen en spraaksignalen een lijnspectrum hebben waarbij de frequentie-componenten exact dezelfde frequenties hebben, maar waarbij de sterkte (amplitude) van de frequentiecomponenten vooral in de hoge frequenties verschilt. Zoals ook intuïtief goed aanvoelbaar is (snelle veranderingen vereisen

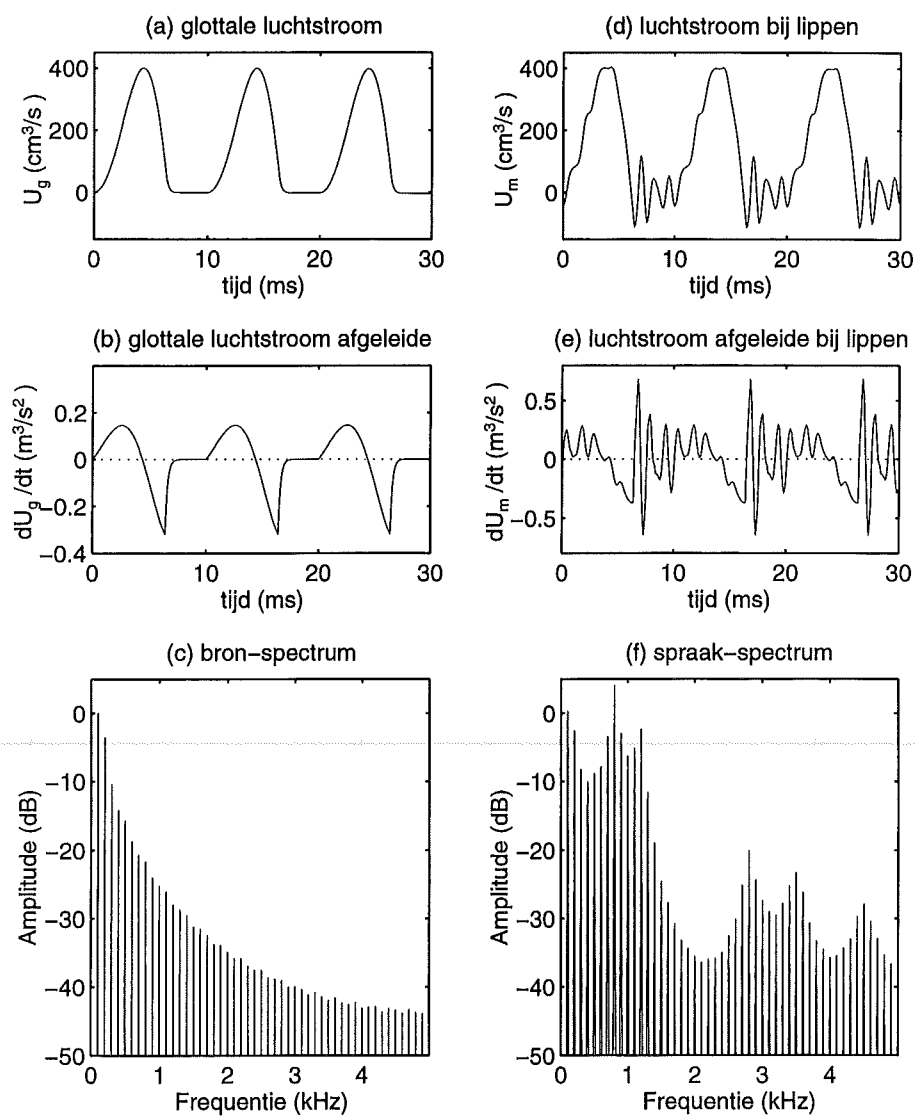


Fig. 2. Met behulp van lineair bron/filter-model gesynthetiseerd voorbeeld van (a) luchtstroom bij de glottis; (b) de eerste afgeleide van de luchtstroom bij de glottis; (c) het bron-spectrum; (d) de luchtstroom aan de lippen; (e) de eerste afgeleide van de luchtstroom aan de lippen (bij benadering equivalent aan het spraaksignaal) en het spraakspectrum. De luchtstroom wordt abrupt onderbroken; de opgewekte formanttrillingen zijn sterk.

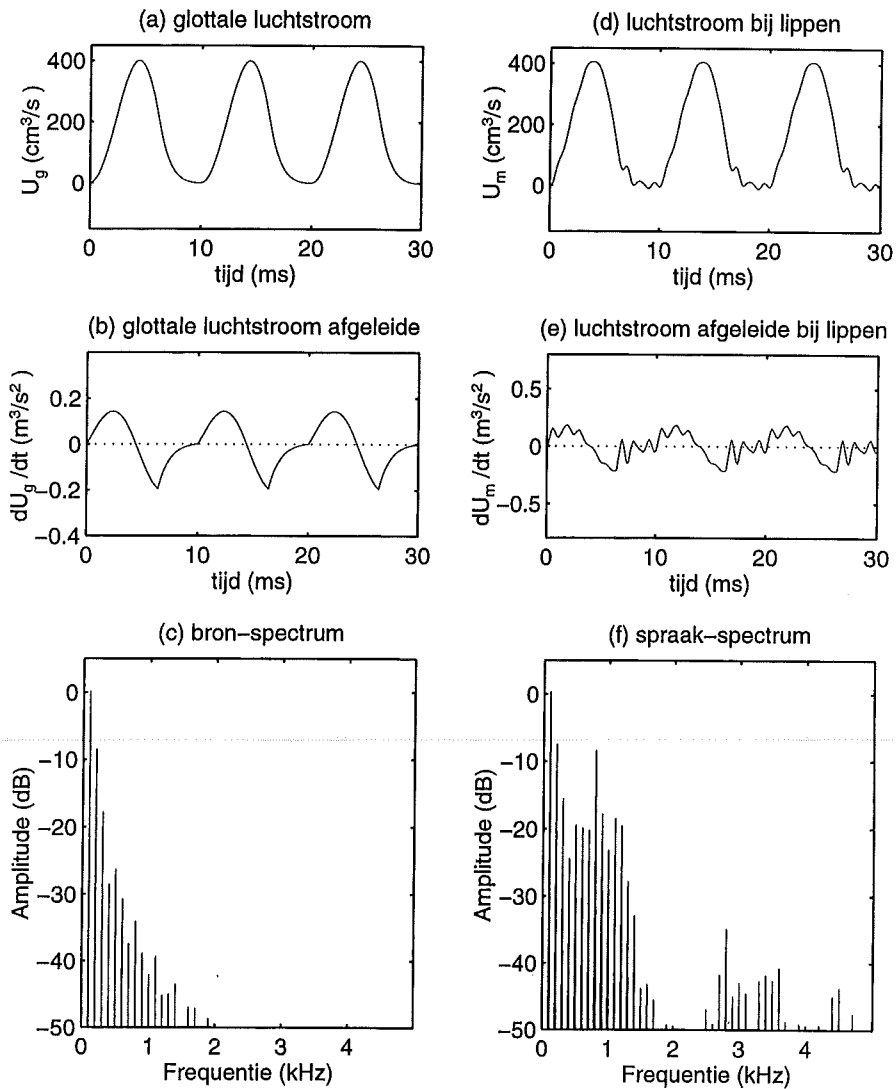


Fig. 3. Met behulp van lineair bron/filter-model gesynthetiseerd voorbeeld van (a) luchtstroom bij de glottis; (b) de eerste afgeleide van de luchtstroom bij de glottis; (c) het bron-spectrum; (d) de luchtstroom aan de lippen; (e) de eerste afgeleide van de luchtstroom aan de lippen (bij benadering equivalent aan het spraaksignaal) en het spraakspectrum. De luchtstroom wordt geleidelijk naar nul gebracht; de formanttrillingen in het spraaksignaal zijn daarom veel zwakker dan in figuur 2.

snel wisselende, dus hoog-frequente componenten) blijkt dat de sterkte van de hoge frequentiecomponenten sterk samenhangen met de abruptheid waarmee de luchtstroom geblokkeerd wordt. Veel energie in de hoge frequenties correspondeert met een heldere of (bij al te veel energie) een scherpe schelle klank, terwijl een spectrum waarbij de energie voornamelijk in de lage frequenties is geconcentreerd overeenkomt met een ronde of zelfs doffe klank.

De stembron die we in het rekenexperiment gebruikten bevatte geen ruis en zal alleen daardoor al een erg kunstmatige indruk teweeg brengen. In de praktijk zal er altijd wel enige mate van turbulentie zijn en deze zal in het spectrum zichtbaar worden als ruisenergie ook op andere frequenties dan een veelvoud van de grondfrequentie (dus ook tussen de harmonischen). Of die ruis-energie ook daadwerkelijk waarneembaar is en dus duidelijk als een bepaalde stemkarakteristiek (bijv. als heesheid), zal naast de absolute sterkte van de ruiscomponent ook afhangen van de mate waarin het ruisgeluid gemaskeerd wordt door de periodieke componenten (de harmonischen) van het stemgeluid. Met name wanneer de glottissluiting erg geleidelijk plaats vindt (de harmonischen in de hoge frequenties zijn dan laag in amplitude), moet men verwachten dat de ruisenergie snel hoorbaar zal worden. We komen hier later nog op terug.

Niet-lineaire interacties

Men dient zich ervan bewust te zijn dat het lineaire bron/filter-model slechts een eerste orde benadering is. Omdat het eindproduct zonder invasieve metingen direct te registreren is (het zij met een microfoon, het zij met een masker waarmee de luchtstroom bij de neus/lippen gemeten wordt), is het in veel gevallen ook in een klinische setting aantrekkelijk om een deel van de diagnose te baseren op een schatting van de luchtstroom bij de stemplooiën verkregen met inverse filtering. Toch dient de interpretatie met de nodige behoedzaamheid te worden gedaan.

Aan de methode van inverse filtering ligt, zoals eerder gezegd, de aanname ten grondslag dat men weet hoe het signaal aan de mond moet worden ontleed in een stembron-component en een aanzetstuk-component. Een veelgebruikte aanname is dat het spectrum van de bron zelf geen noemenswaardige spectrale structuur heeft en dat alle spectrale structuur die men in het mondsignaal ziet moet worden toegerekend aan het aanzetstuk. Wanneer de gemaakte vooronderstellingen geen stand houden, wordt de interpretatie een stuk moeilijker. Het blijkt dat de omstandigheden waaronder dat kan voorkomen samenhangen met de opgewekte drukken en de grootte van de stemspleet. Hieronder zullen we een en ander nader verkennen.

Wanneer men gericht zijn taak wil kunnen uitvoeren om iets te veranderen aan het akoestische eindproduct (het stemgeluid), dient de beschrijving van de luchtstroom tevens een fysiologische interpretatie toe te laten in termen van onderliggende stemplooibewegingen. Immers alleen dan zal men in staat zijn conclusies te trekken over de oorzaak van een eventueel niet optimaal stemgeluid en gerichte pogingen kunnen doen om het trillingsgedrag van de stem-

plooien te beïnvloeden. Wat hiervoor minimaal nodig is, is een model dat beschrijft hoe stemspleet geometrie samenhangt met de resulterende luchtstroom.

Al in de jaren vijftig heeft Van den Berg (1957) laten zien dat zolang de luchtstroom stationair verondersteld kan worden, geldt dat de luchtstroom door de glottis (U_g) evenredig is met de oppervlakte van de stemspleet (A_g) en de wortel uit de druk die over de stemspleet staat (P_p). In formulevorm:

$$U_g(t) = A_g(t) \sqrt{\frac{P_p(t)}{k\rho}} \quad (2)$$

Hierin is k een empirische konstante en ρ stelt de dichtheid van de lucht voor (en is dus ook als een constante te beschouwen).

Lange tijd heeft men gedacht dat de druk over de stemspleet (P_p) min of meer constant was, zodat men (zie vgl. (2)) kon aannemen dat de luchtstroom op ongeveer dezelfde manier verliep als de oppervlakte van de stemspleet. Latere experimenten hebben echter aangetoond (Cranen & Boves, 1985; Kitzing & Löfqvist, 1975) dat die druk zeer grote schommelingen kan vertonen met name vlak voor en na het moment van stemspleetsluiting. Dit suggereert dat de glottale luchtstroom wel eens helemaal niet zo onafhankelijk zou kunnen zijn van het aanzetstuk-filter als is aangenomen bij het opstellen van het lineaire bron/filter-model. Wat geldt voor normale spraak, geldt in versterkte mate voor zang (Schutte & Miller, 1996, 1987).

Deze luchtdrukschommelingen zijn het gevolg van een combinatie van twee factoren. Ten eerste wordt door de trillende stemplooien het netto luchttransport beïnvloed en volgens de wet van energie-behoud in drukvariaties omgezet. Als de stemplooien open zijn zakt de druk, en als ze dichtgaan stijgt deze weer. Ten tweede is er een akoestisch effect wat optelt bij het eerste. Door de plotseling veranderende luchtstroom (mede door de inertie van de lucht in het aanzetstuk is die verandering met name bij het sluiten van de stemplooien abrupt) worden er staande golven opgewekt in zowel de sub- als supraglottale holten. Hierdoor ontstaan er grote slingeringen in de subglottale en supraglottale druk, en dus in het verschil van die twee: de transglottale druk. De frequentie ervan is bepaald door de formantfrequenties van de genoemde holten (zie figuur 4)

Het gegeven dat het verloop van de luchtstroom bepaald is door de transglottale druk (zie formule (2)), maar dat op zijn beurt het verloop van de transglottale druk weer bepaald is door de manier waarop de resonantie-holten worden aangeslagen door de luchtstroom, maakt dat de simplificatie dat bron en filter onafhankelijk en lineair scheidbaar zijn in sommige gevallen een oversimplificatie is. In die gevallen waar men uit luchtstroomgegevens conclusies zou willen trekken over stemploobewegingen, is een goed inzicht in de omstandigheden waaronder men nog wel en niet van het lineaire bron/filter-model gebruik kan maken van essentieel belang.

Zoals uit echte metingen van de transglottale druk blijkt, zijn de formantgebonden trillingen in de transglottale druk het sterkst gedurende het gesloten

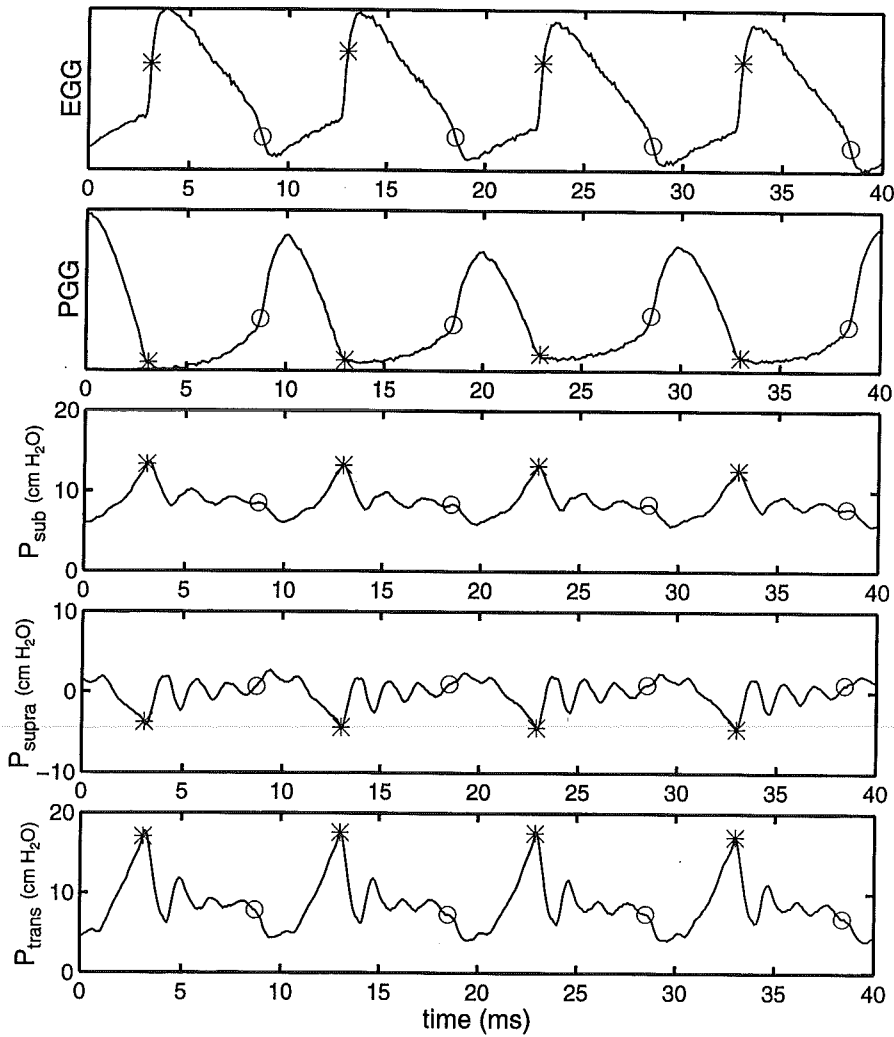


Fig. 4. Wanneer tijdens fonatie de druk gemeten wordt met behulp van een miniatuur drukopnemer, zijn zowel in de subglottale druk (P_{sub}) als in de supraglottale druk (P_{supra}) aanzienlijke variaties waar te nemen die veroorzaakt zijn door akoestische resonantie. Het drukverschil over de stemspleet ($P_{trans} = P_{sub} - P_{supra}$) vertoont zijn piek ongeveer op het moment dat de stemspleet zich sluit; de grootste slingeringen treden op in het interval net erna. De momenten van glottis-sluiting en -opening zijn automatisch afgeleid van het electroglottogram (EGG) en gemarkeerd met een '*' respectievelijk 'o'. Uit het fotoglottogram (PGG) valt te concluderen dat deze momenten redelijk kloppen.

glottis interval (zie figuur 4: de intervallen tussen de '*' en 'o' markeringen die corresponderen met de intervallen waar het fotoglottogram minimaal is). Echter, als de glottis echt dicht is doet het er niet toe wat de druk over de stemspleet is, want door een dichte stemspleet zal zich geen luchtstroom ontwikkelen. In vergelijking (2) wordt dit zichtbaar doordat een eventuele variatie in P_{rr} teniet wordt gedaan door de vermenigvuldiging met $A_g=0$. De stembron blijft in zo'n situatie dus onafhankelijk van het aanzetstuk.

Het wordt anders wanneer de glottis niet compleet sluit (bijvoorbeeld omdat er posterieur een lekopening aanwezig is). In dat geval moet men verwachten (zie formule (2)) dat de luchtstroom die zich in de stemspleet ontwikkelt wel degelijk beïnvloed zal worden door de formant-trillingen in de transglottale druk. Met anderen woorden, de aanname die onderliggend was aan het lineairebron/filter-model, namelijk dat de stembron (lees glottale luchtstroom) zelf geen noemenswaardige eigen spectrale structuur heeft, kan dan in het gedrang komen. *Er is een niet-lineaire interactie tussen bron en filter.*

Of die interactie ook werkelijk optreedt, hangt uiteraard af van het product van A_g en P_{rr} . Als de lekopening heel klein is zal het effect verwaarloosbaar klein zijn en gaat de benadering van een lineair bron/filter-model nog steeds op; als de opgewekte resonantie in sub- en supraglottale holten sowieso niet zo sterk zijn, waardoor de formanttrillingen relatief gering in amplitude zijn, zal de lekopening groter kunnen zijn voordat het lineaire bron/filter-model het opgeeft, dan in een geval waarbij zeer sterke staande golven waargenomen worden.

Modellering van incomplete glottis-sluiting

In een recent onderzoek (Cranen & Schroeter, 1995, 1996) is gepoogd om het belang van de hierboven beschreven niet lineaire interacties op de glottale flow-golfvorm via modelleerwerk wat nauwkeuriger in kaart te brengen.

Centraal in dit werk staat een stembronmodel dat voortbouwt op dat van Titze (1984). Essentieel in die aanpak is dat de stemspleet-oppervlakte niet rechtstreeks wordt voorgesteld als een golfvorm van de tijd, maar dat deze indirect wordt afgeleid door aan te nemen dat de geometrie van de stemplooiën met een beperkt aantal fysiologisch interpreteerbare parameters redelijk adequaat kan worden beschreven (zie figuur 5) en dat trillingspatronen in de stemplooiën bij benadering goed beschreven kunnen worden in termen van eigentrillingen (vgl. ook de paragraaf over 'normal modes' in het boek van Titze (1994, pp. 97–100)). Voor de modellering van de akoestische/aerodynamische aspecten wordt nog steeds gebruik gemaakt van de relatie die beschreven wordt door formule (2), of om precieser te zijn, een variant daarvan die meer recht doet aan het feit dat de glottis over het algemeen geen uniforme doorsnede heeft en die beschreven is in Ishizaka en Flanagan (1972)

Uit de klinische praktijk is bekend dat het vaak voorkomt dat het cartilagineuze deel van de stemspleet open staat terwijl het membraneuze deel redelijk per-

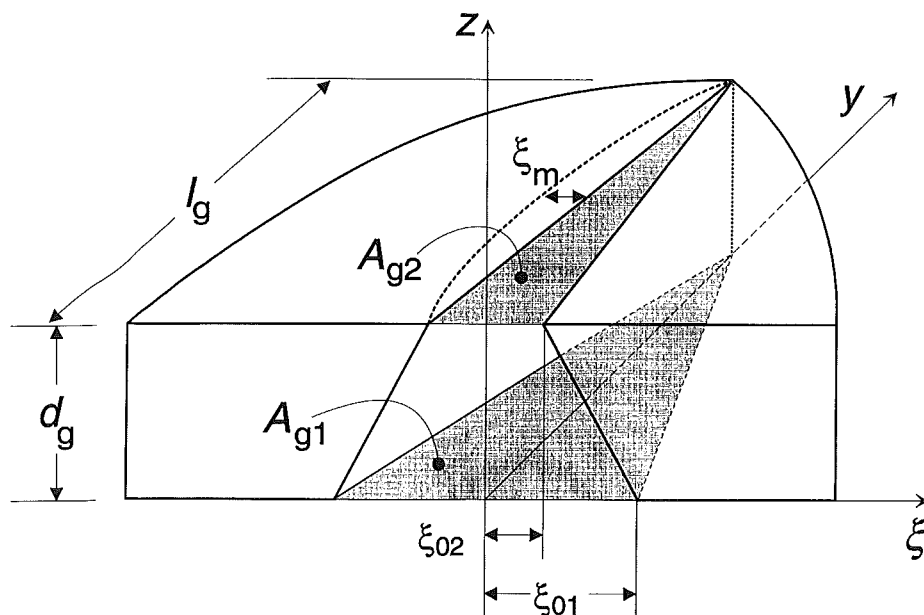


Fig. 5. Parametrisatie van de glottis geometrie volgens Titze (1984). A_{g1} en A_{g2} stellen de oppervlak ten van de glottis aan onder- en bovenzijde voor. De bewegingen van de stemplooien worden geacht plaats te vinden volgens de simpelste eigentrilling: het grootst in het midden en nul aan de voor en achterzijde. In de tijd wordt de beweging gemodelleerd als een sinus met amplitude ξ_m .

fect sluit. Er lijkt weinig consensus te bestaan over de mate waarin dit soort insufficiënte glottissluiting bestreden moet worden. Feit is in ieder geval dat de negatieve invloed op de stemkwaliteit van een glottaal lek niet proportioneel is met de oppervlakte van het lek en bij de mildere vormen soms zelfs niet of nauwelijks waarneembaar. Ons modellerwerk biedt wellicht een handvat om dit te verklaren en om criteria aan te leggen onder welke omstandigheden therapie moet worden nagestreefd.

We hebben al eerder gezien dat het vanuit akoestisch oogpunt belangrijk is om de luchtstroom zo abrupt mogelijk te onderbreken. Aangezien de luchtstroommodulatie voornamelijk plaats vindt in het membraaneuze deel van de stemspleet en er volgens de body-cover theorie (zie Titze, 1994) een staande golf in het membraaneuze weefsel optreedt waarvan er altijd een knoop (een punt waar het weefsel vrijwel in rust is) ligt aan de twee uiteinden wordt die abruptheid vooral beïnvloed door de mate van adductie. Bij stemplooien waarbij de processen vocalis elkaar niet raken, vindt de sluiting tamelijk gradueel plaats; de glottis wordt als door een onzichtbare rits van voor naar achter gesloten. Compleet geadduceerd, sluiten de membraaneuze delen van de stemplooien zich vrijwel instantaan

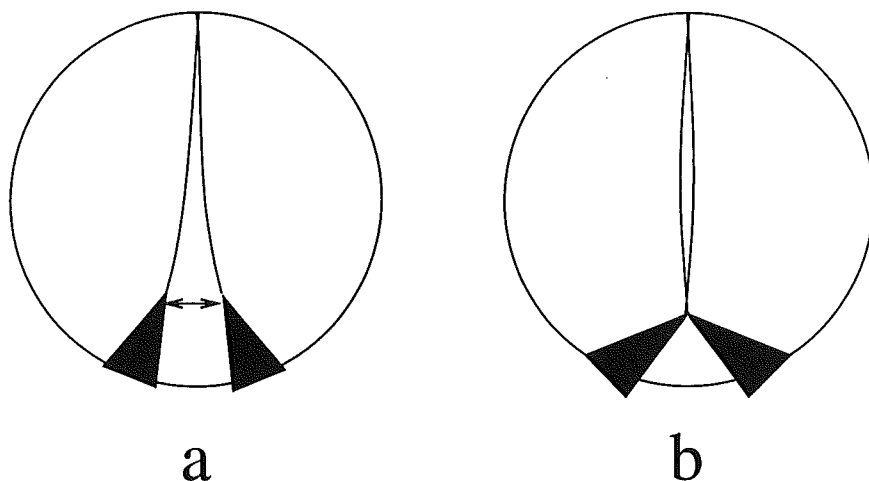


Fig. 6. Door de vrijheidsgraden van beweging die de arytenoiden hebben, kan de glottis verschillende vormen aannemen. (a) Bij incomplete adductie is er een deel van de glottis dat nooit sluit en dat een integraal onderdeel vormt van het membraneuze deel van de stemspleet (gekoppeld lek). (b) Wanneer de stemplooien compleet geadduceerd zijn, kan er nog steeds een lek in de achterste commissuur bestaan dat dan vrijwel onafhankelijk is van de membraneuze glottis (parallel lek).

over de volledige lengte. Uiteraard is de luchtstroommodulatie die teweeg gebracht wordt door de trillende stemplooien gesuperponeerd op de lekstroom die door een eventuele lekopening in de achterste commissuur nog weet te ontsnappen.

In principe kunnen we dus twee soorten glottale lekken onderscheiden waarvan we mogen verwachten dat het akoestische effect totaal anders is (zie figuur 6). Het ene soort lek dat min of meer een onafhankelijk kanaal vormt dat parallel staat aan het membraneuze deel van de stemspleet, zullen we gemakshalve aanduiden met *parallele lekopening* (figuur 6b). De lekopening die ontstaat door incomplete adductie (en die dus integraal onderdeel uitmaakt van het membraneuze deel van de stemspleet) zullen we aanduiden met *gekoppelde lekopening* (figuur 6a). Zonder op de mathematische details in te gaan (die komen uitgebreid aan de orde in Cranen & Schroeter, 1996), presenteren we hier twee simulatiere-sultaten. In figuren 7 en 8 staan achtereenvolgens afgebeeld de golfvormen van de glottale flow en de eerste afgeleide daarvan en het bijbehorende spectrum. Merk op dat in beide figuren de dc-offset in de luchtstroom (de constante luchtstroom die permanent weglekt) ongeveer gelijk is. Dit betekent dus dat ook de lekopening in cm^2 ongeveer gelijk was in beide situaties. In figuur 7 is de lekopening gecreerd door abductie (een gekoppeld lek). In figuur 8 werd de lekopening gevormd door een parallel lek terwijl het membraneuze deel van de stemspleet compleet sloot.

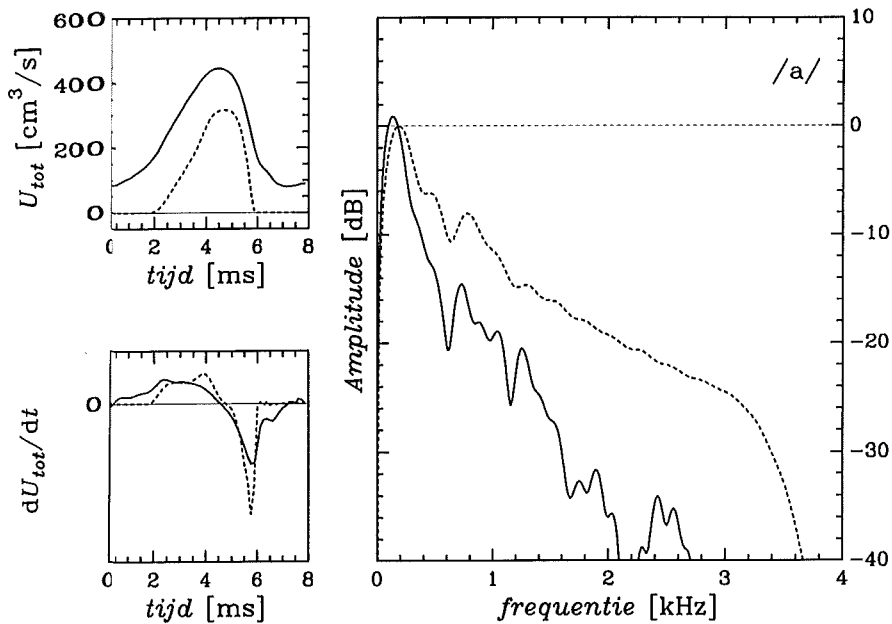


Fig. 7. In een model-experiment blijkt een stemapparaat waarbij de processen vocalis elkaar niet volledig raken (gekoppeld lek) een akoestisch stembronsignaal te genereren waarbij de hoge frequenties veel minder energie hebben dan in volledig geaduceerde toestand zonder lek (stippellijn). Dit wordt voornamelijk veroorzaakt door de geleidelijke sluiting. Daarnaast blijkt akoestische interactie op te treden tussen bron en filter. Dit manifesteert zich in de glottale luchtstroom en zijn afgeleide als rimpels en in het spectrum als dips in de buurt van de formantfrequenties.

Duidelijk is te zien dat de sluitingsfase in figuur 7 geleidelijk verloopt en dat bijgevolg het spectrum sterk afvalt. Als referentie zijn de curves getekend van een situatie zonder lek. Omdat de akoestische energie in de hoge frequenties zo laag is, mag verwacht worden dat de eventuele turbulentieruis die kan worden opgewekt door de lekstroom goed waarneembaar zal worden. De stem loopt een gereede kans het stempel "goed waarneembare wilde lucht" te krijgen.

De situatie bij een parallel lek is compleet anders. Uit figuur 8 blijkt duidelijk dat het spectrum in de hoge frequenties nagenoeg gelijk is aan dat van een stemapparaat zonder lek (vgl. de stippellijn). Die stem zal dus ook even krachtig en helder zijn als wanneer er geen lek geweest zou zijn. Eventuele turbulentieruis die wordt opgewekt door de lekstroom heeft grote kans gemaskeerd te worden door de akoestische energie die afkomstig is van de periodieke luchtstroomimpuls. Ondanks het feit dat er een aanzienlijke lekstroom is, zullen de perceptieve gevolgen verwaarloosbaar zijn.

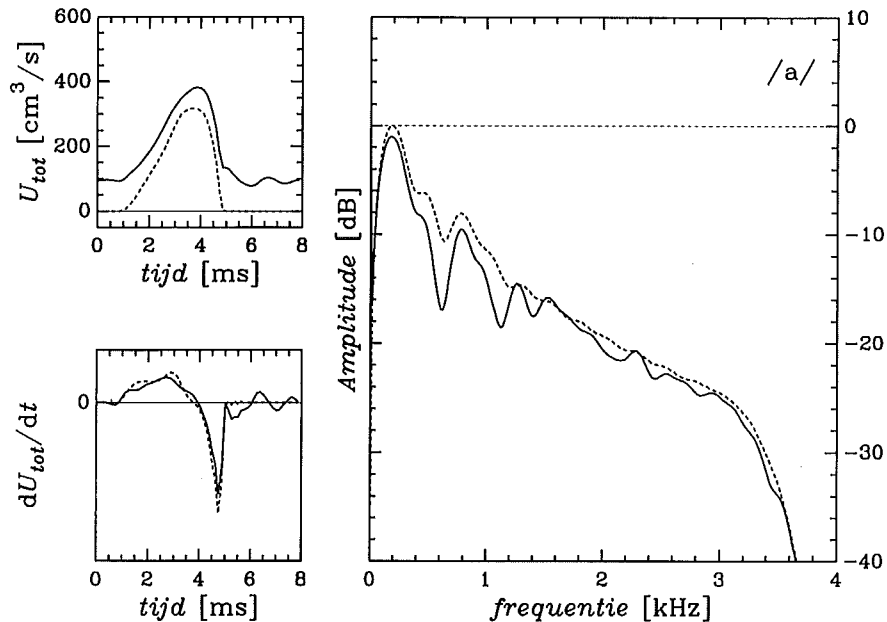


Fig. 8. In een model-experiment blijkt een stemapparaat waarbij de processi vocalis elkaar volledig raken maar waarbij er een lek in de achterste commissuur overblijft (parallel lek), een akoestisch stembronsignaal te genereren waarbij de hoge frequenties bijna even sterk zijn als in volledig geadduceerde toestand zonder lek (stippellijn). Daarnaast blijkt er akoestische interactie op te treden tussen bron en filter. Dit manifesteert zich in de tijdsignalen o.a. als rimpelingen vlak na het moment van sluiten en in het spectrum als dips in de buurt van de formantfrequenties. Doordat de luchtstroom abrupter ophoudt in vergelijking met figuur 7, is ook de akoestische excitatie van de trilhouten sterker en dus ook de interactie-rimpel.

Wat het parallelle lek met name bewerkstelligt is dat de luchtstroompuls een aantal rimpelingen vertoont (gedurende het open glottis interval, maar ook vlak na het moment van sluiten). In het frequentiedomein correspondeert dit met een stembronspectrum dat niet langer een glad afvallend verloop heeft, maar een aantal dips vertoont in de buurt van de formantfrequenties. Wat men hier ziet is de spectrale consequentie van de niet-lineaire interactie tussen bron en aanzetstuk via de transglottale druk.

Discussie en conclusies

Er zijn in dit artikel een aantal fysische principes onder de loep genomen die een rol spelen bij het produceren van stemgeluid. De primaire bron van het stemge-

luid blijkt de pulserende luchtstroom bij de stemplooien te zijn. De manier waarop de luchtstroom door de trillende stemplooien onderbroken wordt, het tijdsverloop van de luchtstroom, bepaalt hoe het stemgeluid klinkt. Als de luchtstroom in een glottale cyclus abrupt onderbroken wordt, levert dat een stembronsignaal op waarbij de hoge frequenties meer energie bezitten dan wanneer de luchtstroom geleidelijk naar zijn minimum gebracht wordt.

Inzicht in de fysische principes die het luchtstroomverloop bepalen, helpt bij het zoeken naar geschikte parametrisaties om dat stemgeluid mee te karakteriseren. Wanneer het om een puur akoestische parametrisatie gaat, is het voldoende om te kijken welke golfvorm-aspecten een belangrijke bijdrage leveren aan het spectrum. Wanneer een parametrisatie gewenst is die ook een interpretatie toelaat in termen van het onderliggende productieorgaan, wordt het belangrijk om het tijdsverloop van de stemspleet-oppervlakte er bij te betrekken.

Een model-experiment waarbij een glottaal lek op twee manieren werd gesimuleerd, suggereert dat het wel eens onvoldoende zou kunnen zijn om alleen het tijdsverloop van de stemspleetoppervlakte te kennen, maar dat men daarnaast ook moet rekening houden met de vorm ervan. Vanuit het oogpunt van stemkwaliteit lijkt het belangrijker om te kijken of het tijdvariante deel van de stemspleet voldoende abrupt sluit dan om te kijken naar de grootte van de lekopening.

Idealiter zou een stembronmodel niet alleen een verklaring moeten leveren voor het resulterende stemgeluid, maar ook van de condities waaronder de stemplooien in trilling geraken en blijven en hoe dit samenhangt met het druk- en luchtstroomverloop bij de stemplooien. Bij dit ideaalbeeld hoort een model waarvan men alle parameters van het model zodanig kan bijstellen dat men voor het stemapparaat van elk individu een computernabootsing zou kunnen maken. Dit ideaalbeeld is nog ver weg en de ontwikkeling en toetsing van dit soort zelfoscillerende modellen is voorlopig voornamelijk kwalitatief van aard. Met name voor wat betreft de weefselmechanica bestaan er nog grote onzekerheden. Voor meer kwantitatief georiënteerde modellering blijkt het aantal parameters in dit soort modellen meestal veel groter dan het aantal metingen dat men mogelijk kan doen om de meerderheid er van vast te leggen. Het gevolg daarvan is dat men onmogelijk kan vast stellen of een goed gelijkend simulatieresultaat wel bereikt is met een modelinstelling die in alle aspecten recht doet aan de werkelijkheid of dat de goede gelijkenis puur toevallig veroorzaakt is door een vreemde instelling van ongecontroleerde parameters.

Zelfs bij de categorie modellen waarbij alleen de akoestiek onderwerp van modellering is, maar waarbij men als een eerste poging om tot fysiologische interpretatie te komen het oppervlak van de stemspleet in de parametrisatie betreft, zou men kunnen zeggen dat de toetsbaarheid een probleem vormt. Immers, zolang er geen manieren zijn om op een betrouwbare manier en met voldoende tijdsresolutie de geometrie van de stemspleet vast te leggen, kan men weinig anders doen dan een paar goed gefundeerde aannamen maken over hoe de stemplooien zullen bewegen en schattingen maken over het verloop van de stemspleetgeometrie. Het nabootsen van de stem van een specifiek individu is pas

mogelijk als men alle parameters uit echte observaties heeft kunnen afleiden. Een haalbare techniek die het dichtst bij een kwantitatieve vorm van toetsing komt, is er een waarbij gepoogd wordt om onder fysisch realistisch ogende condities voldoende spraakmateriaal van een spreker voldoende nauwkeurig na te bootsen [analyse door synthese (Cranen & Schroeter, 1995, 1996)] met een model waarvan men *zoveel mogelijk* parameters een waarde heeft gegeven die uit simultane metingen zijn afgeleid. Als dit soort simulaties lukt en men niet teveel vrije parameters in het model gestopt heeft en testen van het model bovendien uitwijzen dat kleine veranderingen van details in de stemspleet-geometrie geen noemenswaardige effecten op het stemgeluid hebben, kan men aannemelijk maken dat het model voldoende realiteitswaarde heeft.

Het veel gebruikte lineaire bron/filter-model representeert het amplitudespectrum (in dB) van het spraakgeluid als de som van het bronspectrum (het spectrum van de eerste afgeleide van de glottale luchtstroom) en de overdrachtskarakteristiek van het spraakkanaal. Deze benadering gaat slechts beperkt op en wordt slechter naarmate de interactie tussen stembron en aanzetstuk sterker wordt, d.i. naarmate er meer formantenergie van de transglottale druk aan de luchtstroom wordt overgegeven. In heel veel gevallen zal de luchtstroom bij de stemplooiën niet merkbaar beïnvloed worden door de opgewekte staande golven in het aanzetstuk en kan het lineaire bron/filter-model gewoon toegepast worden, maar men dient alert te zijn op situaties waarbij dat niet het geval is. Men dient zich daarbij te bedenken dat het lineaire bron/filter-model geen ingebouwd mechanisme heeft dat alarm kan slaan wanneer het model niet meer adequaat is.

Modellerwerk met meer geavanceerde modellen geeft een indruk onder welke omstandigheden de interactie-effecten een rol van betekenis gaan spelen, nl. als het product van akoestisch geïnduceerde transglottale drukvariaties en glottale opening niet meer voldoende klein is. In welke praktijksituaties die voorkomen zal middels onderzoek dienen te worden vastgesteld.

Een nogal onverwachte gevolgtrekking die men uit dit modellerwerk moet maken is dat de veelgebruikte inverse filter techniek om de luchtstroom bij de glottis te meten in sommige situaties een stuk minder betrouwbaar zou kunnen zijn dan gedacht. Het signaal dat door de procedure wordt opgeleverd zou kunnen afwijken van de echte luchtstroom bij de stemplooiën want er wordt immers hard gebruik gemaakt van het lineaire bron/filter-model. Verder lijkt uit de modelstudies naar voren te komen dat criteria die vaak worden aangelegd om het succes van inverse filtering te beoordelen (zo weinig mogelijk rimpelingen in het gesloten glottis interval) wel eens niet adequaat zouden kunnen zijn. Alleen al het feit dat zeer veel van de in de open literatuur gepubliceerde invers filter resultaten een rimpeling laten zien net na het moment van sluiten kan worden geïnterpreteerd als een indicatie dat de aanname dat stembron en aanzetstuk in isolatie bestudeerd kunnen worden in veel meer gevallen dan tot nu toe gedacht werd niet echt opgaat. Een direct gevolg daarvan is weer dat het hierboven gesignaleerde probleem van de toetsbaarheid van ontwikkelde modellen nog verder verergerd wordt. Immers als een deel van de metingen waarop de model-

len gebaseerd zijn onbetrouwbaar geworden zijn omdat het meetprincipe niet deugt, hoe kan dan verder gebouwd worden aan meer realistische modellen?

Met de huidige stembronmodellen is een theoretisch kader geschapen waarin op een zinnige en fysiologisch georiënteerde manier over stem gedacht en gepraat kan worden. Ook al heeft de toetsing tot op heden vaak allen op een relatief globaal niveau plaats gevonden, ze hebben laten zien waar de zwakke punten in onze kennis zitten. Hoewel er nog geen modellen zijn waarvan alle parameters simpel zo kunnen worden gekozen dat nabootsing van de stem van een willekeurig individu voor allerlei verschillende fonatiecondities simpel mogelijk is, lijken de meest belangrijke fysische principes toch begrepen. Zoals met veel onderwerpen het geval is, blijkt een vergroot begrip de beschrijving van de werkelijkheid nauwelijks simpeler te maken. Sleutelwoord bij de interpretatie van gemeten luchtstroomgolfvormen is niet-lineaire akoestische interactie.

De gesignaleerde problemen met het lineaire bron/filter-model betekenen overigens geenszins dat het in een klap waardeloos is geworden. Een sceptische houding echter t.a.v. de meettechniek waarmee luchtstroom bij de stemplooiën meestal gemeten wordt en die de basis vormt onder vrijwel alle parametrische beschrijvingen, is zeker in gevallen die afwijken van de "normale" spraakproductie conditie (zoals luide fonatie en zang) op zijn plaats.

Zodra onderkend wordt dat akoestische interactie-effecten niet verwaarloosbaar zijn, leidt dit ook onvermijdelijk tot de conclusie dat vernieuwde aandacht gewenst is voor de karakterisering van glottale lekken. Aangezien energie in de transglottale druk alleen kan worden doorgegeven aan de luchtstroom als er een glottale opening is, maakt meting van alleen dc-offset flow als een karakterisering van lek zeker onvoldoende.

Literatuur

- Berg, J.W. van den, Zantema, J.T., & Doornenbal, P. (1957). On the air resistance and the Bernoulli effect of the human larynx. *Journal of the Acoustical Society of America*, 29, 626-631.
- Cranen, B., & Boves, L. (1985). Pressure measurements during speech production using semiconductor miniature pressure transducers - Impact on models for speech production. *Journal of the Acoustical Society of America*, 71, 1543-1551.
- Cranen, B., & Schroeter, J. (1995). Modeling a leaky glottis. *Journal of Phonetics*, 23, 165-177.
- Cranen, B., & Schroeter, J. (1996). Physiologically motivated modelling of the voice source in articulatory analysis/synthesis. *Speech Communication*, 19, 1-19.
- Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 4, 1-13.
- Ishizaka, K., & Flanagan, J.L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, 51, 1233-1286.
- Kitzing, P., & Löfqvist, A. (1975). Subglottal and oral pressure during phonation- Preliminary investigation using a miniature transducer system. *Medical & Biological Engineering*, 13, 644-648.

- Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- Rothenberg, M.R. (1977). Measurement of airflow in speech. *Journal of Speech and Hearing Research*, 20, 155-176.
- Schutte, H.K., & Miller, D.G. (1986). The effect of F0/F1 coincidence in soprano high notes on pressure at the glottis. *Journal of Phonetics*, 14, 385-392.
- Schutte, H.K., & Miller, D.G. (1987). Transglottische drukwisselingen bij de sopraanstem. *Tijdschrift voor Logopedie en Foniatrie*, 59, 34-38.
- Titze, I.R. (1984). Parameterization of glottal area, glottal flow, and vocal fold contact area. *Journal of the Acoustical Society of America*, 85, 901-906.
- Titze, I.R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.