

Meten aan stemgeving

R. Speyer en P.H. Dejonckere

Academisch Ziekenhuis Utrecht, Afdeling Foniatrie

Dit artikel beschrijft een onderzoeksprotocol voor de klinische praktijk van sempathologie ten behoeve van onder meer effectmeting van stemtherapieën. Het protocol zal worden geplaatst binnen het kader van methodologische problemen en praktische beperkingen. Daarnaast zal aandacht worden geschonken aan de complexe relatie tussen akoestische variabelen enerzijds en de perceptie van stemkwaliteit anderzijds. Stemgeving blijkt een multidimensioneel fenomeen en behoeft derhalve multiparametrisch onderzoek.

Inleiding

Binnen het klinisch stemonderzoek kan een aantal hoofddoelen worden nagestreefd; zo wordt beoogd een bijdrage te leveren aan de medisch etiologische diagnostiek en het definiëren van een functionele diagnose. Hierbij wordt onderscheid aangebracht tussen normaal versus abnormaal stemgeluid of stemgedrag en zal – in laatstgenoemd geval – het type en de ernst van de stoornis worden aangeduid. Tevens kan bijzondere aandacht uitgaan naar het bepalen van maximale stemprestaties (bijvoorbeeld in verband met classificatie) of belastbaarheid van de stem alsook naar het vaststellen van prognostische mogelijkheden (Hirano, 1989). Tot slot kunnen het kwantificeren en registreren van stemtherapieën en bijbehorende therapie-effecten nog hoofddoel van onderzoek vormen.

Teneinde deze doelen te bereiken, wordt binnen het stemonderzoek gebruik gemaakt van een scala aan objectieve stemmetingen. Sataloff e.a. (1990) geven bijvoorbeeld de volgende indeling: onderzoek naar stembandtrillingen, aërodynamische metingen, stemfunctieonderzoek, akoestische analyse, laryngale elektromyografie en psychofysische evaluatie. Alhoewel het kwantificeren van laatstgenoemde categorie moeizaam blijkt, wordt binnen deze reeks aan maten het menselijk oor en brein vooralsnog het geschiktste 'instrument' geacht om een stem te kunnen beoordelen. Overige beschrijvingen van de manier waarop data worden vergaard binnen het gebied van de objectieve stemanalyse, worden gegeven door onder meer Hirano (1989) en Jacobson (1994).

In dit artikel zal een onderzoeksprotocol voor de klinische praktijk van stempathologie worden beschreven ten behoeve van onder meer effectmeting van stemtherapieën. Het omvat zowel een beperkt alsook een uitgebreider onderzoeksinstrumentarium en zal worden geplaatst binnen het kader van methodologische problemen en praktische beperkingen. Tevens zal dieper worden ingegaan op de complexe relatie tussen akoestische variabelen enerzijds en de perceptie van stemkwaliteit anderzijds. Stemgeving, zo blijkt, betreft een multidimensioneel fenomeen en behoeft derhalve een multiparametrisch stemonderzoek. Binnen een onderzoek naar effectiviteit van stembehandeling dient echter ook aandacht te worden besteed aan zogenaamde 'quality-of-life'-maten, teneinde in kaart te brengen wat de patiënt zelf als resultaat van een (para)medische behandeling heeft ervaren.

Onderzoeksprotocol

Rekening houdend met methodologische problemen binnen het stemonderzoek en praktische beperkingen – zoals zaken financieel van aard, noodzakelijke tijdsinvestering, beschikbare apparatuur of onderzoeksruimten en voorhanden patiëntenpopulatie –, zou een onderzoeksprotocol voor de klinische praktijk van stempathologie (onder gestandaardiseerde stemgevingscondities) ten minste volgende instrumentele metingen of testen moeten omvatten:

- een gedifferentieerde perceptieve beoordeling (door een groep ervaren luisteraars); inclusief de evaluatie van fluctuatie van stemkwaliteit met de tijd;
- een videolaryngostroboscopisch onderzoek van de stemorganen;
- een akoestische analyse: zowel een (a)periodiciteitsmeting, een hoogfrequente ruismeting alsook een integrerende meting (bijvoorbeeld cepstrummeting).

Metingen worden verricht op gestandaardiseerd spraakmateriaal: een aangehouden [a:] (ten minste twee verschillende luidheden) en een korte, fonetisch geselecteerde standaardtekst (bijvoorbeeld louter stemhebbende fonemen met uitsluiting van fricatieven). Het is echter zinvol bovenstaand protocol nog uit te breiden met de volgende onderdelen:

- een 'quality of life'-maat;
- maximale prestatietaken (bijvoorbeeld bepaling van de maximale fonatieduur of een fonetogram-afname).

Het betreft hier een basisset ten behoeve van de dagelijkse, klinische praktijk van stemonderzoek, die geenszins pretendeert volledig te zijn, maar in de kliniek zijn praktisch nut reeds heeft bewezen. Bovenstaand onderzoeksprotocol wordt eveneens als noodzakelijke basisset beschouwd ten behoeve van het onderzoek naar therapie-effecten.

Een dergelijk onderzoeksprotocol is in de logopedische praktijk echter niet haalbaar noch gewenst, daar bijvoorbeeld het uitvoeren van videolaryngostroboscopisch onderzoek tot de taak van de foniater of KNO-arts behoort en niet binnen het logopedisch werkveld valt. Het hanteren van een weliswaar beperkter onderzoeksprotocol voor de

logopedische praktijk, zal desalniettemin onderzoek en evaluatie van therapieresultaten kunnen objectiveren en zorg dragen voor een stukje professionalisering binnen de logopedie. Het zal beter mogelijk zijn behandelresultaten in kaart te brengen: het volgen van de ontwikkelingslijn van een individu, het effect van een bepaald soort behandeling op een groep stempatiënten of het leggen van relaties tussen therapie-effecten en diagnose of andere beïnvloedende factoren zoals bijvoorbeeld rookgedrag of therapiefrequentie.

In de logopedische praktijk zou volgend protocol kunnen worden gehanteerd: een gedifferentieerde perceptieve beoordeling (bijvoorbeeld het zogenaamde GIBRAS-profiel: Dejonckere e.a., 1996^A), een 'quality-of-life'-maat (bijvoorbeeld de Voice Handicap Index: Jacobson e.a., 1997) en maximale prestatietaken (bijvoorbeeld maximale fonatieduur en fonetogram-afname). Een fonetogram kan betrouwbaar worden afgenomen zowel volledig geautomatiseerd met behulp van de computer alsook handmatig met behulp van een dB-meter en een toongenerator of muziekinstrument (Titze e.a., 1995). Daarnaast zijn inmiddels (eenvoudige) programma's ten behoeve van akoestische signaalverwerking voorhanden die ook binnen het kleinere budget vallen. Dergelijke software-pakketten beschikken soms over interne standaarden of normwaarden voor wat als normale of abnormale stemkwaliteit of akoestische parameterwaarde zou gelden. Alle metingen – waar onder het eigen professionele perceptieve oordeel – dienen te worden meegenomen in het stemonderzoek. De stem is immers een multidimensioneel fenomeen en behoeft derhalve een multiparametrisch onderzoek.

Methodologie

Binnen (stem)onderzoek liggen een aantal algemene impliciete veronderstellingen besloten. Onderzoeksmethoden dienen valide en betrouwbaar te zijn. Dat wil zeggen dat er vanuit wordt gegaan dat het geconstrueerde meetinstrument daadwerkelijk de te meten variabele(n) meet en tevens dat deze meting nauwkeurig en reproduceerbaar is (Janssens, 1985); verschillende auteurs besteden in toenemende mate specifieke aandacht aan het standaardiseren van meetopstellingen, van opnamecondities voor spraak en eisen gesteld aan type spraak- of stemmateriaal. Hiertoe worden in de literatuur nauwgezette richtlijnen gegeven omtrent bijvoorbeeld type microfoon en microfoonafstand of het beperken van achtergrondlawaai (Titze e.a., 1993; Titze, 1995^B; Winholtz & Titze, 1997). Daarnaast wordt het onderzoek klinisch relevant en bruikbaar geacht en dient minimaal invasief te zijn opdat proefpersonen niet meer dan strikt noodzakelijk worden belast en hun stemgedrag niet gaan veranderen of aanpassen aan de onderzoekssituatie of -techniek.

Stemonderzoek kent echter ook een reeks van meer specifieke methodologische problemen die kenmerkend zijn voor dit onderzoeksterrein. Dejonckere (1997) hanteert volgende tweedeling: een onderscheid tussen enerzijds enkelvoudige of univari-

ate metingen en anderzijds complexe of multivariate metingen (Dejonckere, 1997). Onder enkelvoudige metingen wordt bijvoorbeeld een grondtoon- of jitterbepaling verstaan, terwijl de tweede groep van metingen verricht aan complexe visuele of auditieve patronen verwijst naar bijvoorbeeld het bestuderen van videolaryngostroboscopische beelden of perceptieve beoordelingen van stemkwaliteit. Beide vormen van onderzoek leiden tot specifieke (eigen aan desbetreffende onderzoeksvorm) alsook algemene (gemeenschappelijke) methodologische vraagstukken. De belangrijkste vragen die in de literatuur worden aangehaald, zullen hier beknopt worden weergegeven; het is geenszins de bedoeling een uiteindelijke oplossing aan te dragen voor deze methodologische problemen, maar om meer helderheid te verschaffen in zowel de mogelijkheden alsook de beperkingen van het huidige stemonderzoek.

Methodologische overwegingen binnen het stemonderzoek

Selectie van spraakmateriaal

Een zeer belangrijke methodologische keuze betreft het selecteren van spraakmateriaal. Payri (1998) beschouwt de lengte van een spreektaak (bijvoorbeeld klinker versus zin) en de verscheidenheid aan spraakmateriaal (bijvoorbeeld luidop lezen versus spontane spraak) als factoren die onder meer toonhoogte en beklemtoning beïnvloeden. Klingholz (1990) vindt hogere correlaties tussen pathologische laryngale condities – verschillende diagnostische categorieën van stemstoornissen zoals functionele dysfonieën en dysfonieën op basis van poliepvorming, laryngitis of stembandverlamming – en signaal-ruis-verhoudingen gemeten in lopende spraak dan gemeten in aangehouden klinkers. Bij het bepalen van jitter, shimmer of heesheid daarentegen, wordt het belang van het klinkerfragment onderstreept. Hierbij wordt echter als kanttekening de beïnvloeding op deze stemparameters door luidheid en grondtoon geplaatst. Daarnaast kan het zinvol zijn voor bepaalde meetdoelen het spraakmateriaal naar een bepaald fonetisch principe te selecteren (Dejonckere e.a., 1996^A). Dit behelst bijvoorbeeld uitsluiting van fricatieven en stemloze explosieven en een keuze van geheel stemhebbend spraakmateriaal. Aldus kan turbulente ruis in het akoestisch signaal worden herleid tot op voornamelijk glottaal niveau (luchtverlies ten gevolge van onvolledige stembandsluiting) en niet het gevolg zijn van articulatorische vernauwingen. Tevens kunnen dan alle stemloze segmenten als pathologisch worden bestempeld.

In het geval van perceptieve stembeoordeling prefereert Fex (1992) lopende spraak boven klinkerfragmenten. Onderzoek van de Krom (1994) heeft echter geen grotere betrouwbaarheid of consistentie kunnen aantonen voor perceptieve scores op lopende spraak. Deze auteur benadrukt het belang van de klinkerinzet, daar deze relevante informatie bevat die van belang blijkt bij akoestische perceptie van stemkwaliteit.

Keuze van een adequaat psychofysisch classificatiesysteem

In de literatuur wordt door de verschillende auteurs gebruik gemaakt van diverse perceptieve classificatiesystemen. Hier kan bijvoorbeeld het zogenaamde Vocal Pro-

file van Laver (1980) worden aangehaald: een gedetailleerd classificatiesysteem dat zowel laryngale als supralaryngale 'settings' beschrijft; een aantal neutrale basisinstellingen fungeert als referentiepunt bij de beoordeling van stemkwaliteit op vierpuntsschalen. Supralaryngale kenmerken hebben betrekking op onder meer de positie van lippen, kaak of tong, nasaliteit en supralaryngale spanning, terwijl laryngale kenmerken verwijzen naar onder meer laryngale spanning, larynxpositie en fonatietype (bijvoorbeeld heesheid, fluisterspraak of falsetstem). Tevens is een audioband met voorbeelden van stemmen en bijbehorende perceptieve beoordelingen voorhanden. Het zogenaamde GRBAS-protocol (Hirano, 1981) – een perceptief meetinstrument, door de Japanse Vereniging voor Logopedie en Foniatrie als eerste naar voren gebracht en inmiddels frequent gebruikt –, berust op de volgende stemkwaliteitsparameters: 'grade', 'roughness', 'breathiness', 'asthenic' en 'strained'. Ook hier worden vierpuntsschalen gebruikt. Hammarberg (1998) hanteert een soortgelijk stem-beoordelingsmodel voor pathologische fonatie: SVEC (Stockholm Voice Evaluation Consensus Model). Het stemprofiel is echter uitgebreid tot een twaalfstal parameters die worden beoordeeld met behulp van visueel-analogue schalen: lijnen van gelijke lengte, waarvan alleen de uiteinden zijn voorzien van een label, bijvoorbeeld 'nul' versus 'maximaal'. Een haaks geplaatst streepje op een dergelijke lijn geeft de inschatting weer van het perceptief oordeel over desbetreffende stemkwaliteitsparameter. Wendler e.a. (1986) hebben zich daarentegen beperkt tot drie auditieve klassen: 'roughness', 'breathiness' en 'hoarseness'. Er worden echter nog tal van andere classificatiesystemen in de klinische praktijk gebruikt: bijvoorbeeld het multidimensionele model voor stemproductie van Perkins (1983) of het zogenaamde 'Buffalo Voice Profile' van Wilson (1979).

Luistertraining

Luistertraining verbetert het vermogen tot perceptief oordelen (Hammarberg, 1998). De Krom (1994) acht het zinvoller door middel van dergelijke oefening meer onderlinge overeenstemming tussen de verschillende luisteraars na te streven, dan de consistentie in oordelen van iedere beoordelaar op zich trachten te vergroten. Kreiman e.a. (1990) constateren een verschil in beoordelingswijze tussen onervaren en ervaren luisteraars. Terwijl ongetrainde luisteraars gelijke perceptuele strategieën hanteren, blijken individuele clinici wezenlijk te verschillen in keuze van parameters die zij als belangrijk beschouwen om tot een perceptieve beoordeling te komen. Door gebruik te maken van groepsgemiddelden in plaats van individuele oordelen kunnen derhalve aspecten van het individuele perceptieve gedrag van een ervaren beoordelaar aan het oog worden onttrokken.

Consistentie en betrouwbaarheid: interne en externe standaarden

Om de consistentie en betrouwbaarheid binnen psychofysisch onderzoek te vergroten, worden in de literatuur meerdere suggesties gedaan. Naast een ondubbelzinnige terminologie en definiëring, wordt door Fex (1992) geadviseerd een standaardtape met voorbeelden van diagnostische categorieën van pathologische stemgeving (met gesynthetiseerde ruis) aan te bieden. Deze kunnen als referentiekader fungeren;

Gerratt e.a. (1993) stellen dat luisteraars tijdens het beoordelen van stemkwaliteitsparameters, de geboden stimulus vergelijken met een interne standaard of schaal. Deze interne standaarden zijn niet stabiel en worden beïnvloed door niet alleen het fysische signaal, maar ook door omgevingsfactoren. Terwijl extreme stemkwaliteiten worden beoordeeld aan de hand van relatief stabiele interne standaarden, vertonen beoordelingen van tussenliggende kwaliteitsniveau's beduidend meer variatie. Uit onderzoek blijkt de intra- en interbeoordelaarbetrouwbaarheid beduidend hoger te zijn voor luisterexperimenten waarbij gebruik wordt gemaakt van ankerstimuli ofwel externe standaarden.

Statistisch criterium: overeenstemming of consistentie van psychofysische beoordelingen

Naast het feit dat de richtlijnen summier zijn ten aanzien van het weergeven van de ernst van het totale stembeeld of van een specifieke stemparameter, is geen optimaal statistisch criterium voorhanden voor de mate van overeenstemming of consistentie van perceptieve beoordelingen van het akoestisch signaal (Kreiman e.a., 1993).

Subjectiviteit van terminologieën

Al deze psychofysische evaluatieschalen stuiten echter op hetzelfde probleem: de subjectiviteit van dergelijke meetmethoden. Teneinde stemkwaliteit te beschrijven wordt gebruik gemaakt van een lange lijst van subjectieve terminologieën (Fex, 1992). Er is nauwelijks sprake van enige vorm van standaardisatie. Daarnaast laten sommige termen zich moeilijk vertalen naar (exacte) synoniemen in een ander taal. Sonninen en Hurme (1992) pleiten voor het samenstellen van een meertalige woordenlijst die betrekking heeft op het beschrijven van de menselijke stemgeving. Na het bereiken van consensus over welke begrippen belangrijk zijn, dienen deze zowel fysiologisch als akoestisch te worden gedefinieerd. Opsommingen van basisbegrippen binnen het stemonderzoek met bijbehorende definities kunnen onder meer worden gevonden bij Sonninen en Damsté (1971), Titze (1994^A; 1995^A) en Hammarberg (1998).

Keuze van algoritme

Een ander methodologisch vraagstuk verwijst naar het te gebruiken algoritme voor een bepaalde parameter; keuzes van algoritmen die de stemparameters bepalen, moeten helder zijn en berusten op consensus omtrent definiëring (Titze, 1995^B). Zo gebruiken Scherer e.a. (1995) een signaallengte van honderd opeenvolgende cycli op 200 msec. afstand van de steminzet, teneinde een stabiele perturbatiemaat te verkrijgen.

Bewaken van interne validiteit

Vervolgens zal nog aandacht moeten worden besteed aan de interne validiteit; het is de vraag of bijvoorbeeld in een onderzoek naar effectmeting van stemtherapie een gemeten verandering in stemgedrag daadwerkelijk mag worden beschouwd als resultaat van stemtherapie of zijn meerdere beïnvloedende factoren binnen het experiment aan te wijzen? Enerzijds worden de belangrijkste variabelen in de onderzoekssituatie

zo goed mogelijk gecontroleerd, anderzijds wordt getracht natuurlijke spraak in kaart te brengen. In welke mate echter vormen spreektaken uitgevoerd onder gecontroleerde condities een goede afspiegeling van spontane spraak? Stone en Rainey (1991) beschouwen de variabiliteit van subjecten per trial of dag, een problematische zaak binnen de akoestische analyse van stemgeving. Onderzoekresultaten van Orlikoff en Kahane (1991) duiden op de noodzakelijke controle van stemintensiteit om de validiteit en betrouwbaarheid van perturbatiematen te bewaken. Geringe variaties in luidheid op aangehouden klinkers kunnen reeds significante gevolgen hebben voor de belangrijkste akoestische parameters zoals bijvoorbeeld jitter, shimmer en signaal-ruis-verhouding (Dejonckere, 1998). Huang e.a. (1995) adviseren vanwege gelijksoortige bevindingen het 'vocal effort level' (mate van inspanning tijdens fonatie) gelijk te houden als prestaties van eenzelfde patiënt op verschillende meetmomenten worden vergeleken.

Normwaarden voor stemparameters?

Daarnaast blijkt uit de literatuur dat geen normwaarden voor de verschillende stemparameters bestaan. Het is namelijk over het algemeen niet mogelijk een eenduidig onderscheid tussen wat als normaal of abnormaal zou kunnen gelden voor deze parameters, aan te geven (Hammarberg & Gauffin, 1995); er is sprake van overlap tussen de normaalverdeling (Gausskromme) van 'normale' stemgeving en die van pathologische stemgeving. Bovendien zullen subpopulaties of specifieke sprekersgroepen eigen normwaarden kennen; immers, de eisen die worden gesteld aan de stemgeving van bijvoorbeeld beroepssprekers zullen afwijken ten opzichte van de eisen van minder intensief stemgebruik in het dagelijks (beroeps)leven.

Aantal herhalingen van eenzelfde meting: maximale prestatietaken en bepaling van gemiddelde waarden

Een volgende vraag betreft het vereiste aantal herhalingen van eenzelfde meting in het geval van kwantificatie van maximale prestaties, waarbij leer- en vermoeidheidsverschijnselen hun invloed zullen doen gelden (Dejonckere, 1997). Om bijvoorbeeld de maximale fonatieduur te kunnen bepalen, blijkt een volledige demonstratie door de onderzoeker voorafgaand aan de uitvoer van de spreektaak, onmisbaar. Daarnaast zijn minstens drie herhalingen noodzakelijk om daadwerkelijk maximale duur te behalen (Neiman & Edeson, 1981), hetgeen echter nog onvoldoende blijkt daar waar kinderen deel uitmaken van de patiëntenpopulatie (Lewis e.a., 1982). Volgens Stone (1983) blijken proefpersonen dermate inconsistente prestaties op maximale fonatietaken te leveren – afhankelijk van het aantal trials alsook van de gekozen grondfrequentie –, dat de validiteit van dergelijke metingen ernstig bekritiseerd kan worden; na acht trials blijkt nog geen 50 % van de volwassen proefpersonen een maximale fonatieduur bereikt te hebben. Na dertien trials beftreft dit 75 % van de populatie.

Het aantal herhalingen van eenzelfde meting in het geval van spontane, comfortabele spraak beoogt een gemiddelde waarde te bepalen. Scherer e.a. (1995) tonen aan dat teneinde representatieve gemiddelde perturbatiewaarden te kunnen vaststellen

door middel van akoestische analyse, het aantal vereiste herhalingen zal toenemen naarmate de stem minder stabiel zal blijken.

Spraaklaboratorium: data-afhankelijkheid van apparatuur en protocol

Literatuuronderzoek vermeldt tevens dat onderzoekgegevens veelal verwijzen naar een specifieke opstelling van meetapparatuur en een vast protocol van spreektaken. Hierdoor zijn de data van de verschillende onderzoekscentra niet zonder meer onderling te vergelijken, daar bijvoorbeeld opnamecondities of geanalyseerd spraakmateriaal niet gelijk blijken. De ontwikkeling van normwaarden per spraaklaboratorium zou dit dilemma mogelijk ten dele kunnen oplossen (Jacobson, 1994).

Doel en interpretatie van lange-termijn-metingen

Een volgend punt betreft zogenaamde lange-termijn-metingen. Evenals bij de eenvoudige metingen wordt getracht de betrouwbaarheid van metingen aan complexe patronen te verhogen door het gebruik van zogenaamde ‘averaging’ technieken, zoals bijvoorbeeld ‘long-term average spectral analysis’ (LTAS). Wendler e.a. (1986) zien echter vooralsnog geen directe toepassingsmogelijkheden binnen de klinische praktijk, van verbanden tussen LTAS-data en auditieve beoordelingen van stemkwaliteit. Deze laatste wijze van onderzoek wordt hoogst betrouwbaar geacht, in het bijzonder indien de auditieve (blinde) classificatie wordt uitgevoerd door groepen luisteraars.

Löfqvist en Mandersson (1987) besluiten naar aanleiding van onderzoek naar de spectraalverdeling van spraaksignalen over langere tijd dat de variabiliteit van normale sprekers zo groot blijkt, dat vraagtekens kunnen worden geplaatst bij het criterium ‘normaal’ binnen de akoestische analyse. Pathologische stemmen zijn hier echter buiten beschouwing gelaten. Individuele verschillen kunnen resulteren in grote standaarddeviaties voor groepsmetingen en derhalve in statistisch niet-significante verschillen tussen groepen onderling.

Combinatie van signalen

Het combineren van signalen beoogt het bewerkstelligen van een verbeterde klinische diagnose; Wendler (1992) beschouwt videolaryngostroboscopie – in het bijzonder in het geval van organische stemstoornissen – als een uiterst relevante, kwalitatieve onderzoeksmethode. Gezien de normale variatie in trillingspatroon van de stembanden, blijkt het echter nagenoeg onmogelijk om alleen aan de hand van stroboscopisch materiaal onderscheid te maken tussen normale fonatie enerzijds en functionele dysfonie anderzijds. Combineren van stroboscopie met andere observatie- en analysemethoden is gewenst.

Nieuwe ontwikkelingen

Tot slot wordt nog gewezen op nieuwe (veel belovende) ontwikkelingen: de toepassing van de ‘Self-Organizing Map’ ofwel het algoritme van Kohonen voor artificiële neurale netwerken op akoestische data (Leinonen e.a., 1992) en het aantonen van ‘chaos’ binnen de dynamiek van de stemplooitruilling door gebruik van theorieën uit de niet-lineaire dynamica (Titze e.a., 1993).

Relaties tussen perceptie en akoestiek

Hammarberg (1998) stelt dat perceptieve evaluatie van stemkwaliteit – aan de hand van psychofysische classificatiesystemen zoals eerder beschreven – kan worden gebruikt als standaard waaraan andere metingen kunnen worden gerelateerd. Aangezien perceptieve evaluaties echter omslachtig en tijdrovend werk blijken en subjectieve onzekerheden in de beoordelingen van luisteraars nooit volledig kunnen worden gecontroleerd, wordt ook het belang erkend van kwantificeerbare en objectieve stemfunctiematen afkomstig uit akoestische en aërodynamische analyses alsook uit fysiologische observaties. Verdonck-de Leeuw (1998) concludeert dat zolang de complexiteit van stemgeving in lopende spraak nog niet adequaat door middel van akoestische metingen kan worden geanalyseerd, perceptieve beoordelingen door getrainde luisteraars het uitgangspunt blijven voor de analyse van stemkwaliteit. Titze (1995^B) adviseert inzake de analyse van akoestische signalen (stemgeving) de volgende classificatie te hanteren: signalen die zich quasi-periodiek gedragen (type 1), signalen die onderbrekingen, sterke subharmonischen of modulaties vertonen, waarbij modulatie duidt op de systematische variatie van bijvoorbeeld amplitude of grondtoon over meerdere perioden (type 2), en signalen die zich chaotisch of ‘at random’ schijnen te gedragen (type 3). Afhankelijk van het type signaal worden specifieke analysemethoden aangeraden; in het geval van type-1-signalen wordt perturbatie-analyse bijzonder zinvol en betrouwbaar geacht, terwijl voor type-2-signalen verwezen wordt naar visualisatie ofwel afbeeldingen zoals het spectrogram. Perturbatiematen voor dit signaaltype zijn onbetrouwbaar en bevatten slechts weinig informatie met betrekking tot het trillingspatroon. Tot slot wordt voor type-3-signalen verwezen naar perceptuele beoordelingen van stemkwaliteitsparameters zoals schorheid en andere auditieve manifestaties van aperiodiciteit. Tevens wordt hier het belang van zogenaamde ‘phase-portraits’ aangegeven: een analysemethode waarbij de samenhang tussen twee of meer onafhankelijke dynamische variabelen wordt gevisualiseerd.

Gezien het relatieve belang van perceptieve evaluatie worden door Hammarberg en Gauffin (1995), een aantal maatregelen geformuleerd, om de betrouwbaarheid van dergelijke analysemethoden te vergroten: exacte definiëring van stemkwaliteitsparameters op basis van consensus tussen de luisteraars, het gebruik van – reeds eerder vermelde – zogenaamde ankerstimuli vastgelegd op audioband en het zoeken naar akoestische en fysiologische correlaties met perceptuele parameters. In het onderzoek van Kreiman e.a. (1993) wordt een conceptueel raamwerk geformuleerd waarbinnen de verschillende bronnen van variabiliteit in perceptieve beoordeling in ogenschouw worden genomen. Aandacht gaat onder meer uit naar luisterervaring, vorming van interne standaarden en wijze van scores (meerpuntsschalen versus visueel-analoge schalen).

Alhoewel akoestische metingen betrouwbaarder worden geacht dan perceptieve evaluaties wordt niet zonder meer hieraan de voorkeur gegeven (Rabinov e.a., 1995); onderzoek naar perturbatie in stemgeving (klinkerfragmenten) toont aan dat over het

algemeen de overeenstemming tussen luisteraars gelijk aan of zelfs beter is dan de betrouwbaarheid van 'objectieve' algoritmen van verschillende stemanalyse-systemen. Verder is de tegenstrijdigheid of variatie in perceptieve beoordelingen tussen de luisteraars voorspelbaar, terwijl automatische algoritmen schijnbaar willekeurig verschillen. Tot slot blijkt dat daar waar de betrouwbaarheid van het oordeel van luisteraars toeneemt met de ernst van stempathologie, objectieve analysemethoden stranden. Verdonck-de Leeuw (1998) concludeert dat akoestische analyse van de grondtoon echter de voorkeur heeft boven perceptuele toonhoogtebeoordelingen, daar zowel ongetrainde als getrainde luisteraars in hun toonhoogte-oordeel afhankelijk zijn van stemkwaliteit, in het bijzonder van schorheid.

Er is enorm veel literatuur voorhanden waarin pogingen worden ondernomen correlaties te vinden tussen perceptieve beoordelingen van stemgeving enerzijds en akoestische maten anderzijds; onderzoek naar de relatie tussen pathologisch stemtype en akoestische voorspelling van de ernst van de stemstoornis kan worden gevonden bij Wolfe en Steinfatt (1987) en Martin e.a. (1995). Hirano (1986) geeft een akoestische definiëring van het GRBAS-profiel. Hillenbrand en Houde (1996) beperken zich tot de correlaties tussen akoestische variabelen en heesheid waarbij onderscheid wordt gemaakt in perceptieve evaluatie van klinkerfragmenten versus lopende spraak, terwijl Eskenazi e.a. (1990) concluderen dat toonhoogte-amplitude en signaal-ruis-verhouding de meest geschikte parameters worden geacht om stemkwaliteit op basis van klinkerfragmenten te voorspellen. Ook nieuwe akoestische meetmethoden worden onderzocht; de Krom (1993) acht een signaal-ruis-meting gebaseerd op de zogenaamde cepstrummaat een zinvolle parameter binnen de stemanalyse, alhoewel geen direct verband kan worden gelegd met wat er gebeurt op het niveau van de glottis of met perceptuele kenmerken. Volgens Kempster e.a. (1991) wordt de perceptieve beoordeling van klinkerfragmenten door luisteraars voornamelijk bepaald door de akoestische parameters intensiteit, grondtoon en perturbatie van het dysfone stemgeluid. Onderzoek van Giovanni e.a. (1996) met als doel de ontwikkeling van een objectieve onderzoeksmethode op basis van akoestische en aërodynamische maten, vergelijkbaar met een perceptieve evaluatie door een groep beoordelaren, lukt slechts ten dele. Blijkbaar, zo concluderen de onderzoekers, maken luisteraars gebruik van nog andere stemkwaliteitsparameters die het ontwikkelde objectieve meetinstrument buiten beschouwing laat. (Ook in dit onderzoek zijn klinkerfragmenten als spraakstimuli gebruikt.)

Dit zijn slechts een aantal voorbeelden van de vele onderzoeken naar de relatie tussen akoestiek en perceptie. Hoge correlaties kunnen echter niet worden aangetoond (Dejonckere & Lebacqz, 1996^B). Derhalve worden vaak correlaties gelegd tussen panelbeoordelingen en clusters van akoestische maten, die inderdaad enigszins hoger uitvallen: bijvoorbeeld onderzoek naar correlaties tussen akoestische parameters berekend met behulp van het 'Multidimensional Voice Program' (MDVP) van Kay en perceptieve beoordelingsschalen (Dejonckere e.a., 1996^A; Wolfe e.a., 1997). Inter- en intrabeoordelaarsvariantie dienen hierbij bekend te zijn (Dejonckere e.a., 1993 en 1996^A).

Er blijkt echter nog onvoldoende sprake van internationale standaardisatie van evaluatiemethoden voor stemgeving (Hirano, 1989). Titze (1994^B) formuleert een drietal klinische hoofddoelen van standaardisatie van testmateriaal: nagaan van de invloed uitgaande van stemgeving op verstaanbaarheid en effectiviteit van communicatie, inzicht verkrijgen in de laryngale 'gezondheid' of algemene lichamelijke conditie en het pogen de effectiviteit van stemtherapie te kwantificeren. In ditzelfde licht kan bijvoorbeeld de indeling van de logopedische behandeling in zogenaamde behandel-elementen worden gezien (Dejonckere, 1995): een uniforme terminologie biedt mogelijkheden tot kwantificatie van therapie-inhoud.

Multidimensionele benadering

De stem kent vele facetten: het is een multidimensioneel fenomeen. Perkins (1983) geeft een overzicht van acht dimensies van stemproductie (zoals toonhoogte, luidheid, stemhebbendheid en register), waarmee stemgedrag en de daaruit voortvloeiende stemkwaliteit omschreven kunnen worden. Dit veronderstelt een uitgebreid pakket van onderzoekstechnieken binnen het stemonderzoek. Stemonderzoek dient derhalve multiparametrisch van aard te zijn (Dejonckere en Lebacqz, 1996^B). Hill e.a. (1990) belichten nut en complementerende functie van vier klinische instrumenten, te weten videolaryngoscopie, videolaryngostroboscopie, electroglottografie voor laryngale adductie en akoestische perturbatie-analyse.

In het reeds aangehaalde onderzoek van Wendler (1992) wordt op de normale variatie in stembandtrillingspatroon geduid, hetgeen de toepasbaarheid van laryngostroboscopie in de differentiaaldiagnostiek van stemstoornissen beperkt. Ook de fluctuatie van stemkwaliteit beïnvloedt het multidimensionele stemonderzoek. Het moment van meten kan een methodologisch probleem vormen gezien de individuele variatie op tal van stemparameters binnen een reeds beperkte tijdsspanne (Dejonckere, 1997). Het GRBAS-profiel is derhalve door Dejonckere e.a. (1996^A) uitgebreid met de factor 'Instability' ofwel een maat voor de fluctuatie van stemkwaliteit met de tijd.

Met behulp van factoranalyse worden tot slot hoofdtrends binnen de stempathologie beschreven, waarbij zich anatomische, biomechanische, akoestische en perceptieve eigenschappen of kenmerken aansluiten (Dejonckere, 1997); deze kunnen vervolgens leiden tot de ontwikkeling van zogenaamde 'basissets' van instrumentele metingen of testen. In 1989 formuleert Hirano een minimale, een quasi-minimale en een optionele instrumentatie- en onderzoekset ten behoeve van het stemonderzoek. Daarnaast wordt gebruik van volgende metingen of testen geadviseerd: perceptuele evaluatie, visuele inspectie van het stemorgaan, bandopname van de stemgeving, stroboscopie, testen gerelateerd aan grondtoon, testen gerelateerd aan SPL, aërodynamische testen en akoestische analyse. Een recenter overzicht van klinische taken met betrekking tot stemevaluatie wordt gegeven door Johnson (1994); hierbij wordt het takenpakket nog aan betrokken fysiologische functies gerelateerd. Ook Verdonck-de Leeuw (1998) beschrijft een analyse-protocol in het kader van een multidimensionele benadering

van onderzoek naar stemkarakteristieken (na radiotherapie). Gesteld wordt dat een dergelijk protocol tenminste volgende stemdimensies behoort te omvatten: perceptuele analyse van lopende spraak door getrainde luisteraars (volgens standaardprotocol), akoestische analyses van aangehouden klinkers met behulp van gestandaardiseerde apparatuur en uitgevoerd door een ervaren onderzoeker, videolaryngostroboscopia-opnamen en -evaluaties uitgevoerd door een ervaren medicus en schalen voor zelfbeoordeling voor patiënten (volgens standaardprotocol). Vanzelfsprekend biedt de literatuur nog talloze andere basissets, waarvan bovenstaande overzichten slechts voorbeelden kunnen zijn.

Ramig en Verdolini (1998) bespreken aan de hand van literatuuronderzoek het begrip 'treatment efficacy' in relatie tot de verschillende stemstoornissen. Hierbij wordt onderscheid gemaakt in effectiviteit (werkt therapie?), efficiency (werkt de ene therapie beter dan de andere?) en behandelingseffect (op welke wijze verandert therapie bepaalde gedragingen?). Bij het beantwoorden van dergelijke vragen, dient rekening te worden gehouden met het multidimensionele karakter van stemgeving en derhalve gebruik te worden gemaakt van een multiparametrisch onderzoeksinstrument.

Quality of life

Naast kwantificering van perceptieve stemevaluaties en het vergaren van objectieve akoestische stemmaten, wordt in toenemende mate aandacht besteed aan de gevolgen van de stemstoornis voor privé- en beroepsleven van het individu. Dit hangt samen met het zogenaamde ICDH-concept: International Classification of Impairments, Disabilities and Handicaps (World Health Organisation, 1980). 'Impairment' duidt op de dysfunctie ten gevolge van pathologische verandering in het systeem, 'disability' verwijst naar het gevolg van de beperking op de functionele performance (stoornis op het niveau van de persoon) en 'handicap' heeft betrekking op de nadelen die een individu ervaart als gevolg van bovenstaande begrippen. Handicap geeft dus de interactie tussen omgeving en individu weer, alsook de aanpassing van het individu aan die omgeving (Enderby, 1997).

Teneinde inzicht te verkrijgen in de beperkingen of hinder die een persoon ondervindt als gevolg van ziekte of stoornis, worden verschillende meetinstrumenten ontwikkeld; Scott e.a. (1997) hanteren een vragenlijst voor stempatiënten, die resulteert in een totaaloverzicht van alle mogelijke problemen gerelateerd aan dysfonie. Vervolgens worden de problemen onderverdeeld naar ICDH-concept. Een ander instrument om de psychosociale gevolgen van stemstoornissen voor één individu te kwantificeren is de 'Voice Handicap Index' (VHI) van Jacobson e.a. (1997): dertig vragen onderverdeeld naar functionele, lichamelijke en emotionele problematiek.

'Quality of life'-maten verduidelijken wat de patiënt als resultaat van een medische behandeling heeft ervaren: het belang van de zienswijze van de patiënt wordt erkend in het kunnen bepalen van de effectiviteit van een behandeling (Benninger e.a., 1997). Verdonck-de Leeuw (1998) vindt hoge correlaties tussen scores van stempatiënten op

zelfbeoordelingsschalen voor stemperformance enerzijds en stemfunctiematen als ook perceptuele en akoestische maten van stemkwaliteit anderzijds.

Conclusie

Klinisch stemonderzoek en dus ook onderzoek naar effectmeting van stemtherapieën brengen de nodige methodologische problemen en keuzes met zich mee. Het is echter duidelijk dat stemgeving een multidimensioneel gegeven is en derhalve een multiparametrisch onderzoek veronderstelt. De fysische wetten die ten grondslag liggen aan fonatie leiden vooralsnog niet tot een eenduidig natuurkundige interpretatie samengevat in wiskundige algoritmen ten behoeve van akoestisch onderzoek. Hoge correlaties tussen perceptieve oordelen van spraak en akoestische variabelen blijven uit; oor en brein van de (getrainde) luisteraar alsook van de stempatiënt beoordelen klaarblijkelijk niet volgens de normen van uitgebreide spraaklaboratoria. Aandacht zal moeten (blijven) uitgaan naar verdergaande standaardisatie en uniformisatie binnen het stemonderzoek.

Summary

This article describes an analysis protocol to be used in the clinical practice of voice pathology in order to evaluate, amongst others, the effectiveness of voice therapy. Problems of methodology and practical limitations related to the protocol have been discussed. Furthermore, the complex relation between acoustical parameters on one hand, and the perception of quality of voice on the other hand, have been dealt with. Phonation appears to be a multidimensional phenomena and therefore requires a multiparametric approach.

Literatuur

- Benninger, M.S., Gardner, G.M., Jacobson, B.H., Grywalski, C. (1997²). New dimensions in measuring voice treatment outcomes. In: Sataloff, R.T. (ed). *Professional voice: the science and art of clinical care*. Singular Publishing Group, San Diego, 789-794.
- Dejonckere, P.H. (1998). Effect of slightly louder voicing on acoustical measurements in different etiological categories of dysphonia. In: Krom, G. de (ed). *Voicedata98: Symposium on databases in voice quality research and education*, University of Utrecht.
- Dejonckere, P.H. (1997). De zin en onzin van objectieve metingen in de stempathologie. In: Peters, H. (ed). *Recente ontwikkelingen in diagnostiek en behandeling van stemafwijkingen*. NICO-cursus, Katholieke Universiteit Nijmegen.
- Dejonckere, P.H., Burger, E., Broekhuizen, A., Graetz, P. (1995). Elementen in de logopedische behandeling. *Logopedie en Foniatrie* 5, 116-125.
- Dejonckere, P.H., Lebacqz, J. (1996^B). Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *Journal of Oto-Rhino-Laryngology and its Related Special-*

- ties 58, 326-332.
- Dejonckere, P.H., Obbens, C., Moor, G.M. de, Wieneke, G.H. (1993). Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatrica* 45, 76-83.
- Dejonckere, P.H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L., Millet, B. (1996^A). Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Revue de Laryngologie Otologie Rhinologie* 117, 219-224.
- Enderby, P., John, A. (1997). *Therapy Outcome Measures (TOM): speech and language therapy. Technical manual*. Singular Publishing Group, San Diego.
- Eskenazi, L., Childers, D.G., Hicks, D.M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research* 33, 298-306.
- Fex, S. (1992). Perceptual evaluation. *Journal of Voice* 6, 155-158.
- Gerratt, B.R., Kreiman, J., Antonanzas-Barroso, N., Berke, G.S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research* 36, 14-20.
- Giovanni, A., Robert, D., Estublier, N., Teston, B., Zanaret, M., Cannoni, M. (1996). Objective evaluation of dysphonia: preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements. *Folia Phoniatrica et Logopedica* 48, 175-185.
- Hammarberg, B. (1998). Perception and acoustics of voice disorders - a combined approach. In: Krom, G. de (ed). *Voicedata98: Symposium on databases in voice quality research and education*. University of Utrecht.
- Hammarberg, B., Gauffin, J. (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In: Fujimura, O., Hirano, M. (eds). *Vocal fold physiology, voice quality control*. Singular Publishing Group, San Diego.
- Hill, D.P., Meyers, A.D., Scherer, R.C. (1990). A comparison of four clinical techniques in the analysis of phonation. *Journal of Voice* 4, 198-204.
- Hillenbrand, J., Houde, R.A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research* 39, 311-321.
- Hirano, M. (1981). *Clinical examination of voice*, Springer-Verlag, Wien.
- Hirano, M. (1989). Objective evaluation of the human voice: clinical aspects. *Folia Phoniatrica* 41, 89-144.
- Hirano, M., Hibi, S., Terasawa, R., Fujiu, M. (1986). Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. *Journal of Phonetics* 14, 445-456.
- Huang, D.Z., Minifie, F.D., Kasuya, H., Lin, S.X. (1995). Measures of vocal function during changes in vocal effort level. *Journal of Voice* 9, 429-438.
- Jacobson, B.H. (1994). Objective voice analysis: the clinical voice laboratory. In: Benninger, M.S., Jacobson, B.H., Johnson, A.F. (eds). *Vocal arts Medicine. The care and prevention of professional voice disorders*, 135-152.
- Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M.S., Newman, C.W. (1997). The Voice Handicap Index (VHI): development and validation. *American Journal of Speech-Language Pathology* 6, 66-70.
- Janssens, J.M.A.M. (1985³). *'-Ogen' doen onderzoek: een inleiding in de methoden van sociaal-wetenschappelijk onderzoek*. Swets & Zeitlinger, Lisse.
- Johnson, A.F. (1994). Clinical voice assessment. In: Benninger, M.S., Jacobson, B.H., Johnson, A.F. (eds). *Vocal Arts Medicine. The Care and prevention of professional voice disorders*. Thieme Medical Publisher, New York, 103-108.
- Kempster, G.B., Kistler, D.J., Hillenbrand, J. (1991). Multidimensional scaling analysis of

- dysphonia in two speaker groups. *Journal of Speech and Hearing Research* 34, 534-543.
- Klingholtz, F. (1990). Acoustic recognition of voice disorders: a comparative study of running speech versus sustained vowels. *Journal of the Acoustical Society of America* 87, 2218-2224.
- Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., Berke, G.S. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research* 36, 21-40.
- Kreiman, J., Gerratt, B.R., Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research* 33, 103-115.
- Krom, G. de (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research* 36, 254-266.
- Krom, G. de (1994). *Acoustic correlates of breathiness and roughness: experiments on voice quality*. Doctoral dissertation, University of Utrecht.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press, Cambridge.
- Leinonen, L., Kangas, J., Torkkola, K., Juvas, A. (1992). Dysphonia detected by pattern recognition of spectral composition. *Journal of Speech and Hearing Research* 35, 287-295.
- Lewis, K., Casteel, R., McMahon, J. (1982). Duration of sustained /_/ related to the number of trials. *Folia Phoniatica* 34, 41-48.
- Löfqvist, A., Mandersson, B. (1987). Long-time average spectrum of speech and voice analysis. *Folia Phoniatica* 39, 221-229.
- Martin, D., Fitch, J., Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research* 38, 765-771.
- Neiman, G.S., Edeson, B. (1981). Procedural aspects of eliciting maximum phonation time. *Folia Phoniatica* 33, 285-293.
- Orlikoff, R.F., Kahane, J.C. (1991). Influence of mean sound pressure level on jitter and shimmer measures. *Journal of Voice* 5, 113-119.
- Payri, B. (1998). A vocal database for the study of voice quality. In: Krom, G. de (ed). *Voicedata98: Symposium on databases in voice quality research and education*. University of Utrecht.
- Perkins, W.H. (1983). Quantification of vocal behavior: a foundation for clinical management of voice. In: Bless, D.M., Abbs, J.H. (eds). *Vocal fold physiology*, College-Hill Press, San Diego.
- Rabinov, C.R., Kreiman, J., Gerratt, B.R., Bielałowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research* 38, 26-32.
- Ramig, L.O., Verdolini, K. (1998). Treatment efficacy: voice disorders. *Journal of Speech and Hearing Research* 41, S101-S116.
- Sataloff, R.T., Spiegel, J.R., Carroll, L.M., Darby, K.S., Hawkshaw, M.J., Rulnick, R.K. (1990). The clinical voice laboratory: practical design and clinical application. *Journal of Voice* 4, 264-279.
- Scherer, R.C., Vail, V.J., Guo, C.G. (1995). Required number of tokens to determine representative voice perturbation values. *Journal of Speech and Hearing Research* 38, 1260-1269.
- Scott, S., Robinson, K., Wilson, J.A., Mackenzie, K. (1997). Patient-reported problems associated with dysphonia. *Clinical Otolaryngology* 22, 37-40.
- Sonninen, A., Damsté, P.H. (1971). An international terminology in the field of logopedics and phoniatrics. *Folia Phoniatica* 23, 1-32.
- Sonninen, A., Hurme, P. (1992). On the terminology of voice research. *Journal of Voice*. 6,

188-193.

- Stone, R.E. (1983). Issues in clinical assessment of laryngeal function: contraindications for subscribing to maximum phonation time and optimum fundamental frequency. In: Bless, D.M., Abbs, J.H. (eds). *Vocal fold physiology*, College Hill Press, San Diego.
- Stone, R.E., Rainey, C.L. (1991). Intra- and intersubject variability in acoustic measures of normal voice. *Journal of Voice* 5, 189-196.
- Titze, I.R. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of speech and Hearing Research* 36, 1177-1190.
- Titze, I.R. (1994^A). *Principles of voice production*. Prentice-Hall, New Jersey, Glossary 330-337.
- Titze, I.R. (1994^B). Toward standards in acoustic analysis of voice. *Journal of Voice*. 8, 1-7.
- Titze, I.R. (1995^A). Definitions and nomenclature related to voice quality. In: Fujimara, O., Hirano, M. (eds). *Vocal fold physiology, voice quality control*, Singular Publishing Group, San Diego.
- Titze, I.R. (1995^B). *Workshop on acoustic voice analysis: summary statement*. Denver.
- Titze, I.R., Baken, R.J., Herzog, H. (1993). Evidence of chaos in vocal fold vibration. In: Titze, I.R. (ed). *Vocal fold physiology. Frontiers in basic science*. Singular Publishing Group, San Diego, 143-188.
- Titze, I.R., Wong, D., Milder, M.A., Hensley, S.R., Ramig, L.O. (1995). Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *Journal of Speech and Hearing Research* 38, 526-535.
- Verdonck-de Leeuw, I.M. (1998). *Voice characteristics following radiotherapy: the development of a protocol*. Doctoral dissertation, University of Amsterdam.
- Wendler, J. (1992). Stroboscopy. *Journal of Voice* 6, 149-154.
- Wendler, J., Rauhut, A., Krüger, H. (1986). Classification of voice qualities. *Journal of Phonetics* 14, 483-488.
- Wilson, D.K. (1979²). *Voice problems of children*. Williams & Wilkins, Baltimore.
- Winholtz, W.S., Titze, I.R. (1997). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice* 11, 417-421.
- Wolfe, V., Fitch, J., Martin, D. (1997). Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatrica et Logopedica* 49, 292-299.
- Wolfe, V.I., Steinfatt, T.M. (1987). Prediction of vocal severity within and across voice types. *Journal of Speech and Hearing Research* 30, 230-240.
- World Health Organisation (1980): *International classification of impairments, disabilities and handicaps*. WHO, Geneva.