

De Dysphonia Severity Index: een objectieve maat voor de stemkwaliteit

F.L. Wuyts, M.S. De Bodt en P.H. Van de Heyning

Dienst Neus, Keel, Oorzakten, Universitair Ziekenhuis Antwerpen

De Dysphonia Severity Index (DSI) modelleert de stemkwaliteit van een patiënt, met het oog op het weerspiegelen van de gepercipieerde stemkwaliteit door middel van objectieve en kwantitatieve metingen. De DSI is gebaseerd op een gewogen combinatie van geselecteerde stemparameters: de hoogste frequentie (Fo-High in Hz), de laagste intensiteit (I-Low in dB), de maximale fonatietijd (MPT in s) en de jitter (%). Deze DSI is het resultaat van een multivariate analyse van stemmen van 387 personen en wordt als volgt berekend (Wuyts, F. L. et al., 2000): $DSI = 0.13 * MPT + 0.0053 * Fo-High - 0.26 * I-Low - 1.18 * Jitter (%) + 12.4$.

De DSI voor perceptueel normaal klinkende stemmen is +5 en voor ernstig gestoorde stemmen -5. Hoe negatiever een index is, des te slechter is de stemkwaliteit van de patiënt. Daardoor is de DSI een instrument dat zich goed leent voor het evalueren van een therapie bij een patiënt met een hese stem. Daarnaast bestaat er een goede correlatie met de Voice Handicap Index score, die de psychosociale impact van het stemprobleem zoals door de patiënt wordt ervaren, weerspiegelt.

Inleiding

De toename van multidisciplinair onderzoek met betrekking tot het evalueren en behandelen van stemproblemen, gaf gedurende het laatste kwartaal van de 20^{ste} eeuw, aanleiding tot de ontwikkeling van een veelheid aan technieken en parameters voor het meten van stemkwaliteit. Dit resulteerde niet altijd in een uniforme en transparante benadering van het fenomeen stemkwaliteit.

De medische diagnose van een stemploopathologie is voornamelijk gebaseerd op een endoscopisch onderzoek van de stemplooien en de bovenste luchtwegen. Anderzijds wordt heesheid beschreven met perceptuele schalen en objectieve parameters zoals akoestische of aërodynamische metingen. Desalniettemin is perceptuele evaluatie een zeer controversiële methode. Er bestaat een grote variëteit aan schalen en de betrouwbaarheid fluctueert van studie tot studie (zie hiervoor De Bodt et al in ditzelfde

nummer). Tot nu toe is er nog geen internationaal aanvaard perceptueel evaluatieprotocol, maar de GRBAS schaal (Hirano, M., 1981), blijkt mondiaal één van de meest gebruikte te zijn. Een poging om de schaal te verfijnen door er een visueel analoge schaal van te maken faalde (Wuyts, F. L., De Bodt, M. S., & Van de Heyning, P. H., 1999). Derhalve propageren de auteurs de 4-punten versie zoals oorspronkelijk vooropgesteld door Hirano (1981), met de onderverdelingen 0=normaal, 1=licht, 2=matig, 3=ernstig.

Naast de perceptuele evaluatie hebben een reeks van stemkwaliteitsparameters hun intrede gedaan, zoals de maximale fonatietijd, het fonatiequotiënt, de airflow, de subglottische druk, enz. De computer heeft in de laatste decennia eveneens een belangrijke rol gespeeld. Akoestische analyses werden hierdoor beschikbaar voor menig clinicus.

De relatie tussen de perceptuele beoordeling en de akoestische meting is onderwerp van diverse studies, waarbij de meeste onderzoekers zich toelagen op de kwantitatieve correlatie tussen afzonderlijke variabelen zoals jitter, shimmer of harmonics/noise ratio met de perceptuele schaal (Dejonckere, P. H. et al., 1996; Dejonckere, P. H. & Lebacqz, J., 1996; Crevier-Buchman, L. et al., 1998; Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L., 1980; Hillenbrand, J., Cleveland, R. A., & Erickson, R. L., 1994; Rabinov, C. R., Kreiman, J., Gerratt, B. R., & Bielamowicz, S., 1995; Kreiman, J., Gerratt, B. R., & Berke, G. S., 1994; Askenfeld, A. G. & Hammarberg, B., 1986). Andere auteurs hanteren multiple lineaire regressie om deze relatie te onderzoeken (Wolfe, V., Fitch, J., & Cornell, R., 1995; Eskenazi, L., Childers, D. G., & Hicks, D. M., 1990). Wolfe et al. (1995) geven aan dat geen enkele van de akoestische variabelen goed correleerde met de heesheidsscore, en evenmin gaf een combinatie van variabelen een zinvolle predictie van de heesheid. Dit in tegenstelling tot Eskenazi e.a. (1990) die wel een redelijke predictie van de heesheid beschrijven op basis van pitch amplitude en harmonics/noise ratio. Analyse van een groep van 494 patiënten toonde aan dat de beste Spearman Rank correlatie tussen elk van de GRBAS schaal elementen en de akoestische variabelen van het Multi Dimensional Voice Program (Kay Elemetrics) niet groter is dan 0.53 (Wuyts, F. L., De Bodt, M. S., Bruckers, L., & Molenberghs, G., 1996). De jitter bijvoorbeeld, had als correlatiecoëfficiënten voor de 5 elementen van de GRBAS schaal respectievelijk: 0.46, 0.37, 0.31, 0.17 en 0.23. Al deze correlaties waren significant ($p < 0.001$) maar op basis van dergelijke correlaties kunnen geen zinvolle klinische conclusies getrokken worden.

Significante verschillen tussen normaal en pathologisch of tussen diverse vormen van pathologie onderling kunnen bekomen worden indien de aantallen van de groepen voldoende groot zijn, omdat dan de standaardfout (standard error = se) bepaalt of het verschil statistisch significant is of niet. Aangezien de standaardfout omgekeerd evenredig is met de wortel van het aantal metingen ($se = \text{standaard afwijking} / \sqrt{N}$), kan men voor grote aantallen wel verschillen vinden

Wanneer men echter de waarde gevonden bij een patiënt wil vergelijken van normale waarden, dan is het zinvol om het predictie interval (PI) te beschouwen, dat gedefinieerd wordt als [gemiddelde - 2s, gemiddelde + 2s]. Dit interval bevat 95% van de totale waarden, en indien een bepaalde meting van een patiënt buiten dit interval ligt

kan men met grote zekerheid (95%) stellen dat dit afwijkend is. Echter is het zo dat biomedische grootheden die de stem karakteriseren, gekenmerkt worden door een grote variabiliteit – weergegeven door de standaardafwijking (s). Dit maakt dat waarden al erg extreem moeten zijn om buiten het “brede” predictie-interval te liggen. Derhalve is het dus zeer moeilijk om hieruit klinische conclusies te trekken. Dit is vermoedelijk een van de redenen waarom er tot nu toe geen algemene maat voor heesheid is beschreven. Het nadeel van een dergelijke “univariate” benadering kan omzeild worden door de toepassing van multivariate statistische methoden. De essentie van een multivariate benadering is vervat in de mogelijkheid om een bepaald gegeven (bijv. heesheid) te beschrijven door een combinatie van verschillende variabelen. Dan kan de informatie van diverse variabelen gecombineerd worden. Een dergelijk gebruik van variabelen is niet nieuw in de geneeskunde. Met name de Body Mass Index (BMI) (gewicht/lengte²) is een typisch voorbeeld van het combineren van twee aparte grootheden, die ogenschijnlijk onafhankelijk zijn. De combinatie van een kleine persoon met een groot gewicht, resulteert in een afwijkende waarde voor de BMI (BMI >30 is obees), alhoewel op zich de lengte en het gewicht van deze persoon volledig binnen de normale intervallen kunnen liggen. Gewogen sommen worden geregeld gebruikt binnen de multivariate statistische methoden zoals discriminantanalyse, logistische regressie of multi-pele lineaire regressie. Daarnaast kunnen de huidige computers krachtige software pakketten aan welke “stapsgewijze” analyses mogelijk maken. Hierbij vertrekt men van een groot aantal variabelen voor een, bij voorkeur, groot aantal personen, en laat de softwareprocedure de optimale combinatie van variabelen zoeken die het best de “outcome” beschrijft. Een dergelijke techniek was reeds succesvol bij de beschrijving van stemplooi-knobbels aan de hand van fonetogramgegevens (Heylen, L. et al., 1998). Ook Piccirillo et al (1998), hebben een index gedefinieerd die gebaseerd is op aërodynamische en fonetogramgegevens. Zij hebben echter enkel die parameters weerhouden die bij univariate vergelijkingen tussen een normale en dysfone populatie significant verschilden. Desalniettemin is het net de combinatie van variabelen die gevoelig is voor verschillen tussen groepen; de keuze van variabelen dient bepaald te worden door de stapsgewijze statistische methoden, zoals bij logistische regressie.

Een andere benadering voor het opstellen van een globale stemkwaliteitsmaat werd beschreven door de groep van Callan et al (1999), waar “self organizing maps” worden gehanteerd voor de classificatie van een stem als normaal of pathologisch. Er worden aërodynamische, noch fonetografische gegevens gebruikt, enkel een bepaalde set van akoestische variabelen. Het is echter onze overtuiging dat het in de eerste plaats de taak is van de clinicus om een oordeel te vellen of een stem pathologisch is of normaal.

Algemeen kunnen we besluiten dat er behoefte is aan een globale stemkwaliteitsmaat die bij voorkeur het multidimensionele karakter van de stem weerspiegelt. Deze maat dient robuust te zijn, wat impliceert dat verschillende onderzoekers tot dezelfde uitkomst dienen te komen voor eenzelfde stem. Aangezien de perceptuele beoordeling de gouden standaard is voor de evaluatie van stemkwaliteit, omdat ze gebaseerd is op het oordeel van de “expert”, dient deze stemkwaliteitsmaat tevens zoveel mogelijk

overeen te stemmen met dit oordeel. Individuele metingen zoals akoestische, aërodynamische of fonetografische voldoen niet aan deze criteria voor het beschrijven van de globale stemkwaliteit. In de hier beschreven studie wordt een multidimensionale index voorgesteld, die een maat is voor de globale stemkwaliteit en die gebaseerd is op zowel aërodynamische, akoestische als fonetografische karakteristieken. Voor het in-calculeren van de perceptuele beoordeling werd deze index geënt op de gepercipieerde ernst van de heesheid (G van de GRBAS schaal), eerder dan op het verschil tussen normaal versus pathologisch. Aangezien het een weerspiegeling is van een objectieve correlatie voor gepercipieerde stemkwaliteit wordt het de Dysphonia Severity Index genoemd, i.e. de DSI (Wuyts, F. L. et al., 2000).

Materiaal en methode

Personen

De methode voor de constructie van de DSI wordt uitvoerig beschreven in Wuyts, F. L. et al., 2000, maar voor de duidelijkheid worden hier de belangrijkste punten herhaald.

De stemmen werden geselecteerd uit een database die geconstrueerd werd in het

Tabel 1. Aantal patiënten per diagnostische categorie die deel uitmaakten van de database van de Belgian Study Group on Voice Disorders.

Diagnose	N
Controle	68
Onvolledige sluiting	48
Stemplooi knobbels	30
Reinke oedeem	30
Chronische laryngitis	27
Excessieve musculaire spanning	26
Paralyse in abductie	23
Tumor	22
Sulcus glottidis en litteken	21
Paralyse in adductie	20
Mucosale cyste	14
Granuloma	10
Acute laryngitis	8
Haemoragie en trauma	7
Spasmodische dysfonie	6
Ventriculaire fonatie	5
Poliep	4
Psychogene afonie	3
Andere	15
Totaal	387

kader van de “Belgian Study Group on Voice Disorders” (Van de Heyning, P. H. et al., 1996). Deze database bevat de gegevens van patiënten die in vier verschillende Belgische universitaire centra werden onderzocht met betrekking tot stemproblemen. Er werden gedurende 16 maanden gegevens verzameld. Voor elke patiënt werd een standaard protocol gevolgd. De methoden werden gedetailleerd elders beschreven (De Bodt, M. S., Wuyts, F. L., Raes, J. P. F., & Qiu, J., 1996). De gegevens van 387 volwassenen (18-80 jaar, 53% vrouwen, 47% mannen) werden gebruikt voor de berekening van de index. De ondergrens werd op 18 jaar bepaald om stemmutaties en stemmaturatie uit te sluiten, zoals gesuggereerd door een aantal auteurs (Lee, P. A., 1980; Thurman, L. & Klitzke, C. A., 1994). Er werden tevens 68 normale personen (43 vrouwen, 25 mannen) geïncludeerd in de studie. Deze waren vrij van enige stemplooiopathologie of stemklacht, en werden door toeval geselecteerd in de diverse deelnemende centra. De overige 319 patiënten hadden diverse soorten pathologie zoals weergegeven in tabel 1.

Het klinisch onderzoek omvatte anamnese en klinische antecedenten, stroboscopische evaluatie, perceptuele beoordeling, akoestische analyse, fonetografie, aërodynamische metingen en een zelfbeoordeling. De criteriumvaliditeit werd onderzocht bij een afzonderlijke groep van 40 patiënten die sequentieel de Stemkliniek van het Universitair Ziekenhuis van Antwerpen bezochten.

Dataregistratie

De akoestische metingen werden opgenomen met het MDVP van Kay Elemetrics. Een mid-vowel segment van de klinker /a/ - uitgesproken bij normale luidheid en toonhoogte- werd gebruikt bij de analyse. Om de index toegankelijk te maken voor onderzoekers die niet beschikken over de geavanceerde apparatuur, werd initieel geopteerd om enkel de variabelen jitter (%), shimmer (dB) en Noise to Harmonics Ratio (NHR) te weerhouden, aangezien deze de meest gebruikte akoestische stemparameters zijn (De Bodt, M. S., 1997a). De aërodynamische variabelen omvatten de maximale fonatietijd (MPT in s), de vitale capaciteit (VC in cc) en de fonatiequotiënt (PQ in cc/s). De fonetogramvariabelen zijn: 1) laagste intensiteit (I-Low); 2) hoogste intensiteit (I-High); en 3) intensiteitsbereik (I-range = I-High - I-Low). Het intensiteitsbereik definieert het totale dynamisch bereik zonder referentie naar de frequentie. De intensiteiten werden gemeten in dB(A). De frequentiekaracteristieken (in Hz) zijn: 1) laagste F_0 (F_0 -Low); 2) hoogste F_0 (F_0 -High); 3) F_0 range (F_0 bereik = F_0 -High - F_0 -Low) en 4) semitoon bereik (ST-range).

De perceptuele beoordelingen werden uitgevoerd met de GRBAS schaal (Hirano, M., 1981), waar G staat voor ernst, R voor ruwheid, B voor wilde lucht, A voor asthenie en S voor gespannenheid.

De aanwezigheid van andere mogelijk beïnvloedende factoren zoals roken, stemmisbruik, allergie, neusobstructie, stress, chronisch obstructieve longaandoening en infectie van de bovenste luchtwegen werd geregistreerd.

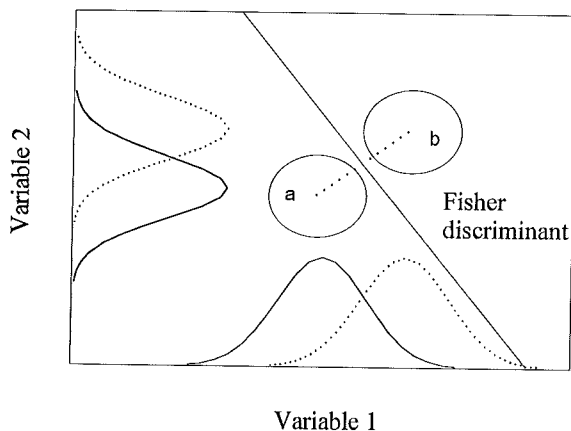
In het kader van de berekening van de index werden alle personen ingedeeld volgens hun G score van de GRBAS schaal, van 0 tot 3. De variabelen die werden opgege-

ven bij de berekening van de index zijn: jitter (%), shimmer (dB), NHR, F_0 -Low (Hz); F_0 -High (Hz), F_0 range (Hz), ST-range, I-High (dB), I-Low (dB), I-Range (dB), MPT (s), VC (cc), PQ (cc/s) en G.

Statistiek

Het normaal gedrag van de distributiefuncties van de diverse variabelen werd onderzocht met behulp van de Kolmogorov-Smirnov test (SPSS V8.0). Vervolgens werd – voor alle variabelen – nagegaan of de varianties gelijk zijn voor de 4 groepen (d.w.z. de patiënten met G0 tot G3) aan de hand van “Generalized Linear Model Anova”. De DSI werd geconstrueerd analoog aan de Fisher’s lineaire discriminant methode (Fisher, L. D. & Van Belle, G., 1993). Dit is een standaardmethode binnen de multivariate statistiek voor het onderscheiden van 2 of meer verschillende groepen op basis van een specifieke combinatie van variabelen. Bij verschillende patiëntenpopulaties wordt een speciale regel geconstrueerd zodat een nieuwe patiënt geklasseerd wordt binnen een van de populaties, op basis van de opgestelde regel. Wanneer het slechts twee populaties betreft, kan de Fisher’s discriminant regel voorgesteld worden door de schuine stippellijn op figuur 1, waarbij de waarden voor de 2 beschouwde variabelen weergegeven zijn door de ovals. De Gauss-distributies op de assen stellen de verdelingen voor bekeken per variabele en per populatie. Enkel de schuine stippellijn, die een maat is voor een lineaire combinatie van de twee beschouwde variabelen, kan beide groepen goed onderscheiden. Deze lineaire combinatie geeft een getal dat indicatief is voor de mate van afwijking.

Aangezien we in onze studie meer dan 2 populaties en variabelen hebben is het meerdimensionale analoge van deze figuur niet te visualiseren. Omdat de variabelen niet normaal verdeeld zijn, werd een semi-parametrische benadering gehanteerd, de



Figuur 1. Illustratie van de verdeling van twee variabelen voor twee populaties (a & b) op twee assen. De Fisher discriminantregel kan beschouwd worden als de schuine lijn die het best beide populaties scheidt. De DSI kan dan gevisualiseerd worden als de stippellijn tussen beide populatie clusters.

“proportional odds logistic regression” (Hosmer, D. W. & Lemeshow, S., 1989). Deze techniek vereist minder aannames dan de Fisher’s lineaire discriminanttechniek. Om de relevante variabelen te selecteren werd een stapsgewijze methode toegepast. Het voordeel van deze techniek is dat er meer dan twee groepen kunnen beschouwd worden. De G score (graad van heesheid) van de GRBAS schaal werd gekozen om de groepen patiënten te identificeren en op die manier het perceptuele karakter van de stem impliciet in de index te incorporeren. De G score werd gekozen omdat deze een globaal perceptueel beeld geeft van de heesheid, in tegenstelling tot de andere GRBAS elementen die slechts bepaalde delen van de stemkwaliteit beoordelen. Door middel van de proportional odds logistic regression procedure werd een discriminantregel opgesteld met slechts enkele variabelen waardoor een nieuwe patiënt geclassificeerd kan worden in één van de groepen G0 tot G3, op basis van de waarden voor de geselecteerde variabelen. De classificatieregule werd vervolgens getransformeerd om aldus een “outcome measure” te bekomen die de DSI werd genoemd. De berekeningen werden uitgevoerd met het statistisch pakket SAS (SAS Institute Inc, Cary, NC).

Om de invloed op de DSI te berekenen die ontstaat door de meetfouten op de individuele metingen van de variabelen, werd bij 6 gezonde personen (3m/3v) op 5 opeenvolgende dagen tweemaal per dag elke variabele gemeten. Aldus werd de standaardafwijking berekend voor elk van de variabelen die gebruikt werd bij het berekenen van de globale fout op de DSI, gebruik makend van de foutenpropagatiemethode (Welsh, A.H., 1996).

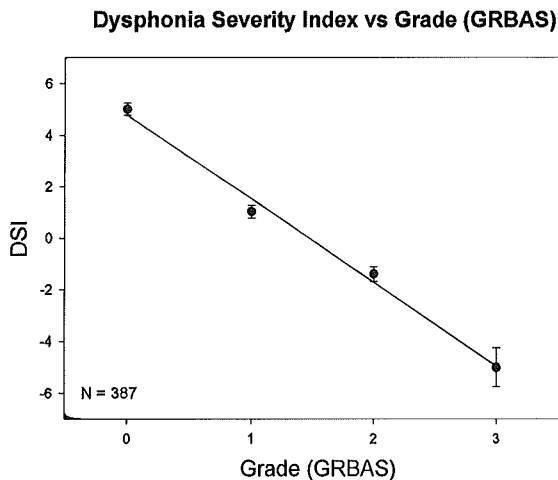
Voice Handicap Inventory

De psychosociale impact van de stemproblemen zoals ervaren door de patiënt zelf werd geëvalueerd met de Nederlandse vertaling van de Voice Handicap Inventory (VHI) (Jacobson, B. et al., 1997; De Bodt, M. S. et al., 2000). Deze vragenlijst bevat 30 items welke vragen naar de impact van de stemstoornis op het dagelijks leven van de patiënt. Het is een statistisch robuuste vragenlijst met een hoge interne consistentie en een goede test-retest betrouwbaarheid (Jacobson, B. et al., 1997). Als dusdanig weerspiegelt de VHI de beleving van de stemstoornis door de patiënt zelf. Het is deze die de aanleiding was voor het bezoek aan de keel-, neus-, oorarts. Een goede correlatie tussen de VHI en de DSI impliceert dus een goede overeenkomst tussen de beleving/perceptie van de stem door de patiënt zelf en de objectieve metingen. Aangezien de perceptuele beoordeling door de clinicus impliciet in de DSI verweven zit, breidt de goede overeenkomst zich ook uit naar de perceptie van de clinicus. Als dusdanig kan een hoge correlatie tussen DSI en VHI geïnterpreteerd worden als een maat voor de criteriumvaliditeit van de DSI.

Bij 40 patiënten die achtereenvolgens de stemafdeling bezochten werd de VHI afgenomen en werd de Pearsons correlatie berekend met de DSI.

Resultaten

Behalve de I-High en de vitale capaciteit was geen van de variabelen normaal ver-



Figuur 2. Dysphonia Severity Index versus graad van heesheid (G van de GRBAS schaal).

deeld, wat het gebruik van de “proportional odds logistic regression” verantwoordt. De volgende variabelen werden geselecteerd door deze statistische procedure, omdat ze, wanneer gebruikt in een specifieke combinatie, het best de G score reflecteren: Fo-High (Hz), I-Low (dB), MPT (s) en jitter (%). De DSI welke de discriminantregel is zoals berekend door de logistische regressie is een lineaire combinatie van deze vier variabelen, waarbij de gewichten door de statistische procedure werden toegekend als (Wuyts, F. L. et al., 2000) $DSI = 0.13 * MPT(s) + 0.0053 * Fo-High(Hz) - 0.26 * I-Low(dB) - 1.18 * jitter (%) + 12.4$.

De DSI is een gewogen combinatie van variabelen zodat de uitkomst het best de mate van heesheid weergeeft, zoals weerspiegeld in de G van de GRBAS schaal.

De relatie tussen de G en de DSI wordt in figuur 2 weergegeven.

Naargelang de DSI negatiever is, des te slechter is de stemkwaliteit van de patiënt. Naarmate de DSI positiever is, des te normaler is de stem. De initieel berekende gewichten werden getransformeerd om een hanteerbare schaal te krijgen waarbij -5 overeenkomt met het gemiddelde van de stemmen uit de groep G3 en +5 met het gemiddelde uit groep G0.

Tabel 2 geeft de DSI voor de verschillende G-scores.

De meetfouten op de individuele variabelen uit de DSI leiden tot een globale meetfout van 0.64 op de DSI. Dit is gebaseerd op de gemeten standaard afwijking van 1.6 s voor MPT, 39 Hz voor Fo-High, 1.7 dB voor I-Low en 0.3 % voor de jitter.

Om de betrouwbaarheid van de DSI te evalueren werd een classificatietabel opgesteld. Deze toont de overeenstemming tussen realiteit en voorspelling door de DSI. Indien er alleen elementen op de diagonaal vallen, dan is er een perfecte overeenkomst. Deze bedraagt voor de DSI 50%. Als zowel de diagonaal als de elementen net naast de diagonaal worden beschouwd, dan is er een overeenkomst tussen realiteit en voorspelling van 98.4%. Zou de overeenkomst puur op basis van toeval zijn, dan ver-

Tabel 2. Gemiddelde waarden voor de DSI (\pm standard fout) bij de personen met G₀ tot G₃-scores.

G score	DSI \pm se
G ₀	5.00 \pm 0.23
G ₁	1.02 \pm 0.235
G ₂	-1.4 \pm 0.2
G ₃	-5.0 \pm 0.28

Tabel 3. Classificatietabel met de geobserveerde en voorspelde deelname in klassen.

		geobserveerd			
		G ₀	G ₁	G ₂	G ₃
voorspeld	G ₀	12	3	0	0
	G ₁	44	91	61	6
	G ₂	0	52	83	20
	G ₃	0	0	8	7

Tabel 4. Gemiddelden (\pm se) voor de DSI en zijn componenten voor mannen en vrouwen met normale stemmen.

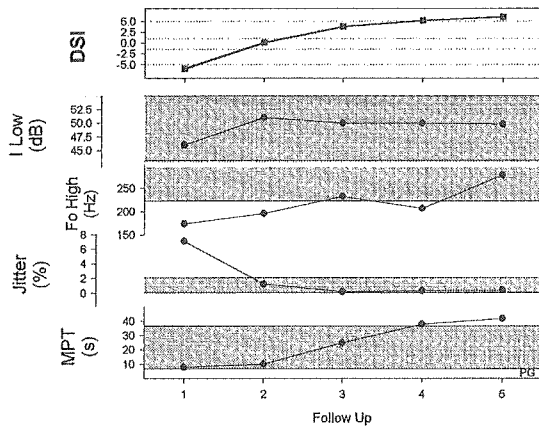
	MPT (s)	F0-High (Hz)	I-Low (dB)	Jitter (%)	DSI
vrouwen (N=43)	16.9 \pm 0.7	905 \pm 31	51.3 \pm 0.2	0.79 \pm 0.10	5.22 \pm 0.26
mannen (N=25)	22.2 \pm 1.7	602 \pm 34	50.4 \pm 0.5	0.63 \pm 0.06	4.7 \pm 0.4

wacht men 62.5 % overeenkomst.

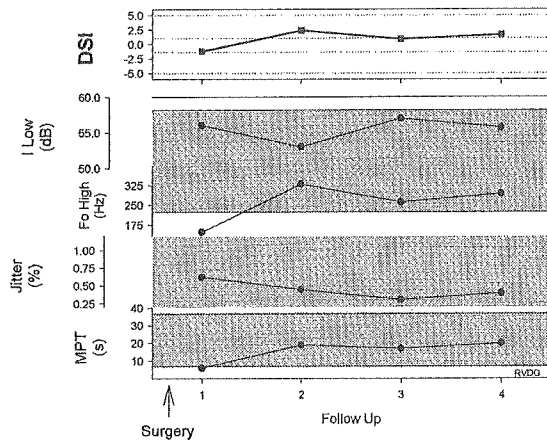
In tabel 4 worden de waarden voor de verschillende variabelen die deel uitmaken van de DSI weergegeven voor mannen en vrouwen met normale stemmen. Het effect van geslacht wordt opgegeven in de DSI omdat de voorfactoren voor MPT en Fo-High hetzelfde teken hebben. Aangezien mannen een hoge MPT maar een lage Fo-High hebben en dit net andersom is bij vrouwen, is het derhalve niet nodig twee aparte geslachtsafhankelijke DSIs te construeren.

Figuren 3, 4 en 5 illustreren de DSI en de vier variabelen voor diverse patiënten die behandeld werden voor hun stemprobleem, hetzij chirurgisch, hetzij logopedisch, hetzij met een combinatie. De veranderingen tijdens de follow-up periode tonen de schommelingen van de individuele variabelen, en het globale effect op de DSI. De grijze schaduwzones indiceren de normale limieten, gebaseerd op de 95% predictie intervallen (Wuyts, F. L., De Bodt, M. S., Bruckers, L., & Molenberghs, G., 1996).

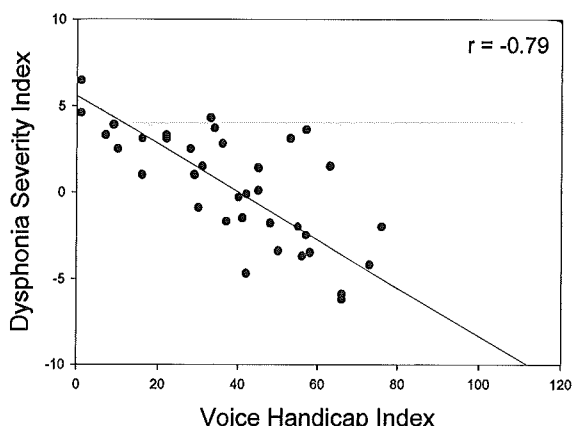
De Pearsons correlatie tussen de DSI en de VHI bedraagt -0.79 ($p < 0.001$). Figuur 6 illustreert deze relatie.



Figuur 3. Follow-up van een 58 jaar oude man met heesheid ten gevolge van een paralyse van de n recurrens na chirurgie van de onderste craniale zenuwen. Stem therapie gedurende 8 maanden: (1): gegevens 15 dagen na de operatie (p.o.). $G_2R_1B_2A_2S_0$, start van de stem therapie; (2): gegevens 5 maanden p.o. $G_2R_1B_1A_1S_0$; (3): 3. gegevens 8 maanden p.o.: einde van de stemtherapie, $G_0R_0B_0A_0S_0$; (4): gegevens 16 maanden p.o.: $G_0R_0B_0A_0S_0$, follow-up evaluatie; (5): gegevens 18 maanden p.o.: $G_0R_0B_0A_0S_0$, follow-up evaluatie. Schaduwzones zijn de 95 % Predictie Intervallen (gemiddelde $\pm 2SD$).



Figuur 4. Follow-up van een 54-jarige patiënt die een chirurgische reductie van een Reinkes oedeem onderging. De patiënt was amateurzanger met een slechte stemhygiëne. (1): Drie weken p.o.: $G_2R_2B_1A_1S_1$, start van de stemtherapie en stemhygiëne; (2): Drie maanden p.o.: $G_1R_1B_0A_1S_1$; (3): Zes maanden p.o.: $G_1R_1B_1A_0S_0$; einde van de stemtherapie. De patiënt volgde amper het hem voorgeschreven programma qua stemhygiëne. Hij bleef tevens roken en zingen; (4): 24 maanden p.o. : $G_1R_1B_0A_0S_1$, status quo zowel qua stemgedrag als qua stemplooi-toestand.



Figuur 5. Follow-up van een 34 jarige vrouw die een stemplooiemialisatie onderging (thyroplastiek, Isshiki II) ten gevolge van een n laryngeus superiorparalyse. (1): Pre operatieve evaluatie; $G_2R_0B_2A_2S_0$; (2): Tien dagen post chirurgie (p.o), verslechtering ten gevolge van oedeem; $G_3R_0B_2A_2S_0$, begin van de stemtherapie; (3): Vier weken p.o.; $G_1R_0B_1A_0S_0$; (4): Zes weken p.o.; $G_1R_0B_1A_0S_0$; (5): Acht weken p.o.; $G_1R_0B_1A_0S_0$; (6): Tien weken p.o.; $G_1R_0B_1A_0S_0$, de patiënt is volledig tevreden met het behaalde resultaat. Einde van de therapie; (7): Acht maanden p.o.; $G_1R_0B_1A_0S_0$. Een controlemeting indiceert een stabiele stemkwaliteit.

Discussie

Akoestische, aërodynamische en fonetogram metingen vallen voor een patiënt vaak binnen de normale grenzen. Dit komt omdat deze grenzen in het algemeen erg breed zijn. Slechts waarden die sterk afwijken van het normale zijn in een klinische context conclusief. In deze studie presenteren we een objectieve maatstaf voor stemkwaliteit waarbij het multidimensionele karakter van de stem werd geïncorporeerd door middel van multivariate statistiek. De DSI bestaat uit een specifieke gewogen combinatie van de hoogste fundamentele frequentie, de laagste intensiteit, de maximale fonatietijd en de jitter. De DSI werd zodanig getransformeerd dat de groep personen met stemmen die gekarakteriseerd worden door een G_0 een DSI hebben die correspondeert met +5. De gemiddelde DSI voor patiënten die perceptueel gemiddeld een G_3 hebben correspondeert met -5. Hoe meer negatief de DSI voor een patiënt, hoe slechter zijn of haar stem. Hoe hoger de DSI, hoe beter de stem. Desalniettemin is de DSI niet beperkt tot het interval +5, -5. In onze klinische praktijk worden soms hogere of lagere waarden gevonden dan respectievelijk +5 en -5. Dit wordt meestal veroorzaakt door een abnormaal hoge waarde voor één van de individuele metingen. Sterke afwijkingen dienen door de onderzoeker kritisch bekeken te worden. De gebruiker dient erop toe te zien dat bij bepaalde dysfone stemmen het niet-periodische karakter van het stemgeluid aanleiding kan geven tot afwijkende en onbetrouwbare akoestische metingen (Titze, I. R., 1995).

De keuze van de variabelen in de DSI wordt volledig bepaald door de stapsgewijze logistische regressieprocedure. Het lijkt desalniettemin logisch dat de hoogste fre-

quentie één van de geselecteerde variabelen is. Het is namelijk zo dat bij vijftig procent van de hese patiënten de stemplooien aangetast zijn en met name een extra hoeveelheid massa bevatten (stemplooiknobbels, oedeem, etc. - zie tabel 1). Deze extra massa, die in het algemeen inhomogeen verdeeld is over de stemplooien, belemmert het trillen bij hogere frequenties, wat zich uit in een verminderde F0-High. Op gelijkaardige wijze zal de aanwezigheid van stemplooiknobbels, oedeem, etc. de glottische weerstand doen toenemen zodanig dat een grotere druk nodig zal zijn voor het initiëren en instandhouden van stemplooitruillingen (Colton, R., 1994). Bijgevolg zal ook de laagste intensiteit toenemen bij patiënten met een hese stem. Gelijkaardige effecten voor F0-High en I-Low werden gevonden in fonetogramstudies bij kinderen (Heylen, L. et al., 1998). Perturbatiemetingen, zoals de jitter, zijn ontwikkeld voor het evalueren van de onregelmatigheid van de stemplooitruillingen. Daarom is het waarschijnlijk dat een hese stem zal gekarakteriseerd worden door verhoogde waarden van perturbatieparameters. De keuze van de maximale fonatietijd als relevante variabele voor de berekening van de DSI kan zijn oorsprong vinden in het feit dat deze parameter de efficiëntie weerspiegelt waarmee het stemapparaat stemgeluid produceert. Factoren zoals luchtweerstand, subglottische druk, en glottissluiting beïnvloeden deze efficiëntie (Hirano, M., 1981).

Piccirillo et al (1998) hebben eveneens een concept voor een vocale functie index ontwikkeld, maar hebben voornamelijk de classificatie tussen normaal en dysfoon beklemd. Met behulp van logistische regressie vonden zij dat een gewogen combinatie van de geschatte subglottische druk, de luchtstroom ter hoogte van de lippen, de vocale efficiëntie en de maximale fonatietijd, discriminatie mogelijk maakte tussen gezonde en pathologische stemmen. Het is vermeldenswaard dat de maximale fonatietijd zowel in hun studie als in ons werk een belangrijke variabele blijkt te zijn voor de algemene beoordeling van stemkwaliteit. Een ander verschil tussen onze werkwijze en deze van Piccirillo et al (1998) ligt in de discriminatie tussen de verschillende gradaties van heesheid (G) enerzijds en tussen normaal en pathologisch anderzijds. Daarbij dient opgemerkt dat de relatie tussen pathologie en heesheid niet triviaal is, aangezien een ernstige pathologie niet noodzakelijk steeds aanleiding geeft tot een ernstige heesheid en vice versa.

De classificatietabel (tabel 3) illustreert de efficiëntie van de toegepaste methode. Een ideale classificatiemethode zou alleen waarden op de diagonaal voortbrengen, wat in de praktijk echter niet wordt gerealiseerd. Het is echter inherent aan een perceptuele beoordeling dat een zekere variabiliteit aanwezig is. Immers, wanneer een groep beoordelaars een aantal patiënten evalueert met behulp van een perceptuele schaal, dan is de interbeoordelaarsovereenstemming op zijn best "goed", maar nooit excellent, zoals weergegeven door de kappa statistiek (De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H., & Croux, C., 1997). Derhalve is het realistisch te veronderstellen dat in sommige gevallen een gecercipieerde G1 evenwel een G2 had kunnen zijn of vice versa. Dit wordt weerspiegeld door de off-diagonal elementen in de classificatietabel. Dit impliceert dat voor een patiënt de gecercipieerde waarde een bepaalde G score kan hebben, daar waar de DSI mogelijk een score net hoger of lager toekent. Dit is inderdaad bij een aantal patiënten het geval. In slechts 6 van de 387 gevallen, klasseert de

DSI echter een patiënt meer dan één scorewaarde afwijkend van de waargenomen waarde. Uiteindelijk dient opgemerkt dat de DSI volstrekt niet tot doel heeft om te classificeren, maar om te graderen.

Wegens het tegenovergestelde gedrag van de variabelen F0-High en MPT voor mannen en vrouwen is er geen behoefte aan een geslachtsafhankelijke DSI.

Om het klinische belang van de DSI en zijn eenvoud in het gebruik te illustreren, stellen we een aantal voorbeelden voor. Figuren 4, 5 en 6 illustreren veranderingen in stemkwaliteit bij drie patiënten onder invloed van diverse therapieën. In sommige gevallen was de pathologie nog steeds aanwezig na therapie, ook al was de stem functie verbeterd, volgens de keel-, neus-, oorarts en de patiënt. In andere gevallen uitte het effect van de therapie zich slechts in de verbetering van één variabele en werden de andere variabelen slechter. De globale trend van de DSI maakt de interpretatie van dergelijke tegenstrijdige effecten aanzienlijk eenvoudiger.

Het duurt slechts 10 tot 15 minuten om bij een patiënt de diverse klinische parameters voor de DSI te meten en de index te berekenen volgens de hoger vermelde formule. Daarvoor kan een rekenmachine of een spreadsheet¹ op PC gebruikt worden. De ankerpunten +5 en -5 vergemakkelijken daarenboven de interpretatie van de DSI wanneer het een geïsoleerde meting betreft. In het appendix wordt een protocol beschreven voor het afnemen van de DSI.

Tot slot dienen nog de inhoud- en de criteriumvaliditeit van de DSI besproken te worden. Inhoud ("content") validiteit heeft betrekking op de vraag of de index effectief weergeeft wat bedoeld wordt, met andere woorden: dekt de naam de lading. Is de DSI inderdaad een maat voor de stemkwaliteit? De vier variabelen die deel uitmaken van de DSI zijn individueel allemaal duidelijke indicatoren van heesheid, aangezien de gemiddelde waarden significant verschillen tussen normale en hese stemmen (Wuyts, F. L., De Bodt, M. S., Bruckers, L., & Molenberghs, G., 1996). Daarnaast is de selectie van deze variabelen gebaseerd op een stapsgewijze statistische methode welke het opstellen van een rekenregel beoogt voor het classificeren van stemmen volgens hun perceptuele G-score. Zo wordt intrinsiek in de DSI de perceptuele beoordeling van de stemspecialist geïncorporeerd. Derhalve kunnen we concluderen dat de DSI voldoet aan de criteria van inhoudvaliditeit.

Criteriumvaliditeit slaat op de betrouwbaarheid van de DSI. Hoe verhoudt de DSI zich tot "de gouden standaard"? Perceptuele beoordelingen worden typisch aanzien als de finale scheidsrechter in het management van een ziekte en worden daarenboven vaak als standaard gebruikt waartegen instrumentele metingen worden afgewogen (Kent, R. D., 1996). Zoals reeds aangehaald is een dergelijke perceptuele beoordeling, met name de G van de veel gebruikte GRBAS-schaal, in de DSI geïncorporeerd. Om de DSI met een externe meting te vergelijken, werd hij gecorreleerd aan de Voice Handicap Index, die per definitie een onafhankelijke maat is. De hoge correlatie tussen beide maten onderbouwt sterk de criteriumvaliditeit van de DSI. Een goede overeenkomst tussen de DSI en de VHI wijst erop dat de DSI niet alleen een goede maat is voor de stemkwaliteit maar eveneens op een goede wijze de handicap ten gevolge van de heesheid weerspiegelt, zoals deze door de patiënt zelf wordt ervaren.

Twee jaar dagelijks gebruik heeft de auteurs geleerd dat de DSI een praktisch hulp-

middel is voor het beschrijven van stemkwaliteit. Het speelt een belangrijke rol in de globale benadering van de patiënt met een stemprobleem. Algemeen kunnen we stellen dat een model voor de stemevaluatie bestaat uit vier componenten (De Bodt, M. S., 1997b): een laryngoscopische inspectie van de stemplooien, een perceptuele evaluatie van de heesheid (met bv. de GRBAS), een subjectieve evaluatie door de patiënt zelf (met bv. de VHI), en de DSI. Het is onze overtuiging dat een dergelijk vier-componentenmodel voldoende is om in een klinische omgeving binnen een beperkt tijdsbestek toch op een wetenschappelijke wijze de stem te evalueren.

De parameters die deel uitmaken van de DSI zijn algemeen toegankelijke parameters die werden overgenomen door de Werkgroep Fonochirurgie van de European Laryngological Society (Graz, 9/10 oktober 1999). Aangezien het objectief meetbare grootheden zijn, is de DSI niet onderhevig aan subjectieve invloeden, wat bij perceptuele beoordelingen wel het geval is. De betrekkelijk kleine meetfout, 0.6 op een schaal van 10, en het feit dat de DSI gebaseerd is op een multicenterstudie onderbouwt zijn robuustheid. Daarnaast draagt het feit dat in de DSI zowel aërodynamische, akoestische als fonetografische parameters vervat zijn, ertoe bij om de DSI te beschouwen als een multidimensionele maat voor stemkwaliteit. Het voorziet de onderzoeker van een maat voor de stemkwaliteit, zonder invloed van tijd, verwachtingspatroon, of andere invloeden waarvan het effect op perceptuele beoordeling eerder werd aangetoond door De Bodt et al elders in dit nummer. Daarenboven kan de onderzochte stem geplaats worden binnen een geheel van hese stemmen. Bij een universeel gebruik laat de DSI toe verschillende therapeutische benaderingen te vergelijken. De DSI kan zich derhalve ontplooien tot zowel een wetenschappelijk onderzoeksinstrument als een methode voor dagelijks gebruik bij de klinische evaluatie van de stemkwaliteit.

Noot

De benodigde Excel spreadsheet is gratis verkrijgbaar bij wuyts@uia.ua.ac.be

Referenties

- Andrews, M. L. (1995). *Manual of Voice Treatment. Pediatrics Through Geriatrics*. San Diego: Singular Publishing Group, Inc.
- Askenfeld, A. G. & Hammarberg, B. (1986). Speech Waveform perturbation analysis a perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research*, 29, 50-64.
- Callan, D. E., Kent, R. D., Roy, N., & Tasko, S. M. (1999). Self-Organizing Map for the Classification of Normal and Disordered Female Voices. *Journal of Speech, Language and Hearing Research*, 42, 355-366.
- Colton, R. (1994). Physiology of Phonation. In M.S.Benninger, B. H. Jacobson, & A. F. Johnson (Eds.), *Vocal Arts Medicine. The care and prevention of Professional Voice Disorders* (pp. 30-60). Stuttgart: George Thieme Verlag.
- Crevier-Buchman, L., Laccourreye, O., Wuyts, F. L., Monfrais-Pfauwadel, M., Pillot, C., &

- Brasnu, D. (1998). Comparison and evolution of perceptual and acoustic characteristics of voice after supracricoid partial laryngectomy with cricothyroidopiglottopexy. *Acta Otolaryngologica*, 118, 594-599.
- De Bodt, M. S. (1997a). *A framework for voice assessment. The relation between subjective and objective parameters in the judgement of normal and pathological voice*. University of Antwerpen.
- De Bodt, M. S. (1997b). *Een onderzoeksmodel voor stemevaluatie. De relatie tussen subjectieve en objectieve parameters in de beoordeling van de normale en pathologische stemfunctie*. Universiteit Antwerpen.
- De Bodt, M. S., Jacobson, B. H., Musschoot, S., Zaman, S., Heylen, L., Mertens, F., Van de Heyning, P. H., & Wuyts, F. L. (2000). De Voice Handicap Index: Een instrument voor het kwantificeren van de psychosociale consequenties van stemstoornissen. *Logopedie*, 13.
- De Bodt, M. S., Wuyts, F. L., Raes, J. P. F., & Qiu, J. (1996). Research work of the Belgian Study Group on voice Disorders 1996. Materials and methods. *Acta oto-rhino-laryngologica belgica*, 50, 325-329.
- De Bodt, M. S., Wuyts, F. L., Van de Heyning, P. H., & Croux, C. (1997). Test-Retest Study of the GRBAS Scale: the Influence of Experience and Professional Background on Perceptual Rating of Voice Quality. *Journal of Voice*, 11, 74-80.
- Dejonckere, P. H. & Lebacqz, J. (1996). Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *ORL*, 58, 326-332.
- Dejonckere, P. H., Remacle, M., Fresnel-Elbaz, M., Woisard, V., Crevier-Buchman, L., & Millet, B. (1996). Differential perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Review Laryngology Otolaryngology Rhinology*, 117, 219-224.
- Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, 33, 298-306.
- Fisher, L. D. & Van Belle, G. (1993). *Biostatistics. A Methodology for the Health Sciences*. New York: A Wiley-Interscience Publication. John Wiley & Sons, Inc.
- Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica*, 90, 441-451.
- Heylen, L., Wuyts, F. L., Mertens, F., De, B. M., Pattyn, J., Croux, C., & Van-de, H. P. (1998). Evaluation of the vocal performance of children using a voice range profile index. *Journal of Speech Language and Hearing Research*, 41, 232-238.
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech and Hearing Research*, 37, 769-778.
- Hirano, M. (1981). *Clinical Examination of Voice*. Springer Verlag: New York.
- Hirano, M. (1990). Clinical Applications of Voice Tests. In NIDCD (Ed.), *Assessment of Speech and Voice Production* (pp. 196-203). Maryland: NIDCD.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.
- Jacobson, B., 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.
- Kent, R. D. (1996). Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, 5, 7-23.
- Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of Acoustical Society of America*, 96, 1291-1301.
- Lee, P. A. (1980). Normal ages of pubertal events among American males and females. *Journal of Adolescent Health Care*, 1, 26-29.

- Piccirillo, J. F., Painter, C., Fuller, D., & Fredrickson, J. M. (1998). Multivariate analysis of objective vocal function. *Annual Otolaryngology Rhinology Laryngology*, 107, 107-112.
- Rabinov, C. R., Kreiman, J., Gerratt, B. R., & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research*, 38, 26-32.
- Thurman, L. & Klitzke, C. A. (1994). Voice Education and Health Care for Young Voices. In M.S. Benninger, B. Jacobson, & A. Johnson (Eds.), *Vocal Arts Medicine: The care and prevention of professional voice disorders* (pp. 226-268). New York: Thieme Medical Publishers, Inc.
- Titze, I. R. (1995). *Workshop on Acoustic Voice Analysis. Summary Statement*. Denver: National Center for Voice and Speech.
- Van de Heyning, P. H., Remacle, M., Van Cauwenberge, P., De Bodt, M. S., Wuyts, F. L., Moerman, M., Van Lierde, K., Raes, J. P. F., Heylen, L., Sasserath, M., Henri de Hassonville, A., Clement, P. A. R., Qiu, J., Molenberghs, G., Bruckers, L., Mertens, F., Pattyn, J., & Millet, B. (1996). Research work of the Belgian Study Group on Voice Disorders 1996. *Acta oto-rhino-laryngologica belgica*, 50, 321-386.
- Welsh, A. H. (1996). *Aspects of Statistical Inference*. New York: Wiley.
- Wolfe, V., Fitch, J., & Cornell, R. (1995). Acoustic prediction of severity on commonly occurring voice problems. *Journal of Speech and Hearing Research*, 38, 273-279.
- Wuyts, F. L., De Bodt, M. S., Bruckers, L., & Molenberghs, G. (1996). Research work of the Belgian Study Group on voice Disorders 1996. Results. *Acta oto-rhino-laryngologica belgica*, 50, 331-341.
- Wuyts, F. L., De Bodt, M. S., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., Raes, J. P. F., & Van de Heyning, P. H. The Dysphonia Severity Index: an objective measure of vocal quality based on a multi-parameter approach. *Journal of Speech, Language and Hearing Research* (in press).
- Wuyts, F. L., De Bodt, M. S., & Van de Heyning, P. H. (1999). Is the reliability of a visual analog scale higher than of an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice*, 13, 508-517.

Appendix

Klinisch evaluatieprotocol

1. Perceptuele evaluatie met de GRBAS-schaal.
2. Bepaling van de maximale fonatietijd (in seconden) op een aangehouden /a/ bij habituele luidheid en toonhoogte. De beste MPT van drie pogingen wordt genoteerd.
3. Bepaling van de hoogste frequentie (Hz) en de laagste intensiteit (dB SPL) door middel van een fonetogram. De patiënt mag glijdende tonen produceren voor het bereiken van de hoogste toon.
4. Bepaling van de jitter (%) met behulp van bij voorkeur het MDVP programma van Kay Elementics op een ± 3 seconden lang aangehouden /a/.
5. De berekening van de DSI wordt dan:

$$DSI = 0.13 * MPT + 0.0053 * Fo-High - 0.26 * I-Low - 1.18 * Jitter (%) + 12.4$$
6. Indien de DSI voor een normale stem systematisch afwijkend is van +5, dan is dit voornamelijk te wijten aan de methode voor de I-Low meting. Deze kan systema-

tisch te laag of te hoog zijn. Vergelijk daarom de gemiddelde I-Low voor een aantal normale stemmen en ga na of dit gemiddelde systematisch afwijkt van de gemiddelde waarde van 51 dB ($se=0.5$). Dit is de gemiddelde waarde voor I-Low zoals gebruikt voor onze studie. Als dit verschil te groot is dan zullen de DSI waarden systematisch te hoog of te laag zijn. Het is in dat geval zinvol om te corrigeren voor deze systematische afwijking. Stel $\langle I-Low \rangle^a$ is het gemiddelde bekomen in de lokale setting voor een normale groep. Het volstaat dan het verschil ($\langle I-Low \rangle^a - 51$) op te tellen bij de I-Low bekomen voor een individuele patiënt om een DSI te bekomen waarvan de schaal +5, -5 volledig toepasbaar is.