

Beoordeling van fonatiewijze

M. Vredenberg en R. Buekers

Afdeling KNO, Academisch Ziekenhuis Maastricht

In dit artikel worden de resultaten besproken van het onderzoek naar de betrouwbaarheid van de diagnostiek van fonatiewijze. Twee groepen logopedisten (wel of niet gespecialiseerd in stem) en twee andere groepen (semi- of niet deskundig op het gebied van stem) werd gevraagd om een beoordeling van de fonatiewijze te geven bij verschillende proefpersonen. Naar aanleiding van eerdere vaststellingen en onderzoeken bleek dat de logopedische beoordeling van stemtechniek tamelijk subjectief is. Dit onderzoek tracht een antwoord te geven op de vragen hoe overeenkomstig logopedisten de fonatiewijze van verschillende proefpersonen beoordelen, en waaruit de professionaliteit van logopedisten bij de diagnostiek van (verkeerde) stemtechniek blijkt. De conclusies zijn dat bij de vrije beoordelingen er weinig overeenstemming is tussen de beoordelaars, tevens wordt het tekort aan een eenduidige terminologie aangetoond. Bij de categorale beoordeling aan de hand van een voorgevormd formulier is er ook weinig overeenstemming, zelfs binnen de groepen beoordelaars. De professionaliteit van de logopedisten blijkt niet uit een grotere overeenstemming, noch wat betreft wijze van beoordelen noch wat betreft gebruikte terminologie.

Inleiding

Op het gebied van stemdiagnostiek wordt door logopedisten aandacht geschonken aan de bepaling van het stemgeluid of de stemkwaliteit en aan het vaststellen van de (verkeerde) wijze van stemgeving. Over de beoordelingen van het stemgeluid wordt al enige jaren onderzoek verricht naar de beste wijze, naar de overeenstemming tussen meerdere beoordelaars en naar de overeenkomst met de resultaten van akoestische analyses (De Moor, 1992, De Bodt et al., 1996, Dejonckere et al, 1997). Beoordeling van stemgedrag en stemtechniek is een belangrijk onderdeel van het stemdiagnostisch onderzoek (Laver, 1980, Stes, 1994). Toch wordt deze methodiek zelden beschreven in handboeken en is deze 'subjectieve' beoordeling weinig nader onderzocht. In het verleden is vaker getracht een eenduidig onderzoekformulier te ontwikkelen, maar dit blijft een ontwerp van slechts enkele personen of stemtherapeuten van een

bepaalde school. De opleidingen slagen er ook niet in op dit gebied een standaard formulier of beoordelingswijze te ontwikkelen. Kooyman (1995) liep al tegen dit probleem aan en vermeldt "grote verschillen in inzicht en het ontbreken van consensus ten aanzien van de terminologie" als de belangrijkste redenen. De Moor (1992) stelt dat de mate van afwijking verschillend beoordeeld wordt en dat de stembeoordeling vaak niet plaats vindt volgens een standaard systematiek. Kooijman (1995) bevestigt dit beeld: "Vrijwel elke logopedist, zelfstandig of in dienstverband, blijkt een eigen vorm van stemonderzoek met eigen terminologie, eigen specifieke aandachtsvelden en eigen vormen van ernst-inschattingen te hanteren, [...] normen of referentiepunten zijn er echter niet of nauwelijks binnen de stemdiagnostiek." Auteurs die tot soortgelijke conclusies komen zijn Smeets e.a. (1988), De Bekker en Moes (1988) en Kreiman et al. (1992).

De aanleiding tot dit onderzoek was de ervaring dat het bij een stembelastingsexperiment niet mogelijk was de proefpersonen in te delen naar goede of foutieve stemtechniek. Hierbij moesten meerdere beoordelaars (logopedisten) het onafhankelijk van elkaar eens zijn over de stemtechniek van kandidaten en toen bleek dat er vaak geen eenvormig oordeel bereikbaar was. Gezocht werd naar verklaring en verbetering hiervan, maar dit leverde weinig op. Bovendien bleek uit de publicatie van Kreiman et al. (1992) dat dit niet zo verwonderlijk was, want zij hadden vastgesteld dat de verschillen in de beoordelingen groter waren, naarmate de beoordelaars meer ervaring en specialisatie in stem hadden.

Los van het feit dat dit niets zegt over de validiteit van de stemdiagnostiek en de waarde van de stemtherapie als dusdanig, blijven grote beoordelingsverschillen tussen deskundigen toch vragen oproepen. Zo staan in dit onderzoek onder meer ter discussie:

- de overeenstemming van logopedische beoordelingen bij de diagnostiek van de fonatiewijze
- de professionaliteit van logopedisten met betrekking tot de diagnostiek van stemstoornissen.

Onderzoek

Voor het onderzoek zijn de beschreven knelpunten als volgt vertaald:

- Wat is de mate van overeenstemming bij het stellen van een diagnose over de fonatiewijze binnen verschillende groepen beoordelaars, onderscheiden naar deskundigheid?
- Wat is de mate van overeenstemming bij het stellen van een diagnose over de fonatiewijze tussen de groepen beoordelaars?
- Is er verschil in de mate van interne overeenstemming binnen een beoordelaarsgroep bij het stellen van een diagnose over de fonatiewijze?

Aan het onderzoek, dat in twee fasen is uitgevoerd, lagen de volgende hypothesen ten grondslag, die zonder meer aannemelijk genoemd mogen worden:

1. Naarmate de deskundigheid van een groep beoordelaars toeneemt, neemt ook de mate van interne overeenstemming over de fonatiewijze toe. Met andere woorden, we veronderstellen dat logopedisten die gespecialiseerd zijn op het gebied van stemstoornissen, het vaker met het elkaar eens zullen zijn bij het stellen van een diagnose, dan logopedisten zonder deze specialisatie, die op hun beurt meer overeenstemming bereiken dan ondeskundigen.
2. Door het verschil in deskundigheid van beoordelaars, en logopedisten in het bijzonder, zal er verschil optreden in de beoordeling van de fonatiewijze. We veronderstellen dat logopedisten met genoemde specialisatie een andere diagnose zullen stellen bij een willekeurige proefpersoon, dan logopedisten zonder deze specialisatie of dan ondeskundigen.

Methode en proefpersonen

In de eerste fase werd een videoband samengesteld met 12 proefpersonen. Deze groep bestond uit 4 personen met een geschoolde stem, 4 personen zonder stemklachten en zonder stemscholing en 4 personen met stemklachten. Van elk werd in studiocondities een video opname gemaakt van telkens ongeveer vijf minuten. Hierop voeren zij de volgende opdrachten uit: lezen van tekst, zingen van glijtonen, zingen van een liedje en een stukje (spontaan) gesprek.

Deze videoband werd voorgelegd aan een groep van 24 beoordelaars, gecategoriseerd naar de mate van deskundigheid. Zo werden vier groepen van zes beoordelaars onderscheiden:

1. zeer deskundigen, ofwel de 'gouden logopedisten' hoofdzakelijk werkzaam met stempatiënten en hiervoor extra geschoold;
2. deskundigen, ofwel de logopedisten (niet werkzaam met stempatiënten, maar wel hiervoor opgeleid);
3. semi-deskundigen, waaronder een zangpedagoog en fysiotherapeut;
4. niet-deskundigen (mensen die hier beroepshalve niet mee in aanraking komen, bv. verpleegkundige).

De beoordelaars kregen de videoband, een instructieformulier en 12 blanco formulieren toegezonden met het verzoek de fonatiewijze van elke proefpersoon te beoordelen. Het doel van de blanco beoordeling was inzicht te krijgen in welke aspecten van de fonatiewijze beoordelaars zouden beschrijven, zonder beïnvloeding vooraf.

De resultaten van de blanco beoordeling zijn verwerkt door te turven hoe vaak een bepaalde term voorkwam, welke ernst-aanduidingen en welke opmerkingen per categorie beoordelaars en per proefpersoon werden gemaakt.

In de tweede fase, na het retourneren van de blanco beoordeling, kregen de beoordelaars een zelf ontwikkeld scoreformulier toegezonden, waarbij gevraagd werd opnieuw de band te bekijken en voor elke proefpersoon afzonderlijk een waarde van 1 tot en met 5 toe te kennen op elk item van het scoreformulier. Deze categorale beoordeling van de fonatiewijze op het scoreformulier is ingedeeld naar de volgende aandachtsgebieden:

1. Houding (1 item)
2. Adem (2 items)
3. Stemgeving (5 items)
4. Resonantie (2 items)
5. Articulatie (1 item)
6. Globale conclusie in een cijfer tussen 1 en 5 (1 item)

Bij deze twaalf items moest de beoordelaar telkens scoren op de vijfpuntschaal. De waarde 1 representeert de meest gewenste situatie, de waarde 5 de meest afwijkende situatie. De gebruikte terminologie van het scoreformulier werd toegelicht teneinde de kans op verschillende interpretaties te verkleinen. De beoordeling met behulp van een scoreformulier maakte het mogelijk gegevens statistisch te bewerken met het programma SPSS (Statistical Package for Social Sciences).

Resultaten eerste fase

Het kwalitatieve onderzoek (blanco) en het kwantitatieve onderzoek leidden tot ongelijksoortige resultaten die niet vergelijkbaar zijn, maar elkaar aanvullen en het trekken van conclusies vergemakkelijken. Eerst zullen de resultaten van de blanco-beoordeling worden besproken.

Hier deden zich problemen voor die, volgens de mening van de onderzoekers, onder meer de onbruikbaarheid van een turflijst als meetinstrument voor overeenstemming aantonen. Niet enkel de methodiek, maar ook de beoordeling zelf legden de volgende – nogal problematische – resultaten aan de dag:

1. Er werd door beoordelaars gebruik gemaakt van verschillende terminologie en ook de omvang van de beschrijvingen liep erg uiteen.
2. Een aantal beoordelaars benadrukte dat houding en adem slecht tot niet te beoordelen waren: het kinetisch aspect kan immers met een video-opname onvoldoende worden beoordeeld. Deze items werden dan wel genoemd, maar niet beoordeeld. Anderen waren daarentegen wel in staat dit te beschrijven.
3. Eenduidigheid in de door logopedisten en 'gouden logopedisten' gebruikte terminologie ontbrak. Dikwijls wordt een afwijking niet gepreciseerd of wordt de ernst niet of onvoldoende aangegeven (bv. verkeerde ademhaling).
4. Een aantal beoordelingen spreekt elkaar duidelijk tegen (bv. ijle, zwakke stemgeving tegen volle, heldere stemgeving en goede resonans tegen geknepen stem).
5. De semi- en niet-deskundigen passen een zeer beschrijvende en vergelijkende beoordeling van de fonatiewijze toe, waaruit de onbekendheid met de logopedische terminologie blijkt (een stem als Kermit, een schurende stem). Zij beoordelen toch vooral het stemgeluid en de articulatie. Evenwel liepen binnen alle groepen de beoordelingen uiteen van enkele aspecten tot alle aspecten van de fonatiewijze.

De hier genoemde problemen dwingen de onderzoeker ertoe in de interpretatie van de individuele terminologie te treden. Echter, indien daartoe zou worden overgegaan, zou door hem in het onderzoek worden geïntervenieerd. Derhalve werd in dit

onderzoek alleen een lijst gemaakt van alle termen die gehanteerd werden binnen alle groepen beoordelaars. Al met al leverde dit toch een beeld op volgens welke aandachtsgebieden de beoordeling van fonatiewijze plaatsvond.

Resultaten tweede fase

De kwantitatieve gegevens, verkregen met behulp van het scoreformulier, werden verwerkt door vergelijking met een maat voor overeenstemming. Als maat voor overeenstemming werd gekozen voor de statistische grootte variantie. Dit geeft de mate weer waarin zekere scores verder van of dichterbij het gemiddelde liggen.

Als eerste stap werd voor elke groep de variantie voor 12 proefpersonen op onderstaande items vastgelegd:

1. Fonatiewijze
2. Ademtype
3. Constrictie
4. Stemkwaliteit

Om vast te stellen of tussen de groepen duidelijke verschillen zijn in de gemiddelde mate van overeenstemming, zijn vervolgens twee vormen van variantieanalyse uitgevoerd. De F-toets voor de vergelijking tussen alle (4) groepen en de T-toets voor een paarsgewijze vergelijking.

Behalve een antwoord op de vraag of er verschil is in de mate van interne overeenstemming, dient gekeken te worden naar de overeenstemming in diagnostiek tussen de groepen, ofwel de externe overeenstemming. Deze analyse dient zich te beperken tot een vergelijking van de gemiddelde waarden die door de groepen beoordelaars werden toegekend aan één willekeurige proefpersoon. Indien alle twaalf proefpersonen in de analyse worden betrokken, worden juist de persoonlijke verschillen in fonatiewijze in het onderzoek betrokken, waardoor een vergelijking van appels en peren zou plaatsvinden. Door de computer is daarom één willekeurige proefpersoon geselecteerd om een vergelijking te kunnen maken van de gemiddelde waarden die door de groepen beoordelaars aan deze proefpersoon werden toegekend. De waarden zijn eveneens onderworpen aan de F-toets en de T-toets.

Bij het onderzoek naar de interne overeenstemming werden geen waarden gevonden waarbij sprake was van een significant verschil tussen de groepen. Uit tabel 1 blijkt, dat voor alle groepen beoordelaars een min of meer gelijke mate van interne overeenstemming gevonden werd (gemiddeld over 12 proefpersonen, zie tabel 1 op de volgende pagina).

De paarsgewijze vergelijking die het verschil in interne overeenstemming tussen telkens twee groepen vaststelt werd uitgevoerd middels de T-toets. Ook dit had geen resultaten tot gevolg die de hypothese konden bevestigen. Elke paarsgewijze vergelijking met betrekking tot elk item leidde tot p-waarden die ver boven het gestelde significantieniveau van 0.05 (of 5%) lagen. Er kan met andere woorden gesteld

Tabel 1 Gemiddelde interne overeenstemming bij 12 proefpersonen, uitgedrukt in de gemiddelde variantie. (in geval van volledige overeenstemming binnen een groep zou de variantie de ideale waarde van 0 hebben; naarmate de waarden in de tabel hoger zijn, is de overeenstemming dus geringer)

	ademtype	constrictie	fonatie	stemkwaliteit
gouden logopedisten	,87	,89	,62	,61
logopedisten	,99	,74	,41	,45
semi-deskundigen	1,10	,90	,59	,77
ondeskundigen	,84	,80	,37	,78
p-waarden (F-toets)	,65	,85	,29	,39

De meningen verschillen. Beoordeling van fonatiewijze.

Tabel 2 Gemiddelde toegekende waarden aan geselecteerde proefpersoon, per groep. (de waarden betreffen de gemiddelde scores op de 5-puntschaal, waarbij 1 de meest gewenste en niet afwijkende situatie representeert terwijl 5 een duidelijk afwijkend beeld is.)

	ademtype	constrictie	fonatie	stemkwaliteit
gouden logopedisten	2,6	3,2	2,8	2,3
logopedisten	3,5	3,2	2,7	1,8
semi-deskundigen	3,2	3,5	2,8	2,3
ondeskundigen	2,7	2,8	2,8	2,3
p-waarden (F-toets)	,32	,67	,95	,65

De meningen verschillen. Beoordeling van fonatiewijze.

worden dat ook tussen twee specifieke groepen beoordelaars geen verschillen zijn vast te stellen met betrekking tot de mate van interne overeenstemming, en dat de deskundigheid er in dat opzicht ook niet toe doet.

Met betrekking tot de externe overeenstemming – bij de beoordeling van 1 proefpersoon – zijn wederom de F-toets en de T-toets uitgevoerd. De resultaten van de eerste toets worden in tabel 2 weergegeven.

Deze analyse wees uit dat groepen een gelijklopend oordeel hadden over de desbetreffende proefpersoon of dat kleine verschillen teniet werden gedaan door een geringe interne overeenstemming op elk item.

Conclusies

Aan de hand van het bovenstaande kunnen wij een voorlopig antwoord formuleren op de vragen omtrent de overeenstemming in diagnostiek van fonatiewijze en de professionaliteit van logopedisten ten aanzien van die diagnostiek.

De resultaten van zowel de blanco beoordelingen als de beoordelingen met het scoreformulier blijken inderdaad complementair te zijn. De mate van interne overeenstemming binnen groepen beoordelaars loopt zeer uiteen, met andere woorden kunnen

we stellen dat de diagnoses niet eenduidig zijn. Bij de blanco beoordelingen zou het verschil in beoordeling door logopedisten verklaard kunnen worden door het ontbreken van eenduidige observatiepunten en terminologie. Het gebruik van de meest uiteenlopende terminologie door de groepen semi- en niet-deskundigen is minder verbazend en is te verklaren door het feit dat deze groepen weinig tot niet bekend zijn met de specifieke vraagstelling.

De verwachting dat de tweede beoordeling met behulp van een standaardformulier het terminologieprobleem zou ondervangen bleek niet juist. Ook bij deze beoordeling was een geringe mate van interne overeenstemming vast te stellen. Anders gezegd was er niet een groep beoordelaars die het intern meer met elkaar eens was dan een andere groep beoordelaars. Dit leidde er vervolgens toe dat niet was vast te stellen of groepen beoordelaars onderling tot dezelfde diagnose bij één en dezelfde proefpersoon konden komen. We zien dus dat een hogere mate van deskundigheid niet betekent dat de kans groter is dat men tot een gelijke diagnose komt als beoordelaars van dezelfde groep. De professionaliteit blijkt dus niet uit de diagnose zelf, noch uit de gebruikte terminologie of de wijze van beoordeling.

Dit onderzoek bevestigt dat de beoordeling van fonatiewijze door logopedisten subjectief is en daardoor niet altijd betrouwbaar. Hoewel kanttekeningen geplaatst kunnen worden, bv. dat het gebruik van een videoband een gedegen beoordeling bemoeilijkt, kan gesteld worden dat de conclusies van het onderzoek de bevindingen en ervaringen van anderen bevestigen. Om meer overeenstemming in het stemonderzoek te bewerkstelligen is het noodzakelijk dat meer gebruik wordt gemaakt van een standaard onderzoeksprotocol met een eenduidige terminologie en ernstaanduiding. Tevens is het dan van belang dat logopedisten getraind worden in het beoordelen van de fonatiewijze, waarbij aard en ernst van een stoornis vergeleken worden met een ontwikkelde standaard en niet met een subjectieve maat. Zulks kan bereikt worden door middel van een gerichte training in de observatie van fonatiewijze; hiervoor is het echter nodig dat een standaard ontwikkeld wordt en dat 'deskundigen' eerst tot overeenstemming komen.

Noten

1. Dit artikel is een bewerking van de afstudeerscriptie 'Deskundigheid en Diagnostiek' aan de opleiding logopedie Zwolle en maakt deel uit van het onderzoek naar stembelasting en stembelastbaarheid van de afdeling KNO-foniatrie van het academisch ziekenhuis Maastricht.
2. Angela Hoogstrate, mede-auteur van 'Deskundigheid en Diagnostiek', dhr. J. Stolk en dhr. K. van Keulen danken wij voor hun medewerking aan het onderzoek, alsmede de proefpersonen en beoordelaars.

Summary

This article presents the results of a research item about the reliability of evaluation of phonation technique. The perceptual evaluation of videotaped subjects is made by 2 groups of speech therapists and 2 groups of lay men. These twenty four judges participated in two evaluation sessions, first on a blanco formular and second on a preprinted interval scale. The first evaluation showed the heterogeneity in terminology and the lack of a severity score as well for speech therapists as for lay men. The second evaluation illustrated no better agreement for speech therapists as for lay men. The evaluation and diagnosis of vocal technique, made by speech therapists are rather subjective. Development of standardisation and evaluation according to protocol is needed for an approved expertise.

Literatuur

- De Bekker, M. en H. Moes, (1988). *Het stemonderzoek: zo helder als wat*. Eindscriptie Hogeschool Nijmegen, studierichting logopedie.
- De Bodt, M., P. van de Heyning, F.L. Wuyts, L. Lambrechts, (1996). The perceptual evaluation of voice disorders. *Acta Oto-Rhino-Laryngologica Belgica*, 50, 283- 291.
- Dejonckere, P.H., M. Remacle, E. Fresnel, V. Woisard, L. Crevier, (1997). Perceptieve stembeoordeling. *Logopedie en Foniatrie*, 67, 68-75.
- Hoogstrate, A. en M.A.C. Vredenberg, (1997). *Deskundigheid en Diagnostiek*. Eindscriptie Christelijke Hogeschool Windesheim Zwolle, studierichting logopedie.
- Kooijman, P.G.C., (1995). Logopedisch perceptueel stemonderzoek. *Logopedie en Foniatrie*, 67, 62-69.
- Kreiman, J, B.R. Gerratt, K. Precoda, G.S. Berke, (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research*, 35, 512-520.
- Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: University Press.
- De Moor, G.M., (1992). Perceptieve stembeoordeling beoordeeld. *Logopedie en Foniatrie*, 64, 267-273.
- Smeets, J.A.M.W. , J.B. Uytterlinde, D.M.P.H. Bastin, H. Reinold, (1988). Verkijken op stem. *Logopedie en Foniatrie*, 60, 249-251.
- Stes R. (1994). *Stemstoornissen. Stemmisbruik en verkeerd stemgebruik*. Leuven/Amersfoort: Acco.