Samenvatting: Automatische beoordeling van stem en spraakverstaanbaarheid na behandeling voor hoofd-halskanker

Achtergronden Kanker in het hoofd-halsgebied en de behandeling daarvan kunnen een negatief effect hebben op iemands stem en spraak. Voor de logopedist vormt het beoordelen van de verstaanbaarheid en kwaliteit van de spraak een belangrijk onderdeel van de behandeling van de patiënt omdat het een indicatie geeft van de ernst van de pathologie en de vooruitgang die de patiënt reeds maakte. Het objectief beoordelen van iemands spraak of stem is echter een subjectief gegeven dat wordt beïnvloed door tal van factoren, zoals familiariteit van de logopedist met de patiënt en de test items. Luisteraars beoordelen vaak ook op een inconsistente manier. Een computer is van nature objectief en wel consistent in het uitvoeren van deze taak.

In dit onderzoek werden automatische voorspellingsmodellen ontwikkeld voor de beoordeling van spraakverstaanbaarheid en stemkwaliteit van twee groepen van sprekers die behandeld werden voor hoofd-halskanker. De eerste groep, besproken in deel I van dit proefschrift, betreft patiënten met voortgeschreden kwaadaardige tumoren in het hoofd-halsgebied. Deze patiënten kregen een niet-chirurgische kankerbehandeling, namelijk concomitante chemoradiotherapie (CCRT). Bij deze behandeling worden de chemotherapie en bestraling gelijktijdig toegediend. Een dergelijke behandeling kan iemands stem en spraak nadelig beïnvloeden. De tweede groep wordt gevormd door patiënten die werden behandeld voor voortgeschreden kwaadaardige tumoren van het strottenhoofd. Hiervoor ondergingen deze patiënten een totale laryngectomie (TL), een operatie waarbij het strottenhoofd en dus ook de stembanden worden verwijderd. Na een TL is spreken weer mogelijk met behulp van een stemprothese. Dit hulpmiddel bevat een éénrichtingsklep, waardoor weer stemvorming mogelijk is omdat de uitademingslucht langs trillende structuren in de slokdarm naar de mondholte kan stromen. Deze zogenaamde tracheoesofageale spraak klinkt duidelijk anders dan de spraak vóór de operatie.

Overzicht van het proefschrift Het Nederlands Kanker Instituut heeft twee grote collecties spraakopnames van deze twee groepen sprekers, waarin ze een kort verhaal voorlezen. Deze collecties worden beschreven in hoofdstuk 2 (CCRT) en hoofdstuk 5 (TL). Al deze spraakopnames zijn door logopedisch geschoolde luisteraars beoordeeld op stemkwaliteit en spraakverstaanbaarheid. In hoofdstukken 3, 4 en 6 presenteren we een aantal manieren om automatisch te voorspellen hoe een luisteraar de opnames zal beoordelen. Daarvoor gebruiken we verschillende vormen van spraaktechnologie die zijn gebaseerd op het feit dat de computer herkent wat de spreker gezegd heeft in termen van individuele klanken, op de kenmerken van het spraakgeluid, en/of op basis van akoestische informatie.

Met behulp van computers kunnen we voorspellingsmodellen ontwikkelen op basis van informatie uit de spraakopnames zelf en/of door het vergelijken van wat er gezegd zou moeten zijn met wat de computer 'gehoord' heeft. In een van de studies onderzochten we ook hoe goed de voorspellingsmodellen waren wanneer er een mismatch was tussen de taal waarmee ze waren getraind en de taal van de spreker. In ons geval was de computer getraind om Vlaamse spraakopnames te analyseren, maar voerden wij Nederlandse opnames in. We vonden dat modellen, die wat er daadwerkelijk gezegd was vergeleken met wat had moeten worden gezegd, in deze situatie minder goed presteerden dan modellen, die het door de computer "gehoorde" niet hoefden te vergelijken met de doeltekst maar gewoon de variatie van het spraakgeluid analyseerden.

In onze zoektocht naar de optimale modellering van de voorspellingen, vonden we dat wanneer het model meer en verschillende soorten sprekerkenmerken ter beschikking heeft, het verschil tussen de beoordelingen door de computer en de luisteraar kleiner wordt. Sommige van de beste modellen presteren op een niveau dat vergelijkbaar is met dat van een gemiddelde luisteraar.

In de hoofdstukken 4 en 7 konden we laten zien dat vergelijkbare resultaten kunnen worden verkregen voor het voorspellen van beoordelingen van *Articulatiekwaliteit* en *Accent* (zie hoofdstuk 4) en dat de modellen ook gebruikt kunnen worden voor TL-sprekers (zie hoofdstuk 7).

Een belangrijk aspect van computermodellen is uiteraard de betrouwbaarheid ervan. In hoofdstuk 7 vonden we dat de prestaties van het voorspellingsmodel betrouwbaar worden wanneer de computer voldoende voorbeelden van elke klank ziet. We vonden dat aan dit criterium voldaan is wanneer het spraakmateriaal tenminste 100 lettergrepen lang is.

Veel akoestische metingen correleren, vaak sterk, met Acoustic Signal Typing (*AST*), een classificatie van de spraak van *TL*-patiënten. In hoofdstuk 5 laten we zien dat slechts twee soorten akoestische informatie nodig zijn om een stem te kunnen indelen in één van de vier AST groepen. De eerste is de **stemhebbendheid**, gerepresenteerd door de *Voiced Fraction* (*VF*) en/of de

maximale fonatieduur (*Maximum Voicing Duration*; *MVD*). De tweede is de **ruis** in de stem, gerepresenteerd door de *Harmonics to Noise Ratio* (*HNR*). Dit betekent dat de stemhebbendheid, fonatieduur en hoeveelheid ruis de belangrijkste aspecten zijn van de AST.

De relatie tussen de AST classificatie en beoordelingen van spraakkwaliteit door luisteraars is onderzocht in hoofdstuk 6. De resultaten ondersteunen het gebruik van AST als onderdeel van een ruimere evaluatie van de stem. Hoewel er een statistisch significant verband tussen de twee maten, levert de AST in zijn huidige vorm echter slechts in beperkte mate prognostische informatie over spraakkwaliteit.

Conclusies De resultaten beschreven in dit proefschrift tonen aan dat het goed mogelijk is om computer-gebaseerd automatische beoordelingen van spraakverstaanbaarheid en stemkwaliteit te gebruiken bij sprekers die behandeld zijn voor hoofd-halskanker en dat computer-gebaseerde automatische beoordelingen een clinicus bij zijn/haar evaluaties kunnen helpen. Let wel dat deze modellen niet ontwikkeld zijn met als doel de clinicus te vervangen. De clinicus blijft de belangrijkste beoordelaar van hoe een spreker praat en klinkt, maar deze computermodellen kunnen gezien worden als een handige, objectieve en accurate toevoeging aan het oordeel van de expert.