

SAMENVATTING

Spraaktechnologie is de afgelopen decennia snel ingeburgerd: Spraakassistenten, spraakbiometrie, geautomatiseerde callcenters, autotune en Vocaloïde idolen maken nu deel uit van ons dagelijkse leven. De snelle groei van deze technologieën is mogelijk gemaakt door doorbraken op het gebied van machinaal leren met neurale netwerken, het zogenaamde "Deep Learning". Deze doorbraken hebben patroonherkenning op grote schaal mogelijk gemaakt met behulp van ongekende hoeveelheden data. Deze spraaktechnologieën zijn nuttig voor iedereen, maar ze werken minder goed, of zelf helemaal niet, voor bepaalde groepen in de samenleving, zoals atypische en pathologische sprekers. Het is met name wrang dat mensen met een spraakbeperking deze technologieën minder of helemaal niet kunnen gebruiken omdat zij vaak ook fysieke beperkingen hebben waardoor juist zij baat zouden hebben bij, b.v. spraakassistenten. Naast spraakassistenten zijn er verschillende andere soorten spraaktechnologieën waarbij de belangrijkste potentiële gebruikers een spraakbeperking hebben. Dit proefschrift bevat een reeks studies die tot doel hebben om spraaktechnologie toegankelijk te maken voor sprekers met een spraakbeperking. De studies hebben betrekking op drie toepassingen van spraaktechnologie: automatische spraakherkenning voor atypische en pathologische spraak, automatische evaluatie van de ernst van de spraakbeperking voor patiënten die behandeld zijn voor mondholtekanker en pathologische spraakconversie.

Hoofdstuk 2: Op weg naar inclusieve automatische spraakherkenning

Hoofdstuk 2 onderzoekt hoe toegankelijk de automatische spraakherkenning (ASR) is voor atypische sprekers van het Nederlands en het Mandarijn Chinees. Om dit te onderzoeken, introduceren we de term bias-metrick, die we berekenen als de laagste (= beste) foutscore (WER) van elke atypische groep sprekers. De studie vindt een aanzienlijke verslechtering bij de herkenning van geaccentueerde spraak van mannelijk sprekers (Nederlands), spraak van kinderen, de spraak van oudere volwassenen en niet-moedertaal sprekers (Nederlands en Mandarijn). Wij concluderen dat het trainingsmateriaal, de uitspraak, de architectuur van de spraakherkenner en de taal allen een bron van bias kunnen zijn bij de herkenning.

Hoofdstuk 3: Automatische spraakherkenning en foutanalyses met weinig spraakdata van spraak na behandeling voor mondholtekanker

Patiënten met mondholtekanker krijgen vaak chemo-radiotherapie tijdens hun behandeling. Als een bijwerking van die behandeling kunnen patiënten tijdelijk aan huis gebonden zijn met een beperkte mobiliteit. Deze patiënten zouden gedurende deze moeilijke periode geholpen kunnen zijn met spraakassistenten. Echter, automatische spraakherkenning werkt momenteel niet goed voor sprekers die behandeld zijn voor mondholtekanker. **Hoofdstuk 3** ontwikkelt en vergelijkt verschillende systemen voor de

herkenning van de spraak van patiënten die behandeld zijn voor mondholtekanker. Als onderdeel van onze studie presenteren wij een spreker-adaptief gebaseerd systeem voor de herkenning van deze spraak dat een absolute verbetering bereikt van 7,7% boven het referentiesysteem getraind op spraak van gezonde sprekers. Uit de studie blijkt ook dat de herkenning van plosieven en sommige klinkers ("aa" en "oe"), waarvan bekend is dat ze beïnvloed worden door behandeling voor mondholtekanker, problemen opleverden voor de ontwikkelde spraakherkenners.

Hoofdstuk 4: Automatische evaluatie van spontane spraak na mondholtekanker met behulp van beoordelingen van naïeve luisteraars

Na een behandeling voor mondholtekanker krijgen veel patiënten problemen met spreken. Voor deze spraakstoornissen krijgen patiënten vaak spraaktherapie aangeboden. Om de effectiviteit van de logopedie te bepalen is een manier nodig om het verloop van de spraakstoornis te volgen tijdens de therapie. De ernst van de spraakstoornis wordt nu nog beoordeeld door de logopedist, maar een dergelijke subjectieve beoordeling heeft nadelen. In **Hoofdstuk 4** worden de spraakherkenningssystemen van **Hoofdstuk 3** vergeleken met methoden gebaseerd op akoestisch features en systeem-vergelijkingen om te onderzoeken of het mogelijk is om de beoordelingen van de logopedisten automatisch te voorspellen op grond van opnamen van spontane spraak van patiënten na behandeling voor mondholtekanker, en welke methode het beste resultaat geeft. De beste technieken in de studie die gebruik maakten van verklaarbare regressiemodellen correleren sterk met de beoordelingen van de expert. Tegelijkertijd verzamelden we beoordelingen van de ernst van de spraakstoornis door naïeve luisteraars en laten we zien dat hun beoordelingen sterk correleren met die van de expert.

Hoofdstuk 5: Pathologische spraakaanpassing met spraakconversie op basis van een autoencoder

Patiënten die wachten op een operatie voor mondholtekanker hebben vaak zorgen en angst voor de toekomst, die worden verergerd door onzekerheid over hoe hun leven er na de operatie uit zal zien. Het is belangrijk om de zorgen van patiënten te verminderen, omdat studies aantonen dat onzekerheid en verkeerde verwachtingen een negatieve invloed hebben op de kwaliteit van leven, zelfs na de operatie. Een bron van onzekerheid voor patiënten zijn de spraakstoornissen die kunnen optreden na hun operatie en de ernst van de stoornissen. Het is belangrijk om patiënten meer informatie te kunnen geven over hun toekomstige spraak en de verwachte problemen daarmee. Een mogelijke manier om hen beter te informeren zou zijn om hen computer gegenereerde spraak van hun mogelijke toekomstige spraakstoornissen te laten horen. Hiervoor zou het nuttig zijn om een hulpmiddel te ontwikkelen dat kan laten horen hoe patiënten kunnen gaan klinken na behandeling voor mondholtekanker. Een mogelijke manier om dit te bereiken zou gebaseerd kunnen worden op spraakconversie, het belangrijkste onderwerp van **Hoofdstukken 5 en 6**. **Hoofdstuk 5** stelt een systeem voor stemconversie voor die pathologische spraak omzet naar de spraak die klinkt als gesproken door een andere pathologische spreker. Het voorgestelde conversie framework produceert voorbeelden van pathologische spraak die klinken als de doelspreker, maar dan met meer of minder

ernstige spraakstoornissen. Dit systeem is echter niet in staat om willekeurige zinnen te produceren wat de toepassingen beperkt.

Hoofdstuk 6: Naar behoud van spreker identiteit bij een stemconversie van normale naar dysartrische spraak

Om deze beperking aan te pakken, stelt **Hoofdstuk 6** een tweetraps stemconversie-aanpak voor. In de eerste stap wordt een sequentie-naar-sequentie (seq2seq) model gebruikt om gezonde spraak om te zetten in pathologische spraak. De eerste stap maakt synthese van willekeurige uitingen mogelijk, maar de identiteit van de gezonde spreker blijft niet behouden tijdens de conversie. Voor dit doel gebruiken we een tweede stap die is gericht op het weer converteren naar de oorspronkelijke identiteit op basis van het model dat wordt voorgesteld in **Hoofdstuk 5**. De voorgestelde tweetraps benadering controleert de ernst van de spraakafwijking volgens drie objectieve meetmethoden. Er wordt een redelijke mate van natuurlijkheid bereikt, maar de huidige techniek levert nog onvoldoende gelijkenis met de oorspronkelijke spraak.

Discussie en slotopmerkingen

Het onderzoek beschreven in de **Hoofdstukken 2** tot en met **6** wijzen er ook op dat standaard meetmethoden in spraaktechnologie, zoals de woord foutscore om automatische spraakherkenners te evalueren en de gemiddelde opiniescore om de natuurlijkheid van gegenereerde spraak te evalueren, een aantal tekortkomingen hebben. De woord foutscore geeft niet expliciet de prestatie weer van automatische spraakherkenners bij ondervertegenwoordigde groepen in de resultaten. Wij raden aan om meer diverse sprekersgroepen op te nemen in automatische spraakherkenningstests en het kwantificeren van de bias te gebruiken als een nieuwe evaluatiestandaard. De gemiddelde opiniescore is niet alleen gevoelig voor natuurlijkheid, maar ook voor andere factoren zoals de ernst van de spraakafwijkingen, leeftijd, geslacht en de mate waarin beoordelaars openstaan voor toepassingen van spraaktechnologie. Wij stellen voor om objectieve natuurlijkheidsmaten te ontwikkelen om de tekortkomingen van de gemiddelde opiniescore van luisteraars te verlichten. Over het algemeen dringen we aan op eerder testen met een diversere groep gebruikers van spraaktechnologie om de toegankelijkheid van spraaktechnologie voor alle gebruikers te vergemakkelijken, inclusief pathologische sprekers.