

Nederlandse samenvatting

Tekstuele ondersteuning ter verbetering van het verstaan van spraak in moeilijke luistersituaties

Het belang van communicatie voor de deelname aan de westerse maatschappij is erg groot (Ruben, 2000). Problemen met het verstaan van spraak hebben een grote invloed op activiteiten die afhankelijk zijn van communicatie, zoals telefoneren, vergaderen, werk en informele afspraken.

Omgevingsfactoren die het verstaan van spraak beïnvloeden zijn bijvoorbeeld achtergrondruis, visuele informatie (bijvoorbeeld het kunnen zien van het gezicht van de spreker), en linguïstische context. Interne factoren die het verstaan van gesproken taal beïnvloeden omvatten onder andere gehoorverlies en cognitieve vaardigheden die van belang zijn voor de verwerking van gesproken informatie (Humes, Burk, Coughlin, Busey, & Strauser, 2007). Leeftijd beïnvloedt zowel het gehoor als cognitieve vaardigheden (Wingfield, Poon, Lombardi, & Lowe; 1985).

Problemen met het verstaan van spraak zouden deels verholpen kunnen worden door het presenteren van ondersteunende visuele informatie, in het bijzonder in situaties waarin de luisteraar de spreker niet kan zien, zoals tijdens telefoongesprekken. Het doel van het onderzoek dat in dit proefschrift wordt beschreven was tweeledig: ten eerste richtte het onderzoek zich op de ontwikkeling van een test waarmee de bijdrage van modaliteits-specifieke factoren (zoals cognitief vermogen) aan het verstaan van spraak kan worden gemeten. Ten tweede richtte het onderzoek zich op het meten van de bijdrage van ondersteunende tekstuele informatie aan het verstaan van spraak in achtergrondruis. Hierbij ging het specifiek om de potentiële ondersteuning van tekst afkomstig uit een automatische spraakherkenner (Automatic Speech Recognition [ASR] systeem). Het idee dat ten grondslag lag aan dit onderzoek is dat slechthorenden profijt kunnen hebben van een ASR systeem dat de spraak direct herkent en weergeeft, zodat de slechthorende de resulterende tekst kan gebruiken ter ondersteuning van het verstaan van de spraak.

Hoofdstuk 2 beschrijft de ontwikkeling van de Text Reception Threshold (TRT) test, een visuele analogoog van de Speech Reception Threshold (SRT) test (Plomp & Mimpen, 1979). Met behulp van de TRT test kan de bijdrage van modaliteits-specifieke processen aan het verstaan van spraak worden onderzocht. Uit het onderzoek bleek dat een aanzienlijk deel (ongeveer 30%) van de vaardigheid om spraak in achtergrondruis te kunnen verstaan samenhangt met

cognitieve vaardigheden die van belang zijn voor het aanvullen van incomplete linguïstische informatie.

Hoofdstuk 3 beschrijft de bijdrage van het presenteren van deels gemaskeerde tekst (TRT test) aan het verstaan van spraak in achtergrondruis (SRT test). Het aanbieden van relatief weinig visuele informatie (een grotendeels gemaskeerde, tekstuele weergave van de gesproken zinnen) droeg aanzienlijk bij aan het verstaan van de spraak in moeilijke luisteromstandigheden (ruis). De additionele tekst gaf extra linguïstische context waarmee de luisteraar een gedeelte van de woorden die niet konden worden verstaan, alsnog kon aanvullen. De audiovisuele prestatie was beter dan dat voorspeld door een model (Blamey, Cowan, Alcantara, Whitford, & Clark, 1989) dat veronderstelt dat de audiovisuele informatie verkeerd wordt waargenomen wanneer zowel de auditieve als de visuele informatie verkeerd worden waargenomen. Dit verschil tussen de werkelijke audiovisuele prestatie en de door het model van Blamey voorspelde audiovisuele prestatie werd kleiner wanneer bij de voorspelling van de audiovisuele prestatie rekening werd gehouden met het aantal onafhankelijke delen per zin.

De studies die worden beschreven in Hoofdstuk 4 en 5 van dit proefschrift onderzochten de ondersteuning van ondertiteling afkomstig van een ASR systeem aan het verstaan van spraak in achtergrondruis. Aangezien automatische spraakherkenning nog niet foutloos werkt en bovendien tijd kost, werd de invloed van het aantal fouten in de tekst en de vertraging van de tekst ten opzichte van de spraak onderzocht. Het profijt van de tekst was groter wanneer er minder fouten in de tekst stonden en wanneer de tekst minder vertraagd was ten opzichte van de spraak. Het profijt van automatische woordherkenning was groter dan het profijt van de automatische fonemherkenning. Daarnaast is in het algemeen de prestatie van ASR systemen beter voor woordherkenning dan voor fonemherkenning, dus op basis van deze resultaten lijkt woordherkenning beter geschikt voor het ondersteunen van het verstaan van spraak dan fonemherkenning. Wanneer ongeveer 75% van de gesproken woorden goed was herkend door de spraakherkenner, was het profijt in SRT ongeveer 3 dB in signaal-ruis verhouding. De mate waarin het verstaan van de spraak verbeterde door het presenteren van de tekst was gerelateerd aan de mate waarin men op basis van *alleen* de (deels foutieve) tekst kon herleiden wat er was gezegd. Het profijt van de tekst bleek niet afhankelijk te zijn van leeftijd en personen met gehoorverlies hadden ongeveer evenveel profijt van de tekst als personen zonder gehoorverlies. Het kostte normaalhorende ouderen echter relatief veel moeite om de tekst met de spraak te combineren. Personen met een grotere werkgeheugencapaciteit hadden minder moeite met

het combineren van de tekst en de spraak.

In tegenstelling tot het objectief gemeten profijt van de tekstuele ASR output, gaven de resultaten van de studies in Hoofdstuk 5 en 6 aan dat er *subjectief* geen profijt is van de tekst tijdens het verstaan van spraak. Hoofdstuk 6 beschrijft de evaluatie van een mogelijk toekomstig hulpmiddel dat automatisch ondertiteling tijdens telefoongesprekken genereert met behulp van ASR. Slechthorenden gaven aan dat het moeilijk is om deels incorrecte tekst te gebruiken bij het verstaan van telefoon-spraak. Slechthorenden die beter presteerden op de TRT test hadden minder moeite om de vertraagde en deels incorrecte tekst te gebruiken tijdens het verstaan van de spraak. De resultaten geven aan dat het verwerken van de tekst extra inspanning kost, waardoor men minder hulp ervaart van de tekst. Dit onderstreept het belang van het door slechthorenden laten evalueren van toekomstige communicatie hulpmiddelen voor deze doelgroep.

In moeilijke luistersituaties hangt het verstaan van spraak af van cognitieve vaardigheden waarmee de onverstaanbare gedeeltes in de spraak kunnen worden aangevuld. De capaciteit van cognitieve vaardigheden zoals werkgeheugen is echter beperkt en het gebruik maken van deze top-down processen kost moeite. Wanneer het subjectieve profijt van de tekstuele informatie niet opweegt tegen de moeite die het kost om de visuele informatie te verwerken, dan zullen luisteraars wellicht niet overtuigd zijn van het voordeel van de extra visuele informatie. Het verlagen van de complexiteit van de audiovisuele taak door het verminderen van het aantal fouten in de tekst of het verkleinen van de vertraging van de tekst ten opzichte van de spraak maakt de taak makkelijker. Het beroep dat wordt gedaan op de cognitieve vaardigheden wordt hierdoor kleiner en hierdoor wordt het subjectief waargenomen profijt van de visuele informatie groter. De studie die wordt beschreven in Hoofdstuk 6 gaf echter aan dat een bijna foutloze ondertiteling die niet vertraagd is ten opzichte van de spraak, nog geen garantie is voor de succesvolle toepassing van ASR in een communicatie hulpmiddel.

Toekomstige studies naar de toepassing van ASR technologie als hulpmiddel om het verstaan van spraak te verbeteren moeten zich richten op methoden die de audiovisuele taak minder complex maken. Naast het verminderen van het aantal fouten in de tekst en het verkleinen van de vertraging van de tekst, kan er hierbij gedacht worden aan het selectief herkennen van de relevante woorden in de spraak (Rose, 1995). Hierbij worden dus niet alle woorden door de spraakherkenner herkend, maar alleen de relevante woorden. Dit leidt tot minder fouten in de tekst. Het kan echter moeilijk zijn om van te voren te bepalen

welke woorden relevant zullen zijn in een gesprek. Daarnaast is het interessant om na te gaan of de weergave van de ondertiteling geoptimaliseerd kan worden om zo het profijt van de tekst te vergroten. Hierbij kan gedacht worden aan het aangeven van de pauzes in de spraak of het markeren van woorden die duidelijk werden benadrukt door de spreker. Verder zouden er strategieën ontwikkeld kunnen worden die de gebruikers van het hulpmiddel (zowel de slechthorende luisteraar als de spreker) beter voorbereiden op het toepassen van het systeem in dagelijkse gesprekken. Dergelijke strategieën zouden de prestatie van de spraakherkenner moeten verbeteren en de slechthorende meer tijd moeten geven om de audiovisuele informatie te verwerken. Zoals blijkt uit het onderzoek dat in dit proefschrift wordt beschreven, zijn deze factoren cruciaal voor het objectieve en subjectieve profijt van de door ASR systemen gegenereerde tekst ter verbetering van het verstaan van spraak in moeilijke luistersituaties.