

## ONDERZOEKSVOORSTEL

# Hoe kan het herkennen van verbale antwoorden door spraakherkenning software, de VR ervaring van een slecht nieuws gesprek realistischer doen aanvoelen.

Bachelorproef, 2022-2023

Jona De Neve

E-mail: [jona.deneve@student.hogent.be](mailto:jona.deneve@student.hogent.be)

Co-promotor: L. De Mol (HOGENT, [lana.demol@hogent.be](mailto:lana.demol@hogent.be))

## Samenvatting

Het zorglab van de Hogeschool Gent biedt studenten, aan de hand van interactieve video's met Virtual Reality, kansen om in realistische situaties een slecht nieuws gesprek in te oefenen. Per scenario bestaat er een film-pje dat in fragmenten is opgedeeld maar om naar het volgende fragment te gaan moet de gebruiker zelf zijn antwoord nog manueel aanduiden. Dit neemt weg van de immersie. Daarom wordt er in deze bachelorproef informatica gezocht naar een oplossing om dit te verhelpen. Hiervoor wordt er met een literatuur studie en een vergelijking, gezocht naar een goede Spraak naar tekst software en een Artificiële Intelligentie die met de gegenereerde tekst het juiste fragment selecteert. Uit de resultaten wordt verwacht dat er een software wordt gevonden die de immersie van de gebruiker kan behouden. Hiervoor wordt gekeken naar een AI gelijkaardig aan degene van 'Dimensions in Testimony' waar bezoekers van het museum vragen konden stellen aan opnames van Holocaust overlevenden.

**Keuzerichting:** AI & Data Engineering

**Sleutelwoorden:** AI, VR, Onderwijs

## Inhoudsopgave

1	Introductie . . . . .	1
2	State-of-the-art . . . . .	2
3	Methodologie . . . . .	2
4	Verwacht resultaat, conclusie . . . . .	3
	Referenties . . . . .	3

## 1. Introductie

Je hebt waarschijnlijk de reclame over de Meta-verse 'The Impact Will Be Real' van Meta (2022) gezien. Daarin tonen ze hun visie over de rol die Virtual Reality (VR) speelt in de toekomst. Zo laten ze verschillende toepassingen ervan in het onderwijs zien. Jammer genoeg staat onze technologie nog niet zo ver als wat er in de reclame gezien kan worden maar ook nu al vind VR zich een baan in verschillende opleidingen.

De Hogeschool van Gent maakt ook gebruik van VR om studenten de kans te geven in meer realistische situaties te oefenen. Hiervoor zijn twee verschillende technieken gebruikt. Allereerst heb je het renderen van een omgeving. Dit laat de gebruiker een interactieve wereld van 3D modellen ontdekken. Zo bestaan er drie virtuele kamers waarin de student kan oefenen. De andere manier is aan de hand van een 360° opname die wordt gemaakt aan de hand van een 360° camera. Omdat dit een opname van de werkelijkheid neemt, ziet deze methode er realistischer uit.

Dit is waar er op het probleem wordt gestoten.

Aangezien de tweede methode werkt met een opname moet er naar verschillende fragmenten gesprongen worden naar gelang het antwoord dat de gebruiker ingeeft. Dit wordt handmatig gedaan via een meerkeuzevraag wat de immersie van de VR weg haalt. Daarom wordt in deze bachelorproef toegepaste informatica onderzocht hoe het overschakelen anders kan aangepakt worden zodat het voor de gebruiker realistischer aan voelt. Hiervoor kijken we richting Artificiële Intelligentie (AI).

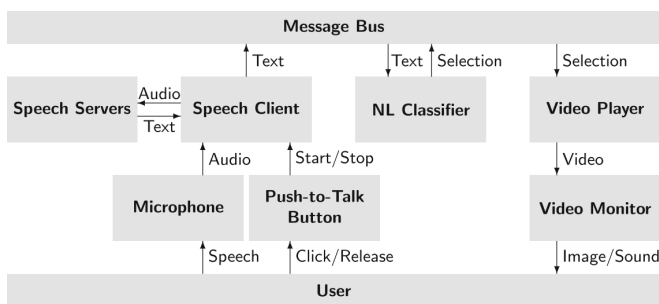
Alle grote IT bedrijven hebben wel een departement die zich bezig houdt met AI. Ze zien allemaal de mogelijkheden die het biedt. Van jobs gemakkelijker te maken tot het zelf creëren van kunst, de toepassingen zijn oneindig. Daarom zal in functie van het zorglab op zoek worden gegaan naar een Spraak naar tekst (STT) en een natuurlijke taalverwerking software (NLP).

De STT zal worden gebruikt om het manueel ingeven van het antwoord te vervangen. In plaats daarvan zal de gebruiker gewoon zijn antwoord luidop kunnen geven en zal de STT dit in tekst omzetten. Dit alleen zal natuurlijk geen volgend fragment voor de gebruiker kunnen kiezen en daarom hebben we de NLP nodig. Deze zal de gegenereerde tekst omzetten naar kernwoorden en aan de hand daarvan bepalen welk fragment als volgende geschikt is.

## 2. State-of-the-art

Om de immersie in de simulatie te vergroten moet ervoor gezorgd worden dat het overschakelen naar een volgend fragment kan gebeuren zonder de gebruiker uit de illusie te halen. Hiervoor zouden verbale antwoorden kunnen gegeven worden.

Een mooi voorbeeld van zo een interactie is het project 'Dimensions in Testimony' van de USC Shoah Foundation (2020). Daar kunnen bezoekers vragen stellen aan een overlevende van de Holocaust die op voorhand een volledig interview heeft afgelegd. Dit werd mogelijk gemaakt dankzij het systeem dat erachter zit (Figuur 1). Dit systeem is opgebouwd uit volgende componenten: een software voor spraakherkenning (ASR) die de gebruikers verbale vraag in tekst omzet, een Natural Language Classifier (NLC) die op basis van de gegenereerde tekst een antwoord via een audio/video fragment voorziet en een mediaspeler die de fragmenten kan afspelen met tussendoor een inactieve animatie (Traum e.a., 2015).

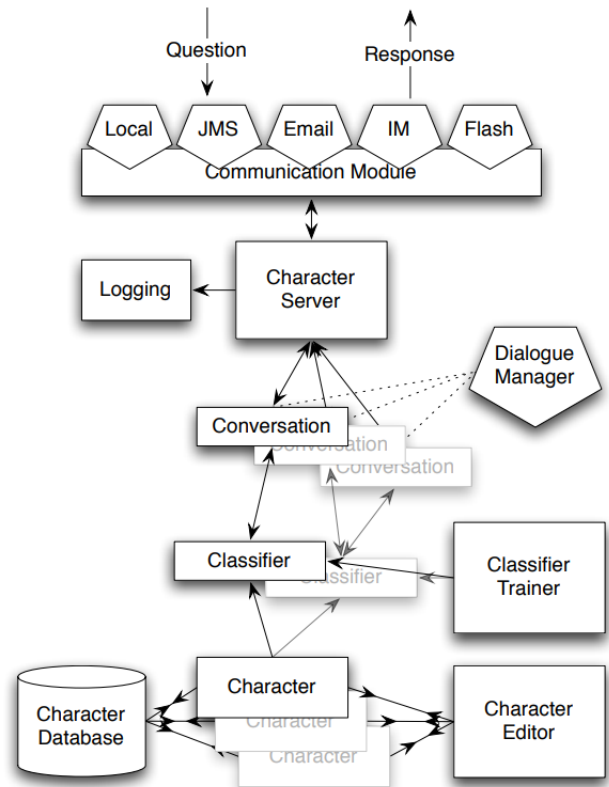


Figuur 1: Dimensions in Testimony - Systeem Architectuur (Traum e.a., 2015)

Spraakherkenning is de laatste jaren veel vooruit gegaan, dit komt omdat er meer data beschikbaar is, de algoritmes beter (deep neural networks) en de computers sneller zijn. Maar perfect zullen ze nooit zijn zegt van Hessen (2020). Dit is geen verrassing aangezien mensen zelf vaak nog problemen hebben bij het verstaan van een andere. De meest voorkomende fouten van een ASR zijn herkenningfouten. Deze kunnen ontstaan door slechte kwaliteit van de opname maar ook van de manier waarop woorden worden uitgesproken. De andere fouten ontstaan omdat ASR zijn vocabulaire niet uitgebreid genoeg is. Dit noemend we Out-of-vocabulary-fouten (OOV) (van Hessen, 2020).

Naast de ASR hebben we ook de natuurlijke taalverwerking (NLP). In 'Dimension in Testimony' maakten ze hiervoor gebruik van NCPeditor (Figuur 2). De classifier in dit systeem berekent welke antwoorden op de ingegeven tekst kunnen gegeven worden door de taalmodellen van beide te vergelijken en de antwoorden te rangschikken. Zolang de ingegeven tekst een foutmarge lager

dan 50% blijft, zal het antwoord ongewijzigd blijven (Leuski & Traum, 2010).



Figuur 2: NCPeditor - Systeem Architectuur (Leuski & Traum, 2010)

## 3. Methodologie

Om te beginnen zal er worden samengezeten met het team van het zorglab. Zo kan worden neergeschreven wat de problemen zijn waar ze tegenaan lopen. Ook wordt besproken wat er verwacht wordt van een oplossing, hoe moet deze in zijn werk gaan? Op basis daarvan zal de literatuurstudie uitgebreid worden met onderzoek naar oplossingen die dit probleem kunnen verhelpen. Hieronder valt onder andere een onderzoek naar ASR. Er zal gekeken worden welke ASR's er al verkrijgbaar zijn en welke er in het Nederlands werken. Alleen een ASR hebben is niet genoeg om de immersie te behouden dus zal er ook worden gekeken naar hoe, aan de hand van de gegenereerde tekst, het volgende fragment kan getoond worden.

Wanneer er een uitgebreide literatuurstudie is uitgevoerd, zal de software die gevonden is op de proef gesteld worden. Hier wordt gekeken hoe correct de ASR's inkomende audio kunnen transcriberen. Dit wordt getest bij mensen zonder spraakaandoeningen als bij degene met een spraakaandoening zoals bijvoorbeeld stotteren. Daarnaast zal ook getest worden of er aan

de hand van kernwoorden een juist fragment kan gekozen worden. Door de AI verschillende prompts te geven en te kijken of hij naar de juiste tijdsaanduiding in de video gaat.

Als laatste is het de bedoeling dat beide software samen in de context van het zorglab getest wordt.

#### 4. Verwacht resultaat, conclusie

Uit deze bachelorproef wordt verwacht dat er een oplossing op het immersie probleem gevonden wordt. Hiervoor zal de Spraakherkenning software gekozen worden die het beste resultaten geeft samen met een natuurlijke taalverwerking software. Het liefst zou deze al geïmplementeerd en wekende zijn in het zorglab. Als het niet zover komt dan zou er ten minste bestaan uit een goed uitgeschreven werking zodat de het kan geïmplementeerd worden in het zorglab.

#### Referenties

- Leuski, A., & Traum, D. (2010). NPCEditor: A Tool for Building Question-Answering Characters. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Verkregen 15 december 2022, van [http://www.lrec-conf.org/proceedings/lrec2010/pdf/660\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/660_Paper.pdf)
- Meta. (2022, juni 15). *The Impact Will Be Real*. Verkregen 11 december 2022, van <https://www.youtube.com/watch?v=80IIEnSNwQc>
- Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S., & Swartout, W. (2015). New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling. In H. Schoenau-Fog, L. E. Bruni, S. Louchart & S. Baceviciute (Red.), *Interactive Storytelling* (pp. 269–281). Springer International Publishing.
- USC Shoah Foundation. (2020, april 3). *Dimensions in Testimony | USC Shoah Foundation*. Verkregen 5 december 2022, van <https://www.youtube.com/watch?v=nGzAc9mIoTM>
- van Hessen, A. (2020, augustus 1). *Automatische spraakherkenning - Hoe kun je het inzetten voor onderwijsmateriaal?* (Onderzoeksrap.). SURF.