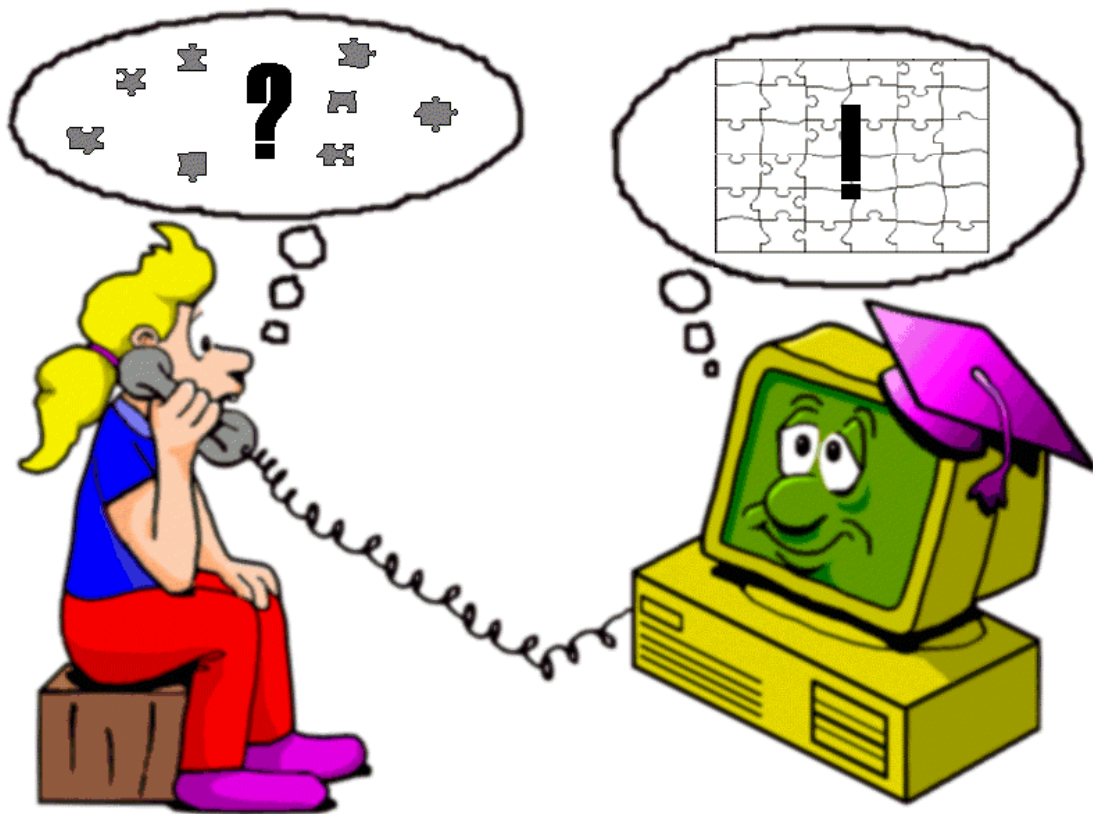


De Digitale Telefoniste

*Gebruikersacceptatie & Evaluatie van
het VIOS-spraakherkenningssysteem*



Voorwoord

Dit verslag is geschreven in het kader van de onderzoekstaak van mijn afstuderen aan de Technische Universiteit Delft bij de leerstoel 'Kennisgestuurde systemen'. Tijdens dit onderzoek is een spraakherkenningssysteem, dat reisinformatie over binnenlandse treinreizen verstrekt, getest door gebruikers. Zij hebben een vragenlijst ingevuld en daarvan zijn de resultaten geëvalueerd. Het spraakherkenningssysteem dat is gebruikt is beschikbaar gesteld door openbaar vervoer reisinformatie (OVR).

Verder wil ik de honderd mensen bedanken die mee gewerkt hebben aan dit onderzoek. Zonder hen was het niet mogelijk geweest. Ook wil ik Dr. Drs. L.J.M. Rothkrantz bedanken voor zijn hulp en medewerking.

Alexandra Peters
Delft 1997

Inhoudsopgave

Voorwoord	i
1. Inleiding	1
2. Experiment	2
2.1 <i>Onderzoeksuitvoering</i>	2
2.2 <i>Proefpersonen</i>	2
2.3 <i>Controlegroep</i>	5
3. Onderzoeksmateriaal	6
3.1 <i>Reisopdrachten</i>	6
3.2 <i>Vragenlijst</i>	6
3.3 <i>Interview</i>	8
4. Nader onderzoek vragenlijst	9
4.1 <i>Keuze onderzoeksmethode</i>	9
4.2 <i>Model en structuur van de vragenlijst</i>	9
5. Onderzoek gesloten vragen van de vragenlijst	12
5.1 <i>Onderzoek vragenscores deel 1 van de vragenlijst</i>	12
5.1.1 <i>Frequentieverdeling van de vragenscores</i>	12
5.1.2 <i>Verschillen in vragenscores OVR- en 'random' - populatie</i>	14
5.1.3 <i>Onderlinge samenhang vragen</i>	17
5.2 <i>Onderzoek vragenscores deel 2 van de vragenlijst</i>	19
5.2.1 <i>Frequentieverdeling van de vragenscores</i>	19
5.2.2 <i>Onderlinge samenhang vragen</i>	20
5.3 <i>Samenhang vragen uit het eerste en tweede deel van de vragenlijst</i>	21
6. Onderzoek achtergrondvragen en open vragen van de vragenlijst	23
6.1 <i>Frequentieverdeling van de vragenscores van de achtergrondvragen</i>	23
6.2 <i>Frequentieverdeling van de vragenscores van de open vragen</i>	23

7. Appreciatiescore	27
8. Betrouwbaarheid, homogeniteit en validiteit van de vragenlijst	30
8.1 <i>Betrouwbaarheid</i>	30
8.1.1 Betrouwbaarheid van de vragenlijst	30
8.1.2 Positieve en negatieve vragen	30
8.2 <i>Homogeniteit</i>	32
8.3 <i>Validiteit</i>	33
8.3.1 Een gewogen of een ongewogen somscore?	33
8.3.2 Dubbelzinnige vragen	33
9. Evaluatie van de interviews	35
9.1 <i>Scoreverdeling van de vragen</i>	35
9.2 <i>Opmerkingen</i>	42
9.3 <i>Conclusie</i>	43
10. Samenvatting resultaten en aanbevelingen	44
10.1 <i>Algemeen</i>	44
10.2 <i>Onderzoeksvragen</i>	44
Literatuurlijst	46

1. Inleiding

Openbaar Vervoer Reisinformatie (OVR) exploiteert een telefonische informatielijn (06-9292) via welke reizigers informatie kunnen krijgen over reizen met het openbaar vervoer. Deze informatie wordt verstrekt door meer dan 400 informatrices. Het aantal verzoeken om reisinformatie is de laatste jaren fors gestegen en de maximale capaciteit is op dit moment bereikt. Om toch de groeiende vraag naar informatie te kunnen verwerken zoekt OVR naar alternatieven voor deze informatieverstrekking. Automatische spraakherkenning lijkt een veelbelovend alternatief. In opdracht van OVR heeft PTT-Telecom een geautomatiseerd informatiesysteem (VIOS) ontwikkeld, gebaseerd op automatische spraakherkenning. In [Pijpers96] is dit systeem nader bekeken. Het systeem verstrekt informatie over vertrek- en aankomsttijden van treinen. Het systeem is sprekersafhankelijk en is geschikt voor continue, spontane spraak. OVR overweegt het systeem te integreren in de dienst. Voor een weloverwogen besluit is het voor OVR en de Vennoten noodzakelijk te beschikken over gegevens hoe VIOS gewaardeerd wordt door de gemiddelde OVR-gebruiker en welke eventuele verbeteringen op de korte of lange termijn noodzakelijk zijn.

In het najaar van 1995 heeft de TU Delft onderzocht welke factoren de acceptatie en appreciatie van automatische spraakherkenningssystemen (ASS) beïnvloeden. In een Wizzard of Oz experiment [Manintveld97], waaraan 400 proefpersonen deelnamen, is nagegaan hoe een aantal gesimuleerde automatische spraakherkenningssystemen werden gewaardeerd. Aan de proefpersonen werd gevraagd aan de hand van reisopdrachten reisinformatie aan te vragen zowel bij OVR als via een van de gesimuleerde systemen. De reisopdrachten en de verschillende systemen werden systematisch op een aantal aspecten gevarieerd. Hun meningen, ervaringen en waardering konden de proefpersonen geven aan de hand van een vragenlijst met gesloten vragen. Sommige proefpersonen zijn nog aanvullend geïnterviewd.

Deze studie is een voortzetting van genoemde evaluatiestudie. In plaats van een van de gesimuleerde spraakherkenningssystemen is nu VIOS gekozen. In totaal zijn 100 personen bereid gevonden om reisinformatie aan te vragen m.b.v. VIOS en na afloop een evaluerende vragenlijst in te vullen. Een beperkt aantal personen zijn na afloop nog aanvullend geïnterviewd. De doelstellingen van het onderzoek waren:

- na te gaan in welke mate VIOS door cliënten geapprecieerd wordt
- de voornaamste wensen, klachten en problemen van gebruikers m.b.t. VIOS te inventariseren en voorstellen tot verbetering te formuleren
- de resultaten van deze appreciatiestudie te vergelijken met de resultaten van een vrijwel dezelfde enquête die onder OVR-vennoten en OVR-telefonistes gehouden is
- de gebruikte vragenlijst en evaluatieprocedure kritisch door te lichten en met name de betrouwbaarheid en validiteit van de vragenlijst te bepalen

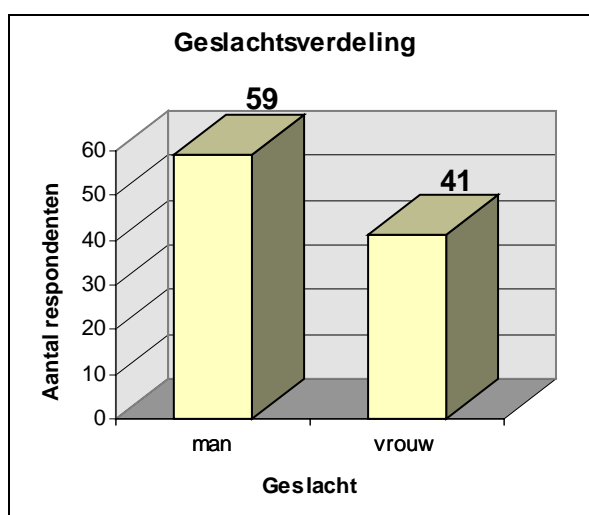
2. Experiment

2.1 Onderzoeksuitvoering

Het onderzoek is uitgevoerd als een postenquête in de periode juli en augustus 1996. De respondenten zijn mondeling benaderd om mee te doen aan het experiment. Bij instemming werd een envelop uitgereikt met een brief waarin de achtergronden en instructies van het experiment werden toegelicht, twee reisopdrachten uit de eerste respectievelijk tweede categorie en tenslotte de vragenlijst. In de brief werd gevraagd eerst VIOS te bellen, reisinformatie te vragen eventueel m.b.v. de bijgevoegde reisopdrachten, vervolgens de vragenlijst in te vullen en deze per post terug te sturen naar de TU Delft. De deelnemers aan het onderzoek kregen na afloop een klein presentje als blijk van waardering. Daarnaast zijn 20 van de 100 respondenten door de onderzoeker geïnterviewd. Deze personen zijn willekeurig gekozen uit de 100 respondenten. Ook hiervoor is men mondeling benaderd waarna een afspraak is gemaakt. Het interview vond plaats op de werkplek van of thuis bij de geïnterviewde of bij de onderzoeker thuis.

2.2 Proefpersonen

De onderzoeker heeft in totaal 112 personen benaderd uit haar familie- en kennissenkring. Deze personen zijn afkomstig uit verschillende regio's en verschillende bedrijfstakken. Uiteindelijk hebben 100 respondenten meegedaan aan het experiment en de vragenformulieren geretourneerd. De respondenten zijn niet aselekt maar wel betrekkelijk willekeurig gekozen. De ingevulde formulieren zijn volledig anoniem verwerkt. Aan de respondenten werd ook gevraagd enige biografische gegevens te verstrekken waarmee de groep van respondenten is te karakteriseren. In onderstaande figuren worden enige karakteristieken van de respondenten weergegeven.



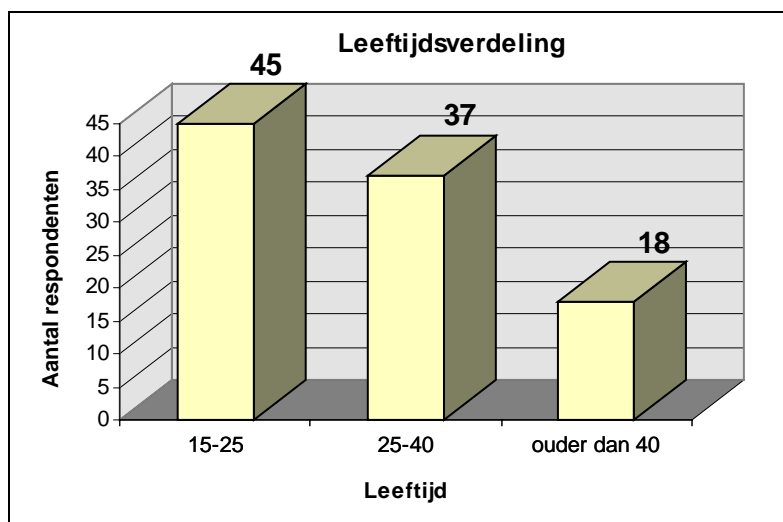
Figuur 1 Verdeling van geslachten

Figuur 1 geeft aan hoeveel mannelijke en vrouwelijke respondenten aan het onderzoek hebben meegewerkt.

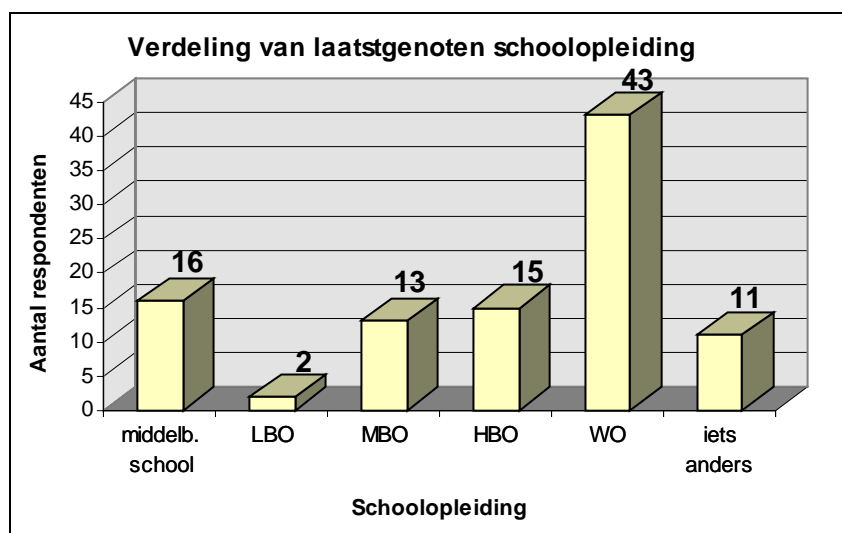
We constateren dat uiteindelijk meer mannen dan vrouwen aan het onderzoek hebben deelgenomen.

Deze grafiek geeft de verdeling van de leeftijden onder de respondenten aan.

In de grafiek is te zien dat er meer jongeren dan ouderen aan het onderzoek hebben meegewerkt. Dit wordt waarschijnlijk veroorzaakt door het feit dat de onderzoeker zelf jong is en dat ouderen vaak, vanwege hun onbekendheid met computers, niet mee wilden werken aan dit onderzoek.



Figuur 2 Verdeling van de leeftijden

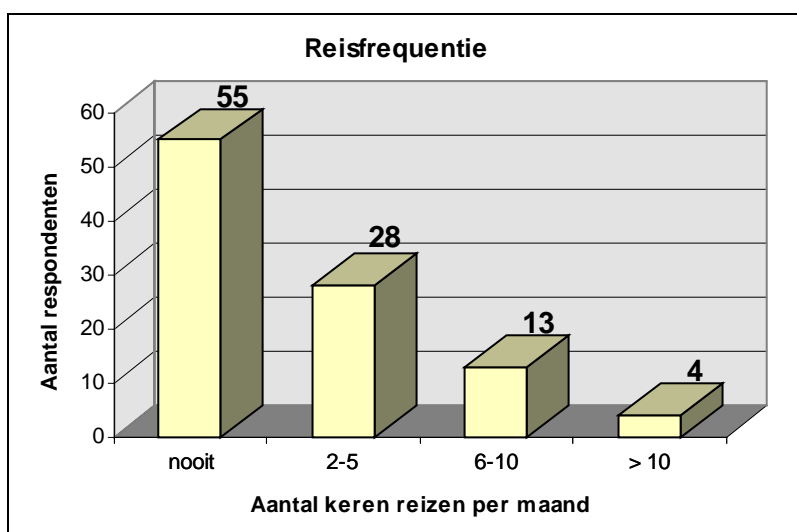


Figuur 3 Schoolopleidingen die de respondenten genoten hebben

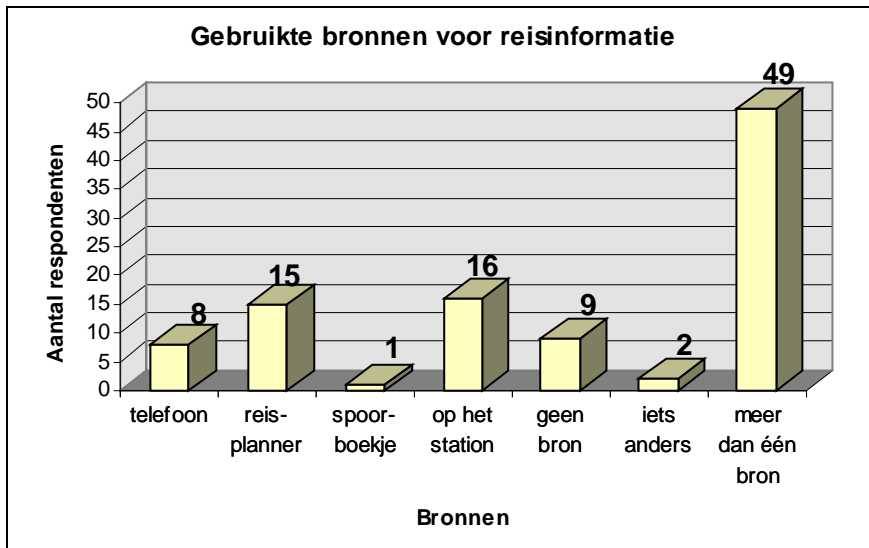
In deze grafiek is te zien welke schoolopleiding de deelnemers als laatste hebben gevolgd.

Het valt meteen op dat de hoog opgeleiden in de meerderheid zijn. Dit komt waarschijnlijk doordat de onderzoeker zelf ook hoog opgeleid is.

Figuur 4 geeft aan hoe vaak de respondenten gemiddeld per maand met de trein reizen. We zien dat meer dan de helft (55%) van de respondenten zelden of nooit met de trein reist.



Figuur 4 Het gemiddelde aantal reizen per maand met de trein

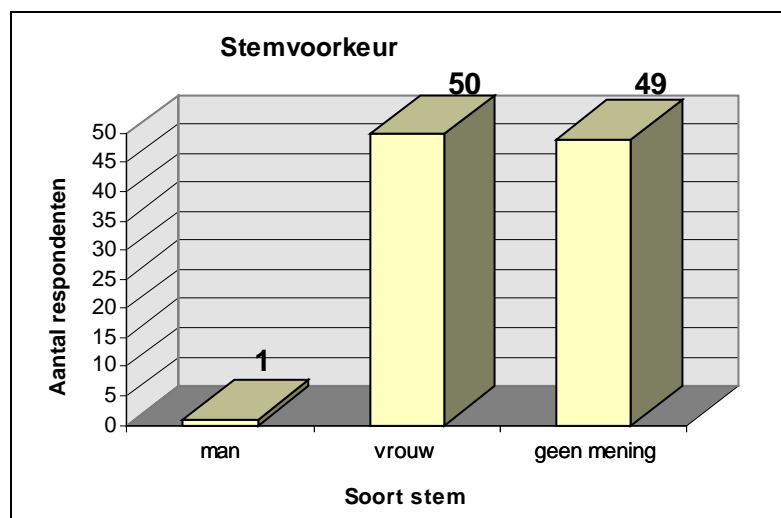


Deze grafiek laat zien hoe de respondenten nu de reisinformatie opvragen. Bijna de helft (49%) van de respondenten maakt gebruik van meerdere reisinformatiebronnen. Ongeveer 20% van de respondenten heeft OVR wel meerdere keren gebeld.

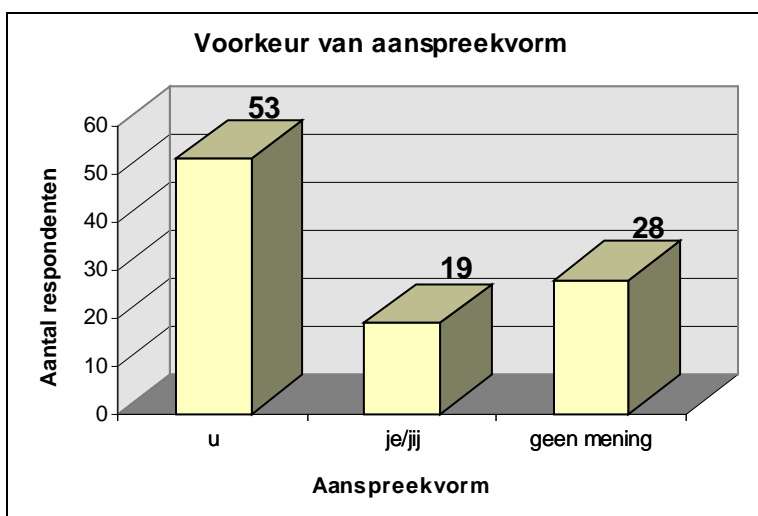
Figuur 5 Gebruikte informatiebronnen door respondenten

Figuur 6 geeft de soort stem aan die de respondenten prefereren voor een spraakherkenningssysteem.

Er is een duidelijke voorkeur voor een vrouwenstem. Dit komt waarschijnlijk doordat men hieraan gewend is omdat de meeste informatiesystemen op dit moment ingesproken zijn door een vrouw.

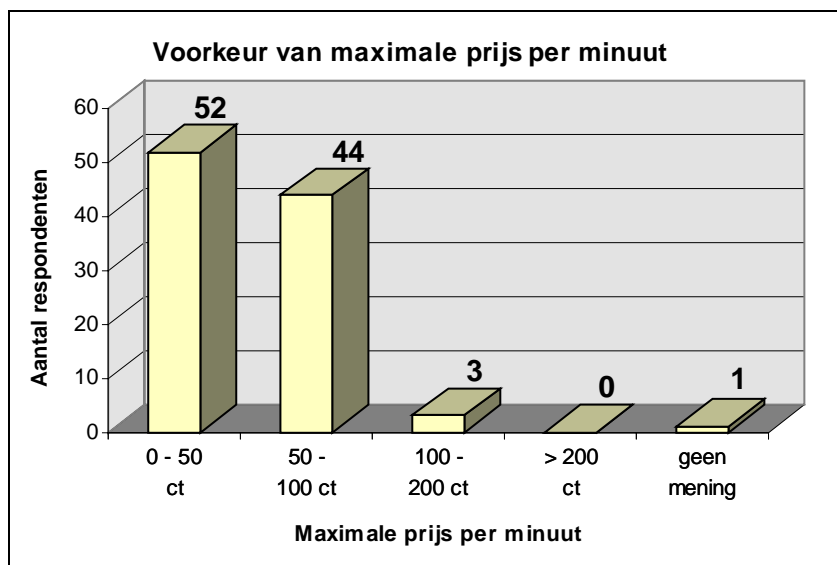


Figuur 6 Verdeling van stemvoorkeur



De respondenten hebben ook een duidelijke voorkeur voor de manier waarop ze worden aangesproken. Men wordt het liefst niet getutoyeerd. De oorzaak hiervan kan zijn dat men vindt dat het systeem te onpersoonlijk overkomt.

Figuur 7 Verdeling van gewenste aanspreekvorm



Figuur 8 Maximale prijs die men wil betalen voor een reisinformatiesysteem

Deze laatste grafiek laat zien welke maximale prijs de respondenten willen betalen voor een reisinformatiesysteem.

De meeste respondenten hebben voorkeur voor een lage prijs per minuut.

2.3 Controlegroep

In tegenstelling tot laboratoriumexperimenten is het bij veldexperimenten en postenquêtes moeilijk alle variabelen onder controle te houden. Uiteindelijk mag je als onderzoeker alleen maar hopen dat de ingevulde vragenlijst gebaseerd is op een informatiedialoog die de betreffende respondent heeft gevoerd met VIOS. Om iets meer zekerheid te verkrijgen zijn de volgende controle-experimenten uitgevoerd:

- Van alle respondenten is nagegaan of het eerste en tweede deel van de vragenlijst op de overeenkomstige vragen consistent is ingevuld.
- In totaal 20 van de respondenten zijn aanvullend geïnterviewd.
- De antwoorden op het eerste deel van de enquête zijn vergeleken met de antwoorden op dezelfde vragenlijst van een groep van 56 proefpersonen geselecteerd uit het bestand van OVR-vennoten en OVR-informatrices. Aan deze personen is gevraagd VIOS te bellen aan de hand van zelf bedachte reisopdrachten en na afloop de vragenlijst in te vullen.

3. Onderzoeksmateriaal

3.1 Reisopdrachten

In totaal zijn 10 reisopdrachten gebruikt. Deze opdrachten waren verdeeld in twee categorieën. In de eerste categorie zijn de simpele reisopdrachten geplaatst. Hierbij gaat het om reizen waarbij vrijwel alle relevante informatie in de reisopdracht aanwezig is. Bij de reisopdrachten in de tweede categorie zijn niet alle gegevens in de opdracht aanwezig en moet de reiziger via probleemoplossend gedrag de ontbrekende gegevens zelf verzamelen of vragen. Om de respondenten geen woorden in de mond te leggen en parafaseren zoveel mogelijk te beperken zijn de reisopdrachten zoveel mogelijk in non-verbale termen gesteld. Hieronder wordt van beide categorieën een voorbeeld gegeven.

SCENARIO

Scenarionummer: 6

Dit is een voorbeeld. U kunt ook uw eigen reis samenstellen.

Lees het onderstaande scenario.

U wilt morgen naar het Dolfinarium in Harderwijk reizen waarbij u gebruik maakt van de trein. U vertrekt vanaf een station in Delft. Op klok 2 kunt u zien hoe laat de middagvoorstelling begint. Vraag de reisinformatie aan.



Bel het systeem op en vraag (*in eigen woorden*) om de gewenste informatie.

Figuur 9 Scenario uit de eerste categorie

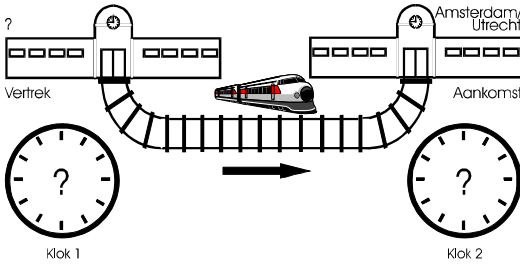
SCENARIO

Scenarionummer: 3

Dit is een voorbeeld. U kunt ook uw eigen reis samenstellen.

Lees het onderstaande scenario.

U heeft morgen vrij genomen en u wilt met de trein naar Amsterdam of Utrecht gaan (maak zelf een keuze). U vertrekt van het dichtstbijzijnde station bij uw eigen woonplaats. Vraag de reisinformatie aan.



Bel het systeem op en vraag (*in eigen woorden*) om de gewenste informatie.

Figuur 10 Scenario uit de tweede categorie

3.2 Vragenlijst

De vragenlijst bestond uit drie onderdelen:

In deel 1 wordt aan de respondent een aantal meningen en opvattingen voorgelegd. Op een 5-puntsschaal kunnen de respondenten aangeven in hoeverre ze het wel of niet eens zijn met deze beweringen. De beweringen zijn op verschillende wijzen geformuleerd. In totaal 12 vragen zijn geformuleerd in de "ik vind dat..."-stijl en 9 vragen zijn onzijdig geformuleerd zoals "het systeem is..". De beweringen zijn weergegeven in tabel 1.

Tabel 1 Beweringen uit deel 1

1) Dit systeem kwam vriendelijk en beleefd over.
2) Dit systeem maakt storende fouten.
3) Dit systeem dacht goed mee bij het plannen van mijn reis.
4) Dit systeem was verwarrend in het gebruik.
5) Dit systeem stelde geen overbodige vragen.
6) Dit systeem was te traag.
7) Dit systeem drong mij een onprettige dialoog op.
8) Dit systeem leverde de door mij gevraagde informatie.
9) Dit systeem was niet flexibel.
10) Ik zou deze dienst graag opnieuw gebruiken.
11) Ik moest me erg concentreren tijdens dit gesprek.
12) Ik heb het idee dat dit een betrouwbaar systeem was.
13) Ik wist helemaal niet hoe ik dit systeem moest gebruiken.
14) Ik vond het vervelend dat ik dit systeem niet kon onderbreken.
15) Ik werd verward door dit systeem.
16) Ik vind de communicatie met dit systeem prettig.
17) Ik denk dat dit systeem sterk verbeterd moet worden.
18) Ik vond het nuttig, dat ik de verbinding nog een keer kon horen.
19) Ik voelde me onder druk staan tijdens dit gesprek.
20) Ik kreeg de indruk dat dit systeem begreep wat ik vroeg.
21) Ik vond dit systeem te onpersoonlijk.

Dit onderdeel van de vragenlijst is een Nederlandse vertaling van een vragenlijst die in het kader van het ARISE project is gebruikt en ontwikkeld door onderzoekers van de Universiteit van Edinburgh. De bedoeling van de ontwerpers was het spraakherkenningssysteem te beoordelen op de volgende aspecten:

- gebruikersvriendelijkheid
- gebruikersgemak
- efficiëntie en effectiviteit
- dialoogvoeringen
- of via het systeem al dan niet het juiste reisadvies kon worden verkregen.

De vragen hebben betrekking op een of meerdere aspecten

(zie hoofdstuk 4: 'Nader onderzoek vragenlijst').

In deel 2, het interviewgedeelte wordt een aantal vragen aan de respondent gesteld over globaal dezelfde aspecten als in het eerste deel van de vragenlijst. De respondenten kunnen deze vragen met 'ja', 'nee' of 'geen mening' beantwoorden. De vragenlijsten 1 en 2 verschillen dus voornamelijk in formaat. Het tweede deel van de vragenlijst is voornamelijk bedoeld om een 'betrouwbaarheidsanalyse' te kunnen uitvoeren. Verder levert het tweede deel aanvullende informatie om de antwoorden uit het eerste deel goed te kunnen interpreteren. In de tabel hiernaast worden de vragen getoond.

Tenslotte bevatte de vragenlijst nog een aantal vragen met betrekking tot biografische gegevens van de respondent en enige open vragen. Deze vragen staan in tabellen 3 en 4. De antwoorden op de open vragen (tabel 4) staan vermeld in de bijlage.

Tabel 2 Interviewvragen uit deel 2

a) Vindt u de manier waarop het systeem de vragen stelt prettig?
b) Vindt u de volgorde van de vragen logisch?
c) Vindt u de vragen die het systeem stelt duidelijk?
d) Vindt u het vervelend dat het systeem het initiatief heeft?
e) Is het voor u duidelijk wat u tegen het systeem kan zeggen?
f) Moest u veel herhalen?
g) Heeft u voordat u belt al bedacht of opgeschreven wat u gaat zeggen?
h) Vindt u het systeem eenvoudig in gebruik?
i) Vindt u dat het systeem veel fouten maakt?
j) Vindt u het vervelend dat alles herhaald wordt?
k) Vindt u dat het systeem te snel spreekt?
l) Maakt het systeem snel fouten?
m) Is het systeem eenvoudig te corrigeren?
n) Vindt u het gesprek te lang duren?
o) Vond u dat u lang moest wachten tijdens het gesprek?
p) Zou u het prettig vinden om de verkregen informatie te kunnen zien? (bv. via teletekst)

Tabel 3 Biografische vragen uit deel drie

• Wat is uw geslacht?
• Wat is uw leeftijd?
• Hoe vaak reist u gemiddeld per maand met de trein?
• Welke bronnen gebruikt u nu om informatie te krijgen?
• Wat is uw laatst genoten schoolopleiding?
• Prefereert u een mannenstem of een vrouwenstem?
• Wilt u met 'u' of met 'je/jij' aangesproken worden?
• Wat is de maximale prijs per minuut die u over heeft voor een reisinformatie systeem? (bv. 50 c/pm)

Tabel 4 Open vragen uit deel drie

• Hoe vond u dit experiment?
• Hoe staat u tegenover techniek in het algemeen en tegenover computers en spraakherkenningssystemen in het bijzonder?
• Op dit moment kunt u ook reisinformatie verkrijgen met behulp van telefonistes. Wanneer het automatische spraakherkenningssysteem goed zou werken en beschikbaar zou zijn, waar zou u voorkeur dan naar uitgaan, naar telefonistes of het automatische spraakherkenningssysteem? Kunt u dit toelichten?
• Eventuele op- of aanmerkingen

3.3 Interview

Het interview bestaat uit 16 vragen. Deze staan afgebeeld in tabel 5. De vragen zijn opgesteld door de onderzoeker zelf. Een aantal van de eerder genoemde aspecten zijn in het interview verwerkt. Wanneer het nodig was, zijn de vragen toegelicht tijdens het interview. De gegeven antwoorden zijn tijdens het interview genotuleerd door de onderzoeker en daarna uitgewerkt tot een verslag. Dit laatste is weergegeven in de bijlage.

Tabel 5 Vragen gesteld tijdens het interview

1) Vindt u een spraakherkenningssysteem prettig om mee te werken?
2) Zou u het handig vinden om een spraakherkenningssysteem te gebruiken voor andere informatie? Zo ja, welke informatie?
3) Vindt u de telefoon een handig medium of prefereert u een ander medium zoals teletekst of een computerprogramma?
4) Ziet u dit systeem als een mens/telefoniste?
5) Vindt u het vervelend wanneer de telefoniste u niet zou verstaan?
6) Vindt u het systeem gebruiksvriendelijk?
7) Wat zijn uw positieve ervaringen met het systeem? of Wat vond u goed aan het systeem?
8) Wat zijn uw negatieve ervaringen met het systeem? of Wat vond u slecht of vervelend aan het systeem?
9) Had u snel door hoe het systeem werkt?
10) Is het voor u duidelijk wanneer u tegen het systeem kan spreken?
11) Verstond het systeem u goed? Zo nee, vond u het vervelend dat u veel moest herhalen?
12) Luistert u naar de vraag en geeft u precies antwoord of probeert u maar wat? Bv. veranderen van vertrektijd in aankomsttijd. (d.w.z. wanneer het systeem u er niet naar vraagt zegt u het ook niet?)
13) Vindt u de stem vervelend?
14) Zou u zelf andere vragen prefereren? Alles in één vraag of vindt u het juist fijn dat de vragen in stukjes zijn opgedeeld?
15) Zou u liever de controle hebben zoals bij een telefoniste?
16) Wat is uw algemene oordeel: Positief of negatief?

4. Nader onderzoek vragenlijst

4.1 Keuze onderzoeksmethode

Om de cliëntappreciatie van produkten te bepalen zijn er verschillende methoden en technieken beschikbaar. Bij het onderhavige onderzoek is gebruik gemaakt van vragenlijsten en een aanvullend interview. Het voordeel van vragenlijsten met gesloten vragen is dat het een betrekkelijk goedkope en objectieve methode van dataverzameling is en dat de verkregen data zich eenvoudig statistisch laat verwerken. Een probleem bij vragenlijsten is dat het weliswaar een objectief instrument is maar in feite worden subjectieve meningen en opvattingen gepeild. De vraag is of een individuele respondent zijn eigen opvattingen en meningen voldoende genuanceerd kan weergeven op de vragenlijst. Om aan deze problemen tegemoet te komen zijn een aantal open vragen aan de enquête toegevoegd. Verder zijn nog een aantal respondenten aanvullend geïnterviewd. Het doel van het interview en de open vragen is na te gaan of er bij de vragenlijst wezenlijke facetten ontbreken. Om een vergelijk tussen het interview en de vragenlijst mogelijk te maken heeft het interview een gestructureerd karakter d.w.z. de interviewer gaat uit van een vaste verzameling vragen.

4.2 Model en structuur van de vragenlijst

De belangrijkste doelstelling van het huidige onderzoek was na te gaan hoe goed VIOS door cliënten geapprecieerd wordt. Het was tevens de bedoeling na te gaan in hoeverre dit appreciatie-onderzoek kon plaatsvinden middels vragenlijsten. De onderzoeker beschikte over een vragenlijst van onderzoekers van de Universiteit van Edinburgh die recentelijk was gebruikt bij een vergelijkbaar appreciatie-onderzoek van spraakherkenningssystemen. Een probleem was dat de vragenlijst niet was gebaseerd op een expliciet beschreven model. Om de vragenlijst te kunnen evalueren is het noodzakelijk over een dergelijk model te beschikken.

Het was de bedoeling een vragenlijst te ontwerpen waarmee een overall appreciatiescore berekend kon worden. Verondersteld wordt dat aan de vragenlijst slechts een factor ten grondslag ligt. Deze ene factor heeft een aantal facetten. Met de scores op de diverse facetten wordt de totaalscore berekend. Deze ordinale score loopt van een lage, middelmatige tot en met een hoge appreciatiescore.

In de literatuur worden diverse aspecten beschreven waarop spraakherkenningssystemen in het algemeen beoordeeld worden. Het betreft hierbij voornamelijk aspecten waar de ontwerpers van spraakherkenningssystemen rekening mee dienen te houden. Het ligt voor de hand om ook de beoordeling hierop te richten. VIOS is een specifieke applicatie. Bij de appreciatiestudie hebben we hier rekening mee gehouden. In een eerder uitgevoerd Wizzard of Oz experiment [Manintveld97] zijn de effecten onderzocht van een aantal specifieke factoren op de appreciatie van spraakherkenningssystemen. We formuleren eerst een aantal uitgangspunten voor de appreciatiestudie:

- De vragenlijst is bedoeld voor modale OVR-gebruikers, dit wil zeggen dat er geen specifieke deskundigheid van spraakherkenningssystemen wordt verondersteld.
- De vragenlijst is bedoeld voor gebruikers na een korte kennismaking met VIOS; de gebruiker moet het systeem drie tot tien keer geraadpleegd hebben. Het gaat dus om een peiling van de eerste indrukken en bevindingen.
- Het invullen van de vragenlijst mag niet meer dan 5 minuten tijd vergen.

- De vragenlijst moet vooral vragen bevatten die de verschillen evalueren tussen een geautomatiseerd informatiesysteem en een systeem met menselijke operators. VIOS is een alternatief voor het huidige OVR-informatiesysteem.
- De vragenlijst is gericht op de introductie van een nieuw systeem kort voor of na de introductie en niet de evaluatie van een bestaand systeem.

Bij de keuze van de vragen is vooral gelet op de verschillen tussen mens-mens en mens-machine systemen. In zijn algemeenheid worden spraakherkenningssystemen minder gewaardeerd vanwege de volgende redenen:

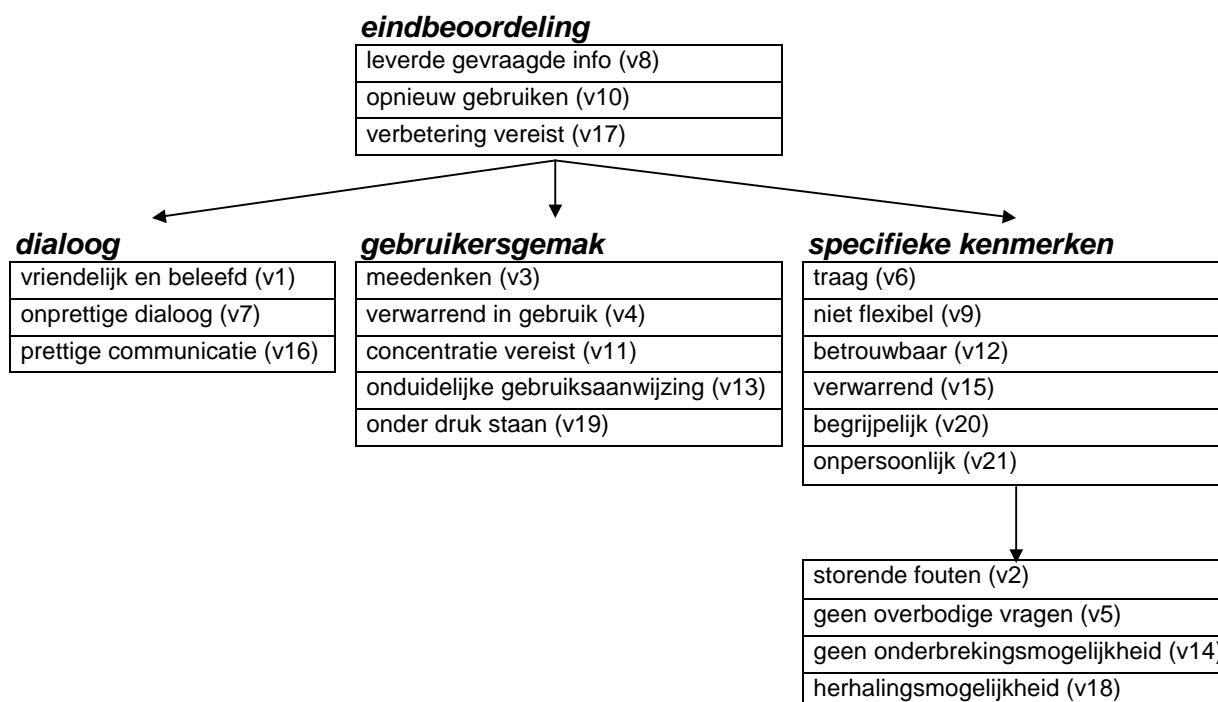
Spraakherkenningssystemen verstaan en begrijpen niet alles wat een gebruiker zegt. Dit betekent dat systemen fouten maken. Verder bevatten deze systemen voortdurend vragen om bevestigingen en herhalingen. Dergelijke systemen zijn in de regel strak geprogrammeerd, dus minder flexibel en met een directieve dialoogvoering. Dergelijke systemen worden vaak als onpersoonlijk ervaren omdat ze mechanisch overkomen.

In feite wordt vooral geëvalueerd in hoeverre spraakherkenningssystemen de menselijke systemen benaderen. We realiseren ons dat spraakherkenningssystemen geen getrouwe kopie hoeven te zijn van menselijke systemen om positief gewaardeerd te worden. Spraakherkenningssystemen hebben ook een aantal typisch positieve aspecten zoals 24-uurs-bereikbaarheid, dergelijke systemen raken niet geïrriteerd en zijn in hoge mate voorspelbaar. Deze aspecten zijn pas goed te evalueren als VIOS daadwerkelijk in gebruik genomen is en door gebruikers getest in het gebruik van alledag en niet alleen in een experimentele situatie na kortstondig gebruik. Mogelijk dat deze facetten in deel 2 van de vragenlijst en in het interview door gebruikers aan de orde worden gesteld. We geven een opsomming van de respectievelijke facetten en geven hierbij tevens aan welke vragen uit deel 1 van de vragenlijst bij een bepaald facet behoren. De codering staat tussen haakjes achter ieder genoemde vraag. Deze codering is ook te bekijken in bijlage C 'Het codeboek'.

- ◆ **Eindbeoordeling**
Verondersteld wordt dat een gebruiker van spraakherkenningssystemen meer tevreden zal zijn als de interactie met het systeem de gevraagde informatie oplevert (v8), hij zal het systeem vervolgens in de toekomst opnieuw gebruiken (v10) en verbeteringen van VIOS minder noodzakelijk vinden (v17).
- ◆ **Dialoog**
Verondersteld wordt dat een gebruiker van spraakherkenningssystemen meer tevreden zal zijn als hij beleefd en vriendelijk wordt toegesproken (v1) en wanneer hij de communicatie (v16) en dialoog (v7) prettig vindt verlopen.
- ◆ **Gebruikersgemak**
Verondersteld wordt dat een gebruiker van spraakherkenningssystemen meer tevreden zal zijn als het systeem een duidelijke gebruiksaanwijzing heeft (v13), het systeem meedenkt (v3), minder verwarrend is in het gebruik (v4), weinig concentratie vereist (v11) en de gebruiker niet onder druk zet (v19).
- ◆ **Specifieke kenmerken**
Verondersteld wordt dat een gebruiker van spraakherkenningssystemen tevredener is als het systeem betrouwbaar is (v12), geen storende fouten maakt (v2) en de cliënt goed begrijpt (v20), geen overbodige vragen stelt (v5), vlot (v6), flexibel (v9) en niet verwarrend is (v15), persoonlijk is (v21) en de mogelijkheid voor interruptie (v14) en herhaling van het reisadvies (v18) heeft.

We merken op dat de vragen behorende bij de verschillende facetten onderling sterk afhankelijk kunnen zijn. Zo is het evident dat de gebruikers die vinden dat het systeem veel fouten maakt, traag en niet flexibel is, dat deze gebruikers van mening zijn dat het systeem sterk verbeterd moet worden. In feite zijn alle vragen in het facet eindbeoordeling

gecorrleerd met vragen in de overige facetten. Verder merken we op dat de vragen verschillende reikwijdte hebben. De vragen in het facet 'eindbeoordeling' zijn zeer algemeen, de vragen in het facet 'specifieke kenmerken' specifiek. Maar ook binnen een facet kunnen de vragen verschillen naar reikwijdte. De vragen *storende fouten* (v2) en *geen onderbrekingsmogelijkheid* (v14) zijn expliciet geformuleerd en daarmee ook specifiek. De vragen *verwarrend* (v15) en *niet flexibel* (v9) zijn niet specifiek en laten allerlei subjectieve interpretaties toe. We kunnen de vragen en facetten hiërarchisch ordenen zoals in figuur 11 weergegeven.



Figuur 11 Hiërarchisch overzicht van facetten en vragen

5. Onderzoek gesloten vragen van de vragenlijst

We beschikken over 100 ingevulde vragenlijsten. Deze respondenten noemen we de 'random'-populatie. Daarnaast is door de 'OVR-populatie' een vrijwel dezelfde enquête¹ als deel 1 van de vragenlijst ingevuld. De grootte hiervan is 56 personen. In dit hoofdstuk gaan we op de eerste plaats na hoe we met deze gegevens de eerste twee onderzoeksvragen kunnen beantwoorden:

- 1) In welke mate wordt VIOS door cliënten geapprecieerd?
- 2) Wat waren de voornaamste klachten en problemen van de VIOS-gebruikers?

Het eerste deel van de enquête bestond uit 21 vragen waarin de mening van de respondent werd gevraagd over bepaalde facetten van VIOS. Een eerste indruk van de waardering van VIOS kan verkregen worden door bij iedere vraag na te gaan hoeveel personen een positief, negatief dan wel neutraal antwoord gaven. De negatieve antwoorden kunnen bij de meeste vragen opgevat worden als klachten over het systeem. In paragraaf 5.1 en 5.2 geven we de resultaten. Meerdere vragen hebben betrekking op hetzelfde facet. Naast de waardering van de diverse VIOS-facetten zijn we geïnteresseerd in de waardering van VIOS als totaal systeem. In feite willen we de appreciatie uitdrukken in een rapportcijfer. Dit zullen we nader onderzoeken in hoofdstuk 7.

5.1 Onderzoek vragenscores deel 1 van de vragenlijst

We starten de analyse met een commentaar op de scoreverdeling van de vragen.

5.1.1 Frequentieverdeling van de vragenscores

De frequentieverdeling van de vragenscores wordt weergegeven in tabel 6. Deze staat ook vermeld in de bijlagen. De nummers 1 t/m 5 staan voor de antwoorden die de respondenten konden kiezen, deze zijn:

1. volledig mee eens
2. mee eens
3. neutraal
4. mee oneens
5. volledig mee oneens

We bespreken nu per facet de verschillende vragen.

◆ Eindbeoordeling

Meer dan de helft van de respondenten vindt dat het systeem de gevraagde informatie levert en 30% vindt van niet (v8). Ongeveer 40% zou de dienst graag opnieuw gebruiken en 33% niet (v10). Meer dan de helft vindt dat het systeem verbeterd moet worden en ongeveer 15 % vindt van niet (v17).

Conclusie:

Ongeveer 40% van de respondenten is positief en 30% is negatief te oordelen over VIOS. Waarschijnlijk oordelen er meer respondenten positief wanneer het systeem verbeterd wordt.

¹ De verschillen zitten in de vragen 5 en 14. Vraag 5 is verschillend gesteld en daarom niet te vergelijken. Vraag 14 is positief en negatief gesteld t.o.v. de appreciatie en wel te vergelijken doordat de frequentieverdeling van één van de twee populaties is omgekeerd.

Tabel 6 Frequentieverdeling vragenscores deel 1 van de vragenlijst

Vraag	Omschrijving	1	2	3	4	5
v1	vriendelijk en beleefd	27	60	10	3	0
v2	storende fouten	22	28	16	25	9
v3	meedenken	7	23	40	18	12
v4	verwarrend in gebruik	8	18	16	42	16
v5	geen overbodige vragen	25	49	9	9	8
v6	traag	1	6	26	48	19
v7	onprettige dialoog	8	10	22	45	15
v8	leverde gevraagde info	26	36	9	9	20
v9	niet flexibel	15	30	23	25	7
v10	opnieuw gebruiken	11	32	24	16	17
v11	concentratie vereist	3	24	25	39	9
v12	betrouwbaar	8	44	23	15	10
v13	onduidelijke gebruiksaanwijzing	3	9	14	40	34
v14	geen onderbrekingsmogelijkheid	18	47	18	15	2
v15	verwarrend	6	17	22	41	14
v16	prettige communicatie	4	24	39	26	7
v17	verbetering vereist	25	33	26	15	1
v18	herhalingsmogelijkheid	30	46	19	3	2
v19	onder druk staan	2	10	17	47	24
v20	begrijpelijk	10	35	24	17	14
v21	onpersoonlijk	7	15	35	33	10

◆ Dialoog

Een ruime meerderheid van de respondenten vindt VIOS vriendelijk en beleefd (v1). Meer dan de helft vindt de dialoog prettig (v7) en 28% vindt de communicatie met VIOS prettig en 33% vindt dat niet.

Conclusie:

De dialoog met VIOS wordt overwegend positief gewaardeerd. Eenderde van de respondenten vindt het niet prettig te communiceren met VIOS. Vermoedelijk vinden deze respondenten de communicatie met een verbeterde versie van VIOS ook niet prettig verlopen, omdat deze personen waarschijnlijk aversie tegen automatische spraakherkenningssystemen in het algemeen hebben.

◆ Specifieke kenmerken

Een meerderheid vindt dat het systeem geen overbodige vragen stelt (v5). De helft vindt dat het systeem storende fouten maakt (v2). Een ruime meerderheid vindt het systeem niet te traag (v6). Bijna de helft vindt het systeem niet flexibel en 32% vindt van wel (v9). Een ruime meerderheid vindt het vervelend dat het systeem niet onderbroken kan worden en 15% vindt dat niet (v14). Een enkeling vindt het niet nuttig dat de verbinding nog een keer gehoord kan worden en een ruime meerderheid vindt van wel (v18). Bijna de helft voelt zich begrepen door het systeem en 31% voelt zich niet begrepen (v20). De helft vindt het een betrouwbaar systeem en 25% vindt van niet (v12). Iets meer dan de helft wordt niet verward door het systeem (v15) en 43% vindt het systeem niet te onpersoonlijk (v21).

Conclusie:

De voornaamste klachten zijn dat het systeem storende fouten maakt, dat het systeem niet kan worden onderbroken en dat het systeem niet voldoende flexibel is.

◆ Gebruikersgemak

Een meerderheid vindt VIOS niet verwarrend in het gebruik (v4). Bijna de helft hoefde zich niet sterk te concentreren (v11). Ongeveer drie kwart wist hoe ze het systeem moesten gebruiken (v13). Eveneens een ruime meerderheid voelde zich niet onder druk staan (v19). Ongeveer 30% van de respondenten vindt dat VIOS goed resp. niet goed meedacht (v3).

Conclusie:

VIOS-gebruikers oordelen overwegend positief over het gebruikersgemak. Toch zijn er voldoende klagers om aandacht te besteden aan het gebruikersgemak. Het blijkt dat bellers een verkeerde indruk hebben van de gebruikersmogelijkheden. Uit de open vragen destilleren wij de volgende specifieke klachten:

- *Wanneer het systeem gegevens foutief bevestigt en tegelijkertijd een nieuwe vraag stelt, weten bellers niet hoe te reageren.*
- *Het systeem vraagt alleen naar vertrektijden en niet naar aankomsttijden. Onduidelijk is hoe te handelen als de beller wil weten hoe laat hij moet vertrekken om op een bepaalde tijd aan te komen.*
- *Sommige prompts zijn ambigu, zoals "Wanneer wilt u vertrekken?" (dag/tijd) en "Wilt u nog meer informatie?" (bedoeld wordt over vertrek- en aankomsttijden).*
- *De open window structuur werkt soms verwarrend. De ontwerpers van VIOS gingen er van uit dat net als bij mens-mens dialogen, bellers niet altijd een antwoord geven op de vraag. Het systeem probeert uit ieder antwoord van de beller maximale informatie te halen. Op het moment dat het systeem fouten maakt bij duidelijke antwoorden leidt dit tot onbegrip en geeft dit een negatieve waardering.*

5.1.2 Verschillen in vragenscores OVR- en 'random' - populatie

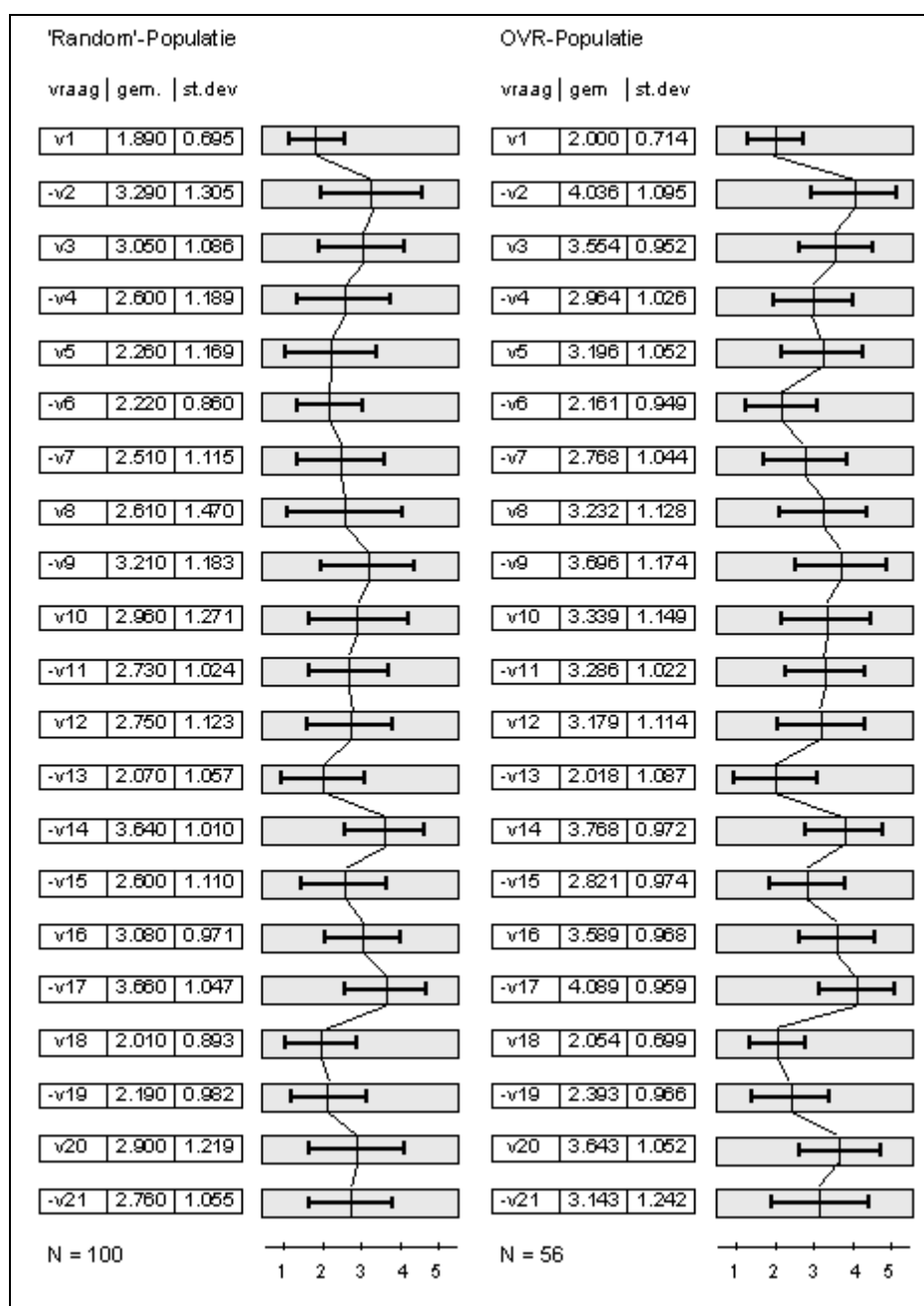
Automatische spraakherkenningssystemen zijn (nog lang) niet gelijkwaardig aan systemen met menselijke operatoren. De OVR-populatie bestond voornamelijk uit 'operatoren'. Het is te verwachten dat zij VIOS kritischer zullen benaderen. Deels omdat zij als geen ander het huidige OVR-systeem kennen en constateren dat automatische spraakherkennings-

Tabel 7 Frequentieverdeling vragenscores OVR-populatie

Vraag	Omschrijving	1	2	3	4	5
v1	vriendelijk en beleefd	21.4	60.7	14.3	3.6	0
v2	storende fouten	37.5	44.6	8.9	1.8	7.1
v3	meedenken	0	16.1	28.6	39.3	16.1
v4	verwarrend in gebruik	3.6	32.1	28.6	28.6	7.1
v5	efficiënt	1.8	30.4	25	32.1	10.7
v6	traag	1.8	8.9	16.1	50	23.2
v7	onprettige dialoog	7.1	17.9	25	44.6	5.4
v8	leverde gevraagde info	3.6	30.4	17.9	35.7	12.5
v9	niet flexibel	28.6	37.5	12.5	17.9	3.6
v10	opnieuw gebruiken	3.6	21.4	33.9	19.6	21.4
v11	concentratie vereist	10.7	35.7	26.8	25	1.8
v12	betrouwbaar	3.6	30.4	23.2	30.4	12.5
v13	onduidelijke gebruiksaanwijzing	5.4	5.4	10.7	42.9	35.7
v14	geen onderbrekingsmogelijkheid	21.4	48.2	17.9	10.7	1.8
v15	verwarrend	3.6	25	25	42.9	3.6
v16	prettige communicatie	0	14.3	32.1	33.9	19.6
v17	verbetering vereist	39.3	39.3	14.3	5.4	1.8
v18	herhalingsmogelijkheid	17.9	62.5	16.1	3.6	0
v19	onder druk staan	1.8	14.3	19.6	50	14.3
v20	begrijpelijk	1.8	16.1	19.6	41.1	21.4
v21	onpersoonlijk	19.6	19.6	21.4	33.9	5.4

systemen met betrekking tot functionaliteit en dialoogvoering nog te kort schieten vergeleken met het huidige OVR-systeem. Deels omdat automatische spraakherkenningssystemen een bedreiging voor de werkgelegenheid kunnen zijn. Vergeleken met de respondenten uit de 'random'-populatie vinden significant meer OVR-respondenten dat. In tabel 7 is de frequentieverdeling van de OVR-populatie weergegeven.

In figuur 12 (linkerkolom) wordt voor alle vragen van deel 1 het gemiddelde en de standaarddeviatie grafisch weergegeven. Om de vragen te kunnen vergelijken zijn de negatieve vragen omgezet zodat het antwoord 'volledig mee eens' voor iedere vraag een positieve waardering over het systeem geeft. Anders gezegd, de negatief gestelde vragen zijn omgepoold. Wanneer dat is gebeurd is dat te zien aan het minteken dat voor de code van de vraag staat.



Figuur 12 Overzicht van de gemiddelden en standaarddeviaties van deel 1

We constateren dat bij de 'random'-populatie het gemiddelde varieert tussen '1.9-3.7' en de standaarddeviatie tussen '0.7-1.5'. Ook zijn de gemiddelden en de standaarddeviaties van alle vragen van de OVR-populatie te zien (rechterkolom). Hier varieert het gemiddelde tussen '2.0-4.1' en de standaarddeviatie tussen '0.7-1.2'.

We zijn per vraag nagegaan waar de scores van beide populaties significant verschillen.

- Dit systeem maakt storende fouten (v2)
82% van de OVR-populatie is het hiermee eens en maar 50% van de 'random'-populatie. Deze laatste zijn dus iets positiever.
- Het systeem dacht goed mee bij het plannen van mijn reis (v3)
Ook hier is de OVR-populatie negatiever, 55% vindt dat het systeem niet goed meedenkt. Bij de andere populatie is dit 30%.
- Dit systeem was verwarrend in het gebruik (v4)
58% van de 'random'-populatie vindt het systeem niet verwarrend en 25% vindt dat wel. Daarentegen is 35% van OVR het er wel mee eens en ook 35% vindt het systeem niet verwarrend. Dus deze zijn weer negatiever.
- Het systeem leverde de door mij gevraagde informatie (v8)
Dit was bij de 'random'-populatie bij 62% het geval en maar bij 33% van de OVR-populatie.
- Het systeem was niet flexibel (v9)
Hierbij is het verschil niet zo groot tussen de beide populaties. Maar de OVR-populatie is ook hier negatiever.
- Ik zou deze dienst graag opnieuw gebruiken (v10)
Hierbij is de 'random'-populatie in de meerderheid. 43% zou het systeem graag opnieuw gebruiken t.o.v. 25% bij de OVR-populatie. De meerderheid (41%) bij deze populatie wil het systeem niet opnieuw gebruiken.
- Ik moest me erg concentreren tijdens dit gesprek (v11)
45% (OVR) is het hiermee eens terwijl dat bij de 'random'-populatie 27% is.
- Ik heb het idee dat dit een betrouwbaar systeem was (v12)
Een kleine meerderheid van de OVR-populatie vindt het systeem onbetrouwbaar. Bij de 'random'-populatie vindt 52% het systeem juist betrouwbaar.
- Ik vind de communicatie met dit systeem prettig (v16)
De communicatie met het systeem wordt door 53% van de OVR-populatie niet prettig gevonden. Een kleine meerderheid (33%) van de 'random'-populatie is het hiermee eens. De meeste respondenten (39%) waren hier neutraal over.
- Ik denk dat dit systeem sterk verbeterd moet worden (v17)
Beide populaties zijn het erover eens dat het systeem verbeterd moet worden. Bij de OVR-populatie is 78% daarvan overtuigd en 58% van de 'random'-populatie. Bij deze laatste populatie zijn de meningen meer verspreid. Ook hier is de OVR-populatie negatiever.
- Ik kreeg de indruk dat dit systeem begreep wat ik vroeg (v20)
Hier zijn de meningen verdeeld. De OVR-populatie (62%) vindt dat het systeem niet begrijpelijk overkwam, terwijl 45% van de 'random'-populatie het systeem wel begrijpelijk vond.

We constateren dat de OVR-respondenten negatiever over VIOS oordelen, geheel conform onze verwachting. In figuur 12 is duidelijk te zien dat de OVR-populatie inderdaad negatiever over VIOS is dan de 'random'-populatie. De gemiddelden liggen bijna bij iedere vraag hoger dan bij de 'random'-populatie. En hoe hoger het gemiddelde hoe negatiever men over het systeem is.

5.1.3 Onderlinge samenhang vragen

Het is interessant te weten of en hoe de verschillende vragen onderling samenhangen. Anders geformuleerd zijn er groepen van respondenten die op verschillende vragen vergelijkbare antwoorden hebben gegeven. Er wordt verondersteld dat er een dominante factor aan de grondslag ligt. Een manier om dit te testen is factor analyse. Uit tabel 8 kunnen we aflezen dat er inderdaad een dominante factor is. De vragen, waarbij de factor dichtbij 1 of -1 ligt, bepalen in grote mate de appreciatie van het systeem. De vragen met een lage factor dragen niet veel bij in de appreciatie van het systeem. De vragen die het sterkst corresponderen met de factor en onderlinge samenhang vertonen, zijn:

- storende fouten (v2)
- leverde gevraagde info (v8)
- opnieuw gebruiken (v10)
- betrouwbaar (v12)
- prettige communicatie (v16)
- verbetering vereist (v17)
- begrijpelijk (v20)

Tabel 8 Factorscores

Vraag	1 ^e factor
v1	-0.25
v2	0.70
v3	-0.60
v4	0.67
v5	-0.43
v6	0.46
v7	0.66
v8	-0.72
v9	0.56
v10	-0.77
v11	0.49
v12	-0.72
v13	0.27
v14	0.49
v15	0.65
v16	-0.73
v17	0.73
v18	-0.19
v19	0.57
v20	-0.83
v21	0.51

De samenhang laat zich als volgt verklaren: Een systeem dat storende fouten maakt, komt niet begrijpelijk over, heeft geen prettige communicatie, levert vaak niet de gevraagde info, waardoor men het systeem niet opnieuw wil gebruiken en er duidelijk verbetering gewenst is.

Tabel 9 Correlatiematrix v1v21-v1v21

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21
v1	1	.19	.05	.13	.10	.16	.32	.04	.03	.16	.30	.03	.22	.10	.07	.22	.01	.00	.25	.21	.21
v2	.19	1	.35	.50	.10	.30	.47	.46	.45	.50	.32	.48	.18	.34	.39	.44	.63	.13	.37	.53	.18
v3	.05	.35	1	.34	.18	.29	.39	.46	.45	.53	.08	.43	.02	.31	.31	.48	.45	.03	.17	.42	.24
v4	.13	.51	.34	1	.29	.28	.58	.35	.46	.38	.47	.35	.15	.28	.55	.39	.47	.08	.37	.41	.24
v5	.10	.10	.18	.29	1	.34	.35	.33	.03	.32	.19	.31	.02	.22	.32	.32	.16	.09	.27	.36	.18
v6	.16	.30	.29	.28	.34	1	.38	.27	.26	.28	.25	.28	.03	.27	.08	.27	.26	.08	.33	.30	.10
v7	.32	.47	.39	.58	.35	.38	1	.30	.39	.34	.31	.37	.09	.26	.40	.49	.39	.04	.42	.49	.25
v8	.04	.46	.46	.35	.33	.27	.30	1	.37	.60	.16	.61	.18	.34	.42	.44	.57	.16	.25	.73	.30
v9	.03	.45	.45	.46	.03	.26	.39	.37	1	.36	.26	.37	.17	.27	.26	.36	.46	.01	.27	.37	.11
v10	.16	.50	.53	.38	.32	.28	.34	.60	.36	1	.25	.57	.19	.39	.38	.63	.60	.19	.37	.67	.45
v11	.30	.32	.08	.47	.19	.25	.31	.16	.26	.25	1	.20	.39	.24	.43	.31	.23	.04	.51	.29	.21
v12	.03	.48	.43	.35	.31	.28	.37	.61	.37	.57	.20	1	.13	.39	.45	.44	.54	.23	.34	.66	.30
v13	.22	.18	.02	.15	.02	.03	.09	.18	.17	.19	.39	.13	1	.04	.24	.07	.12	.01	.38	.23	.17
v14	.10	.34	.31	.28	.22	.27	.26	.34	.27	.39	.24	.39	.04	1	.21	.22	.28	.10	.21	.34	.26
v15	.07	.39	.31	.55	.32	.08	.40	.42	.26	.38	.43	.45	.24	.21	1	.40	.46	.18	.37	.49	.39
v16	.22	.44	.48	.39	.32	.27	.49	.44	.36	.63	.31	.44	.07	.22	.40	1	.52	.23	.37	.66	.43
v17	.01	.63	.45	.47	.16	.26	.39	.57	.46	.60	.23	.54	.12	.28	.46	.52	1	.17	.22	.61	.25
v18	.00	.13	.03	.08	.09	.08	.04	.16	.01	.19	.04	.23	.01	.10	.18	.23	.17	1	.56	.13	.17
v19	.25	.37	.17	.37	.27	.33	.42	.25	.23	.37	.51	.34	.38	.21	.37	.27	.22	.06	1	.37	.43
v20	.21	.53	.42	.41	.36	.30	.49	.73	.37	.67	.29	.66	.23	.34	.49	.66	.61	.13	.37	1	.47
v21	.21	.18	.24	.24	.18	.10	.25	.30	.11	.45	.21	.30	.17	.26	.39	.43	.25	.17	.43	.47	1

5.2 Onderzoek vragenscores deel 2 van de vragenlijst

De vragen in deel 2 van de vragenlijst (het interviewgedeelte) waren bedoeld als een controle van c.q. uitbreiding op de vragen van deel 1 van de vragenlijst. We hebben het bestand van vragenscores met een aantal statistische technieken geanalyseerd. We vermelden de resultaten van deze analyses.

5.2.1 Frequentieverdeling van de vragenscores

In tabel 10 is de verdeling aangegeven van de antwoorden over de verschillende categorieën van iedere vraag. Dit zijn de volgende categorieën:

1. ja
2. geen mening
3. nee

We constateren het volgende:

De duidelijkheid van de vraagstelling (v28), de logische volgorde van de vragen (v27) en de manier waarop het systeem de vragen stelt (v26) worden door het overgrote deel van de respondenten positief gewaardeerd. Ze vinden bovendien dat het systeem eenvoudig in het gebruik is (v33). Dat het systeem alles herhaald vinden de meesten niet vervelend (v35). Het systeem spreekt ook niet te snel (v36). De meeste respondenten vinden ook dat ze niet lang moeten wachten tijdens het gesprek (v40) en ze vinden het gesprek ook niet te lang duren (v39).

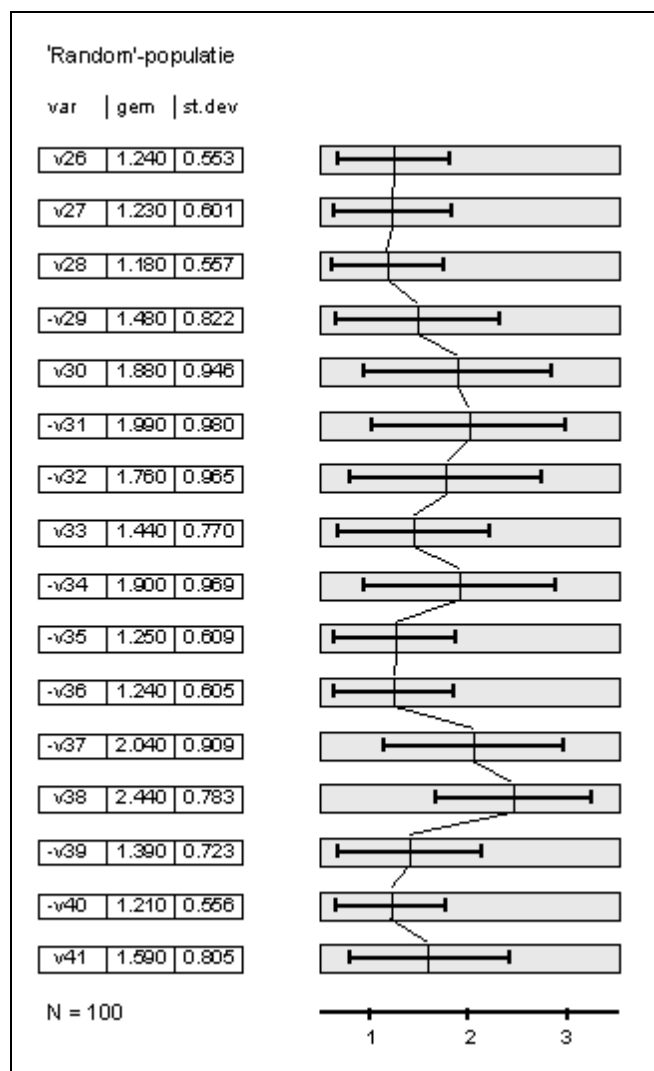
Tabel 10 Frequentieverdeling vragenscores deel 2

Code	Vragen	1	2	3
v26	Vindt u de manier waarop het systeem de vragen stelt prettig?	82	12	6
v27	Vindt u de volgorde van de vragen logisch?	86	5	9
v28	Vindt u de vragen die het systeem stelt duidelijk?	90	2	8
v29	Vindt u het vervelend dat het systeem het initiatief heeft?	21	6	73
v30	Is het voor u duidelijk wat u tegen het systeem kan zeggen?	51	10	39
v31	Moest u veel herhalen?	47	5	48
v32	Heeft u voordat u belt al bedacht of opgeschreven wat u gaat zeggen?	37	2	61
v33	Vindt u het systeem eenvoudig in gebruik?	73	10	17
v34	Vindt u dat het systeem <i>veel</i> fouten maakt?	42	6	52
v35	Vindt u het vervelend dat alles herhaald wordt?	9	7	84
v36	Vindt u dat het systeem te snel spreekt?	9	6	85
v37	Maakt het systeem <i>snel</i> fouten?	43	18	39
v38	Is het systeem eenvoudig te corrigeren?	18	20	62
v39	Vindt u het gesprek te lang duren?	14	11	75
v40	Vond u dat u lang moest wachten tijdens het gesprek?	7	7	86
v41	Zou u het prettig vinden om de verkregen informatie te kunnen zien? (bv. via teletekst)	61	19	20

De respondenten waren kritischer met betrekking tot de volgende aspecten. Een kwart van de respondenten vindt het vervelend dat het systeem het initiatief heeft (v29). Voor ongeveer 40% van de respondenten is het niet duidelijk wat ze tegen het systeem kunnen zeggen (v30). Bijna 40% had vooraf bedacht wat ze tegen het systeem gingen zeggen. (v32) Verder had meer dan 60% van de respondenten behoefte aan aanvullende visuele informatie (v41).

Bijna de helft zegt dat ze veel moesten herhalen (v31). Iets meer als 40% vindt dat het systeem *veel* fouten maakt (v34). Eveneens vindt 40% dat het systeem *snel* fouten maakt (v37). Ruim 60% vindt dat het systeem niet eenvoudig is te corrigeren (v38).

In figuur 14 is voor iedere vraag het gemiddelde en de standaarddeviatie aangegeven. Ook hier zijn de negatieve vragen omgezet zodat bij alle vragen '1' het meest positieve antwoord is. Wat onmiddellijk opvalt is de scheve verdeling en geringe spreiding van de antwoorden bij een aantal vragen. Het gemiddelde varieert van '1.2' tot '2.4' en de standaarddeviatie tussen '0.6' en '1.0'. De respondenten zijn positief gestemd over het systeem aangezien het gemiddelde bij de meeste vragen dichtbij 1 ligt.



Figuur 14 Overzicht van gemiddelden en standaarddeviaties van deel 2

5.2.2 Onderlinge samenhang vragen

Het is interessant te weten of en hoe de verschillende vragen onderling samenhangen. Bij het samenstellen van deel 2 van de vragenlijst was het uitgangspunt dat er minstens drie groepen respondenten zouden zijn namelijk degenen met een hoge, een middelmatige en een lage VIOS-appreciatie. Maar mogelijk zijn er meer onderling onafhankelijke ordinale schalen waarlangs de respondenten geordend kunnen worden. Om dat nader te onderzoeken passen we factoranalyse toe op de dataset van vragenscores. We vinden twee groepen van onderling samenhangende vragen en een aantal min of meer losse vragen.

Groep 1:

- herhaling gebruiker (v31)
- eenvoudig in gebruik (v33)
- *veel* fouten (v34)
- *snel* fouten (v37)
- eenvoudige correctie (v38)

Tabel 11 Factorscores deel 2

Vraag	1 ^e factor	2 ^e factor
v26	-0.18	-0.38
v27	-0.20	-0.62
v28	-0.15	-0.06
v29	0.44	0.32
v30	-0.18	-0.22
v31	0.79	0.06
v32	0.38	0.18
v33	-0.67	-0.13
v34	0.82	0.12
v35	-0.00	0.76
v36	0.08	0.31
v37	0.81	-0.01
v38	-0.59	0.01
v39	0.35	0.64
v40	0.05	0.71
v41	0.03	0.12

Bij de appreciatie van VIOS laten sommige klanten hun positieve of negatieve waardering blijken door een positieve resp. negatieve score op genoemde vragen. De samenhang tussen de vragen laat zich als volgt interpreteren:

Een systeem wat wel/geen fouten maakt bevat wel/geen herhalingen, bovendien zijn gemaakte fouten wel/niet moeilijk te corrigeren waardoor het systeem wel/niet eenvoudig is in het gebruik.

De tweede groep bestaat uit de volgende onderling samenhangende vragen:

- logische volgorde van de vragen (v27)
- herhaling systeem (v35)
- gespreksduur (v39)
- wachttijden tijdens gesprek (v40).

De onderlinge samenhang laat zich als volgt interpreteren:

Door de vele herhalingen gaat de informatiedialoog langer duren, moet de cliënt telkens wachten en is de volgorde der vragen niet altijd de meest logische.

Zoals reeds opgemerkt zijn de vragen in genoemde twee groepen gedeeltelijk onafhankelijk van elkaar. De eerste groep vragen hangt duidelijk samen met de appreciatiefactor uit deel 1 van de vragenlijst. De vragen uit groep twee zijn hiervan min of meer onafhankelijk. De overige vragen hebben ook geen dominante invloed op de appreciatiegraad maar bevatten interessante gegevens. Zo is het interessant dat bijna de helft van de respondenten, voordat ze een informatiedialoog starten, in hun hoofd hebben wat ze gaan zeggen. Verder wil meer dan de helft van de respondenten aanvullende visuele reisinformatie, b.v. via teletekst.

5.3 Samenhang vragen uit het eerste en tweede deel van de vragenlijst

In tabel 12 vinden we een overzicht van de correlaties tussen de vragen uit het eerste deel (v1 t/m v21) en het tweede deel van de vragenlijst (v26 t/m v41). We bespreken de vragen die onderling sterk samenhangen.

- De vragen v26, v27 en v28 betreffen de manier waarop het systeem vragen stelt. De antwoorden op deze vragen laten weinig spreiding zien. Vrijwel alle respondenten antwoorden positief. Bijgevolg correleren deze vragen nauwelijks met v1-v21.
- Vraag v29 (systeem heeft initiatief in dialoog) correleert matig met v16 (prettige communicatie).
- Tegen de verwachting in correleert v30 (wat er gezegd kan worden is duidelijk) nauwelijks met v13 (onduidelijke gebruiksaanwijzing).
- Vraag v31 (herhaling) correleert hoog met v8 (leverde gevraagde info), v10 (opnieuw gebruiken), v16 (prettige communicatie), v17 (verbetering vereist) en v20 (begrijpelijk). Blijkbaar oordelen de respondenten die vonden dat ze veel/weinig moesten herhalen negatief respectievelijk positief.
- Vraag v32 (van te voren bedacht wat te zeggen) verschilt inhoudelijk sterk van v1-v21 en correleert nauwelijks met deze vragen.

Tabel 12 Correlatiematrix v26v41-v1v21

	v26	v27	v28	v29	v30	v31	v32	v33	v34	v35	v36	v37	v38	v39	v40	v41
v1	.15	.01	.08	.25	.15	.16	.15	.26	.10	.10	.14	.02	.16	.15	.02	.19
v2	.07	.06	.16	.07	.12	.44	.10	.22	.53	.03	.18	.51	.42	.06	.05	.06
v3	.13	.21	.09	.20	.17	.36	.30	.32	.43	.21	.09	.42	.31	.27	.07	.20
v4	.27	.05	.31	.10	.20	.28	.14	.25	.39	.01	.05	.33	.34	.10	.10	.03
v5	.14	.10	.09	.17	.02	.16	.16	.19	.17	.09	.07	.09	.18	.17	.26	.10
v6	.19	.14	.15	.01	.03	.22	.03	.17	.22	.18	.06	.14	.19	.32	.39	.22
v7	.23	.12	.03	.20	.15	.37	.10	.25	.41	.02	.01	.39	.32	.18	.05	.05
v8	.08	.24	.10	.12	.10	.49	.29	.40	.52	.09	.04	.44	.33	.23	.26	.05
v9	.11	.13	.26	.02	.19	.18	.25	.22	.27	.00	.04	.31	.42	.10	.07	.03
v10	.24	.20	.15	.26	.13	.50	.35	.51	.60	.26	.20	.56	.38	.25	.10	.21
v11	.37	.17	.07	.36	.14	.19	.09	.27	.24	.13	.04	.23	.18	.14	.19	.09
v12	.21	.13	.12	.27	.04	.41	.21	.35	.49	.08	.10	.48	.32	.15	.07	.11
v13	.06	.05	.12	.21	.17	.11	.03	.24	.14	.05	.04	.18	.07	.09	.04	.01
v14	.12	.17	.01	.15	.01	.20	.01	.31	.23	.20	.06	.26	.24	.11	.14	.09
v15	.13	.18	.35	.23	.17	.34	.18	.31	.39	.10	.02	.40	.37	.02	.06	.04
v16	.30	.23	.16	.39	.10	.50	.31	.41	.49	.14	.16	.47	.41	.36	.06	.09
v17	.14	.16	.16	.10	.04	.48	.23	.40	.68	.06	.00	.64	.41	.26	.12	.07
v18	.06	.15	.02	.10	.03	.29	.08	.20	.18	.22	.28	.12	.20	.29	.18	.23
v19	.31	.06	.04	.14	.16	.19	.16	.22	.24	.09	.06	.29	.17	.05	.15	.14
v20	.22	.21	.12	.38	.02	.63	.30	.49	.62	.06	.05	.58	.41	.30	.16	.02
v21	.38	.20	.05	.38	.04	.26	.23	.34	.18	.13	.12	.28	.17	.03	.04	.20

- Vraag v33 (eenvoudig in gebruik) correleert nauwelijks met v13 (onduidelijke gebruiksaanwijzing), maar wel met v8 (leverde gevraagde info), v10 (opnieuw gebruiken), v16 (prettige communicatie), v17 (verbetering vereist) en v20 (begrijpelijk). Een systeem dat wel/niet eenvoudig is in gebruik heeft een hoge/lage appreciatie.
- Vraag v34 (veel fouten) correleert met v2 (storende fouten) en met v8 (leverde gevraagde info), v10 (opnieuw gebruiken), v16 (prettige communicatie), v17 (verbetering vereist) en v20 (begrijpelijk), kortom met de vragen die de uiteindelijke positieve of negatieve appreciatie bepalen.
- De meeste respondenten vonden de herhaling van het systeem (v35) niet vervelend. Door de geringe spreiding van de antwoorden correleert deze vraag nauwelijks met v1-v21.
- Bij vraag v36 (snel spreken) is er eveneens weinig spreiding in de antwoorden.
- Vraag v37 (systeem maakt snel fouten) correleert hoog met v2 (storende fouten), v10 (opnieuw gebruiken), v12 (betrouwbaar), v17 (verbetering vereist) en v20 (begrijpelijk). Kortom met de vragen die de appreciatie bepalen.
- Vraag v38 (eenvoudig te corrigeren) correleert met v2 (storende fouten), v9 (niet flexibel), v10 (opnieuw gebruiken), v16 (prettige communicatie), v17 (verbetering vereist) en v20 (begrijpelijk). Kortom de appreciatie bepalende vragen. Correctie eenvoud is mede bepalend voor de appreciatie.
- Bij vraag v39 (lange gespreksduur) is er matige correlatie met v6 (traag).
- Vraag v40 (wachtijd tijdens gesprek) correleert met v6 (traag).
- Vraag v41 is inhoudelijk sterk afwijkend van v1-v21 en correleert nauwelijks.

We constateren dat de samenhang tussen de vragen uit deel 1 en deel 2 conform onze verwachtingen is. Er waren in ieder geval geen onverwachte hoge respectievelijk lage samenhangende vragen.

6. Onderzoek achtergrondvragen en open vragen van de vragenlijst

Naast de vragen die direkt betrekking hadden op de appreciatie van VIOS bevatte de vragenlijst een aantal algemene vragen met betrekking tot spraakherkenning en vragen die betrekking hadden op karakteristieken van de respondent (achtergrondvragen). En tenslotte bevatte de vragenlijst een aantal open vragen.

6.1 Frequentieverdeling van de vragenscores van de achtergrondvragen

De frequentieverdelingen van de achtergrondvragen zijn al eerder weergegeven in paragraaf 2.2 (figuren 1 t/m 8). Met betrekking tot de persoonskarakteristieken merken we op dat de mannen ruimer vertegenwoordigd waren in de steekproef. Minder dan de helft (45%) was jonger dan 25 jaar en slechts 18% was ouder dan 40 jaar. Bijna de helft heeft een WO-opleiding gevolgd. Opvallend was dat meer dan de helft zelden of nooit met de trein reisde en dat vrijwel iedereen meerdere informatiebronnen raadpleegde bij het opzoeken van reisinformatie.

We constateren dat de helft van de respondenten een voorkeur heeft voor een vrouwenstem. Er was slechts één voorkeurstem voor een mannelijke stem. En 49% had geen voorkeur. Verder wilde ongeveer de helft met 'U' aangesproken worden. Iets meer dan de helft wil niet meer dan f0.50 betalen per minuut. Ruim 40% wilde wel meer betalen, maar niet meer dan een gulden.

6.2 Frequentieverdeling van de vragenscores van de open vragen

We bespreken vervolgens de antwoorden op de open vragen. Bij iedere vraag hebben we de antwoorden verdeeld in een aantal categorieën. De frequentieverdeling van de categorieën wordt hieronder per vraag weergegeven. De antwoorden op de open vragen met bijbehorende categorie staan vermeld in de bijlage. Het frequentie overzicht van de laatste vraag (v48) staat in bijlage H.

♦ Vraag 1a: 'Hoe vond u dit experiment?' (v22)

De meeste respondenten lieten zich in positieve bewoordingen uit over het experiment. Maar het is niet altijd even duidelijk of ze nu het experiment beoordeelden en/of het systeem. Een aantal respondenten maakte een duidelijk onderscheid tussen de beoordeling van het experiment en van het systeem. Bijgevolg ordenen we de antwoorden in een + of - categorie, waarbij, wanneer dat mogelijk is, vooral gelet wordt op de mening over het systeem.

+	leuk, prima, fantastisch, geslaagd, werkte prima, verrast door hoge kwaliteit
+/-	wel leuk, aardig, redelijk
-	waardeloos, niet geslaagd, slecht, het systeem begreep mij niet, systeem is nog lang niet perfect
?	niets ingevuld

In de volgende tabel geven we de frequentieverdeling van de hierboven gedefinieerde antwoordcategorieën.

Tabel 13 Frequentieverdeling vraag 1a

Categorie	Aantal respondenten
+	52
+/-	18
-	19
?	11

Een duidelijke meerderheid (52%) is positief over het experiment. Maar nogmaals het was niet altijd duidelijk of de respondenten het systeem en/of het experiment beoordeelden. Wanneer men 'prima' als reactie gaf, is dat in de '+-categorie' gescoord, maar dat wil niet zeggen dat men positief gestemd is over het systeem.

- ◆ *Vraag 3: 'Hoe staat U tegenover techniek in het algemeen en tegenover computers en spraakherkenning in het bijzonder?'*(v42)

Vrijwel alle respondenten staan positief ten opzichte van techniek, computers en spraakherkenning. We scoren de antwoorden in drie categorieën, waarbij we vooral letten op het standpunt met betrekking tot automatische spraakherkenningssystemen.

+	zonder meer positief
+/-	positief maar met bedenkingen en voorwaarden
-	overwegend negatief, of heel forse bezwaren en bedenkingen en aanvullende voorwaarden.
?	niets ingevuld

In tabel 14 geven we het frequentie overzicht.

Tabel 14 Frequentieverdeling vraag 3

Categorie	Aantal respondenten
+	60
+/-	22
-	14
?	4

Ook bij deze vraag zijn de meeste respondenten positief. De personen die negatief geantwoord hebben, waren dat vooral met betrekking tot het spraakherkenningssysteem. Over techniek in het algemeen waren ze meestal positief.

- ◆ *Vraag 5: 'Waar zou uw voorkeur naar uitgaan, naar telefonistes of automatische spraakherkenningssystemen? Kunt u dit toelichten?'*(v48)

De frequentieverdeling van de voorkeuren van de respondenten staat in tabel 15. De meeste respondenten (44%) kiezen toch voor de telefoniste. Het valt ook op dat eenderde van de respondenten geen voorkeur hebben voor één van de twee.

Tabel 15 Frequentieverdeling vraag 5

Voorkeur	Aantal respondenten
telefoniste	44
ASS	23
geen voorkeur	33

De respondenten kiezen om uiteenlopende redenen voor telefonistes of automatische spraakherkenningssystemen (ASS). We hebben de verschillende redenen gegroepeerd in een aantal categorieën. De categorieën zijn:

- positieve redenen voor telefonistes
- negatieve redenen voor telefonistes
- positieve redenen voor automatische spraakherkenningssystemen
- negatieve redenen voor automatische spraakherkenningssystemen
- redenen waarom men geen voorkeur heeft voor één van de systemen

In onderstaande tabellen geven we een frequentie overzicht van de respectievelijke categorieën. In iedere tabel staat aangegeven welke voorkeur de respondenten hebben. Het aantal ingevulde respondenten komt niet overeen met 100, omdat men meerdere redenen heeft gegeven. Iedere reden die is genoemd is dus gescoord. De laatste kolom ('?') wil zeggen dat men geen voorkeur heeft. Het kan dus voorkomen dat men gekozen heeft voor de telefoniste, maar toch positieve redenen heeft genoemd voor het automatische spraakherkenningssysteem.

Tabel 16 Positieve redenen voor telefoniste

Reden	telefoniste	ASS	?
telefoniste is sneller	9		
telefoniste is goedkoper	2		
telefoniste is persoonlijker	21		1
telefoniste geeft meer info	4		5
telefoniste is vriendelijker/gezelliger	4		
telefoniste kan complexe problemen aan	1		1
telefoniste is begripvoller	5		
telefoniste is makkelijker te onderbreken	3		
telefoniste is flexibeler	9		
telefoniste is effectiever			1
beller kan initiatief overnemen	1		
telefoniste kan fouten herstellen	1		
duidelijk wanneer telefoniste bezig	1		
telefoniste is nauwkeuriger	2		

De meeste respondenten hebben als positieve reden aangegeven dat de telefoniste persoonlijker is. Flexibiliteit en snelheid zijn daarna de meest genoemde redenen waarom men voor de telefoniste kiest. Er waren geen respondenten die voorkeur hadden voor het ASS en toch positieve redenen voor de telefoniste noemden. Van de mensen die geen voorkeur hebben, zijn er vijf die genoemd hebben dat ze het positief vinden dat de telefoniste meer informatie kan verstrekken. Er zijn veel minder negatieve redenen genoemd dan positieve redenen voor de telefoniste. De meeste negatieve redenen zijn gegeven door mensen die de voorkeur geven aan ASS. Twee mensen die voor de telefoniste gekozen hebben, geven als nadeel dat er vaak een grote wachtrij is.

Tabel 17 Negatieve redenen voor telefoniste

Reden	telefoniste	ASS	?
telefoniste is onvriendelijk/ongeduldig		5	
telefoniste heeft grote wachtrij	2	1	2
telefoniste is onpersoonlijk			1
geen behoefte aan live-gesprek		1	

In tabel 18 staan de positieve redenen voor ASS. Het valt op dat de meeste respondenten die de voorkeur geven aan ASS, als reden geven dat ASS sneller en anoniemer is. Ook hier zijn er mensen die de voorkeur aan iets anders geven en toch positieve redenen voor ASS geven. Zoals diegenen die voor de telefoniste kozen geven ook aan dat het ASS

Tabel 18 Positieve redenen voor automatische spraakherkenningssystemen (ASS)

Reden	telefoniste	ASS	?
ASS is sneller		7	1
ASS is goedkoper	2		1
ASS geeft minder tijdsdruk/rustiger		2	
ASS vermindert de wachtrijen	1	2	1
ASS is gestandaardiseerd		4	2
ASS is gemakkelijk in het gebruik		1	1
ASS is anoniem		6	
ASS heeft lagere drempel			1
ASS is altijd beschikbaar		1	1
ASS is effectief			2
meer raadplegingen tegelijk		1	

goedkoper is en de wachtrijen vermindert. In vergelijking met de positieve redenen voor telefonistes zijn er minder respondenten die positieve redenen voor ASS hebben genoemd, maar dit komt ook overeen met de eerder getoonde frequentieverdeling in tabel 15. Er zijn duidelijk meer negatieve redenen voor ASS genoemd dan voor de telefoniste. Het merendeel van de negatieve redenen is gegeven door de respondenten die de voorkeur geven aan de telefoniste.

Tabel 19 Negatieve redenen voor automatische spraakherkenningssystemen (ASS)

negatieve redenen ASS	telefoniste	ASS	?
ASS is niet gebruikersvriendelijk	1		
ASS moeilijk te onderbreken	2		
ASS is minder flexibel	1		1
ASS maakt fouten	6		
ASS is duurder	1		
ASS duurt langer	2		1
ASS is onpersoonlijk			1
ASS geeft minder info	3		
ASS is moeilijk te corrigeren	4		1

In tabel 20 staan de redenen waardoor de voorkeur voor één van de systemen wordt bepaald. De meeste respondenten die geen voorkeur hadden vinden het belangrijkste dat er goede informatie verstrekt wordt. Door middel van welk systeem dit dan gaat maakt ze niets uit.

Tabel 20 Redenen waarom men geen voorkeur heeft voor één van de systemen

overwegingen	telefoniste	ASS	?
voorkeur wordt bepaald door goede info			10
voorkeur wordt bepaald door snelle info			7
voorkeur wordt bepaald door goedkoopste oplossing	2		7
voorkeur wordt door kortste wachttijd			1
liever persoonlijk contact			1
als er maar een back up telefoniste is			1

7. Appreciatiescore

In hoofdstuk 4 werd er verondersteld dat er een dominante factor aan de vragenlijst ten grondslag ligt en corresponderend met deze ene factor een meetschaal waarmee de waardering voor VIOS gemeten kan worden. Uit tabel 8 in paragraaf 5.1.3 kunnen we aflezen dat er inderdaad een dominante factor is. De vragen die het sterkst corresponderen met de factor zijn:

- storende fouten (v2)
- leverde gevraagde info (v8)
- opnieuw gebruiken (v10)
- betrouwbaar (v12)
- prettige communicatie (v16)
- verbetering vereist (v17)
- begrijpelijk (v20)

en in mindere mate:

- meedenken (v3)
- verwarrend in gebruik (v4)
- onprettige dialoog (v7)
- verwarrend (v15)

Gelet op de inhoudelijke betekenis van genoemde vragen kunnen we de dominante factor interpreteren als de appreciatiefactor. Als alle vragen in meer of mindere mate de appreciatie van VIOS meten, kan een betrouwbare maat verkregen worden door de scores van de diverse vragen bij elkaar op te tellen. Tellen we de scores van alle vragen van deel 1 bij elkaar op dan krijgen we een somscore die we interpreteren als de appreciatiescore. Rekening houdend met de oriëntatie van de variabelen is de appreciatiescore als volgt gedefinieerd:

$$\text{appsc} = v1 - v2 + v3 - v4 + v5 - v6 - v7 + v8 - v9 + v10 - v11 + v12 - v13 - v14 - v15 + v16 - v17 + v18 - v19 + v20 - v21$$

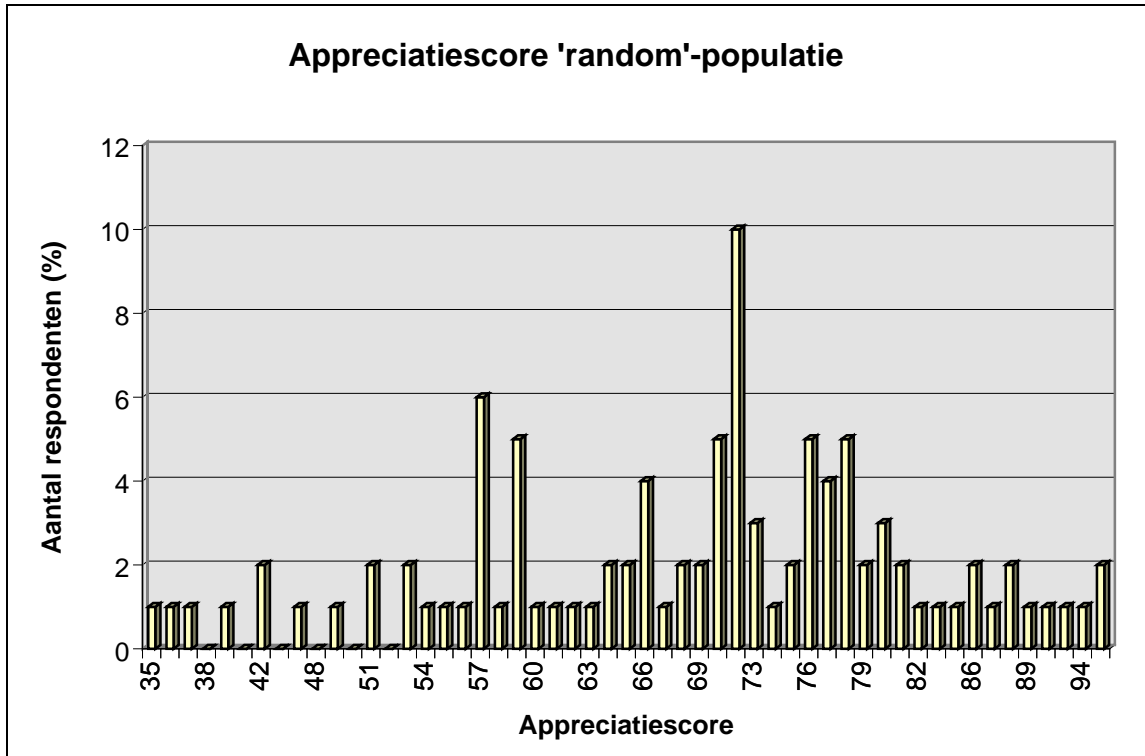
Personen met een lage resp. hoge appreciatiescore blijken inderdaad negatieve resp. positieve meningen, opvattingen en waarderingen te hebben over VIOS. Het blijkt verder dat deze somscore een goede benadering is van de factorscore, waarbij de scores op de respectievelijke vragen met ongelijke gewichtscoefficienten worden opgeteld.

In hoofdstuk 4 presenteerden we ons model van de vragenlijst. De vragen in de categorie 'eindbeoordeling' werden beschouwd als een omvattende appreciatiemaat. Tot onze grote voldoening zijn het juist deze vragen die het sterkst samenhangen met de zojuist gedefinieerde appreciatiemaat (zie figuur 13 paragraaf 5.1.3). Dit resultaat is een bijdrage aan de validatie van de vragenlijst en appreciatiemaat.

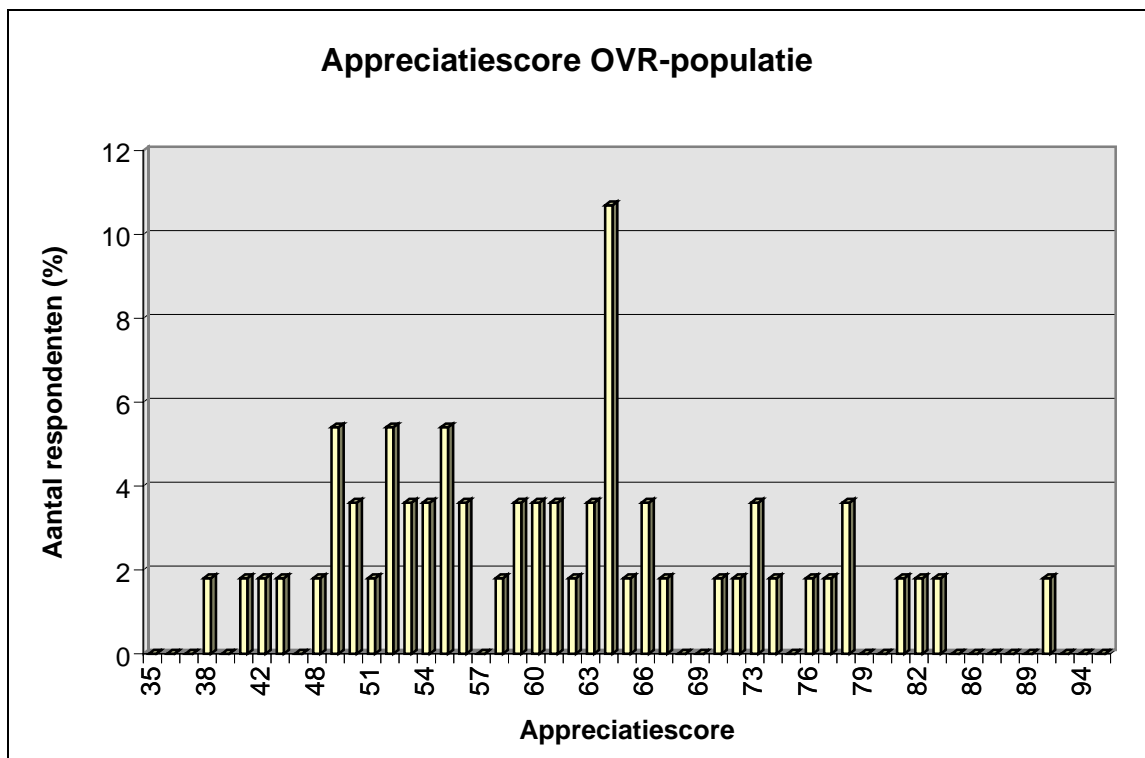
We hebben de appreciatiescore voor alle respondenten berekend. Er geldt: Hoe hoger de appreciatiescore hoe positiever men het systeem waardeert. Het is interessant om de groep met een hoge respectievelijk lage appreciatiescore nader te karakteriseren. We hebben de samenhang onderzocht tussen de appreciatiescore en de achtergrondvragen. Het blijkt dat mannen een iets hogere appreciatiescore hebben dan vrouwen. Jongeren hebben een lagere appreciatiescore dan ouderen.

In figuur 15 zien we dat de appreciatiescore bij benadering een trichotome verdeling heeft met een middengroep van gematigden en een groep van positieve en negatieve appreciatiescoorders. De grenzen tussen de groepen zijn nu nog niet betrouwbaar vast te stellen. Immers de vragen van de enquête werden gescoord op een ordinale en geen absolute schaal. Bijgevolg is ook de appreciatiescore een ordinale schaal.

We hebben ook de appreciatiescore berekend van de respondenten uit de controlegroep.

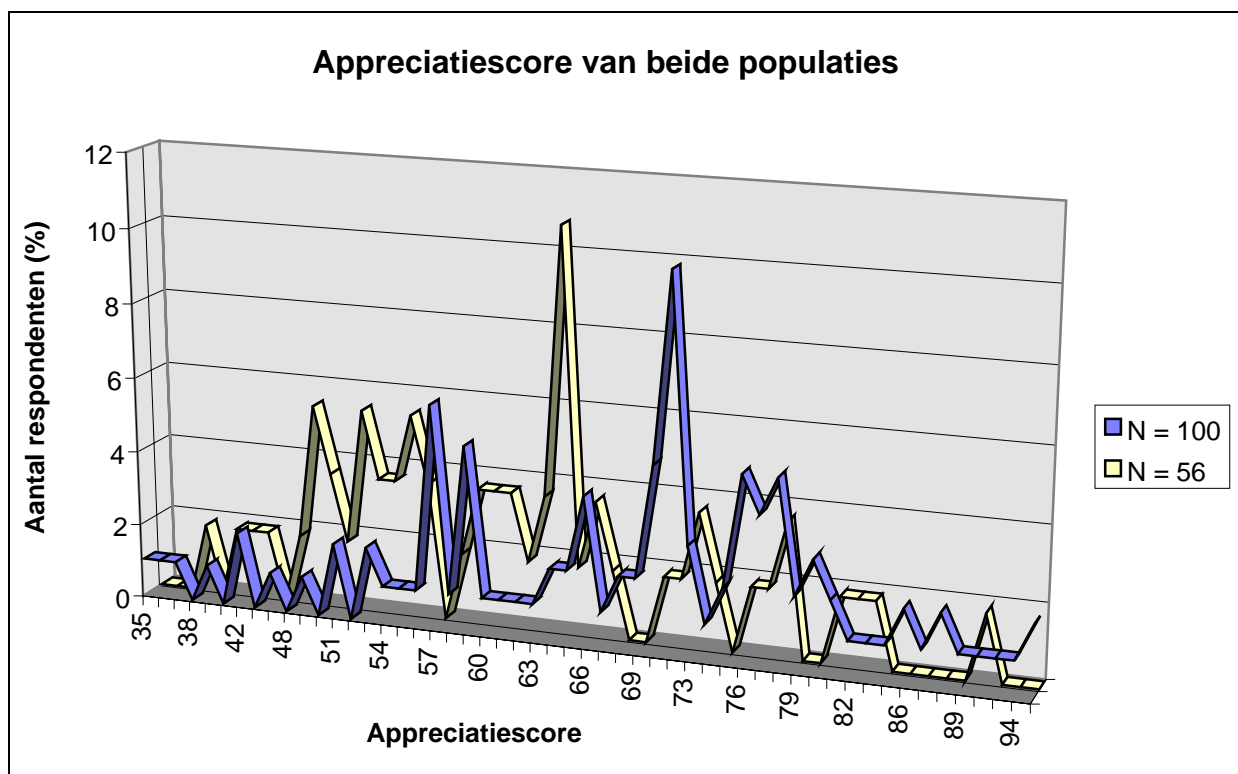


Figuur 15 Verdeling appreciatiescore van 100 respondenten



Figuur 16 Verdeling appreciatiescore van 56 OVR-respondenten

Deze is weergegeven in figuur 16. Opnieuw zien we drie groepen verschijnen van lage, middelmatige en hoge scoorders. Een afwijkend verschil is dat de groep van sterk negatieve scoorders groter is vergeleken met de testgroep. Dit is ook te zien in figuur 17 waar de piek van de OVR-populatie meer naar links ligt (lagere score) dan de piek van de 'random'-populatie. Verder valt het op dat de OVR-populatie als geheel negatiever scoort terwijl de meerderheid bij de 'random'-populatie positief is. Kortom, de testgroep is positiever over het systeem dan de controlegroep.



Figuur 17 Appreciatiescores van OVR- en 'random'-populatie naast elkaar

8. Betrouwbaarheid, homogeniteit en validiteit van de vragenlijst

In dit hoofdstuk wordt de vragenlijst nader bekeken. We gaan ervan uit dat de respondenten de vragenlijst niet willekeurig hebben ingevuld maar dat de antwoorden de meningen en opvattingen van de respondent goed weergeven. Wanneer de respondenten de vragenlijst nog eens zouden invullen, dan verwachten we dat zij dat vrijwel op dezelfde wijze doen. Eventuele afwijkingen worden veroorzaakt door toeval en zijn niet afhankelijk van bijvoorbeeld het latere tijdstip, een andere lokatie of een andere context.

8.1 Betrouwbaarheid

Een vragenlijst is betrouwbaar wanneer de vragen maar op één manier te interpreteren zijn. Om de betrouwbaarheid van de vragenlijst te meten, zou dezelfde vragenlijst op een ander tijdstip of andere plaats nog een keer ingevuld moeten worden door dezelfde respondenten. De correlatie tussen de antwoorden van de test en de hertest is dan een indicatie van de betrouwbaarheid. Hoe hoger de correlatie, hoe betrouwbaarder de vragenlijst.

8.1.1 Betrouwbaarheid van de vragenlijst

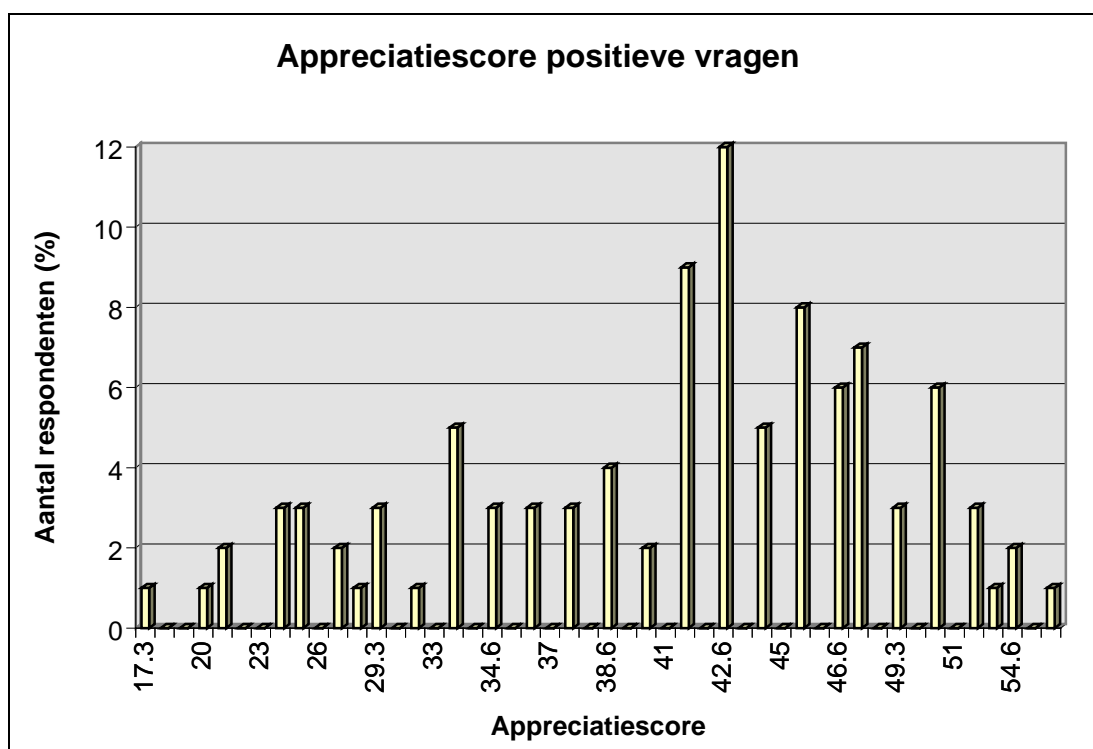
Om de respondenten van dit onderzoek twee keer een (min of meer) gelijke vragenlijst te laten invullen, was onder andere door de anonimiteitseis moeilijk realiseerbaar. Vandaar dat de betrouwbaarheid in dit onderzoek slechts globaal is onderzocht. De betrouwbaarheid is onderzocht door te bepalen of er enige samenhang was tussen min of meer gelijke vragen uit de delen 1 en 2 van de vragenlijst. In paragraaf 5.3 zijn de correlatiescores van die vragen bepaald en is de verwachte samenhang gevonden. De vragenlijst is redelijk betrouwbaar op basis van de vragen die onderzocht zijn. Wanneer de appreciatiescores van alle vragen bij het betrouwbaarheidsonderzoek betrokken hadden kunnen worden zou er een betere betrouwbaarheidsindicatie zijn bereikt. Statistisch gezien is aan te geven dat het meten van de betrouwbaarheid over meerdere antwoorden een beter resultaat geeft. Veronderstellen we dat er een factor aan de vragenlijst ten grondslag ligt, die zich laat interpreteren als een appreciatiescore, dan is iedere afzonderlijke vraag te beschouwen als een appreciatiescore en de somscore een gemiddelde score over alle 21 vragen. De standaarddeviatie wordt hierdoor met een factor $\sqrt{21} = 4.583$ gereduceerd.

8.1.2 Positieve en negatieve vragen

Om te voorkomen dat respondenten alle vragen identiek scoren zijn sommige vragen positief en anderen negatief gesteld met betrekking tot de appreciatiescore. We berekenen de somscore van de positieve en negatieve vragen dat wil zeggen:

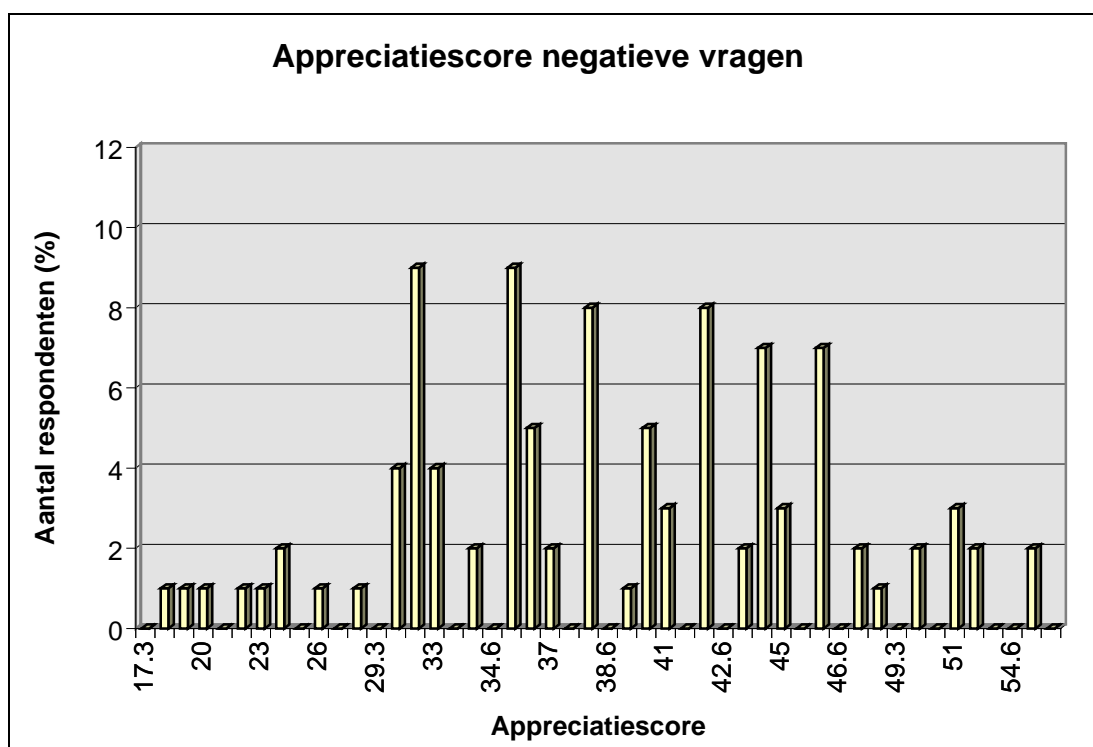
positieve vragen	= v1+v3+v5+v8+v10+v12+v16+v18+v20
negatieve vragen	= v2+v4+v6+v7+v9+v11+v13+v14+v15+v17+v19+v21
appreciatiescore	= positieve vragen - negatieve vragen (zoals in hoofdstuk 7 is vermeld)

In de figuren 18 en 19 zijn de positieve en negatieve vragen met hun appreciatiescore uitgezet. We constateren dat de negatieve vragen meer spreiding te zien geven. Er zijn respondenten die positief dan wel negatief hebben geantwoord op de negatieve vragen. De reden is dat deze vragen de kritische aspecten van VIOS aan de orde stellen. De positieve vragen worden door de meeste respondenten positief benaderd, m.a.w. er zijn minder negatieve uitschieters.



Figuur 18 Appreciatiescores die de respondenten gescoord hebben bij de positieve vragen

In de totale appreciatiescore (figuur 15 in hoofdstuk 7) komt dit verschil tussen de positieve en negatieve vragen niet naar voren. Daar valt vooral op dat men het systeem redelijk positief waardeert. Het feit dat de vraag positief dan wel negatief is gesteld, beïnvloed de appreciatie. Hierdoor is de vragenlijst minder betrouwbaar.



Figuur 19 Appreciatiescores die de respondenten gescoord hebben bij de negatieve vragen

8.2 Homogeniteit

We veronderstellen dat elk van de vragen v1-v21 in meer of mindere mate de gebruikersappreciatie van VIOS scoren. Om het scoren minder afhankelijk te maken van toevalligheden is de somscore bepaald met behulp van alle vragen. De bijdrage van iedere vraag aan de uiteindelijke, totale appreciatiescore varieert. In tabel 21 geven we de samenhang van iedere vraag met de appreciatiescore. Wanneer de correlatie hoog is, is de vraag een maat voor de appreciatie. Wanneer de correlatie laag (dichtbij 0) is, meet de vraag iets anders of zit er te weinig spreiding in de antwoorden op de vraag (aangezien er in de appreciatiescore veel spreiding zit).

Het is te verwachten dat de samenhang hoger wordt als we vragen weglaten die matig correleren met de totale somscore. De vragenlijst wordt dan homogener, maar mogelijk ook minder breed met alle facetten. Een maat voor de homogeniteit van de vragenlijst is Cronbach's Alpha coëfficiënt. Wanneer de coëfficiënt de waarde 1 heeft, is de vragenlijst homogeen. De homogeniteitscoëfficiënt van de vragenlijst gebruikt voor dit onderzoek is '0.9033'. De vragenlijst is dus redelijk homogeen. In tabel 22 geven we voor deel 1 van de vragenlijst de homogeniteitscoëfficiënten wanneer we één van de vragen weglaten. Wanneer we bijvoorbeeld 'v1' weglaten, wordt de waarde van Alpha '0.9046'. Uit de tabel is af te leiden dat het weglaten van vragen leidt tot een geringe verhoging van de Alpha. In dit geval heeft het weglaten van vragen weinig gevolgen voor de homogeniteit van deze vragenlijst.

Tabel 21 Correlatie appreciatiescore

Vraag	Correlatie	Vraag	Correlatie
v1	0.27	v26	0.31
v2	-0.70	v27	0.25
v3	0.58	v28	0.16
v4	-0.66	v29	-0.33
v5	0.45	v30	0.17
v6	-0.46	v31	-0.56
v7	-0.65	v32	-0.29
v8	0.71	v33	0.52
v9	-0.56	v34	-0.63
v10	0.77	v35	-0.16
v11	-0.51	v36	-0.11
v12	0.72	v37	-0.60
v13	-0.32	v38	0.50
v14	-0.50	v39	-0.28
v15	-0.65	v40	-0.20
v16	0.72	v41	0.13
v17	-0.71		
v18	0.22		
v19	-0.59		
v20	0.82		
v21	-0.52		

Tabel 22 Cronbach's Alpha Coëfficiënt

Vraag	Alpha
v1	0.9046
v2	0.8958
v3	0.8991
v4	0.8967
v5	0.9029
v6	0.9016
v7	0.8971
v8	0.8954
v9	0.8998
v10	0.8934
v11	0.9008
v12	0.8952
v13	0.9057
v14	0.9010
v15	0.8972
v16	0.8957
v17	0.8956
v18	0.9066
v19	0.8988
v20	0.8917
v21	0.9006

8.3 Validiteit

Wanneer een vragenlijst meet wat deze verondersteld is te meten, is de vragenlijst valide. In dit geval is de onderliggende factor de VIOS-appreciatie. We veronderstellen dat de gedefinieerde appreciatiescore een betrouwbare maat voor de appreciatie is. Zoals te zien was in tabel 21 is de ene vraag een betere maat voor de appreciatie dan de andere. Maar de vragen meten de appreciatie wat betekent dat de vragenlijst redelijk valide is.

8.3.1 Een gewogen of een ongewogen somscore?

Bij de appreciatiescore worden de vragen simpelweg bij elkaar opgeteld. Een somscore waarbij de 'betere' vragen een hogere wegingsfactor krijgen dan de 'slechte' vragen zou mogelijk tot een 'verbeterde' appreciatiescore leiden en zo tot een meer valide vragenlijst. Berekenen we met Principal-Components Analysis de 1-dimensionale factor oplossing dan constateren we dat de lading van de vragen op deze ene factor aanzienlijk verschillen (zie tabel 8 paragraaf 5.1.3). Dit betekent inderdaad dat vragen met een hoge lading een betere maat zijn voor de appreciatie dan vragen met een lagere lading.

Bij dit onderzoek zijn de gewogen en ongewogen somscore niet met elkaar vergeleken, omdat het niet mogelijk was de data te splitsen (vanwege te weinig data). Maar uit de literatuur is bekend dat de gewogen somscore de ongewogen score nauwelijks verbeterd.

8.3.2 Dubbelzinnige vragen

In een vragenlijst kunnen dubbelzinnige vragen voorkomen, die verkeerd begrepen kunnen worden door de respondenten. Omdat je niet weet of je enige waarde kunt hechten aan deze vragen, wordt de vragenlijst minder betrouwbaar. Deze vragen dragen dan niet bij aan de validiteit van de vragenlijst. Bij de controlegroep, de OVR-populatie, is bij iedere gesloten vraag (v1-v21) om toelichting en commentaar gevraagd. Uit de commentaren blijkt of de respondenten de vragen goed begrepen en geïnterpreteerd hebben. Bij de 'random'-populatie is niet om toelichting en commentaar gevraagd. Dus een extra evaluatie van deel 1 van de vragenlijst is niet mogelijk. Deel 2 van de vragenlijst is opgesteld door de onderzoeker zelf. De onderzoeker heeft de gesloten vragen (v26-v41) daarom zelf op basis van observeringen en opmerkingen van respondenten geëvalueerd.

OVR-populatie

Bij de meeste vragen worden er bepaalde termen gebruikt zoals 'vriendelijk', 'een storende fout', 'efficiënt', 'verwarrend', 'traag', 'prettig' of 'flexibel'. Iedere respondent zal deze termen hetzelfde interpreteren, maar het gewicht wat men aan de term hangt is voor iedere respondent verschillend. De ene respondent zal bijvoorbeeld al bij één gemaakte fout aangeven dat het systeem storende fouten maakt, terwijl de ander dat pas bij 16 gemaakte fouten doet. Bijgevolg moet de schaal als een ordinale schaal en niet als een metrische schaal geïnterpreteerd worden. Maar dit heeft geen grote invloed op de validiteit, want het blijft de mening van de respondent. Het verkeerd begrijpen van de vraag wel. Bij het bekijken van de commentaren bleek dat de meeste respondenten de vragen goed begrepen hebben. Toch zijn er een paar vragen die op meerdere manieren geïnterpreteerd kunnen worden.

Bij *Dit systeem kwam vriendelijk en beleefd over (v1)* is het niet helemaal duidelijk wat er als vriendelijk en beleefd beoordeeld moet worden. Uit de commentaren blijkt dat hiermee de stem en de dialoog (met welke bewoording geeft het systeem antwoord) bedoeld kunnen worden. Dit geldt ook voor *Dit systeem drong mij een onprettige dialoog op (v7)*. Men beoordeelt de dialoog op woordgebruik of stemgebruik.

Ook bij *Dit systeem was verwarrend in het gebruik (v4)* is er een dubbelzinnigheid. Reageert het systeem verwarrend of is het verwarrend voor de gebruiker wat te doen?

Een derde voorbeeld is *Dit systeem was te traag (v6)*. Men heeft het 'oplezen van de verbinding' of 'de dialoog als geheel' op traagheid beoordeeld. Dit staat lijnrecht tegenover elkaar. Men vindt het systeem traag wanneer de dialoog te lang duurt, maar te snel wanneer de verbinding wordt opgelezen.

Ook *Ik ben tevreden over de mogelijkheden tot herhaling (v18)* kan gezien worden vanuit twee standpunten. Gaat het om het herhalen van de verbinding door het systeem of het herhalen van de antwoorden van de gebruiker (het corrigeren)? Deze vraag is in de vragenlijst van dit onderzoek al anders gesteld, namelijk als volgt: *Ik vond het nuttig, dat ik de verbinding nog een keer kon horen*. Hierdoor wordt het meteen duidelijk welke herhaling er bedoeld wordt.

Het viel op dat één persoon nogal geïrriteerd reageerde op een aantal vragen. Door irritatie kan een vraag ook verkeerd begrepen worden. Bij deze persoon was dat niet het geval. De irritatie was niet ontstaan vanwege de vragen, want uit de commentaren bleek dat diegene in het algemeen niet tevreden was over het systeem.

'Random'-populatie

Ook hier is er bij de meeste vragen sprake van termen waarvan de gewichten door iedere respondent persoonlijk geïnterpreteerd worden. Het gaat hier om de termen 'prettig', 'logisch', 'duidelijk', 'veel', 'eenvoudig', 'veel fouten', 'vervelend', 'snel', 'snel fouten' en 'lang'. De enige vraag waarbij is opgemerkt dat die dubbelzinnig opgevat kan worden is *Is het voor u duidelijk wat u tegen het systeem kan zeggen? (v30)*. Men weet vaak wel *wat* er geantwoord kan worden, maar niet welke woorden (*hoe*) ervoor gebruikt kunnen worden. De andere vragen zijn volgens de onderzoeker niet van twee kanten te interpreteren.

Wat de respondenten wel aangegeven hebben over de vragenlijst is dat men het vervelend en verwarrend vond dat er positieve en negatieve vragen door elkaar gesteld werden. Bv. *Het systeem is niet flexibel*. Wanneer het woordje 'niet' weggelaten zou worden zou de vraag al beter te begrijpen zijn. Dit zou een verklaring kunnen zijn voor het eerder genoemde verschil tussen de positieve en negatieve vragen. De antwoordmogelijkheden in de vorm van 'volledig mee eens' tot 'volledig mee oneens' maken het er ook niet makkelijker op bij negatief gestelde vragen. Verder werd er door een aantal respondenten opgemerkt dat er geen verschil is tussen 'volledig mee eens' en 'mee eens'. Zij vonden een 5-puntsschaal niet nodig.

9. Evaluatie van de interviews

Aan 20 van de 100 respondenten zijn nog extra vragen gesteld in de vorm van een interview. In dit hoofdstuk worden de vragen van het interview geëvalueerd. Aan de eerste 7 personen zijn 15 vragen gesteld. Daarna bleek uit de resultaten dat er behoefte was aan een zestiende vraag, welke om een meer directe conclusie vraagt. In de tweede paragraaf zijn de opmerkingen die tijdens het interview genoemd zijn op een rijtje gezet. En de laatste paragraaf behandelt de conclusie.

9.1 Scoreverdeling van de vragen

In deze paragraaf zullen we laten zien wat de scoreverdeling per vraag is en weergeven wat de respondenten geantwoord hebben. Per vraag zijn de antwoorden ingedeeld in een aantal categorieën en deze zijn aangegeven in de bijbehorende tabellen. Bij de vragen 7 en 8 is geen scoreverdeling gegeven omdat deze vragen niet om een mening vragen, maar om positieve en negatieve ervaringen en deze zijn zo verschillend per persoon dat hier geen eenduidige score aan te hangen is.

1) *Vindt u een spraakherkenningssysteem prettig om mee te werken?*

N = 20	Ja	Nee	geen voorkeur
aantal respondenten	13	2	5

De meerderheid vindt spraakherkenningssystemen prettig om mee te werken. Twee personen vinden dat niet omdat de ene liever direct contact met mensen heeft en de ander vindt een telefoniste fijner onder meer omdat zij meer informatie kan verstrekken en eventueel kan doorverwijzen. Redenen waarom men het wel prettig vindt zijn:

- je krijgt geen chagrijnige mensen aan de telefoon
- het is handig
- het is niet zo persoonlijk
- je wordt niet steeds doorverbonden van de ene telefoniste naar de andere
- geen wachtrij
- efficiënter
- je kan anoniem bellen en fouten maken

Maar men vindt wel dat het systeem optimaal moet werken en verbeterd moet worden voordat men er gebruik van zal maken. Iemand gaf aan het prettiger te vinden wanneer er een normaler gesprek gevoerd zou kunnen worden. Ook is men van mening dat een spraakherkenningssysteem vooral handig is voor het opvragen van informatie en dat het een goede aanvulling is op 06-diensten.

Er zijn ook personen die aangegeven hebben dat het niet uitmaakt welk systeem ze gebruiken, als ze de informatie maar krijgen. Deze staan in de tabel onder het kopje 'geen voorkeur'. Belangrijke factoren die bij de keuze meespelen zijn snelheid, kosten en wachttijd.

2) Zou u het handig vinden om een spraakherkenningssysteem te gebruiken voor andere informatie? Zo ja, welke informatie?

N = 20	Ja	Nee	als het moet
aantal respondenten	17	2	1

Zeventien van de twintig personen vinden het handig om een spraakherkenningssysteem te gebruiken voor andere informatie. Enkele suggesties die genoemd zijn, zijn de volgende:

- opvragen van telefoonnummers
- informatie over films en bioscopen
- informatie over vliegreizen
- opvragen van prijzen van tweedehands auto's
- informatie over wintersportgebieden
- informatie over het weer
- informatie over concerten
- informatie over hypotheek
- informatie over dagjes uit
- gebruiken van de computer als tolk
- opvragen van postcodes
- als aanvulling zodat je terwijl je wacht toch al informatie kunt krijgen
- informatie over theatervoorstellingen of algemene kaartverkoop
- i.p.v. systemen waar je de telefoontoetsen moet gebruiken om te communiceren

Maar men vindt dan wel dat de informatie die gegeven wordt algemeen en afgebakend moet zijn zodat je van te voren precies weet wat je als beller kunt verwachten en kunt vragen. En het mag natuurlijk niet te lang duren. Er is ook een persoon die er op zich geen problemen mee heeft, maar uiteindelijk toch liever met een telefoniste belt.

De twee personen die het niet handig vinden hebben liever menselijk contact en vinden dat het systeem op dit moment nog teveel beperkingen heeft. De persoon die genoemd staat onder het kopje 'als het moet' zou alleen met een spraakherkenningssysteem bellen wanneer er niks anders beschikbaar is.

3) Vindt u de telefoon een handig medium of prefereert u een ander medium zoals teletekst of een computerprogramma?

N = 20	Telefoon	Ander medium	geen voorkeur
aantal respondenten	9	9	2

Hier zijn de meningen gelijk verdeeld. Men kiest voor de telefoon omdat deze voor iedereen snel beschikbaar is, doordat de meeste mensen er een in huis hebben. Verder is de telefoon handig, snel en gemakkelijk. De meeste personen die een ander medium prefereren vinden de computer het makkelijkste en het snelste vanwege het visuele aspect, maar je moet er dan wel een in huis hebben. Enkele personen vinden het vervelend om te bellen en kiezen daarom voor iets anders. De twee personen die geen echte voorkeur hadden laten het vooral afhangen van de situatie waarin ze zich bevinden en welke middelen ze daar tot hun beschikking hebben.

4) Ziet u dit systeem als een mens/telefoniste?

N = 20	mens	computer	allebei
aantal respondenten	3	16	1

De meeste personen zien het systeem als een computer. Er is veelvuldig geantwoord dat het goed te horen is aan de stem, aan bepaalde klanken en aan de woorden doordat het spreken hikkelig gaat, dezelfde intonatie wordt gebruikt en stukken tekst duidelijk hoorbaar aan elkaar worden geplakt. Een andere reden was dat men van te voren al wist dat men met een computer zou praten en daar heeft men dan rekening mee gehouden. Men vindt de stem wel vriendelijk klinken, maar het is goed te horen dat het een computer is. De overige 4 personen weten ook dat het een computer is, maar hebben meer het gevoel dat er een mens achter zit omdat het ook ingesproken is door een mens.

5) Vindt u het vervelend wanneer de telefoniste u niet zou verstaan?

N = 20	Ja	Nee	allebei
aantal respondenten	14	4	2

Zoals te verwachten was, vindt de meerderheid het vervelend wanneer de telefoniste ze niet verstaat. De meningen onder deze 14 personen zijn verdeeld. De een tolereert meer fouten van de telefoniste en de ander meer van een computer. Men tolereert meer van de telefoniste omdat zij zich makkelijker aanpast en je krijgt uiteindelijk toch de informatie. Bij de computer loop je tegen een muur aan. Ook kun je bij de telefoniste reageren op de fouten en haar corrigeren. En als je maar duidelijk praat verstaat ze je uiteindelijk wel. De personen die een hogere tolerantie ten opzichte van het systeem hebben vinden dat omdat het voor hun maar een machine is en die mag fouten maken dit in tegenstelling tot een telefoniste die het gewoon goed moet doen. Maar de computer mag ook maar een beperkt aantal fouten maken. Wanneer er teveel fouten worden gemaakt door de computer en je niets anders meer kan doen dan de hoorn op de haak gooien is dit behoorlijk irritant.

Bij de mensen die aangeven dat ze het niet vervelend vinden wanneer de telefoniste ze niet verstaat ligt de irritatiegrens lager. Zij geven ongeveer dezelfde redenen als hierboven genoemd alleen voor hen is het niet 'vervelend' maar 'minder leuk'. De twee overige personen dulden geen storende fouten van beide systemen, maar vinden wel dat de telefoniste minder snel fouten maakt en dat de interactie beter is.

6) Vindt u het systeem gebruiksvriendelijk?

N = 20	Ja	Nee	redelijk
aantal respondenten	5	9	6

De meningen hierover zijn verdeeld. De personen waarbij het telefoongesprek goed verliep zijn positief over de gebruiksvriendelijkheid. Ze vinden het systeem makkelijk om mee te werken omdat er duidelijke vragen gesteld worden waardoor het meteen duidelijk is wat er van je gevraagd wordt. En je kunt ophangen wanneer je wilt.

Diegenen die het systeem niet gebruiksvriendelijk vinden, zijn van mening dat het de goede kant opgaat maar dat er nog aan gesleuteld moet worden. Er kleven nog teveel mankementen aan zoals het niet kunnen onderbreken, niet flexibel genoeg en moeilijk te corrigeren. En wanneer het systeem fouten maakt, kun je erg vast komen te zitten en kun je vaak niets anders dan de hoorn neerleggen en opnieuw bellen.

Zij die er geen uitgesproken positieve of negatieve mening op na houden vinden het systeem geduldig en handig voor een simpele reis, maar daarentegen is het erg moeilijk te corrigeren, weet men niet goed wat er gezegd kan worden en wanneer er afgeweken wordt van de normale weg kan het systeem niet meer zo vriendelijk zijn. De herhaling van de informatie en het lang kunnen blijven telefoneren wordt weer als prettig ervaren.

**7) Wat zijn uw positieve ervaringen met het systeem? of
Wat vond u goed aan het systeem?**

Positieve aspecten die werden genoemd:

- bespaart arbeid
- alles wordt herhaald
- 24 uur per dag bereikbaar
- systeem is erg geduldig
- herhaling van de verbinding
- kunt blijven doorvragen desnoods de hele dag
- mogelijkheid om eerstvolgende verbinding te vragen
- het systeem vertelt je wat je wilt weten
- systeem is snel
- je hoeft niet te wachten; geen wachtenden
- vriendelijk
- prettige en duidelijke stem
- niet langdradig
- duidelijke vragen
- geeft de informatie die je hebben wilt
- gevoel dat het systeem je begrijpt

**8) Wat zijn uw negatieve ervaringen met het systeem? of
Wat vond u slecht of vervelend aan het systeem?**

De negatieve ervaringen die genoemd zijn, zijn de volgende:

- kunt niet onderbreken
- kunt niet corrigeren. Als je vast zit, zit je vast.
- verstaat de stationsnamen niet goed
- lastig dat de verbinding wordt opgesomd en dat je de verbinding niet kan zien bv. op teletekst
- weet niet wat je tegen het systeem kan zeggen
- kan geen eerdere verbindingen vragen
- veel fouten
- ontbreekt een helpfunctie
- er moeten meer opties in
- duurt te lang
- niet gebruiksvriendelijk
- mogelijkheid om terug te kunnen springen naar het begin
- vervangt werk

- traag bij oplezen van verbinding
- weet niet wat het systeem kan
- dezelfde zinnen met dezelfde intonatie
- geen vraag over 'aankomen'
- gevoel dat het systeem je niet begrijpt
- wanneer je de stationsnamen niet kent, kan het systeem je niet helpen.

9) Had u snel door hoe het systeem werkt?

N = 20	Ja	Nee	gedeeltelijk
aantal respondenten	18	1	1

Deze verdeling laat zeer duidelijk zien dat men weet hoe er met het systeem gewerkt moet worden. Men vindt het duidelijk en niet moeilijk. Het wijst zich eigenlijk vanzelf, omdat het systeem meteen het initiatief neemt door een vraag te stellen waar je dan op moet antwoorden en dat bevestigt het systeem meteen weer. Wat niet duidelijk is, is dat men niet weet wat er tegen het systeem gezegd kan worden, hoe je het systeem kan corrigeren en of je kunt onderbreken. Ook is men van mening dat het scheelt wanneer men ervaring met de reisplanner en het gewone informatienummer heeft omdat hierbij dezelfde gegevens gevraagd worden.

Er is een persoon die de basisdingen wel begrijpelijk vindt, maar het niet duidelijk vindt wat er allemaal met het systeem gedaan kan worden. En bij ingewikkelde vragen is verbeteren moeilijk. De laatste persoon staat huiverig tegenover het werken met een computer en was daardoor van te voren al negatief.

10) Is het voor u duidelijk wanneer u tegen het systeem kan spreken?

N = 20	Ja	Nee
aantal respondenten	19	1

Volgens deze verdeling is het duidelijk wanneer tegen het systeem gesproken kan worden. De meesten vonden het logisch omdat gewoon antwoord gegeven kan worden wanneer de vraag gesteld is en je weet dat, wanneer je netjes opgevoed bent, je niet mag praten wanneer de ander aan het woord is. Omdat het zo duidelijk is wanneer je kunt praten is er geen pieptoon nodig. Diegene die het als enige oneens was met de rest hoort niet zo goed waardoor er langzamer gereageerd werd en er gepraat werd terwijl het systeem ook bezig was waardoor er weer meer herhaald moest worden.

11) Verstond het systeem u goed? Zo nee, vond u het vervelend dat u veel moest herhalen?

N = 20	Ja	Nee	redelijk
aantal respondenten	7	6	7

Uit deze cijfers blijkt dat het systeem eenderde van de respondenten goed verstaan heeft. Deze personen hoefden niets tot weinig te herhalen. Een paar hielden er rekening mee dat ze met een computer zouden bellen en hebben daarom extra duidelijk gesproken en dat is van grote invloed of het systeem je kan verstaan.

Aangezien sommige mensen zowel goed als slecht verstaan werden door het systeem kwalificeerden ze het als 'redelijk'. Ze hebben niet veel hoeven te herhalen want ze werden over het algemeen wel verstaan. Ook hier waren er mensen die er rekening mee hielden dat ze met een computer zouden praten en hebben daarom rustiger en langzamer gepraat.

De personen die 'nee' hebben geantwoord verschillen van mening in hoeverre het systeem hun verstond. De meningen verschillen van 'waardeloos' en 'heel slecht' tot 'net iets te weinig'. Iemand gaf zichzelf zelfs de schuld voor het veel moeten herhalen omdat deze persoon het systeem op de mogelijkheden uitprobeerde. Men vindt het moeten herhalen wel vervelend en frustrerend.

12) Luistert u naar de vraag en geeft u precies antwoord of probeert u maar wat? of Wanneer het systeem u er niet naar vraagt zegt u het ook niet?

N = 20	Geantwoord	Uitgeprobeerd
aantal respondenten	15	5

De meerderheid geeft precies antwoord op de vraag omdat ze niet weten wat het systeem begrijpt en wat je tegen het systeem kan zeggen. Ook is niet duidelijk of je volzinnen kunt gebruiken. Verder denken de meesten er niet bij na en gaan er niet van uit dat het systeem meer kan. Sommige passen hun antwoorden zelfs aan de vragen van het systeem aan. Ze hadden bijvoorbeeld een bepaalde aankomsttijd in gedachten en veranderen dit in vertrektijd omdat het systeem niet naar de aankomsttijd vraagt.

De personen die wat dingen uitgeprobeerd hebben, hebben dat gedaan omdat dat in de scenario's gevraagd werd of omdat ze het systeem wilden uitproberen. Ze gingen ervan uit dat het systeem dezelfde informatie kan geven als een telefoniste. Maar ook voor hen was het een probleem dat niet duidelijk was wat er tegen het systeem gezegd kan worden. Je weet van te voren niet wat je opties zijn en dat wil je wel graag weten omdat je weet dat een computer beperkingen heeft. Een aantal probeerden wat 'woordjes' uit zoals 'stop' en 'help' om te proberen of het systeem daarmee te corrigeren was.

13) Vindt u de stem vervelend?

N = 20	fijne stem	gaat wel
aantal respondenten	16	4

De meeste personen vinden de stem niet vervelend, maar juist heel prettig, duidelijk en vriendelijk. Het is een standaard telefoonstem. Wat ze ook is opgevallen, is dat je duidelijk kunt horen dat het uit stukjes is opgebouwd, waardoor de stem mechanischer klinkt. Het zou mooier zijn wanneer het geheel wat vloeiender dus zonder hakken of stoten zou klinken. Men prefereert een mensenstem boven een computerstem. En dan het liefst een aardige en vriendelijke vrouwenstem. De andere vier vinden de stem niet echt vervelend maar ook niet altijd aangenaam. Ze vinden de stem koel overkomen waardoor de stem af en toe wat minder vriendelijk klinkt. Het kan wat vriendelijker gevraagd worden. In het algemeen is men tevreden over de stem.

14) Zou u zelf andere vragen prefereren? Alles in één vraag of vindt u het juist fijn dat de vragen in stukjes zijn opgedeeld?

N = 20	Stukjes	Één vraag	allebei
aantal respondenten	12	6	2

Diegenen die kiezen voor de huidige opzet, de vraag opgedeeld in stukjes, vinden dit makkelijker, duidelijker, gestructureerder, helder en overzichtelijker. Op deze manier weet je wat je moet zeggen en vergeet je niet bepaalde informatie te zeggen. Het verkleint de kans op fouten, omdat je gestuurd wordt. Het is handig wanneer het systeem naar de informatie vraagt, want anders moet je teveel informatie onthouden.

Het grootste voordeel van alles achter elkaar te kunnen spuien is dat het sneller is. Het systeem moet dan wel alles herhalen en om de ontbrekende informatie vragen. Men vindt het ook een handige en efficiënte manier, omdat dit door hen ook bij de telefoniste wordt gedaan en omdat je de informatie die je wilt zeggen toch wel weet. Dit in tegenstelling tot de anderen die vinden dat je bepaalde informatie snel kunt vergeten. Dat het mogelijk is om alles tegelijk te zeggen moet je dan van te voren wel weten.

Er was ook nog het idee om de keuze aan de gebruiker over te laten. Bijvoorbeeld wanneer men het systeem lang niet gebruikt heeft is het handig dat je gestuurd wordt en wanneer je vaak belt ken je het systeem en kun je efficiënter werken door de informatie in één keer in te geven. Als het dan toch niet blijkt te werken kun je nog altijd overschakelen naar de gestructureerde opbouw.

15) Zou u liever de controle hebben zoals bij een telefoniste?

N = 20	Ja	Nee	gemengd	geen voorkeur
aantal respondenten	4	11	2	3

De meeste respondenten geven er de voorkeur aan om het initiatief bij het systeem te laten, want dat is duidelijker en het is makkelijk wanneer je gestuurd wordt en het systeem kunt volgen. Men verwacht ook dat je vragen moet beantwoorden omdat het een automatisch systeem is en het versterkt het gevoel dat er wordt meegedacht door het systeem. Sommigen denken dat het fout gaat wanneer de controle bij de beller zou liggen.

Snelheid is een belangrijke factor voor de personen die graag zelf het initiatief hebben. Men wil graag meteen alle informatie kwijt zodat de computer enkele stappen over kan slaan en om de ontbrekende informatie kan vragen. Voor degenen die 'geen voorkeur' hebben is het vooral belangrijk dat men de informatie krijgt en wie dan de controle heeft doet er niet toe.

Twee mensen geven aan dat er de mogelijkheid moet zijn om de gebruiker meer vrijheid te geven en tegelijkertijd het systeem gestructureerd te laten werken.

16) Wat is uw algemene oordeel: Positief of negatief?

N = 13	Positief	Negatief	allebei
aantal respondenten	7	3	3

Alle respondenten zijn van mening dat het systeem nu nog niet gebruiksklaar is, maar ze zien er wel een toekomst in. Ze geven alleen een ander oordeel. De een noemt het 'positief' en de ander 'negatief'. Degenen die positief gestemd zijn, oordelen mild omdat het systeem in een testfase zit. Wanneer ze het systeem zouden beoordelen zoals het nu werkt zouden

ze negatief zijn. Er zijn er ook die het systeem gewoon goed vinden, omdat het een snelle en efficiënte manier is. De mensen die niet overwegend positief of negatief zijn vinden de technische kant positief en interessant, maar negatief dat het systeem nu nog veel fouten maakt. Ook ziet men het systeem meer als aanvulling op de bestaande dienst en niet als vervanging. Bij de negatieve personen geldt hetzelfde. Zij vinden het systeem nu nog niet goed genoeg, maar zijn wel positief voor de toekomst.

9.2 Opmerkingen

Er zijn ook een aantal opmerkingen tijdens het interview gegeven door de respondenten om het systeem te verbeteren. De volgende opmerkingen zijn er gedaan:

- meer bedenktijd
- willen weten wat de opties zijn.
- meerdere opties en alternatieven mogelijk zijn zoals 'via' stations, meerdere oplossingen van een verbinding (later/eerder) en de mogelijkheid om naar de telefoniste te gaan
- vervelend dat aankomsttijd niet aan te geven is
- het is mooier en professioneler wanneer de zinnen zonder hakken of stoten worden gezegd.
- Kiest de computer het kortste (in km) of het snelste (in min.) traject?
- er zou een duidelijke correctiemogelijkheid in moeten zitten
- het zou beter zijn wanneer er een prijs per gesprek of per verbinding gehanteerd wordt i.p.v. per minuut
- de opgevraagde informatie per e-mail thuis krijgen of op teletekst laten zien zonder dat de telefoonverbinding moet blijven bestaan
- systeem aanpassen aan de regio waar het telefoontje vandaan komt zodat de daar in de buurt liggende plaatsen als eerste als stationsnamen gebruikt worden
- het kunnen zeggen van plaatsnamen i.p.v. de precieze stationsnamen zodat de computer het dichtstbijzijnde station selecteert
- mogelijkheid om het inleidend praatje over te kunnen slaan en meteen het gesprek te kunnen beginnen
- wanneer het systeem je drie keer niet verstaan heeft zou het je automatisch door moeten verbinden met de telefoniste
- mate van systeembegeleiding aan gebruiker overlaten (keuze uit een expert- of een beginners-interface)
- keuze aan de beller overlaten of de verbinding snel of langzaam wordt voorgelezen
- mogelijkheid om het systeem te kunnen onderbreken met behulp van een bepaald commando en daarmee bijv. naar het begin terug te kunnen gaan.
- mogelijkheid tot online-help die op ieder moment opvraagbaar is

Zoals uit de opmerkingen blijkt, zijn er genoeg ideeën bij de mensen om het systeem gebruikersvriendelijker te maken.

9.3 Conclusie

Helaas was het niet mogelijk om de vragenlijst en het interview per persoon te vergelijken, doordat het onderzoek anoniem was opgezet. Er is wel gekeken of er in de frequentieverdelingen van overeenkomende vragen van de vragenlijst en het interview significante verschillen of nieuwe verrassende afwijkingen voorkomen. Deze zijn niet geconstateerd. Alle overeenkomende vragen zijn op dezelfde manier beantwoord. Dus je kunt je afvragen of een aanvullend interview dan wel zin heeft. Misschien was alleen een vragenlijst ook wel genoeg geweest. Bovendien zijn er door de meeste respondenten veel opmerkingen bij de enquête geplaatst waar ook veel informatie uit te halen is.

Maar tijdens het interview zijn er weer andere ideeën (zie 9.2) naar voren gekomen en de antwoorden zijn beter toegelicht. Bovendien zijn er vragen gesteld die niet in de vragenlijst voorkomen. Ook is gebleken dat men positief is voor de toekomst wat betreft spraakherkenningssystemen, maar dat men op dit moment nog erg negatief is. Terwijl uit de appreciatiescore van de vragenlijst blijkt dat men toch redelijk positief is. Vanwege deze verschillende uitslag is de laatste vraag (*Wat is uw algemene oordeel: Positief of negatief?*) ook noodzakelijk gebleken. Zo is een beter beeld ontstaan van de algemene waardering voor het systeem.

Een ander voordeel van een interview is dat de vragen op de juiste manier geïnterpreteerd worden doordat de interviewer de vragen kan toelichten. Op deze manier zijn de vragen in ieder geval duidelijk voor de respondent. Volgens de onderzoeker is het interview, gezien de resultaten, nuttig gebleken.

10. Samenvatting resultaten en aanbevelingen

In dit hoofdstuk worden de resultaten nog even kort op een rijtje gezet en zullen de doelstellingen van het onderzoek, waaronder ook de aanbevelingen, besproken worden.

10.1 Algemeen

Het is mogelijk gebleken een surveystudie uit te voeren waarbij met behulp van vragenlijsten en aanvullende interviews de appreciatie van een automatisch spraakherkenningssysteem, genaamd VIOS, bepaald is. Het systeem VIOS is ontwikkeld in opdracht van 'Openbaar Vervoer Reisinformatie (OVR)' en er kan automatisch informatie over vertrek- en aankomsttijden van treinen mee verkregen worden. De surveystudie is uitgevoerd onder 100 personen die afkomstig zijn uit de familie- en kennissenkring van de onderzoeker. In totaal hebben 100 van de 112 personen meegewerkt aan het onderzoek. De respondenten zijn persoonlijk benaderd door de onderzoeker en zijn tijdens de onderzoeksperiode aangespoord te reageren. Bij analyse bleek dat de respondenten de vragenlijst zorgvuldig hebben ingevuld, ook de open vragen werden beantwoord en er werden veel aanvullende opmerkingen geplaatst. Onder de respondenten bleken zowel voor- als tegenstanders van automatische spraakherkenningssystemen aanwezig te zijn. Dit zorgde onder andere voor voldoende spreiding van de antwoorden. Verder is uit het onderzoek naar voren gekomen dat de volgende factoren een belangrijke rol spelen bij het bepalen van de appreciatie voor het systeem:

- Maakt het systeem weinig storende fouten?
- Levert het systeem de gevraagde informatie?
- Zal de gebruiker het systeem opnieuw gebruiken?
- Is het systeem betrouwbaar?
- Is de communicatie met het systeem prettig?
- Moet het systeem sterk verbeterd worden?
- Begrijpt het systeem de gebruiker?
- Moet de gebruiker weinig herhalen?
- Is het systeem eenvoudig in het gebruik?
- Is het systeem eenvoudig te corrigeren?

10.2 Onderzoeksvragen

In deze paragraaf zullen de onderzoeksvragen, die in de 'Inleiding' geformuleerd zijn, besproken worden.

In welke mate wordt VIOS geapprecieerd door de respondenten?

Volgens de resultaten oordeelt 40% positief en 30% negatief over VIOS. Dat de meerderheid van de respondenten positief is, is ook naar voren gekomen in de appreciatiescore. De dialoog en het gebruikersgemak worden positief gewaardeerd. Toch geeft minder dan de helft (43%) aan dat ze de dienst zullen gaan gebruiken met daarbij de voorwaarde dat het systeem dan optimaal werkt. Het is voor het merendeel (58%) heel duidelijk dat er veel aan het systeem verbeterd moet worden. Vandaar dat op dit moment 44% kiest voor de telefoniste, met als belangrijke redenen dat een telefoniste persoonlijker is en zij meer informatie kan geven. Maar het systeem heeft toch wel voordelen, zoals de 24-uurs bereikbaarheid, je belt anoniem en er is geen wachtrij. Men ziet het systeem duidelijk als een aanvulling en niet als vervanging van de telefoniste. Maar het is duidelijk dat de meeste respondenten in spraakherkenningssystemen wel een toekomst zien.

Wat zijn de voornaamste wensen, klachten en problemen van de gebruikers?

De voornaamste klachten en wensen die uit dit onderzoek naar voren zijn gekomen, zijn de volgende:

- ☞ Het systeem maakt te veel storende fouten.
- ☞ Het systeem kan niet worden onderbroken.
- ☞ Het systeem is moeilijk te corrigeren.
- ☞ Het systeem is niet flexibel.
- ☞ Men weet niet hoe te reageren wanneer het systeem gegevens foutief bevestigt en tegelijkertijd weer een nieuwe vraag stelt.
- ☞ Het is van te voren niet duidelijk wat het systeem allemaal kan.
- ☞ Men wil graag meer mogelijkheden/opties.
- ☞ Het systeem vraagt alleen naar de vertrektijd en niet naar de aankomsttijd.
- ☞ Het is niet duidelijk wat er tegen het systeem gezegd kan worden, d.w.z. welke woorden of zinnen het systeem begrijpt.
- ☞ Er kan geen eerdere verbinding gevraagd worden zonder de hele dialoog opnieuw te voeren.
- ☞ Het systeem kan je niet helpen wanneer je geen stationsnamen weet.

Voorstellen tot verbetering

- ☞ Betere interruptiemogelijkheid bv. met behulp van een commando.
- ☞ Makkelijker kunnen corrigeren van fouten.
- ☞ Online-help die op ieder moment opvraagbaar is.
- ☞ Van te voren weten wat de mogelijkheden van het systeem zijn.
- ☞ Meerdere opties en alternatieven inbouwen, zoals 'via' stations.
- ☞ Het systeem laten vragen of men op een bepaalde tijd wil *vertrekken* of *aankomen*.
- ☞ De mogelijkheid om naar het begin terug te kunnen springen.
- ☞ Het in kunnen geven van plaatsnamen zodat het systeem het bijbehorende station neemt.
- ☞ Wanneer het gesprek bv. drie keer fout gegaan is, de beller automatisch doorverbinden met de telefoniste.
- ☞ Eventuele visuele presentatie van de verbinding op teletekst.
- ☞ Tijdens het wachten op de telefoniste alvast het spraakherkenningssysteem kunnen raadplegen.

Hoe wordt het systeem gewaardeerd door de OVR-populatie?

Aan de appreciatiescore is te zien dat het systeem door de 'at random' gekozen populatie positiever gewaardeerd wordt dan door de OVR-populatie. Een reden voor het negatievere oordeel van de OVR-populatie zou kunnen zijn dat het automatische spraakherkenningssysteem een bedreiging voor de werkgelegenheid kan vormen.

De vragenlijst

Uit het onderzoek komt naar voren dat de vragenlijst voor dit onderzoek redelijk betrouwbaar, homogeen en valide is gebleken.

Literatuurlijst

- [Pijpers96] Pijpers, P., "*VIOS: een reisinformatie systeem nader beluisterd*", Alparon, TU Delft, 1996, 96-09.
- [Manintveld97] Manintveld, W.A.Th., Rothkrantz, L.J.M., "*The OVR-WOz experiment: Setup and Analysis*", Alparon, TU Delft, 1997, 97-04.
- [SPSS92] Norušis, M.J., "*SPSS for Windows Release 5.0*", SPSS Inc., 1992