

Spraaktechnologie

Verslag van een worsteling

Jan Hemel j.b.hemel@rc.rug.nl

Moderne computerapparatuur is dermate snel en krachtig dat allerlei vormen van spraaktechnologie binnen bereik van de gewone gebruiker komen. Dit biedt niet alleen perspectieven voor visueel gehandicapten en mensen met een muisarm. We kunnen nu onze computer toespreken, maar luistert die ook?

Met het toenemen van de kracht van computers voor persoonlijk gebruik groeit ook het aantal toepassingen dat een zware claim legt op de processor, het geheugen en de schijfomvang. De doorbraak van grafische gebruikersinterfaces zoals Windows is daarvan al een uiting, maar vooral ook het toenemend gebruik van multimedia-applicaties. Het is dan ook niet vreemd dat ook spraaktechnologie zijn intrede doet op de pc. Deze ontwikkeling wordt nog extra bespoedigd door de snelle toename van RSI (muisarm) onder computergebruikers.



Spraaktechnologie

Al geruime tijd worden computers gebruikt om spraak te **genereren** op basis van geschreven tekst. Een primitief voorbeeld hiervan is de gesproken weergave van de juiste tijd via de telefoon. Hierbij is geen sprake van genereren van klanken, maar wordt door de computer slechts een combinatie samengesteld van vooraf ingesproken tijd-teksten.

Interessanter is de toepassing voor visueel gehandicapten waarbij teksten die op het scherm van de computer staan worden voorgelezen door een synthetische stem. Zwakke plek hierbij is de intonatie die doorgaans erg vlak is en daardoor vermoeiend en moeilijk verstaanbaar. Door de applicatie begrip bij te brengen van de betekenis van een zin kan momenteel al enige nuance in de uitspraak worden bereikt. Interpunctie, bijvoorbeeld, doet wonderen:

?Je bent niet lekker!? wordt anders gelezen dan ?Je bent niet lekker??

Een ander probleem vormt het afwisselend gebruik van verschillende talen: in een Nederlandse tekst zullen minder gangbare Engelse woorden op z'n Nederlands worden uitgesproken.

Toepassingen

Spraak**herkenning** kan gebruikt worden voor verschillende doelen: **identificatie** van een onbekende spreker lijkt voorlopig nog (lang) niet binnen bereik, maar **authenticatie**, controleren of de spreker is die hij zegt te zijn, is al wel mogelijk.

Het hoofdprobleem daarbij is de grote mate van natuurlijke variatie van iemands stem, die behalve van de stemming bijvoorbeeld ook van eet- en drinkgedrag en gezondheidstoestand afhankelijk is. Het zal niet lang meer duren of toepassingen hiervan voor het openen van deuren en waarborgen van gesproken bestellingen kunnen hun intrede doen.

Dit artikel echter is gewijd aan spraakherkenning voor het **dicteren** van teksten en het **bedienen** van een computer.

De techniek

Zelfs met optimale apparatuur is de beste programmatuur niet in staat tot bruikbare herkenning van losse woorden en klanken. Het maximaal bereikbare is ongeveer 70% herkenning, vergelijkbaar met de prestaties van een mens, die voor een groot deel van de herkenning ook afhankelijk is van de context, de volgorde van de klanken. Door de waarschijnlijkheid van het voorkomen van een woord mee te nemen in de herkenning kan dit percentage vergroot worden. Moderne spraakherkenningsprogramma's gebruiken frequentiegegevens die afhankelijk zijn van het soort teksten dat de spreker gebruikt, en zelfs van de context in de zin: voorafgaand aan een zelfstandig naamwoord is een lidwoord het waarschijnlijkst, maar ook een bijvoeglijk naamwoord is daar gangbaar.

Naarmate een herkenningsprogramma een groter deel van de zin bij de herkenning betreft zal het trager worden en dus meer computerkracht vereisen. Al met al scoort deze aanpak op een zware pc rond de 95%

herkenning. Dit is voldoende om bruikbaar te zijn.

Nu reeds zijn ontwikkelingen gaande waarbij door middel van een camera ook de gezichtsuitdrukking wordt gevolgd. Hiermee komt herkenning van rond de 99% ?in zicht?. Ook zonder deze laatste toevoeging melden geoefende gebruikers nu al herkenning in de buurt van de 99%. Er is wel zorgvuldigheid en doorzettingsvermogen nodig om dit te bereiken.

Producten

Voor consumenten komen de volgende serieuze spraakherkenningsprogramma's in aanmerking:

- **Dragon Naturally Speaking** van Lernout & Hauspie, onder meer toegepast door Corel in WordPerfect Office.
- **Freespeech** van Philips
- **ViaVoice** van IBM
- **Voice Xpress** van Lernout & Hauspie, toegepast door Microsoft in Office XP

Uit recensies van nieuwsgroepen en elders op het internet ontstaat de indruk dat Freespeech is achtergeraakt op de concurrentie. De beide producten van Lernout & Hauspie zijn al enige tijd in een haasje-over-spel gewikkeld, waarbij de laatst uitgebrachte versie telkens marginaal beter is dan de concurrent.

Een hard voordeel van (het nauwelijks uit te spreken) Dragon Naturally Speaking is de mogelijkheid relatief eenvoudig van taal te wisselen; prijs en verkrijgbaarheid zijn echter wat minder gunstig dan die van Voice Xpress. Overigens is de toekomst van beide producten onzeker, daar Lernout & Hauspie in surseance van betaling verkeert.

Internationaal wordt ViaVoice als het beste pakket geprezen. Dit programma kan vooralsnog niet met Nederlands overweg.

In veel recensies worden de producten vergeleken op basis van de resultaten direct na installatie. Mij is gebleken dat een dergelijke vergelijking zinloos is, omdat de herkenning in hoge mate verbetert nadat spreker en programma grondig aan elkaar gewend zijn. We moeten hier denken aan tientallen uren nauwgezet inspreken en corrigeren. Dit vereist zeker doorzetten. Dit artikel geeft dan ook geen vergelijking, maar een onafhankelijk verslag van het werken met Voice Xpress 5.

Voice Xpress wordt geleverd met een bruikbare headset. U wordt geacht te beschikken over een pc op tenminste 266 MHz met 96 MB RAM-geheugen en 705 MB vrije schijfruimte, en natuurlijk een geluidskaart. Mijn ervaring is dat dit zeer krap is bemeten. Op een pc met Windows NT of Windows 2000 moeten we eerder denken aan 256 MB of 512 MB RAM-geheugen om prettig te kunnen werken, en een Pentium II op 350 MHz blijkt te traag om een hoge herkenningssprezisie te kunnen instellen.

Trainen

Na het installeren en afregelen van positie en gevoeligheid van de microfoon, moet het programma wennen aan je stem. Hiervoor moet je circa een kwartier lang een tekst voorlezen van het scherm. Voor betere resultaten kun je dit enkele malen herhalen. Om onduidelijke redenen mag dit niet vaker dan zesmaal. Vervolgens vraagt het programma om teksten die je eerder hebt getypt. Deze gebruikt het om uw vocabulaire te leren kennen. Vervolgens wordt u uitgenodigd de onbekende woorden stuk voor stuk uit te spreken, opdat het programma ze leert kennen.

Dan kunt u beginnen met dicteren. Voice Xpress maakt automatisch onderscheid tussen gedicteerde tekst en commando's, en herkent ook uitgesproken interpunctie (zoals *punt* en *vraagteken*). Natuurlijk gaat het niet altijd goed. Dan kunt u kiezen tussen corrigeren met gesproken commando's, of even buiten Voice Xpress om wat sjoemelen met het toetsenbord. Het laatste is meestal het snelst, maar als je Voice Xpress wijst op herkenningfouten, zal dat het resultaat voor de toekomst verbeteren. Het vergt oefening om tussen deze twee correctietechnieken een juist evenwicht te vinden.

Een gebruiker krijgt op den duur een zekere feeling voor problematische woorden en constructies en zal deze vermijden, of ze typen in plaats van dicteren..

Het kost ook enige gewenningstijd om de zinnen in je hoofd zo goed voor te bereiden dat je ze in een keer foutloos kunt uitspreken. Ook moet je gevoel krijgen voor de ideale lengte van de zinsfragmenten die je zonder onderbreking dicteert. Hoe langer de fragmenten, hoe sneller je werkt, maar hoe meer tijd ook verloren gaat met benodigde correcties.

Resultaten

Voice Xpress Professional kan bestaande teksten voorlezen met een vriendelijke damesstem. De intonatie wordt hierbij enigszins aangepast aan de interpunctie; niettemin gaat de eentonigheid snel vervelen ook al is

het voorlezen goed te verstaan. Voor mij persoonlijk heb ik voor deze functie nog geen toepassingen kunnen ontdekken.

Algemeen wordt geadviseerd je eerst toe te leggen op het dicteren, en pas in een later stadium de bediening van de computer ook met de stem te doen. Dan komen extra voorzieningen als het opnemen van macro's en het kennen van geavanceerde bedieningsopdrachten van Office goed van pas. Deze snufjes heb ik (na drie maanden) nog nauwelijks gebruikt.

In het algemeen is de herkenning na de gewenningsperiode goed en snel voor beschrijvende teksten met weinig opmaak. De handleiding claimt een maximum van 160 woorden per minuut. Of ik dit gehaald heb weet ik niet, maar het programma kan vast concurreren met mijn typesnelheid.

Het anti-RSI-programma Workspace heeft niet door dat ik niet zelf typ, en vindt de 'type'snelheid' onaanvaardbaar groot.

Speciale teksten zoals programmeercode, po^ㅍㅍe, en tekst met afwisseling van taal, geven meer problemen.

Irritant vind ik het filter waarmee bijgeluiden ('uh', 'hm') van de herkenning worden uitgezonderd. Door dit filter hebben alle woorden met 'uh'- en 'ui'-klanken het moeilijk. Het lijkt niet uit te zetten.

Ook stoorde mij het automatisch afkorten of herschrijven van woorden: *bijvoorbeeld* wordt *b.v.*, *een* wordt nogal eens *1* en *vijf gulden* verandert Voice Xpress in *NLG 5,-*. De mogelijkheden om deze aanpassingen in te stellen functioneren onvoldoende.

Ook waren wij het vaak oneens over het los of aaneen schrijven van woorden. Maar dat ben ik het ook wel met anderen.

Spraakherkenning gebruikt veel geheugen. Voice Xpress raakte de kluts kwijt als het geheugen anders werd ingedeeld doordat ik applicaties sloot en andere opende. In een nieuwsgroep vond ik herhaalde malen het advies de pc opnieuw te starten na surfen op het internet. Misschien speelt dit minder bij meer geheugen; ik heb 256 MB.

Mensen die, zoals ik, de neiging hebben te mummelen, worden sterk gestimuleerd beter te articuleren. Een nuttig bijeffect. Helaas wordt monotoner spreken ook bevorderd.

Conclusie

Spraakherkenning is uiteindelijk een prettig en productieverhogend hulpmiddel voor het invoeren van teksten, dat zeker kan bijdragen aan de bestrijding van RSI.

Om er plezier van te hebben zijn een geleerde werkomgeving en een zware pc noodzakelijk evenals een forse investering in het trainen van het programma en het wennen aan de bediening. De ingevoerde tekst heeft nog wel zorgvuldige controle nodig.

> Waarschuwing

Deze tekst is gemaakt met behulp van spraakherkenningssoftware. Mocht de tekst onbegrijpelijk lijken dan kunt u het volgende doel:

- lees het bericht hardop voor niemand anders;
- zorg dat u hierbij niet in de wacht ziet;
- hoe meer vooral niet te begrijpen wat een lijst;
- vragen aan de Tour de wat denkt het u bedoelt.

Link

www.avt-lh.nl

▲ Begin pagina ▲

[index](#) Pictogram 4