FALLBEISPIELE DER SPRACHEINGABE

H. Mutschler,

Beratungsbüro für Ergonomische Fragen, Linkenheim

1. Einleitung

Die Spracheingabe (SPE) zum Aufruf von Funktionen technischer Geräte oder zur Dateneingabe in einen Rechner hat gegenüber manuellen Eingabeverfahren bestimmte ergonomische Vorteile, denen jedoch technologische Nachteile der Spracherkennung gegenüberstehen. Ein wesentlicher Vorteil der SPE besteht darin, den manuell-visuellen Kanal für andere, gleichzeitig durchzuführende Tätigkeiten freizuhalten, wie es in den Abschnitten 2. - 4. an Beispielen im Bereich des Warenrückflusses, im Kraftfahrzeug sowie bei der militärischen Aufklärung dargestellt wird. Ebenso ist die SPE nützlich, wenn der manuelle oder visuelle Kanal des Menschen durch körperliche Behinderungen gestört ist (Abschnitt 5.). Bei den dargestellten Anwendungen handelt es sich teilweise um Konzepte (2., 3.), teilweise um Erprobungssysteme (4., 5.). Beim Aufruf von Funktionen (3., 5.) oder bei der Eingabe weniger, einzelner Daten (3., 4.) genügt der Einsatz von Einzelworterkennern, bei der Eingabe von mehrziffrigen Zahlen ist aus ergonomischer Sicht eine Wortkettenerkennung notwendig (2.).

2. Spracheingabe beim Warenrückfluß

In einer Großhandlung für pharmazeutische Produkte werden 80000 Artikel über 6 Zweigstellen und 1400 Lieferanten eingekauft und vertrieben. Das Einlagern und die Entnahme der Artikel im Lager erfolgt großenteils automatisch. Die Kunden, d.h. die Apotheken, haben als Serviceleistung der Großhandlung die Möglichkeit, bei falschen Bestellungen oder Überschreiten des Verfalldatums die Waren zurückzugeben, was zu einem Warenrückfluß von 2% der verkauften Waren führt. Diese Waren werden in der Retourenabteilung entgegengenommen, auf Wiederverwendbarkeit kontrolliert und von Hand registriert. Hierzu werden der Lieferschein und die Kundenangaben mit den Waren verglichen, ggfs. korrigiert und mit der Mengenangabe und einem Bewertungsschlüssel versehen. In einem zweiten Arbeitsgang werden an einem anderen Arbeitsplatz die Daten der Retourenabteilung mit dem Rechnungsdurchschlag verglichen und als Eingabe in den Rechner erfaßt. Dazu werden Rechnungsnummer, Menge, Artikelnummer und der Bewertungsschlüssel eingegeben.

Beide Arbeitsgänge sollten nun in einem Arbeitsschritt zusammengefaßt werden, um die doppelte Datenerfassung zu vermeiden. Wegen der manuellen Beanspruchung beim Auspacken und Wenden der Waren mit gleichzeitiger Ablesung und Kontrolle sollte geprüft werden, ob statt einer Tastatur ein Spracherkennungsgerät zur Erkennung von Wortketten verwendet werden kann.

Es wurde folgende Gestaltung der SPE vorgeschlagen: Die zurückgegebenen Waren werden mit Hilfe der SPE über die sukzessive Eingabe der hierarchisch gegliederten Kategorien (Kunden-, Rechnungs- und Artikelnummern) und der zugehörigen Daten erfaßt, wobei pro Kategorie mehrere Subkategorien und mehrere Daten möglich sind. Die Kategorien (Kundennummer etc.) und der Bewertungsschlüssel (z.B. F=Fehlen) werden in Textform, die numerischen Daten (1-7-stellige Nummern) werden als Zahlen eingesprochen. Die gängigen Artikel können auch in der vertrauten Textform eingegeben werden, wenn ein Spracherkennungssystem mit einem entsprechend großen Wortschatz eingesetzt wird. Die Gruppierung der numerischen Daten in 2er-Gruppen erfolgt in Anlehnung an menschliche Stereotypen sowie zur Optimierung der Spracherkennung in 2er-Gruppen. Eine ungeradzahlige Anzahl von Ziffern sowie Nullen zwingen zu einer Aussprache mit 1er- oder 3er-Gruppen. Alle Eingaben werden zur Protokollierung und zur Orientierung des Sprechers in Form eines Dialogprotokolls optisch und persistent auf einem Bildschirm rückgemeldet, wobei bearbeitete Subkategorien lediglich summarisch angegeben werden und die Übersetzung der Nummern in Textform und umgekehrt automatisch erfolgt. Beispiel:

```
Kundennummer    a) 12 34 56  Müller, Karlsruhe    (2 Rechnungen)
Kundennummer    b) 98 76 54  Maier, Ludwigshafen
   Rechnungsnummer  a)  50 56 82                  (3 Artikel)
   Rechnungsnummer  b)  28 65 05
     Artikelnummer       a) 1x Fl Lavandulae Dac 17 90 98 4  P
     Artikelnummer       b) 2x Gelonida Na Tabl  22 40 57 6  F
```

Im Sinne einer flexiblen Syntax kann von jedem Punkt bei der Abarbeitung eines Lieferscheins jeder andere direkt erreicht werden. Das Lokalisieren von Kategorien und Daten im Dialogprotokoll kann wahlweise verbal über die alphabetische Kennzeichnung ("Anton" etc.) oder manuell mit Hilfe eines berührempfindlichen Bildschirmes erfolgen. Die Eingabe einer Date, einer Kategorie oder eines Lieferscheins werden explizit mit "Ende" abgeschlossen. Nach Abschluß eines Dialogschrittes wird die standardmäßig folgende Subkategorie voreingestellt, z.B. nach Abschluß der Artikelnummer die Menge. Datenanzahl, Datenlänge (1-7-stellig), Datenart (Text, numerisch) und soweit möglich Dateninhalt (mit Prüfziffer) werden unter Einbeziehung von pragmatischem Wissen automatisch kontrolliert. Eingabefehler werden mit einem Signalton angezeigt.

3. Spracheingabe in einem Kraftfahrzeug-Navigationssystem

Zukünftige Kfz-Navigationssysteme unterstützen den Kraftfahrer bei der Wahl seiner Fahrtroute, z.B. bei Verkehrsstau. Bei der Benutzung solcher Systeme macht der Fahrer auch während der Fahrt Eingaben, z.B. Positionsangaben oder Abruf gezielter Verkehrsinformationen. Während Tastaturen wegen der Auge-Hand-Koordination stets einen Teil der visuellen Aufmerksamkeit beanspruchen, ermöglicht die SPE in Verbindung mit der Sprachausgabe eine kontinuierliche Beobachtung des Straßenverkehrs. Hierzu wurde einer Automobilzulieferfirma der Einsatz eines Spracherkennungssystemes zur sprecherunabhängigen Erkennung von Einzelwörtern mit einem Geräuschunterdrückungsverfahren vorgeschlagen.

Zunächst wurde der phonetische Gehalt des vorgesehenen Vokabulars analysiert, um die zu erwartende Erkennungssicherheit abschätzen zu können. (3 Vokabulargruppen: Ziffern und "Zehn"; Buchstaben; Navigationsbegriffe und Kommandos, z.B. Autobahn, Eingabe etc.). Dabei wurden die Wörter in die Lautschrift übersetzt und ihr phonetischer Abstand in Anlehnung an das Spracherkennungsverfahren des Dynamic Time Warping (DTW) paarweise berechnet. Abstände zwischen Phonemen ergeben sich durch Addition ihrer unterscheidenden Merkmale, die z.B. bei den Konsonanten durch Artikulationsart, Artikulationsort und Stimmhaftigkeit bestimmt werden. Z.B. unterscheiden sich die plosiven Phoneme B und K hinsichtlich Artikulationsort (bilabial - velar) und Stimmhaftigkeit (stimmhaft - stimmlos), was einen phonetischen Abstand von 2 ergibt. Auf dieser Grundlage ergibt sich der phonetische Abstand zwischen (gleichlangen) Wörtern aus der Summe der phonetischen Abstände der gleichpositionierten Phoneme. Bei unterschiedlich langen Wörtern werden die Vokale einander als "Anker" zugeordnet, z.B. ergibt das Wortpaar "Korrektur" und "Bundesstraße" eine Distanzsumme von 14 (Vokale erhalten doppeltes Gewicht, Auslassungen erhalten "Strafpunkte").

Auf diese Weise wurden die Phonemdistanzen der Vokabulargruppen berechnet, wobei der mittlere Abstand bei den Ziffern 6.1, bei den Buchstaben 2.8 und der Navigationsbegriffe bzw. Kommandos 8.3 betrug. Verbesserungen der Vokabulargestaltung haben also bei den Buchstaben anzusetzen, z.B. in Form des phonetischen Alphabets "Anton".., und bei den einsilbigen, ähnlichen Wortpaaren, z.B. Starten und Uhrzeit statt Start und Zeit.

4. Spracheingabe bei einer Überwachungsaufgabe

Bei der RPV (Remotely Piloted Vehicle) - Aufklärung fliegt ein unbemannter Flugkörper mit elektrooptischen Sensoren über Gelände mit feindlichen, taktischen Zielen und überträgt das Bildmaterial zu einer Bodenstation. Dort untersucht ein Beobachter das Bildmaterial auf einem Bildschirm in Echtzeit und kennzeichnet die Ziele. Da es sich also um eine intensive visuelle Überwachungsaufgabe bei gleichzeitigen Eingaben in einen Rechner handelt, liegt die Anwendung der SPE nahe und es wurde ein experimenteller Vergleich von manuellen und gesprochenen Eingaben durchgeführt. Bei der manuellen Variante wurden Ort und Richtung der Ziele (entsprechend der Windrose) mit einem Lichtgriffel auf dem Bildschirm gekennzeichnet und das Gelände (Feld etc.) sowie Zahl und Typ der Ziele über eine Funktionstastatur eingegeben, im Falle der SPE wurden alle 5 Kategorien eingesprochen. Die Versuchspersonen wurden mit Einblendungen der aktuellen Eingabekategorien sowie mit Rückmeldungen der Eingaben auf dem Bildschirm geführt. Die implizite Einzelverifizierung sah lediglich das Korrekturkommando "Falsch" vor, ansonsten wurde eine richtige SPE angenommen, wenn die nächste Eingabe erfolgte. Das Spracherkennungssystem war sprecherabhängig und erkannte Einzelwörter.

In einem ersten Versuch ohne Zeitbegrenzung wurde bei den räumlichen Informationen (Ort, Richtung) der Lichtgriffel von den allermeisten Versuchspersonen bevorzugt, da die SPE eine

Umcodierung in Koordinaten bzw. in die Himmelsrichtungen der Windrose erforderte und daher erheblich mehr Eingabefehler verursachte. Bei den leicht verbalisierbaren, z.T. gegenständlichen Informationen (Gelände, Zielzahl, -typ) hingegen bevorzugten 2/3 der Versuchspersonen die SPE, obwohl sie insbesondere bei den vielsilbigen Zieltypen und der Antwortzeit des Spracherkennungssystems eine bis doppelt so lange Eingabezeit benötigte. Nur 0.5% der Eingaben wurden falsch erkannt, so daß nur wenig Korrekturen notwendig waren.

In einem zweiten Versuch wurde mit einer Darbietungszeit der Bilder von 5 s der Zeitdruck bei der RPV-Aufklärung simuliert. Hier wurde mit Mehrheit auch bei den gegenständlichen Informationen die Tastatur bevorzugt, da keine Zeit zur Korrektur von Erkennungsfehlern blieb. Selbst bei einer zusätzlichen manuellen Einstellaufgabe zur Bildverbesserung bestand bei Zeitbegrenzung eine geteilte Präferenz zwischen manueller und gesprochener Eingabe.

5. Spracheingabe als technische Hilfe für Behinderte

In der BRD gibt es ca. 40000 motilitätsbehinderte Personen, die ihre Arme bzw. Hände nicht oder nur unkoordiniert bewegen können. Dieser Personenkreis muß zur Steuerung technischer Geräte andere Körperteile heranziehen, wie z.B. Fuß, Kinn, Augen, Zunge etc. Ist die Sprechfähigkeit nicht wesentlich eingeschränkt, ist die SPE allen anderen Hilfsmittel hinsichtlich der Informationsmenge pro Zeiteinheit überlegen. Zur Steuerung von Fernsehgeräten durch hochgradig Querschnittsgelähmte (Tetraplegiker) in einer Rehabilitationsklinik wurde eine Sprachsteuerung entwickelt und erprobt. Hierzu wurde ein preiswertes sprecherabhängiges SPE-Gerät zur Erkennung von Einzelwörtern in einem PC eingesetzt. Zur Benutzung des Systems nicht nur vom Bett, sondern auch vom Rollstuhl aus wurden die erkannten Kommandos in Schaltsignale für eine handelsübliche Fernbedienungseinheit umgesetzt. Das Mikrofon war an einem Stativ befestigt, das zu Beginn des Fernsehens vom Pflegepersonals an den Mund des Behinderten positioniert werden mußte.

Es konnten alle Funktionen des Fernsehgerätes (Lautstärke, Programme etc.) sowie des Spracherkennungssystems (Lernphase etc.) über eine starre Syntax gesteuert werden, wobei wertekontinuierliche Funktionen, z.B. Helligkeit, durch diskrete, vordefinierte Sprünge verändert werden konnten. An den kritischen Stellen des Dialogs, z.B. zum Einschalten oder bei der Funktionswahl, wurde als Aktivierungskommando ein "SSSSSS" gewählt, um Fehlalarme z.B. bei lautem Schließen der Türe zu vermeiden.

Die Fernsehsteuerung wurde von drei jungen, frisch operierten Tetraplegikern während einer Woche abwechselnd vom Bett aus benutzt und bewertet. Alle Personen beurteilten sie als nützlich und bedauerten, daß sie nach Ende der Erprobung wieder abgebaut werden mußte. Als besonders wertvoll wurde die selbständige Benutzbarkeit des Fernsehgerätes genannt. Die insbesondere bei einer Person mit geringer Vitalkapazität häufigen Fehlerkennungen wurden jedoch als störend empfunden.