SPRACHSYNTHESE-ANWENDUNGEN

Ulrich Kordon

Technische Universität Dresden

Für die praktische Anwendung von Sprachsynthesetechnik bietet der Bereich der Behindertenhilfstechnik, vor allem für Blinde oder hochgradig Sehbehinderte, günstige Voraussetzungen. So stellt ein Hilfsmittel mit Sprachausgabe oft die einzige Möglichkeit für den Betroffenen dar, Zugang zu bestimmten Informationen zu erlangen. Dadurch wird häufig eine Akzeptanzschwelle für die Qualität der synthetischen Sprache erreicht, die den Einsatz von low-cost-systemen erlaubt, und damit andererseits bestimmte technische Lösungen von konkreten Hilfsmitteln mit Sprachausgabe überhaupt erst ermöglicht.

Umfragen (z. B. [1]) haben ergeben, daß etwa 60 % der Betroffenen Sprachausgaben befürworten, während sich nur ca. 10 % ausdrücklich gegen synthetische Sprache bei der Informationsübermittlung aussprachen (ca. 30 % keine eindeutige Aussage). Als besonders günstig wurde Sprachsynthesetechnik bei Hilfsmitteln im Bürobereich (Bürorechentechnik, Schreibtechnik) sowie für Alltagshilfen (Thermometer, Uhr, Waage usw.) eingestuft.

An der Technischen Universität Dresden wird seit 1985 an der Realisierung von Blindenhilfsmitteln mit Sprachausgabe gearbeitet, wobei eine enge Kooperation mit dem Blinden- und Sehschwachenverband besteht. Die Arbeiten sind einerseits als Impulse für entsprechende Aktivitäten im industriellen Bereich gedacht, andererseits dienen sie zur praktischen Absicherung grundlegender konzeptioneller Vorstellungen zur Blindenhilfstechnik.

Die nachfolgenden Beispiele sollen den dabei erreichten Stand verdeutlichen. Für die Sprachausgaben wird entweder auf einen 4-Formant-Reihenfiltersynthetisator oder auf einen CVSD-Decoder (bei kleineren, festen Vokabular) zurückgegriffen [2]. Die erforderliche Steuerung des Syntheseteils bzw. die Realisierung der Gerätegrundfunktionen übernimmt ein Ein-Chip-Mikrorechnersystem auf Basis eines UB8830-Prozessors (ähnlich ZILOG Z8).

Wissenschaftlicher Taschenrechner BR 1 mit Sprachausgabe

Der Taschenrechner BR 1 ist ein speziell für Blinde bzw. Sehbehinderte in enger Zusammenarbeit mit dem Blinden und Sehschwachenverband der DDR entwickelter Taschenrechner mit Sprachausgabe.
Alle Eingaben (numerisch, Operationen, Funktionen) und Ergebnisse bzw. Statusinformationen werden sprachakustisch ausgegeben. Zur optischen Informationsausgabe ist ein rot leuchtendes abgesetztes Display (Ziffernhöhe ca. 13 mm) ansteckbar. Über diese Koppelstelle ist auch der Anschluß einer Braille-Zeile möglich.

Zur Sprachausgabe kommt ein speziell entwickeltes Sprachwiedergabeverfahren zur Anwendung. Die synthetische Sprache erreicht Telefonqualität.

Bei Bedarf können individuelle Wünsche zum Sprachinventar wie Charakter der synthetischen Stimme (z. B. männl./weibl. u.a.), Textinhalt oder Fremdsprachen berücksichtigt werden.
Bei Displaybetrieb arbeitet das Gerät mit Netzteil (im Displayteil enthalten), sonst mit Batterie (Intervallbetrieb).

Der Rechner BR 1 verfügt über verschiedene Hilfsfunktionen, die sehgeschädigten Nutzern ein effektives Arbeiten erlauben (z. B. erweiterte Korrekturmöglichkeiten bei Eingabefehlern, akustische Ausgabe bisheriger eingegebener Tastenfolgen zu beliebigen Zeitpunkten u.a.). Mit 10 Speichern besitzt der BR 1 eine wesentlich über das übliche Maß hinausgehende Merkfähigkeit.
Die Ansage numerischer Daten ist in verschiedenen Modi (ziffernweise, Zifferngruppen) möglich. Durch den Einsatz eines frei programmierbaren Mikrorechnersystems im BR 1 sind verschiedene Leistungskategorien durch einfachen Modulwechsel realisierbar. Standardversion ist ein wissenschaftlicher Taschenrechner äquivalent dem in der DDR empfohlenen Schulrechner SR 1.
Die Abmessungen betragen 105 mm x 170 mm x 40 mm ohne Displayteil.
Die großflächigen Tastenköpfe (ca. 10 mm x 10 mm) sind teilweise durch Fühlpunkte markiert und erlauben eine sichere Bedienung. Der eingebaute Lautsprecher wird beim Anstecken eines Kopfhörers abgeschaltet. Die Lautstärke ist in 3 Stufen einstellbar.

Sprachmodul SM als sprechende Korrekturhilfe an der Schreibmaschine

Der Sprachmodul SM kann an der elektronischen Schreibmaschine angeschlossen werden. Er ist in der Lage, bis ca. 1/2 Seite zuletzt geschriebenen Text zu speichern und diesen Bereich vollständig oder in Teilen sprachakustisch auszugeben.
Eine Reihe von Zusatzfunktionen erleichtert die Schreibarbeit für Blinde bzw. stark Sehbehinderte.
Folgende Funktionen können ausgeführt werden:

* Sofortansage aller betätigten Tasten der Schreibmaschine mit/ohne aktives Schreibwerk,
* Ausgabe des Speicherinhalts
  - mit vorheriger Positionierung im Speicher ("vor-/rückwärts")
  - bei gleichzeitiger Positionierung ("vorwärts/rückwärts"),
* Notizfunktionen, (bei Abschaltung des Schreibwerks), Eingabe aktueller "Notizen" über Tastatur der Schreibmaschine in gesonderten Speicherbereich zur späteren akustischen Ausgabe,
* Vorlesen der Speicher der Schreibmaschine,
* Ansage der aktuellen Schreibposition,
* Wiederholung der letzten akustischen Ausgabe.

Die Ausgabe kann zeichenweise buchstabieren, z. B. für orthographische Kontrollen) oder als Text erfolgen.
Der Modul ist auch als Interfaceumsetzer Schreibmaschine - V 24 nutzbar.
Als Sprachausgabe wird ein 4-Formant-Reihenfiltersynthetisator verwendet.

Die hohe Eigenintelligenz des Moduls SM (eigener Mikrorechner) erlaubt weitestgehend ein "Doubeln" aller Schreibwagenfunktionen der Schreibmaschine (z. B. Rückschritte, Leertasten u.a.), so daß z. B. nachträgliche Korrekturen im Modulspeicher mit ausgeführt werden.
Der eingebaute Lautsprecher wird bei Anschluß eines Kopfhörers (Euro- bzw. 3,5-mm-Klinke) abgeschaltet. Die Lautstärke ist über Schiebesteller kontinuierlich einstellbar. Über eine Minimaltastatur (11 Tasten) erfolgt die Bedienung des Moduls.

Sprachspeicher-Modulsystem

Mit dem Sprachspeicher-Modulsystem können bis zu 8 min Sprache in Telefonqualität gespeichert und bei Bedarf wieder ausgeben werden.
Über eine speziell entwickelte Codiereinrichtung wird das Sprachsignal in einen 1-bit-breiten Datenstrom gewandelt, in einem Halbleiterspeicher-Modul (Festwert- oder Schreiblesespeicher) abgespeichert und über einen Decoder wieder rückgewandelt. Der Speicher kann dafür in Segmente von 2, 4, 8, 16, 32, 64, 128 oder 256 Sekunden aufgeteilt werden, auf die wahlfrei bei Speicherung bzw. Wiedergabe zugegriffen werden kann. Damit ist ein freizügiger Einsatz der Sprachspeicherung möglich. Das Steuerwerk übernimmt die Betriebsartenumschaltung, die Speichersteuerung sowie die Nutzerkommunikation (Segmentanwahl, Segmentfolgesteuerung).
Ein prozessorgesteuerter Koppelmodul realisiert die Verbindung des Sprachspeichers zu dem angeschlossenen Gerät (z. B. Umweltmeßgerät) oder auch über ein Standardinterface zu einem PC. Für den Betrieb des Moduls an einem IBM-kompatiblen PC steht ein entsprechendes Steuerprogramm zur Verfügung.

Die Sprachdaten können auch aus einem Festwertspeicher entnommen werden, sodaß vorprogrammierte Ansagen z. B. für eine Uhr, Waage oder Thermometer möglich sind.

Der typische Informationsfluß zwischen Speicher und Codec beträgt 16 kbit/s und kann in Abhängigkeit von der geforderten Sprachqualität zwischen 8 und 48 kbit/s variiert werden, wobei sich entsprechend andere Speicherkapazitäten ergeben.

Auf der Basis dieses Sprachspeichersystems wurde ein vollelektronisches "Notizbuch" realisiert, das 4 Minuten Sprache auf 8 "Seiten" (30 s) speichern, ca. 3 - 4 Tage speichern und bei Bedarf wiedergeben kann. Außerdem wurde der Decoder-Teil des Sprachspeichers im Blindenrechner BR 1 sowie in einer Uhr mit Sprachausgabe eingesetzt.

Zur Generierung der Ansteuerdaten für die Synthesesysteme entstanden verschiedene Entwicklungssysteme, u.a. für den an der TU Dresden entwickelten Formantsynthetisator TUSY.

Spracheditiersystem SPREDI

Das Spracheditiersystem SPREDI ist ein Programmpaket zur Gewinnung und Aufbereitung von sprachlichen Einheiten u.a. für den Schreibmaschinen-Sprachmodul SM.

Die hohe Eigenintelligenz des Moduls SM (eigener Mikrorechner) erlaubt weitestgehend ein "Doubeln" aller Schreibwagenfunktionen der Schreibmaschine (z. B. Rückschritte, Leertasten u.a.), so daß z. B. nachträgliche Korrekturen im Modulspeicher mit ausgeführt werden.
Der eingebaute Lautsprecher wird bei Anschluß eines Kopfhörers (Euro- bzw. 3,5-mm-Klinke) abgeschaltet. Die Lautstärke ist über Schiebesteller kontinuierlich einstellbar. Über eine Minimaltastatur (11 Tasten) erfolgt die Bedienung des Moduls.

Sprachspeicher-Modulsystem

Mit dem Sprachspeicher-Modulsystem können bis zu 8 min Sprache in Telefonqualität gespeichert und bei Bedarf wieder ausgeben werden.
Über eine speziell entwickelte Codiereinrichtung wird das Sprachsignal in einen 1-bit-breiten Datenstrom gewandelt, in einem Halbleiterspeicher-Modul (Festwert- oder Schreiblesespeicher) abgespeichert und über einen Decoder wieder rückgewandelt. Der Speicher kann dafür in Segmente von 2, 4, 8, 16, 32, 64, 128 oder 256 Sekunden aufgeteilt werden, auf die wahlfrei bei Speicherung bzw. Wiedergabe zugegriffen werden kann. Damit ist ein freizügiger Einsatz der Sprachspeicherung möglich. Das Steuerwerk übernimmt die Betriebsartenumschaltung, die Speichersteuerung sowie die Nutzerkommunikation (Segmentanwahl, Segmentfolgesteuerung).
Ein prozessorgesteuerter Koppelmodul realisiert die Verbindung des Sprachspeichers zu dem angeschlossenen Gerät (z. B. Umweltmeßgerät) oder auch über ein Standardinterface zu einem PC. Für den Betrieb des Moduls an einem IBM-kompatiblen PC steht ein entsprechendes Steuerprogramm zur Verfügung.

Die Sprachdaten können auch aus einem Festwertspeicher entnommen werden, sodaß vorprogrammierte Ansagen z. B. für eine Uhr, Waage oder Thermometer möglich sind.

Der typische Informationsfluß zwischen Speicher und Codec beträgt 16 kbit/s und kann in Abhängigkeit von der geforderten Sprachqualität zwischen 8 und 48 kbit/s variiert werden, wobei sich entsprechend andere Speicherkapazitäten ergeben.

Auf der Basis dieses Sprachspeichersystems wurde ein vollelektronisches "Notizbuch" realisiert, das 4 Minuten Sprache auf 8 "Seiten" (30 s) speichern, ca. 3 - 4 Tage speichern und bei Bedarf wiedergeben kann. Außerdem wurde der Decoder-Teil des Sprachspeichers im Blindenrechner BR 1 sowie in einer Uhr mit Sprachausgabe eingesetzt.

Zur Generierung der Ansteuerdaten für die Synthesesysteme entstanden verschiedene Entwicklungssysteme, u.a. für den an der TU Dresden entwickelten Formantsynthetisator TUSY.

Spracheditiersystem SPREDI

Das Spracheditiersystem SPREDI ist ein Programmpaket zur Gewinnung und Aufbereitung von sprachlichen Einheiten u.a. für den Schreibmaschinen-Sprachmodul SM.

Es bietet folgende Funktionen:

* Analyse eines natürlichen, digitalisierten Sprachsignals und Extraktion von zeitlichen Formantdatenverläufen zur Gewinnung von Ansteuerdaten für 4-Formant-Reihenfilter-Synthetisatoren,
* interaktive Optimierung der von der Analyse gelieferten Ansteuerdaten,
* diverse Hilfsfunktionen.

Die Gewinnung der Formantdaten basiert auf dem LPC-Verfahren; für die Ermittlung der Steuerdaten für die Grundfrequenzsteuerung kann wahlweise ein AMDF- oder Cepstrumverfahren genutzt werden.
Das Programm zur Optimierung der Steuerdaten bietet dem phonetisch vorgebildeten Nutzer eine Vielzahl von Möglichkeiten zur interaktiven Arbeit an den Synthesedatenfolgen. Die Arbeit wird optisch durch Darstellung der Steuerdaten-Zeit-Verläufe auf einem Farbdisplay und akustisch durch Ausgabe der Steuerdatenfolgen auf den Synthetisator unterstützt.
Dem Nutzer wird die Möglichkeit geboten, umfassende Manipulationen an den Ansteuerdaten vorzunehmen:

* Zusammenstellen von Experimentierdateien,
* Veränderung von Parameterwerten für Formantmittenfrequenzen, -bandbreiten, Amplitude, Grundfrequenzen (einzeln oder bereichsweise)
* Separieren von Zeitabschnitten mit der Steuerdatenfolge
* Einschieben von leeren Zeitabschnitten
* Tauschen und verdoppeln von Zeitabschnitten.

Mit den Hilfsprogrammen können Dateien mit Ansteuerdaten für Sprachsynthetisatoren umgespeichert, geplottet, gedruckt, sowie Dateien für die EPROM-Programmierung erstellt werden. Ein weiteres Programm ist das Zusammenstellen von Steuerdatenfolgen zur Erzeugung von fließender synthetischer Sprache. Außerdem ist es möglich, Mikrointonationskonturen in Folgen von Ansteuerdaten einzubringen.

Das System läuft auf einem Rechner des Typs K 1630 (16 bit, DEC-kompatibel) unter dem Betriebssystem OMOS. Notwendige Peripherie sind: ADU/DAU, Farbdisplay, Drucker, Plotter.
Es können Sprachsynthetisatoren des Typs TUSY sowie der Schaltkreis MEA 8000 der Firma VALVO angesteuert werden.
Die Optimierung der Synthesedatenfolgen kann auch auf einem IBM-kompatiblen PC vorgenommen werden.

Ein ähnliches System ist für die Aufbereitung von CVSD-Daten verfügbar.

Literatur

[1] Hallensleben, K.: Psychologisch-ergonomische Untersuchungen zur nutzerfreundlichen Gestaltung von Blindenhilfstechnik auf der Basis synthetischer Sprachausgabe.
Diplomarbeit, TU Dresden, Sektion Arbeitswissensch., 1989

[2] Kordon, U.; Seveke, L.: Sprachausgabetechnik für den Einsatz in der Laborautomatisierung.
Nachrichtentechnik-Elektronik, Berlin 40 (1990), i.V.