Spracheingabe-Modul

Mit einem Spracheingabe-Modul 1.40.690032.4 werden die über eine Sprechgarnitur eingegebenen akustischen Signale digitalisiert und im Computer weiterverarbeitet/ausgewertet. Anhand von Sprachmustern können Worte erkannt und damit Aktionen ausgelöst werden. Ca. 50 Worte werden unterschieden.

Das Spracheingabemodul für den Z9001 ist eine gewisse Besonderheit. Für keinen anderen Kleincomputertyp ist mir ein vergleichbares Modul bekannt.



Spracheingabemodul mit Sprechzeug SP75 und Kassette R0113 (AS)

Entwickelt wurde das Spracheingabemodul 1985/1986 und zur Leipziger Frühjahrsmesse 1987 gemeinsam mit dem KC87 erstmals der Öffentlichkeit vorgestellt. Erhältlich war das Modul ab 1987 für industrielle Abnehmer zu einem IAP von 343 Mark.

Eigenschaften des Moduls:

- Wortschatzgröße 50 Wortklassen
- Max. 200 Referenzmuster
- Wortlänge 0,2 bis 1,8 s
- Reaktionszeit nach Wortende: 0,2 s
- Pause zwischen 2 Worten: > 0,2 s
- Erkennungsquote f. geübten Sprecher > 90%
- Rückweisung: ~ 5%
- Fehlererkennung: ~5%

Mein Spracheingabe-Modul stammt direkt aus dem Entwicklernachlass (LE).

Zum Modul gehören eine Kassette R0113 mit der nötigen Software SPRACHE1/2.COM und eine Anleitung zum Modul spracheingabe-modul.pdf.



Spracheingabe-Modul, Kassette R0113

Entwicklung

2007 hatte ich Kontakt zu den Entwicklern des Spracheingabe-Moduls, Dr. U. Kordon und Dr. L. Seveke.

U. Kordon: *„Der Spracheingabenmodul für den Z9001 war damals eine Gemeinschaftsentwicklung des Kombinats Robotron und der TU Dresden. Die Arbeiten erfolgten im Auftrag des damaligen VEB Messelektronik Dresden als eigentlicher Hersteller des Z9001 (Themenleiter Z9001: Dr. Gunter Kleinmichel). Wir als TU waren nur für die Hardware des Spracheingabemoduls zuständig, die gesamte Software dazu wurde von Robotron entwickelt und basierte auf dem (kommerziellen) Spracheingabemodul für das K1520-System K7821 (Basis U880/Z80, Einbau mit Bedieneinheit in den Einzelwort-Spracherkenner K7823). Der damalige Bearbeiter bei Robotron war Dr. Seveke, der ursprünglich aus unserer TU-Arbeitsgruppe kam. … Es gab bei Robotron eine eigene Arbeitsgruppe zur Sprachsignalverarbeitung, die unter anderem für das K1520-System Module zur Einzelworterkennung und Sprachwiedergabe entwickelt hat.“*

L. Seveke: *„Ich war aber auch Entwickler beim ZFTM (Sprach-Ein- und Ausgabe) und ab und zu mal an den Z9001 und später an Dr. Kleinmichel ausgeliehen. Das Kassetten-Programm (Kansas-City-Format) und vor allem die Entwicklung des Spracheingabe-Moduls stammen von mir.“*

Von Dr. Kordon habe ich Unterlagen zum Modul erhalten, unter anderem Listing-Ausdrucke des Softwarekerns für das K1520-System K7821. Damit war es möglich, für die Z9001-Software SPRACHE1/2.COM den Assemblercode zu rekonstruieren.

In der TU wurde ein leicht abweichendes Modul benutzt (ohne LED), aber mit gleichen Portadressen etc.

KC-Modul TU ohne LED, CTC 0e0h..0e3h
KC-Modul MKD mit LED Port 0E4h ein, 0E8h aus , CTC 0e0h..0e3h

Literatur

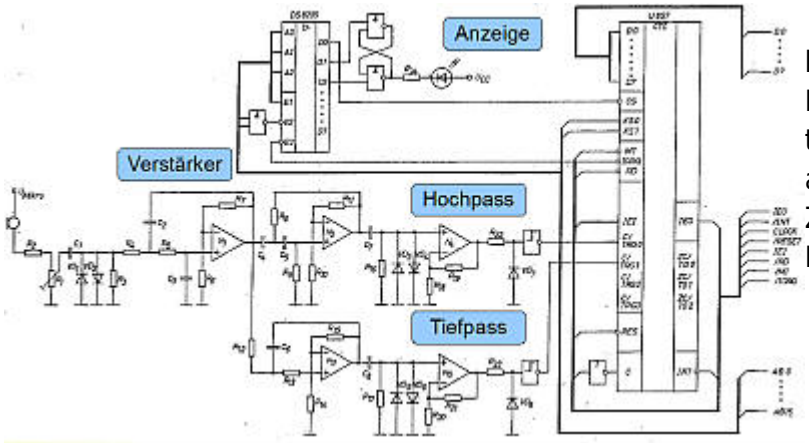
Außerdem gab es zwei ausführliche Veröffentlichungen in der rfe, die die verwendeten Prinzipien der Spracherkennung und die konkrete Umsetzung in Hard- und Software beschreiben.

- Dr.-Ing. Ulrich Kordon: Sprachsignalerkennung und Sprachsynthese, rfe 10/1987, S. 651-655
- Dr.-Ing. Lothar Seveke und Dr.-Ing. Ulrich Kordon: Spracherkenner-Zusatzmodul für U880-Mikrorechner, rfe 11/1987, S. 728-731

Beim KC-Klub-Treffen 2008 habe ich einen Vortrag über die Spracheingabe gehalten und vor allem die Software vorgestellt. Im Archiv gibt es die Anleitung, die komplette Softwarebeschreibung sowie den rekonstruierten Assembler-Quelltext und die Version **sprache1hx**, die das normierte Histogramm YFELD (s.u.) zu Demostrationszwecken anzeigt.

Software

Das Hauptproblem beim Erfassen von Sprachsignalen ist die für 8-Bit-Rechner enorme Datenmenge. Ein Samplen mit 8 Bit und 11 Khz Abtastrate erzeugt pro Sekunde (!) 11 KByte Daten, damit wäre der verfügbare Hauptspeicher des Z9001 mit 1 Aufnahme und 3 Referenzen fast vollständig belegt!



Die Daten müssen also effektiv komprimiert und auf vergleichbare Merkmale reduziert werden. Das gelingt trotz der sehr geringen Hardware durch ausgeklügelte Algorithmen in Zusammenarbeit mit einer speziellen Hardware.

Es ist faszinierend und einen tiefen Blick in die Beschreibung und den Quellcode wert, wie eine so umfangreiche Aufgabe wie die Spracherkennung mit einem 2 MHz-Z80 und nur 2 KByte Software gelöst werden kann!

Die Hardware besteht aus einem Verstärker für das Mikrofonsignal und je einem Tief- und Hochpass (Trennfrequenz ca. 1 kHz), die auf die Triggereingänge einer CTC gehen. CTC-Kanal 0 und 3 arbeiten als Zähler mit Zählwert 1 und werfen beim Nulldurchgang von Tief- und Hochpass einen Interrupt (Programmteile TP-NDG, HP-NDG), der CTC-Kanal 1 arbeitet als Zeitgeber (Analyse- Zeitfenster), durch Abfrage des Zählwerts erhält man die Position innerhalb des Zeitfensters.

Für ein gewisses Zeitfenster werden die Abstände der Nulldurchgänge ermittelt und anschließend klassifiziert. Die Routine NDG werten diese Abstände als Intervallklasse aus. Die Interrupt-Routine des Zeitfensters CTC 1 sammelt die Intervallklassen von Hochpass und Tiefpass in 8 Byte zusammen. Die Pausendetektion (PAUSE) sucht Wortanfang und -Ende, der 8-Byte-Vektor wird 1-bit-quantisiert (Schwellwert SWHL), die 8 Bit werden in den Signalspeicher XFELD hinzugefügt. Durch eine zeitliche Normierung wird der Signalspeicher durch die Routine ZEIN zu einem 16 Byte-Feld YFELD normiert (ähnlich Bild stauchen/strecken). Damit ist jedes erkannte Wort-Signal auf einen 16-Byte-Datensatz reduziert! (Umwandlung des Audiosignals in eine Folge von Merkmalsvektoren)

Weitere Routinen sortieren diese Datensätze in Wortklassen ein (mehrere Refrenzmuster pro Wort). Das erfolgt mit den Routinen VGLEI (minimale Hamming-Distanz berechnen), KNORM (Clusterbewertung normieren), CLST (Cluster streichen), CLUST (Clusterbildung). Der Vergleich mit vorher ermittelten Referenzmustern erfolgt als bitweiser Vergleich mit bereits gespeicherten Referenzmustern (Hammingdistanz = Anzahl der verschiedenen Bits), das Referenzmuster mit geringster Distanz ist das erkannte Wort. Ist die Distanz zu groß, wird das Wort abgewiesen. Andere Routinen dienen zur Einteilung von Wortklassen von Referenzmustern, zur Wortsignaldetektion (Anfang, Ende), dem automatischen Verwerfen falscher Muster und zur Erkennung bereits während

der Signalanalyse.

Zur Benutzung in eigenen Programmen gibt es schließlich Routinen, die das Arbeiten mit Wortklassen ermöglichen: Lernen, Nachlernen, Klassifizieren, Laden/Speichern. Zur praktischen Arbeit wird der gesamte mögliche Umfang in 5 Teilwortschätze mit je max. 50 Klassen, insgesamt max. 200 Referenzmuster, aufgeteilt.

Über die definierte Schnittstelle CALL RECOG → Ergebnis Klasse (0..50) wird die Spracherkennung in eigenen Programmen genutzt. Es wird auf eine Spracheingabe gewartet. Die Routine gibt die erkannte Wortklasse als Zahl zurück.

```
10 TW=1:POKE 80,TW :!Teilwortschatz einstellen
20 ! (nur bei Bedarf)
30 CALL DEEK(76) :!Aufruf des Erkennerprogramms A
40 ER=PEEK(81) :!Erkennungsergebnis
50 IF ER>127 THEN GOTO xxx :!Fehlerbehandlung
```

Im Downloadpaket sind außerdem noch BASIC-Programme der Entwickler, die leider nicht auf der Kassette R0113 enthalten sind, aber gut die Möglichkeiten des Spracheingabe-Moduls demonstrieren:

- SEDEMO.SSS, ein BASIC-Programm für die Leipziger Messe, das die erkannten Worte auf dem Bildschirm als Text ausgibt (VEB,ROBOTRON,MESSELEKTRONIK,'OTTO SCHOEN',DRESDEN,STELLT VOR,EIN,KLEINCOMPUTER,KC 87,MIT,SPRACHEINGABE-MODUL,IN,VERSTEHT,MENSCHLICHE,SPRACHE,WIRKLICH,?)
- AHASPRA.SSS, die erste Version des obigen Programms
- SP-TAFEL.SSS, ein sprachgesteuertes Zeichen-Programm !!

Natürlich müssen zur Nutzung dieser Programme erst die benötigten Wörter gelernt werden!

Sprachausgabe

Interessant ist auch noch folgendes:

U. Kordon: *„Übrigens haben sich auch noch Unterlagen für einen geplanten Sprachausgabemodul zum Z9001 auf der Basis des damaligen Valvo-Schaltkreises MEA8000 (4-Formant-Synthetisator) angefunden. Der Sprachausgabemodul ist bis zum Funktionsmuster entwickelt worden, der dann mit „einheimischen Rohstoffen“, d. h. mit einem Ersatz des Importschaltkreises, realisiert werden sollte. Das haben wir übrigens auch geschafft, allerdings schrieb man da inzwischen das Jahr 1989 … (ohne nostalgische Anwandlungen, einfach als Fakt).“*

Bemerkungen

Vor allem männliche (tiefe) Stimmen reizen den Zählumfang bis 255 oftmals aus, so dass Worte nur schwer zu unterscheiden sind (Die 16-Byte-Datensätze sehen alle aus wie F8FFFFFEF5F8FEFF…). Bei höheren Stimmlagen (Frauen, Kindern) ergeben sich normale Muster und damit eine wesentlich bessere Worterkennung.

Eine entscheidende Rolle spielt dabei auch das Mikrofon! Mit dem regelbaren Widerstand auf der Platine kann die Empfindlichkeit des Mikrofons angepasst werden.

Zur Arbeit mit dem Sprachprogramm ist es sinnvoll, sich eine Wortliste bereitzulegen, die die Klassennummern 1..50(200) und die Wörter enthält. Das Sprachprogramm arbeitet nur mit den Klassennummern und nicht mit Klartexten!

Vielen Dank an Dr. Kordon und Dr. Seveke für die freundliche Zuarbeit und die Überlassung der Unterlagen!

From:
https://hc-ddr.hucki.net/wiki/ - **Homecomputer DDR**

Permanent link:
https://hc-ddr.hucki.net/wiki/doku.php/z9001/module_robotron/spracheingabe?rev=1465546171

Last update: **2016/06/10 08:09**