



Spracherkennung mit dem Z9001



Kommerziell hergestelltes Modul für den
Kleincomputer Z9001
zur
Echtzeitspracherkennung



Spracheingabemodul





Spracheingabemodul



- Entwickelt 1985/1986
- Vorgestellt Leipziger Frühjahrsmesse 1987
gemeinsam mit dem KC87
- Erhältlich 1987
- Produzierte Stückzahl ???



Entwickler



Dr.-Ing. Lothar Seveke

... „Ich war 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 sind auf meinem Mist gewachsen“ ...



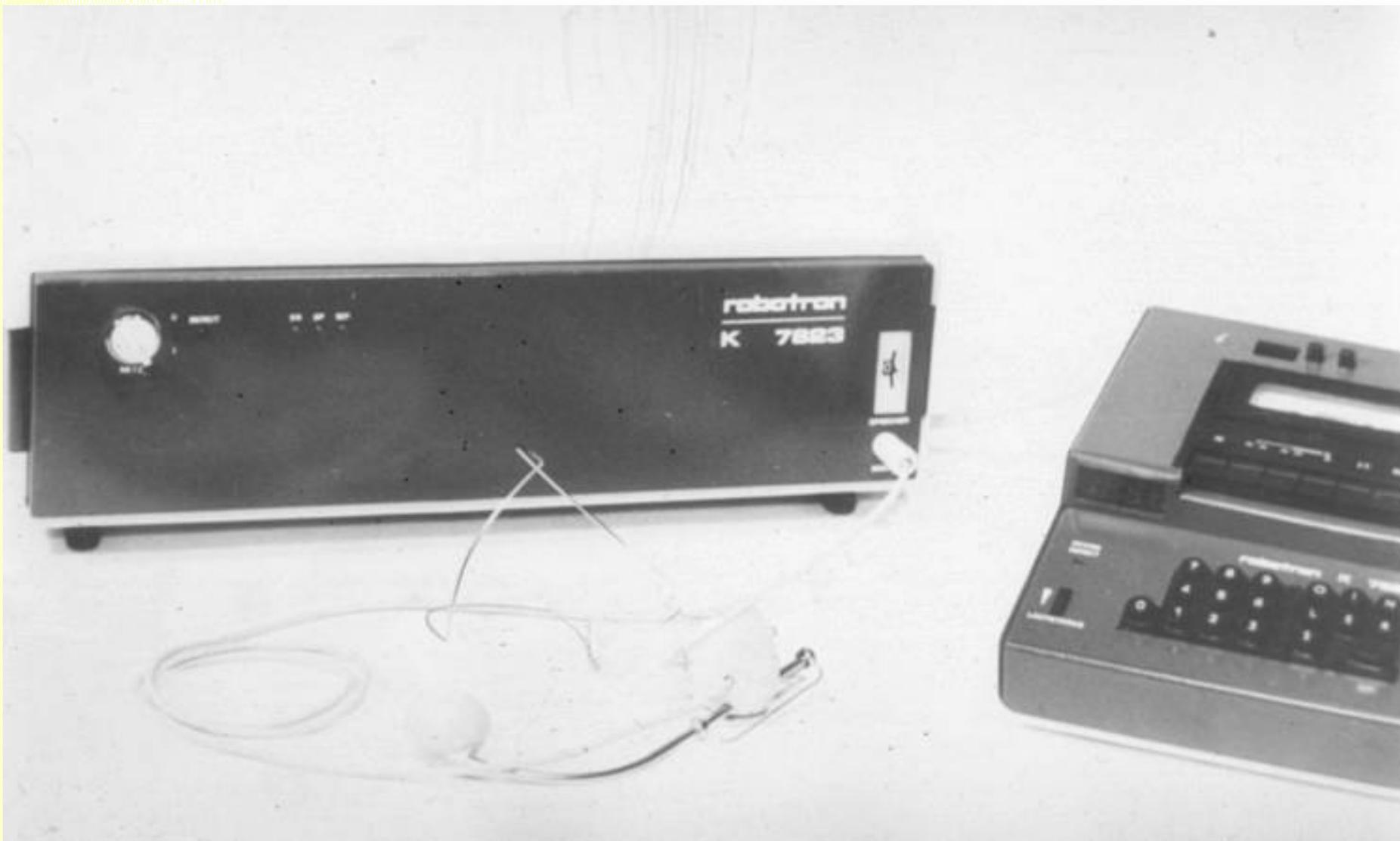
Dr.-Ing. Ulrich Kordon

... „der Spracheingabemodul für den Z9001 war damals eine Gemeinschaftsentwicklung des Kombinars 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“ ...



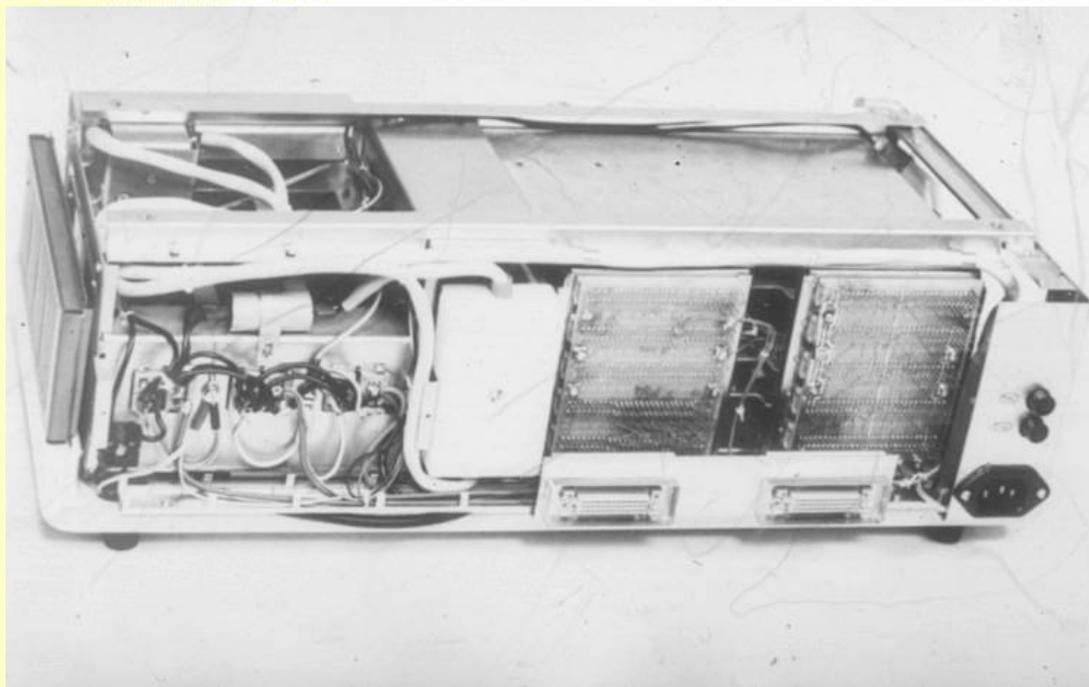


System ESE K7823





System ESE K7823



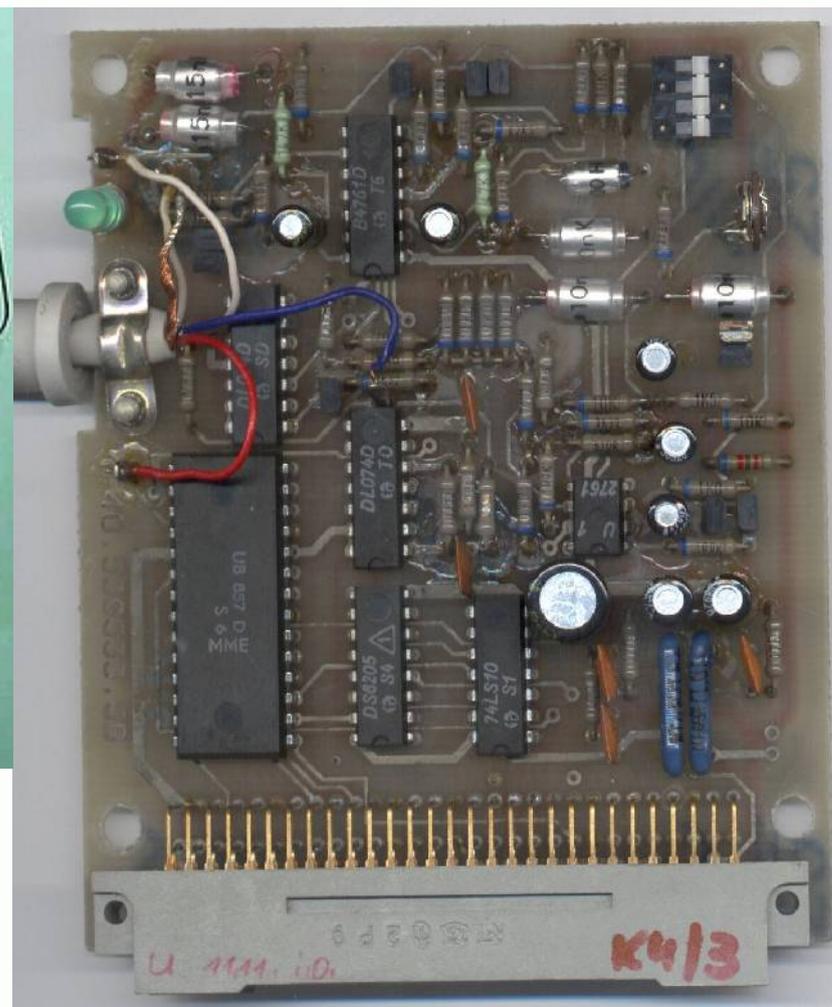
ese K7823 (oben)
bde K7822 (rechts)



Spracheingabemodul

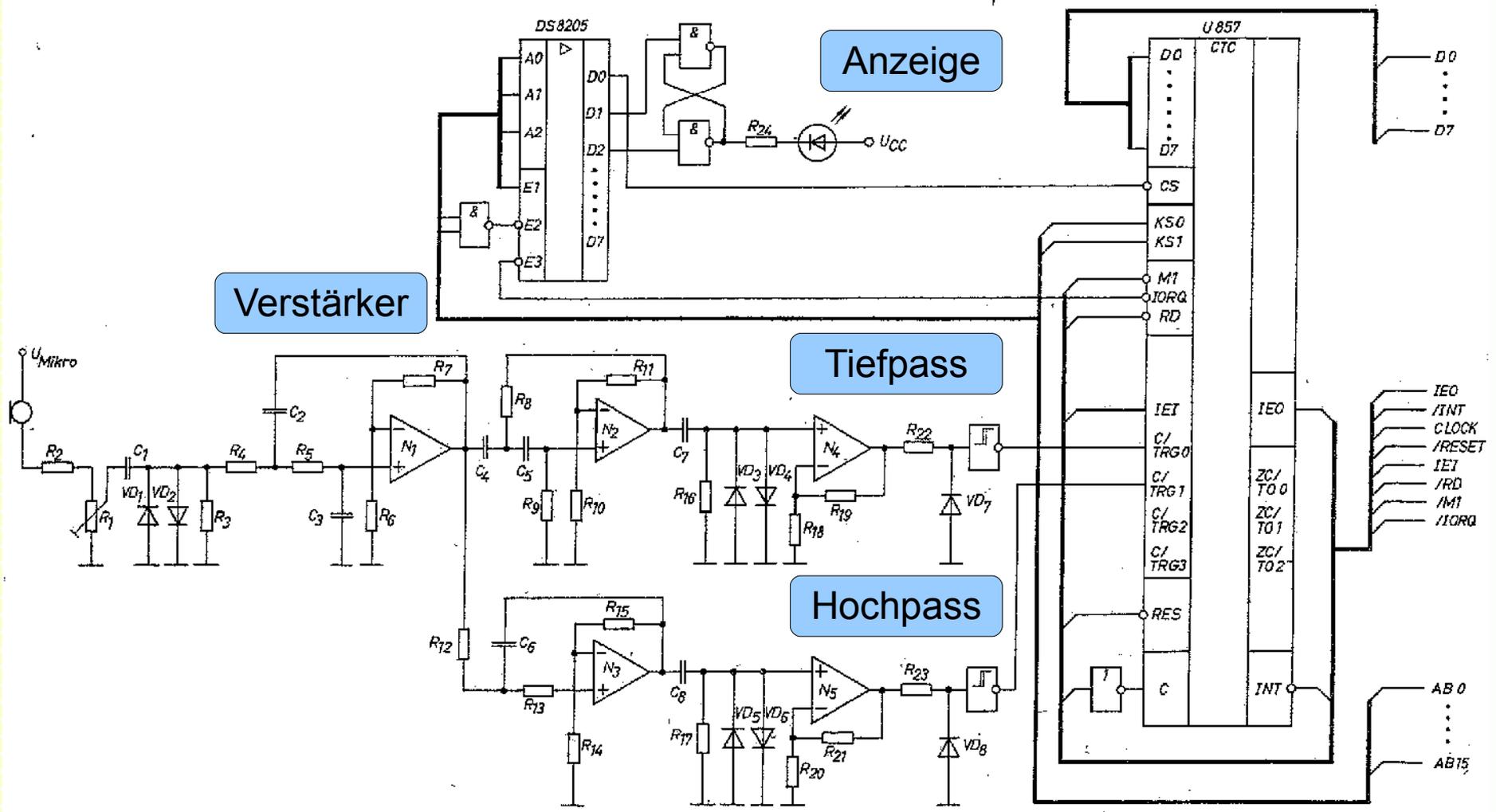


Sprechzeug SP75 (oben)
geöffnetes Modul (rechts)





Schaltplan





Eigenschaften



- 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: $\sim 5\%$
- Fehlererkennung: $\sim 5\%$



Spracherkennung

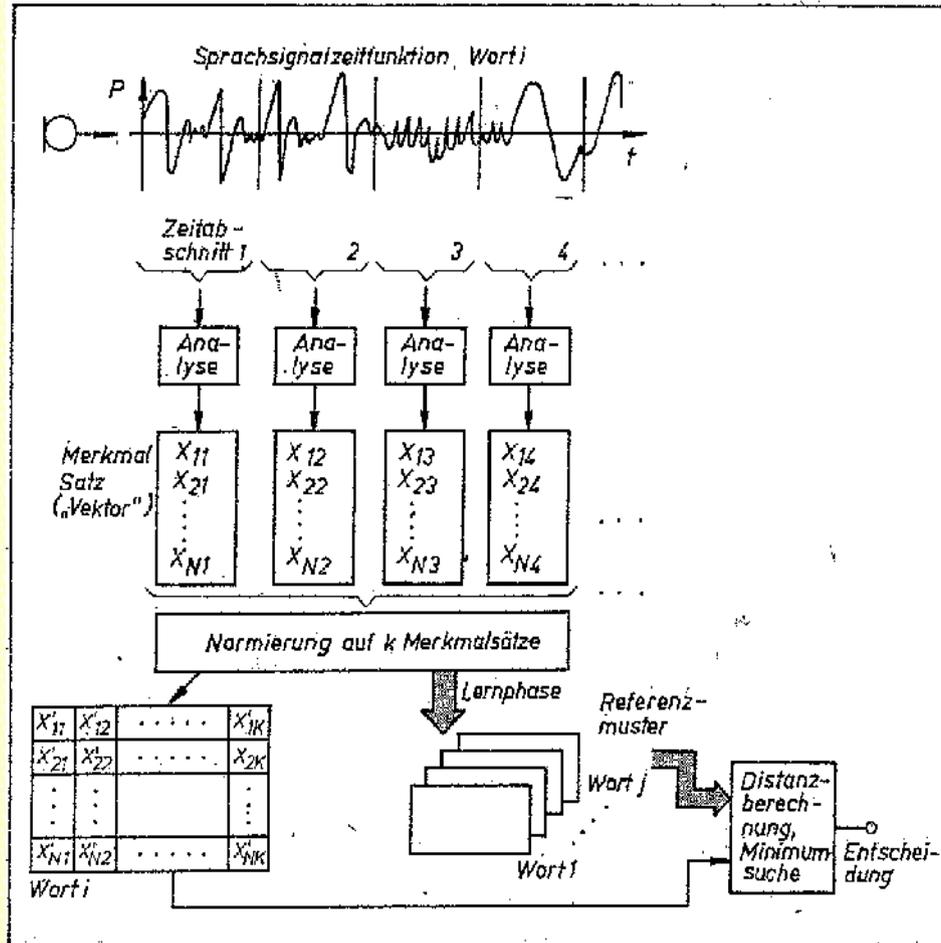


Bild 6: Prinzip einfacher Spracherkennung

1. Umwandlung des Audiosignals in eine Folge von Merkmalsvektoren
2. Normierung
3. Vergleich mit vorher ermittelten Referenzmustern



Das NDG

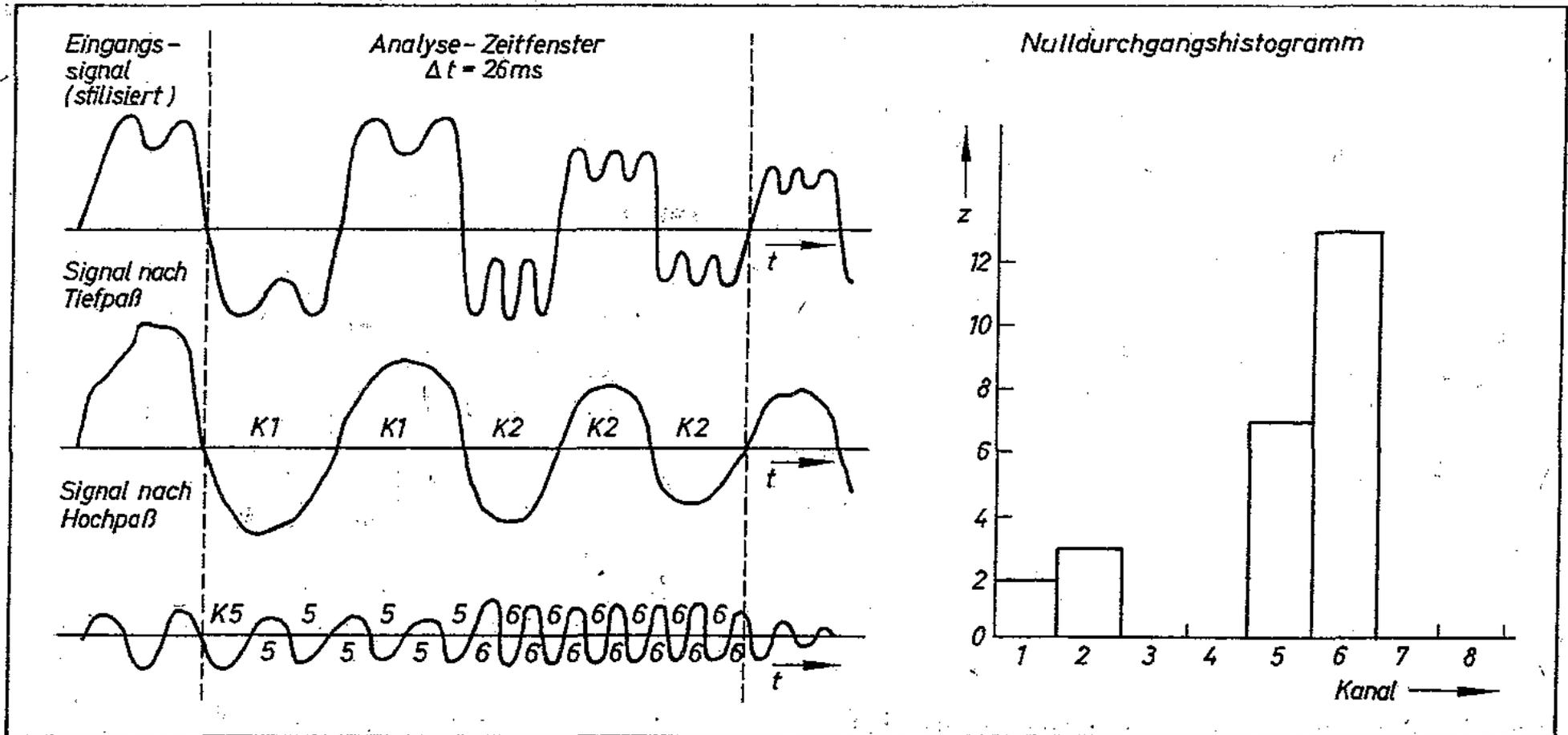
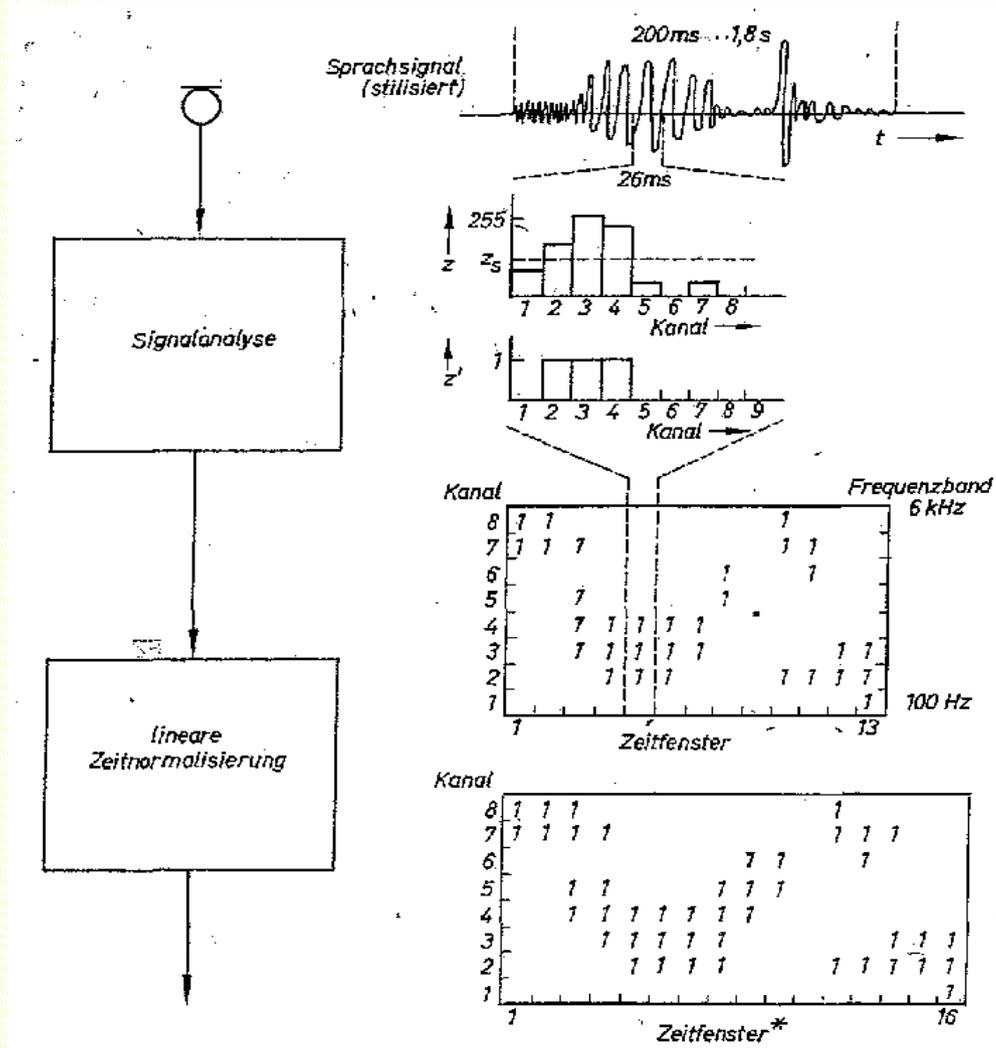


Bild 2: Bildung des Nulldurchgangshistogramms



Bildung der Muster



- ← Zeitfenster
- ← Merkmalsvektor
- ← 1-Bit-Quantisierung
- ← Muster (Folge von 8-Bit-Merkmalsvektoren)
- ← Normiertes Muster (16 Byte)



Beispiele



<p style="text-align: center;">zwei</p> <pre> 8 I **** I I * ** I I **** I I * **** I I **** ** I I **** ** I I * ** I I *** **** I 1 I 1234567890123456 s w a e </pre>	<p style="text-align: center;">sieben</p> <pre> 8 I **** I I * **** * I I **** ** I 1 I 1234567890123456 z i b m </pre>	<p style="text-align: center;">Kanal eins</p> <pre> 8 I I I * I I ** I 1 I 1234567890123456 a e n s </pre>
<p style="text-align: center;">vier</p> <pre> 8 I **** I I **** I I *** **** I I * **** I I * I I ** I I **** ** I I **** ** I 1 I 1234567890123456 f i e </pre>	<p style="text-align: center;">acht</p> <pre> 8 I I I ** I I ** * I I **** I I **** I I **** I I **** I I * * I I * * I 1 I 1234567890123456 a x - t </pre>	<p style="text-align: center;">Kanal fünf</p> <pre> 8 I ** I I *** ** I I **** ** I I * * * I I ** I 1 I 1234567890123456 f y n f </pre>



Erkennen der Wortklasse



Bitweiser Vergleich mit bereits gespeicherten Referenzmustern (Hammingdistanz = Anzahl der verschiedenen Bits)

Referenzmuster mit geringster Distanz =
erkanntes Wort

Distanz zu groß \Rightarrow Abweisung

Verbesserung: Wortklassen von
Referenzmustern



weitere Aufgaben



- Wortsignaldetektion (Anfang, Ende)
- Clusterbildung
- automatisches Verwerfen falscher Muster
- Erkennung bereits während der Signalanalyse



Software

33-1524 (SCPT v 7.8)

LO 0000 LWA PAUSE
 C100 0000 A. (AZUST)
 L1 0000 2. ZURÜ
 A10 0000
 30

KEIN WORTENDE.
 IN DER PROG. WORTENDE
 IN VOR PROG. WORTENDE ODER SICHEREM WORTENDE

ANALYSE: ANSPRECHEN MIT XFELD-UEBERLAUF ODER SICHEREM WORTENDE
 ZURÜ: C100 ENTL0
 C101 ZEIN
 C102 A
 C103 (X0Y0), A
 C104 W0LEI
 C105
 C106
 C107
 C108
 C109
 C110
 C111
 C112
 C113
 C114
 C115
 C116
 C117
 C118
 C119
 C120
 C121
 C122
 C123
 C124
 C125

ALLE CLUSTER BETRACHT
 LAUFENDE NUMMER DER
 VERKANNTEN KLASSE IN
 TABELLE DER ANZEIG
 ANZEIGE SYMBOL DE

CALL BERLIN
 JR ZURÜ

```

0395A 1A
0355A FF 00A
0358A 30 09
035A 45
035B 4F
035C 78
035D 86
035E 77
035F 3E F0
0361 45
0362 6F
0363 08
0364 09
0365 FH
0366 00 00
  
```

```

LO A, (DE)
CP KREN
JR C, APT
K00 A, A
00 A, A JR TP6
LD A, R
400 A, R
LD M, A
LD A, OFOH
A100 L
LD L, A
HPI: EX AF, AF
EJ
EI
RETI
  
```

PAUSE

PAUSEDETEKTION

LAUSPEICHERN EINES A-BYTE-VEKTORS AUS DEM VEKTORPUFFER
 ALS 1-BYTE-VEKTOR AUF DER AKTUELLEN ADRESSE IM SIGNALSPEICHER
 HL=REG.I ADRESSE DES AKT. VEKTORPuffers
 XFELD: REG.I SIGNALSPEICHER
 EXPFL: ENDE DES SIGNALSPEICHERS
 XA: AKTUELLE ADRESSE IN XFELD
 VANZ: ZAHL DER VEKTOREN BIS PROG. WORTENDE
 SW1: MINIMALE KO.-SUMME FUER SIGNALANFANGSERKENNUNG
 T1: MINDESTDAUER FUER SIGNALANFANGSERKENNUNG
 SW2: MAXIMALE KO.-SUMME FUER PAUSEERKENNUNG
 T2: MAXIMALE DAUER FUER PAUSEERKENNUNG
 SW3: MINIMALE KO.-SUMME FUER KNACKERKENNUNG
 T3: MAXIMALE DAUER FUER KNACKERKENNUNG
 TT1: AKTUELLE SIGNALDAUER NACH SIGNALANFANGSERK., T1=T1...0
 TT2: AKTUELLE PAUSEDAUER NACH PAUSEERKENNUNG, T2=T2...-13
 TT3: AKTUELLE KNACKDAUER NACH KNACKERK., T3=T3...0
 AZUST: BIT 0 GESETZT: PROG. WORTENDE
 BIT 1 GESETZT: SICHERER WORTANFANG
 BIT 2 GESETZT: PROG. WORTENDE
 BIT 3 GESETZT: SICHERES WORTENDE
 BIT 4 GESETZT: XFELD-UEBERLAUF

PAUSE: PUSH HL
 PUSH DE
 PUSH RC

1-BYTE-VEKTOR UND KOMPONENTENSUMME ERMITTELN
 LD A, (SWHL)
 LD C, A
 LD E, 0
 LD H, 0
 LD A, M
 LD M, 0
 CP C
 RR D

FC: HL-SCHWELLVERT
 IB: KOMPONENTENZAehler
 IAKTUELLEN VEKTORPuffer LOESCHEN
 ID: 1-BYTE-VEKTOR, HL-SCHWELLE UEBERSCHR., #01

```

0360F ES
0360F DS
036A CS
0360 3A 0000
0360 0F
0360 1E 00
0371 06 00
0373 7E
0379 36 00
037A B9
0377 C0 14
  
```



Software-Umsetzung



Signalanalyse

- CTC-Kanal 0 und 3 arbeiten als Zähler mit Zählwert 1 und werfen beim Nulldurchgang von Tief- und Hochpass einen Interrupt (TP-NDG, HP-NDG)
- CTC-Kanal 1 arbeitet als Zeitgeber (Analyse-Zeitfenster), durch Abfrage des Zählwerts erhält man die Position innerhalb des Zeitfensters.
- Die NDG werten diese Position als Intervallklasse aus



Software-Umsetzung



- Die Interrupt-Routine des Zeitfensters sammelt die Intervallklassen von HP und TP 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.

Zeitliche Normierung

- ZEIN normiert das Feld XFELD zu einem 16 Byte-Feld YFELD (ähnlich Bild stauchen/strecken)



Software-Umsetzung



Weitere Routinen

- VGLEI: minimale Hamming-Distanz berechnen
- KNORM: Clusterbewertung normieren
- CLST: Cluster streichen
- CLUST: Clusterbildung

das alles waren Routinen des
Spracherkenners SE 9001, © PHONEI 1986

Was fehlt nun noch?



Rahmenprogramm



```
KC 85/1 Emulator
Emulator Ansicht Optionen Hilfe

Musteranzahl im Wortschatz: 0
Eingestellter Teilwortschatz: 1

A Wortschatz laden
B Lernen
C Wortschatz abspeichern
D Teilwortschatz aendern
E Klassen umbenennen
F Klassen loeschen
G Wortschatz loeschen
STOP Programmende

Kennbuchstabe:■

50 fps
```

```
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
```

- Lernen
- Nachlernen
- Klassifizieren
- Laden/Speichern

Erweiterung

- 5 Teilwortschätze mit je max. 50 Klassen; insg. max 200 Referenzmuster

Nutzung von BASIC

- über definierte Schnittstelle: **CALL RECOG** ⇒ Ergebnis Klasse (0..50)



homecomputer.ddd-de.vu



Danke!