

# EINIGE KRITERIEN FÜR DIE ERKENNUNG AKUSTISCHER SIGNALE

GEORG F. MEIER—JOHANNES KYRITZ

Die automatische Zeichenerkennung wird derzeit sowohl in biologischen wie auch in technischen Systemen erforscht. Für die Erforschung des technischen Vorganges, z. B. für automatische Diktieranlagen, kann ein rein technischer Weg eingeschlagen werden, d. h. daß die akustischen Parameter von Silben bzw. Lautübergängen und ihre Verarbeitung durch analoge oder digitale Datenverarbeitungsanlagen untersucht werden. Ein anderer Weg, der zugleich den menschlichen Kommunikationsvorgang, besonders die Dekodierung erfassen kann, ist der *bionische Weg*. Besonders der letztgenannte Weg soll nachfolgend behandelt werden. In dieser kurzen Mitteilung sollen nur 3 Probleme und Ergebnisse herausgegriffen werden:

## 1. ROLLE DER SEMANTIK BEI DER SIGNALAUFNAHME

Über die Relevanz oder Redundanz einzelner Elemente (Segmente) eines Textes können einige allgemeine Gesetze aufgestellt werden:

1.1 Jeder Satz hat als Ganzes eine Kontaktbedeutung, die oft nur durch ein einziges Element (Fragepartikel) oder durch wenige seiner Elemente getragen wird (Inversion, Frageton usw.). Man spricht auch vom „kommunikativen Aspekt“.

1.2 Jeder Satz stellt einen Komplex von Bedeutungselementen (Noemen) dar, deren Träger die Lexeme, Lexemgruppen, Positionen der Lexeme, Toneme oder Tonemfolgen und andere Prosodeme sind.

1.3 Die Noeme bestimmter Syntagmen bilden untereinander Durchschnittsmengen aus den Vereinigungsmengen der am Syntagma beteiligten Sememe, wenn zwischen diesen Sememen Verträglichkeitsrelationen bestehen.

1.4 In der Ontogenese verknüpfen sich viele häufig vorkommende Lexeme zu dynamischen Stereotypen, so daß die Auslösung des Anfangsgliedes genügt, um den Rest des Stereotyps abfolgen zu lassen.

1.5 Die Satzgenerierung erfolgt vom Gedanken zur Form: bestimmte semantische Elemente oder Komplexe steuern die formalen Stereotypen. Lückenteste und Ratefolgeteste haben gezeigt, daß es semantische stochastische Prozesse (Markov-Ketten) gibt. Laute, Silben und teils auch Wörter können objektiv entfallen, ohne daß sie beim Perzeptionsprozeß vermißt werden und ohne daß es ein Mißverständnis gibt.

1.6 Wirkliche Oppositionen in der aktuellen Rede gibt es nur, wenn der gesamte Kontext völlig gleich ist und der Text sich nur in dem betreffenden Element unterscheidet.

1.7 Phonetische Stereotypen, denen semantische Folgen zugeordnet sind, sind teilweise redundant. Die Länge des Kontextes ist proportional zur Redundanz einzelner Elemente. Lautverwechslungen, Überschreiten der Invarianzgrenzen und Homophonie erschweren das Verständnis umgekehrt proportional zum Informationswert der Umgebungselemente.

1.8 Die Phonemabgrenzung beruht nicht zwingend auf semantischen Kriterien, sondern auf akustischen (statistischen) Gewohnheiten.

## 2. ROLLE BESTIMMTER ABHÄNGIGKEITEN ZWISCHEN FORM UND FUNKTION

Alle formalen, d. h. grammatischen und prosodischen Erscheinungen müssen auf ihre kommunikative Funktion hin geprüft werden. Als Beispiel sei der Komplex der Intonation herausgehoben: Die physikalische Abhängigkeit eines Tones sei dargestellt durch:  $I = f$  (Formanten, Grundfrequenz, Zeit, Amplitude) =

$$f(p_1, p_2, p_3, p_4).$$

Ergeben mehrere Töne zusammen eine Tonfolge:  $I_1 I_2 I_3 \dots I_n$ , so entsteht eine Tonfolge-Relation  $R_{seq}(I_1, I_2, I_3, \dots, I_n)$ . Für den einfachsten Fall einer Tonfolge mit zwei Tonelementen ergibt sich folgendes Abhängigkeitsbild:

$$\begin{aligned} \{R_{seq}[I_1 = f(p_1, p_2, p_3, p_4), I_2 = f(p'_1, p'_2, p'_3, p'_4)] = \\ = f[R_i(S_1, \#, S_2) + R_{prop}(S, 2S) + R_{struct}(W, S_i) + \\ + R_{struct}(S_i, \text{Syntagma} > W) + \text{voix}(Sp)] = \\ = f\{R_{loc}(Sp, Hr) + R_{psych}(Sp, Hr) + \\ + R_{isom}[(S_1, S_2), (N_1, N_2, \dots, N_m)] + Z_{physiol}(Sp) + \text{Intentio}(Sp) \dots\}. \end{aligned}$$

Dabei bedeutet:  $R$  = Relation,  $t$  = Zeit,  $prop$  = Proportion,  $S$  = Silbe bzw. Tonträger,  $struct$  = Struktur,  $W$  = Lexem,  $voix$  = Stimmlage,  $loc$  = lokal,  $psych$  = psychisch interpersonell,  $Sp$  = Sprecher,  $Hr$  = Hörer,  $isom$  = Zuordnung,  $N$  = Noem,  $Z$  = Zustand

Diese Formel besagt, welche Faktoren bei dem Zustandekommen der Intonation hauptsächlich mitwirken.

Eine wichtige Rolle spielt dabei und bei vielen anderen Abhängigkeiten der Zeitfaktor.

## 3. BESTIMMUNG DES ZEITFAKTORS

Alle Messungen, seien es technische oder bionische, müssen objektiv sein. Ich verweise auf zahlreiche Arbeiten mit Segmentiergeräten und sogenannten Zeitfilteranlagen (wie z. B. von Tscheschner entwickelt). Es geht dabei zunächst um die Ermittlung bestimmter perceptiver Merkmale von Lautsegmenten in Abhängigkeit von ihrer zeitlichen Dauer, um die Feststellung des Einflusses von Verdeckungseffekten zeitlich

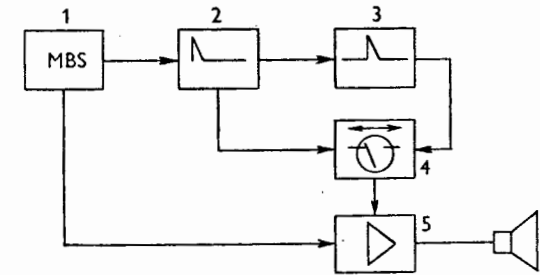


Fig. 1. Prinzipielles Blockschaftbild eines Zeitfilters.

vorausliegender Laute auf das folgende zu untersuchende Segment, um die Ermittlung der zeitlichen Dauer von Lautübergängen und Lauten für verschiedene Zwecke technischer und psycholinguistischer Art, so z. B. auch für die Herstellung des Ausgangsmaterials zur Durchführung von Lückentesten, bei denen die semantische Wirkung des inhaltlichen oder auch lautlichen Kontextes auf den Grad seiner Mächtigkeit, eliminierte Textelemente zu ergänzen, untersucht wird. Unser Mitarbeiter J. Kyritz hat dafür folgende Variante des Segmentierens entwickelt:

Fig. 1 zeigt das grobe Prinzip. Die 5 Funktionseinheiten bedeuten: 1 = Signalpeicher (Tonbandschleifengerät), wobei die Speicherzeit  $T_n$  von der Länge der Bandschleife und der gewählten Bandtransportgeschwindigkeit abhängt. 2 = Startimpulsgeber, der den Beginn jedes neuen Bandschleifenumschlages, also den Zeitpunkt  $t_0$  durch Abgabe eines Impulses signalisiert. 3 = Impulsverzögerung, ein Zeitglied mit der Aufgabe, einen Ausgangsimpuls bei einem frei wählbaren Zeitpunkt  $t_m$  abzugeben. 4 = Schalter, der den Verstärkerweg vom Bandschleifengerät zum Abhörlautsprecher, gesteuert durch die vom Startimpulsgeber und vom Impulsverzögerer einlaufenden Impulse, ein- und ausschaltet bzw. umgekehrt. 5 = Verstärker mit Lautsprecher oder Kopfhörer, der durch den Schalter auf- und zugetastet wird.

Die Untersuchung eines speziellen phonetischen Problems fordert von dem zu verwendenden Zeitfilter höchstmögliche Genauigkeit in bezug auf Festlegung und Rekonstruierbarkeit der Schaltzeitpunkte, die außerdem zeitlich verschiebbar und meßtechnisch eindeutig erfaßbar sein sollen. Diese Forderungen lassen sich durch die Anwendung digital gesteuerter Impulsverzögerer und durch synchrone Speicher

zung der steuernden Impulse zusammen mit dem zu untersuchenden Signal verwirklichen.

Fig. 2 zeigt den detaillierten Aufbau schematisch an.

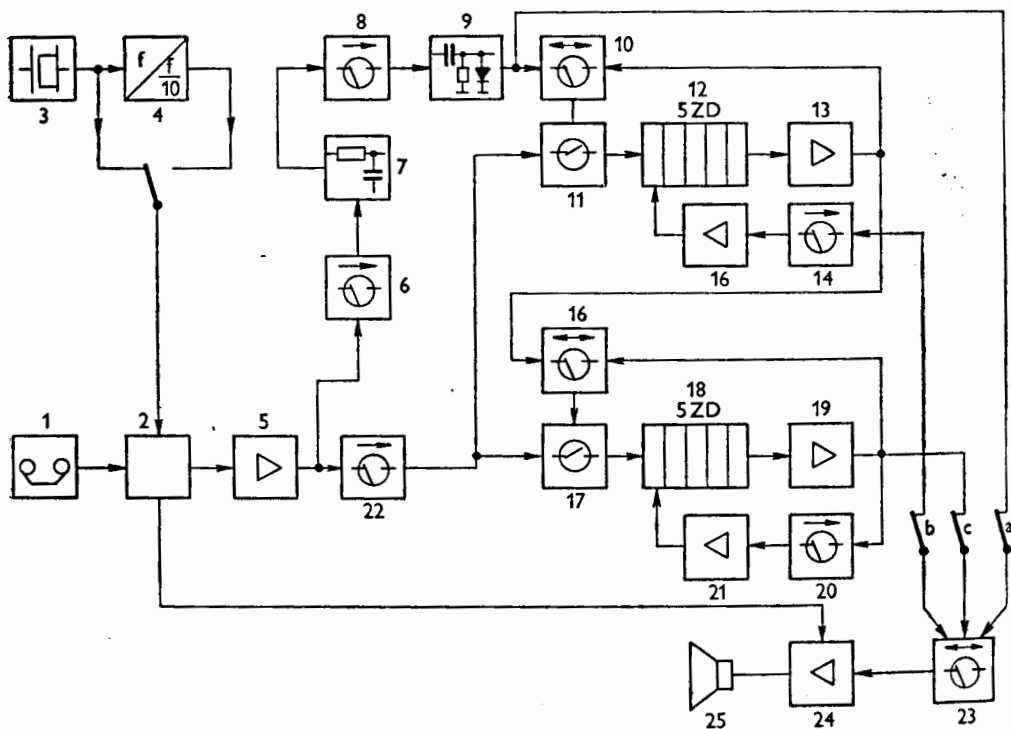


Fig. 2. Funktionsschaltbild eines im Institut für Phonetik und Kommunikationswissenschaft der HU Berlin entwickelten digital gesteuerten Zeitfilters mit optimaler Genauigkeit.

Legende:

1 Tonbandgerät, von dem die zu untersuchende Signalfolge auf das Bandschleifengerät überspielt wird. 2 Bandschleifengerät. 3 Temperaturstabilisierter quarzgesteuerter 10 kHz-Frequenzgenerator. 4 Frequenzuntersetzer 1 : 10. 5 Steuersignalverstärker. 6 Impulsformer (Monovibrator). 7 Integrationsglied. 8 = 6. 9 Differenzierglied. 10 Steuerbivibrator. 11 Torstufe. 12 5 hintereinandergeschaltete Zähldekaden. 13 Ausgangsimpulsverstärker. 14 Impulsformer. 15 Impulsleistungsverstärker. 16 = 10. 17 = 11. 18 = 12. 19 = 13. 20 = 14. 21 = 15. 22 = 14. 23 Schaltbivibrator. 24 Getasteter NF-Verstärker. 25 Lautsprecher.

Bem. Da der Bauelementaufwand bei herkömmlich aufgebauten Zähldekaden relativ hoch ist, wurden zwei Zeitfiltervarianten mit Schältröhren entwickelt. Variante 1 benutzt pro Dekade ein Dekatron mit direkt gekoppelter Ziffernanzeigeröhre (Z 572 S und Z 560 M, Hersteller: VEB Werk für Fernseh elektronik, Berlin, DDR). Variante 2 verwendet gasgefüllte dekadische Kaltkathodenzählröhren mit schräggestellten Kathoden (Typ N 4, Hersteller: Deutsche Glimmlampengesellschaft Pressler, Leipzig).