VERFAHREN DER ELEKTRONISCHEN SPRACHSIGNALVERARBEITUNG FÜR DIE EXPERIMENTALPHONETIK

Manfred Krause,

Technische Universität Berlin

1. Einleitung

In der Experimentalphonetik spielt die Analyse sprachlicher Äußerungen eine große Rolle. Diese konnte in hinreichend vollkommener Form bisher nur mit sehr teuren Geräten und Rechnern durchgeführt werden. Aber auch dann setzte die Aufgabe eine sehr gute Kenntnis der Geräte voraus und die einzeln Analyse konnte nicht über Sequenzen im Sekundenbereich ausgedehnt werden, was den Arbeitsaufwand beträchtlich erhöhte, z.B. bei Gesprächsanalysen oder Dialektuntersuchungen.
Neuere, auch auf PCs verwendbare Signalanalyse-Programme weisen eine Mächtigkeit auf, die weit über die Anforderungen der Phonetik hinausgehen. Damit setzt die Benutzung der Programme sehr gute Software-Kenntnisse voraus und die Arbeitsabläufe werden durch viele Menüabfragen verlangsamt. Im Rahmen von Seminaren zur experimentellen Phonetik wird im Fachgebiet Kommunikationswissenschaft der TU Berlin versucht, ein Programmsystem und die notwendige Hardware zu erstellen, die auf die Belange der Sprachsignalanalyse für Experimentalphonetiker zugeschnitten sind. Die Arbeiten sind noch im Gange und dieser Beitrag soll zugleich Diskussionsgrundlage dafür sein, welche Anforderungen über die bisher üblichen hinaus von Phonetikern an ein modernes Analysesystem gestellt werden.

2. Apparative Anforderungen

Bei unserem Vorhaben sind wir davon ausgegangen, daß als Rechner IBM-kompatible PCs wegen ihrer weiten Verbreitung zweckmäßig sind. Damit ergeben sich bereits bestimmte Konsequenzen für die Hardware- und Software-Komponenten zur Datenaufnahme und -ausgabe. Zur Signalumsetzung genügt für die Sprachanalyse i. allg. ein 12-Bit-Analog/Digital-Umsetzer, wenn eine Aussteuerungskontrolle vorgesehen wird. Eingänge für Mikrofone und hochpegelige Signale sind vorzusehen. Die Abtastrate sollte in weiten Grenzen zwischen 50 kHz und 5 kHz veränderbar sein, wobei der A/D-Umsetzer mit einem softwaregesteuerten Antialiasing-Filter versehen sein sollte. Da in Zukunft auch digitale Tonaufzeichnungsverfahren verwendet werden, muß eine digitale Schnittstelle vorgesehen werden

können, die ggf. die meist vorgegebene 16-Bit-Datenstruktur in die 12-Bit-Datenstruktur umsetzt.
Um längere Sprachsequenzen verarbeiten zu können, muß der einkommende Datenstrom direkt auf die Festplatte übertragen werden. Ein DMA-Programm (DMA = Direct Memory Access) ist daher erforderlich. Entsprechende Anforderungen gelten auch für die Ausgabe von Sprachsignalen, für die die 12-Bit-Datenstruktur ausreichend ist. Diese Leistungsmerkmale weisen nur wenige Programmsysteme auf. Eine Schnittstelle zu digitalen Tonaufzeichnungsgeräten ist wünschenswert.

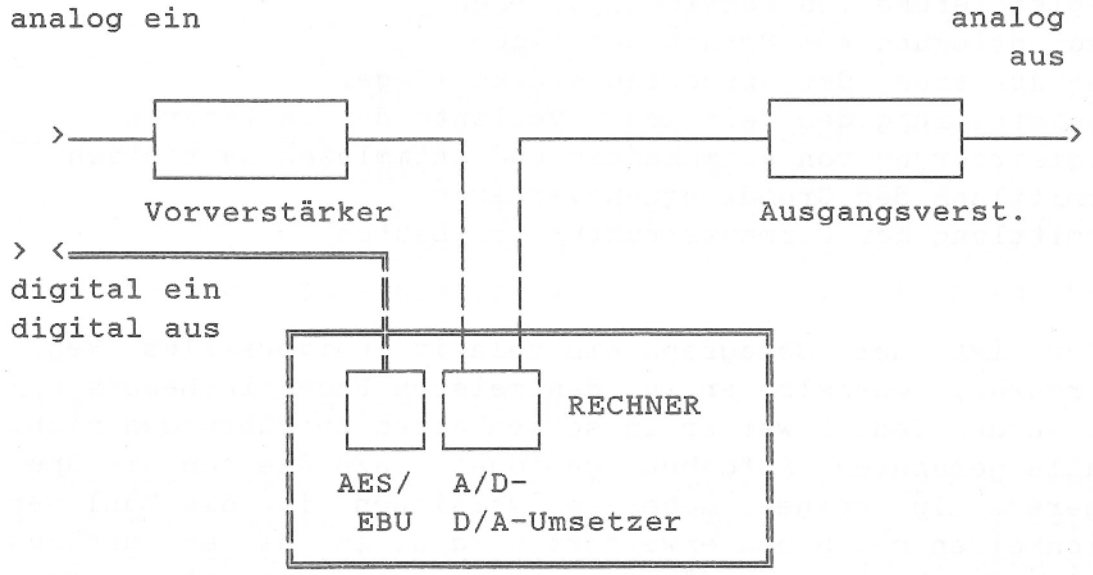


Bild1: Prinzipielle Struktur des Signalverarbeitungssystems für die Experimentalphonetik

Der Rechner selbst sollte eine möglichst hohe Taktfrequenz besitzen und mit einem Koprozessor ausgestattet sein, um einen hinreichend schnellen Arbeitsablauf zu gewährleisten.

2. Grundaufgaben in der experimentellen Phonetik

Wenn man die früher üblichen Werkzeuge des Experimentalphonetikers auflistet, so findet man :

- Magnetbandgerät
- Oszilloskop
- Schwingungsschreiber
- Pegelschreiber
- Sonagraph

und dazu Spezialgeräte, wie:

- Grundfrequenz-Analysator
- Nulldurchgangsraten-Analysator
- Segmentator
- Glottograph

Die Grundaufgaben des Experimentalphonetikers bestehen u.a. in :

- Aufzeichnung der Sprachsignale
- Registrierung von Schwingungsformen
- Segmentierung von Sprachlautfolgen
- Registrierung der Sprechlautstärke (Pegel)
- Registrierung des zeitlichen Verlaufs der Lautstärke
- Registrierung von stimmhaften und stimmlosen Lautfolgen
- Ermittlung des Grundfrequenzverlaufs
- Ermittlung der Formantstruktur von Lauten

Hierfür ist der Sonagraph ein relativ universelles Registriergerät, weshalb er in den meisten Phonetik-Labors benutzt wurde. Jedoch war er in seinen alten Ausführungen nicht für alle genannten Aufgaben geeignet. Dazu dienten die Spezialgeräte. In seinen neuesten Versionen ist die Zahl der Möglichkeiten bedeutend erweitert worden, so daß er durchaus als eine "work station" für den Phonetiker angesehen werden kann. Allein der Preis schränkt seine allgemeine Nutzung ein.

Mit PCs ist dem Sonagraphen eine ernsthafte Konkurrenz erwachsen. Die immer weiter zunehmende Zahl von Signalverarbeitungsprogrammen macht den auch anderweitig nutzbaren PC in Verbindung mit der geeigneten Hardware ebenfalls zu "work stations" für den Experimentalphonetiker. Als Vorteil für den PC mag gelten, daß es leichter möglich ist, eigene Programme zu integrieren, die einer selbst gestellten Aufgabe angepaßt sind.

Von dieser Universalität des PC ausgehend, können Überlegungen angestellt werden, welche Funktionen für die Experimentalphonetik in einem Signalverarbeitungsprogramm verwirklicht werden sollten.

3. Funktionen eines Signalverarbeitungsprogramms für die Phonetik

3.1 Editieren von Sprachlautfolgen

Für die Weiterverarbeitung von Sprachsignalen ist i. allg. eine Sichtung und Ordnung des aufgenommenen Materials erforderlich. Hierbei müssen unbrauchbare Teile, wie Pausen oder Störungen entfernt und ggf. später aufgenommene Sequenzen eingefügt werden. Gegenüber dem früher üblichen Editieren eines Tonbands durch Schneiden, das heute wegen der üblichen Kassettentechnik bzw. bei Digitalaufnahmen nicht mehr praktiziert werden kann, muß ein bequemes Verfahren für Rechner bereitstehen. Dieses Verfahren muß in anschaulicher graphischer Form einen Überblick über den gesamten Verlauf der Signaldatei bieten und unter auditiver Kontrolle des Signals einen manipulierenden Zugriff auf jeden Signalausschnitt erlauben. In der Tonstudiotechnik sind solche Verfahren bekannt.

3.2 Schleifenbildung und Segmentation

Die auditive Analyse ist ein nicht gering einzuschätzender Teil der Experimentalphonetik. Für das Erlernen des Transkribierens in phonetische Schreibung ist das Segmentieren von Sprachlaufolgen wichtig. Das wiederholte ausschnittweise Abtasten der Sprachsignale, früher durch geklebte Tonbandschleifen und durch getriggerte elektronische Schalter erzielt, kann mit Rechnern wesentlich einfacher, schneller und genauer durchgeführt werden. Insbesondere das langsam im Signalverlauf fortschreitende Segmentieren, das genaue Positionieren oder das langsame Vergrößern oder Verkleinern der Schalteröffnungszeit unter auditiver Kontrolle erleichtert die Vorbereitung weiterführender Analysen.

3.3 Ermittlung von prosodischen Merkmalen

Ein Aspekt der Erforschung der Lautsprache ist die Prosodie. Als prosodische Merkmale, die aus Sprachsignalen extrahierbar sind, kommen die Verläufe von Lautstärke (Pegel und Lautheit), Grundfrequenz, Spektra und Laut- bzw. Pausendauerveränderungen in Betracht.

3.3.1 Pegel, Lautstärke und Lautheit

Handelsübliche Signalverarbeitungsprogramme enthalten nur Module zur Ermittlung des Signalpegelverlaufs. In einem Programmsystem für die Experimentalphonetik ist es empfehlenswert, auch Module zur Bestimmung des Lautheitsverlaufs nach ZWICKER /1/ vorzusehen. Solche Programme sind an verschiedenen Stellen entwickelt worden /2/. Mit Lautheitsverläufen wird der subjektiven Empfindung besser Rechnung getragen.

3.3.2 Grundfrequenz

Die Ermittlung des Grundfrequenzverlaufs ist immer noch ein Problem für die Sprachsignalverarbeitung. Es gibt kein Programm, das für alle Stimmen fehlerfrei arbeitet. Daher ist es zweckmäßig, hierfür mehrere verschiedene Algorithmen zur Auswahl zu haben. Bedeutungsvoll ist auch die Erfassung kleiner Schwankungen der Grundfrequenzperioden (jitter), die u.a. zur Natürlichkeit synthetisierter Sprache beitragen /3/. Grundfrequenzschwankungen haben neben der pathologischen auch prosodische Signifikanz.

3.3.3 Spektra und Formantverläufe

Die Darstellung spektraler Zusammenhänge in der Experimentalphonetik ist durch die Sonagraphie seit langem eingeführt. Je nach der Einstellung der Frequenz- bzw. Zeitauflösung kann die Darstellung mehr frequenz- bzw. zeitorientiert sein. Bei den handelsüblichen Signalverarbeitungsprogrammen wird die schnelle Fourier-Analyse (FFT) benutzt, obwohl es auch andere Alternativen für die Frequenzanalyse gibt. Eine davon ist die FTT-Analyse nach TERHARDT /4/, die für die Sprachsignalanalyse wegen der Berücksichtigung von Gehöreigenschaften günstig sein kann. Dies trifft auch auf die Darstellung von Teiltonzeitmustern nach HEINBACH zu /5/. Aus der bekannten Breitband-Darstellung der Sonagramme kann der Experimentalphonetiker mit einiger Erfahrung visuell oder von Hand leicht Formantverläufe extrahieren. Es gibt zwar Programme für das automatische Verfolgen (tracking) von spektralen Maxima; sie erfüllen jedoch noch nicht alle Ansprüche auf Fehlerfreiheit. Ein besonderer Aspekt bei der sonagraphischen Darstellung ist die Skalierung der Frequenzdimension. Für den Experimentalphonetiker kommt es häufig darauf an, Laute, deren Formantlagen nur geringfügig voneinander abweichen, zu unterscheiden. Weder die lineare, die logarithmische noch die in Bark geteilte Frequenzachse sind hierfür vollkommen geeignet. Hierfür sind evtl. andere nichtlineare Skalen notwendig.

4. Einige neue Module für die Experimentalphonetik

Im Rahmen verschiedener Aktivitäten auf dem Gebiet der Experimentalphonetik und der Entwicklung von Artikulationshilfen für Gehörlose wurden von uns in den letzten Jahren verschiedene aus der Literatur bekannte Module für die Sprachsignalverarbeitung erprobt und einige selbst entwickelt. Als Ausgangssoftware standen verschiedene Signalverarbeitungsprogramme zur Verfügung, u. a. ein im Heinrich-Hertz-Institut Berlin entwickeltes Programm SIPRO 23 /6/.
SIPRO 23 enthält Module, die auch Belange der Signalverarbeitung in Tonstudios, d.h. für sehr hochwertige Signale, geeignet sind. Beim Entwurf dieses Programms wurde auf eine sehr elaborierte graphische Oberfläche Wert gelegt, mit der die Vielfalt der Signalverarbeitungsprozesse gestaltet und ausgelöst werden kann.
Wie bei allen bekannten Signalverarbeitungsprogrammen wird von der Mächtigkeit des Programmsystems in der Experimentalphonetik jedoch nur ein kleiner Teil benötigt. Daher erweist sich das System für diesen Zweck als zu umständlich handhabbar.
In unseren Untersuchungen wurde das System als Basis benutzt. Eingebunden wurden im System nicht enthaltene Module für die Umsetzung von Abtastraten, Darstellung von Pegel- und Grundfrequenzverläufen und von Sonagrammen.
Kurz vor dem Abschluß der Entwicklung sind Programme für die dynamische Zeitanpassung und das Formanttracking. Die dynamische Zeitanpassung wird z.B. für Dialektuntersuchungen verwendet, da ein Teil dialektaler Merkmale in der Zeitstruktur der sprachlichen Äußerungen zu finden sind. Für das Formant-Tracking werden die Hüllkurven der Frequenzspektren mit Hilfe der AKIMA-Interpolation ermittelt /7/. Ein Beispiel zeigt Bild 2. Die AKIMA-Interpolation bildet die Hüllkurven, wie sie auf visuellem Weg gewonnen werden, befriedigen gut nach. Die Maxima der Hüllkurven werden als Formantfrequenzen weiter verwertet.

5. Gestaltungsprinzipien für die Benutzeroberfläche

5.1 Menüs und Optionen

Die Gestaltung von Benutzeroberflächen muß sich an der Komplexität der Arbeitsabläufe orientieren, sollte aber die Freiheit des Anwenders, sich neue Arbeitsmöglichkeiten zu schaffen, nicht einschränken.
Zweckmäßig wird für die Programmodule eine Baumstruktur gewählt, deren Tiefe nicht mehr als zwei Stufen umfaßt.

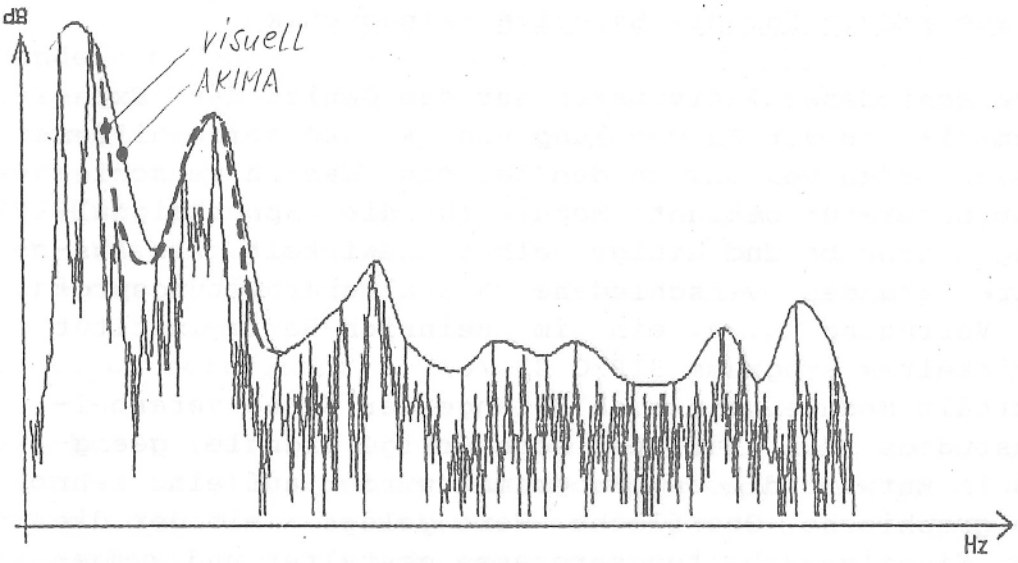
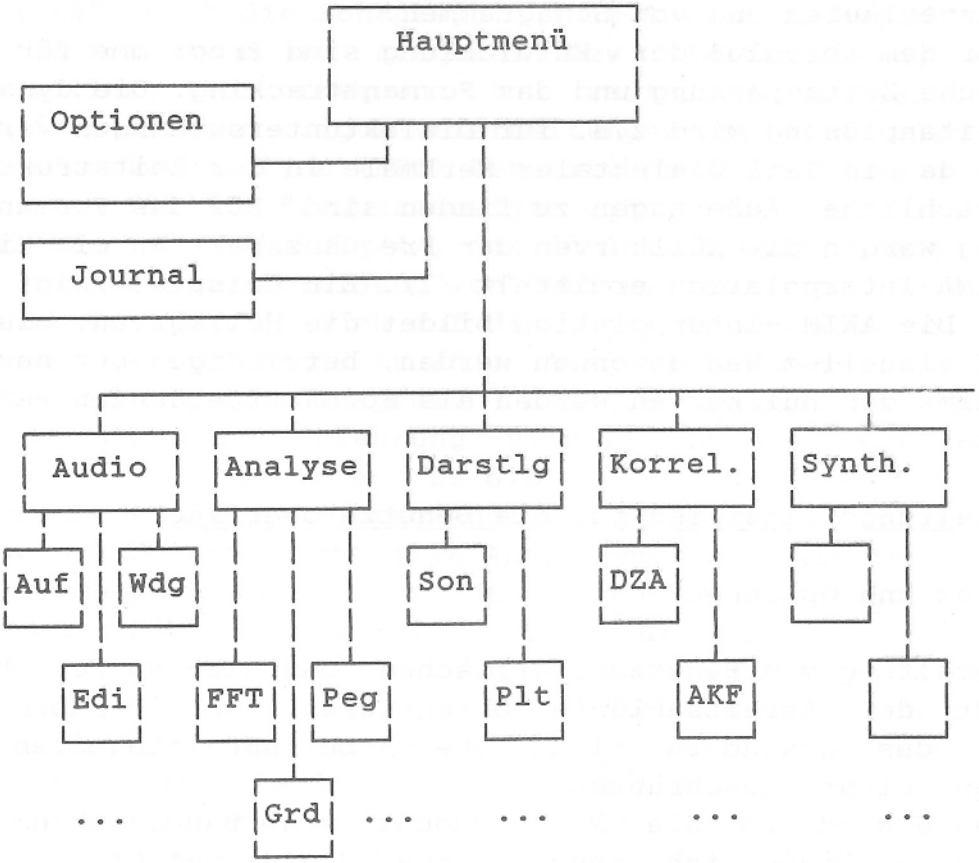


Bild 2: Hüllkurve eines Frequenzspektrums ermittelt mit der AKIMA-Interpolation im Vergleich zu einer visuell ermittelten Hüllkurve /7/

Die Gestaltung der Arbeitsfenster sollte so einheitlich wie möglich sein, was insbesondere für die Hauptsteuerfunktionen gilt, die durch das System leiten. Die Funktionen zur Steuerung des Arbeitsablaufs sollten hierarchisch geordnet sein.



Optionen sollten die für eine Arbeitsphase erforderlichen Einstellungen von Hard- und Software enthalten, wie z.B. Abtastfrequenz, Analyseformate, Darstellungsformate u.s.w. Die Optionen sollten aus jedem Anwendungsbereich heraus veränderbar sein; günstig ist es, mehrere Optionen voreingestellt zur Verfügung zu haben.

5.2 Dateiverwaltung

Bei unseren bisherigen Untersuchungen hat es sich als zweckmäßig erwiesen, alle Dateien automatisch zu verwalten. Neben frei wählbaren Namen ist es angebracht, die einzelnen Anwendungen durch die Extensionen der Dateinamen zu kennzeichnen und an den Namen laufende Unterscheidungszeichen anzubringen. Der gesamte Dateiverkehr sollte in einem Journal verzeichnet werden.

X. Literatur

/1/ E. Zwicker, R. Feldtkeller: Das Ohr als Nachrichtenempfänger, S. Hirzel Verlag Stuttgart, 1967 und DIN 45631

/2/ E. Paulus, E. Zwicker: Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln, Acustica 5, 1972, 253 - 266

/3/ D. G. Childers, Ke Wu: Quality of Speech Produced by Analysis-Synthesis, Speech Communication 9 (1990), 97 - 117

/4/ E. Terhardt: Fourier Transform of Time Signals, Acustica 57, 1985, 242 - 256

/5/ Heinbach: Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster, Dissertation TU München 1988

/6/ T. Selinger: Signalerzeugung und Darstellung von Signal verläufen in zwei- und dreidimensionealer Form an einem AT-System, Magisterarbeit am Fachgebiet Kommunikationswissenschaft der TU Berlin 1988

/7/ R. Wegener: Analyse und Implementierung eines geeigneten Interpolationsverfahrens für die Darstellung der Formantverläufe von Sprachsignalen, Studienarbeit am Fachgeb. Kommunikationswiss. TU Berlin 1990