PHONETISCHE ARBEITSSTATIONEN: HARDWARE UND SOFTWARE ZUR EXPERIMENTELLEN SPRACHSIGNALVERARBEITUNG

Michel T. M. Scheffers, Werner Thon,
Institut für Phonetik und digitale Sprachverarbeitung (*ipds*)
der Universität Kiel

1. Übersicht

Ausgehend von den Erfahrungen mit der digitalen Sprachsignalverarbeitung auf einem Minicomputer mit dem Programmpaket SSP (speech signal processor) des *ipds* wurde ein Netzwerk von 'Phonetischen Arbeitsstationen'(PAS) mit einem Softwarepaket ASSP (advanced speech signal processor) implementiert, das es Phonetikern erlaubt, Aufgaben der phonetischen Experimentsignalverarbeitung rechnergestützt am Arbeitsplatz durchzuführen. Hierzu gehören insbesondere:

* Aufnahme und Speicherung beliebig langer Sprachsignale (begrenzt nur durch die Festplattenkapazität)
* akustische Wiedergabe beliebiger Stücke aus diesen Signalen
* Analyse der Signale hinsichtlich verschiedener Parameter (Formanten, LPC-Koeffizienten, Grundfrequenz, Energie) mit verschiedenen Analysealgorithmen
* Flexible Darstellung von Signalen und Parameterverläufen mit hoher Auflösung
* Interaktive Manipulation von Signalen und Parameterverläufen
* Resynthese von Sprachsignalen aus manipulierten Parameterverläufen
* Durchführung von Hörtests mit gezielt manipulierten Sprachsignalen

2. Hardware

2.1 Struktur der PAS

Die für diese Aufgaben benötigten Hardware-Komponenten sind weitgehend in einem relativ kompakten Gerät untergebracht. Bild 1 zeigt die Hardware-Struktur einer PAS. Der Rechnerkern auf VMEbus-Basis besteht aus einem Einplatinen-Rechner mit einer Motorola 68000 CPU, 1MB RAM, einem National 32081 Gleitkommaprozessor, einem SCSI-Host-Adapter und zwei seriellen Schnittstellen. Er wird ergänzt durch eine Speicherkarte mit 4MB RAM. Über den SCSI-Bus steuert ein SCSI-Controller die Massenspeicher: ein 5.25" Disketten-Laufwerk, ein 60MB-QIC2-Magnetband-Kassetten-Laufwerk und eine 65MB-Festplatte. Eine serielle Schnittstelle wird von der Tastatur belegt, die zweite ist frei für Drucker o.ä. An die Tastatur kann wahlweise eine Maus oder ein Trackball angeschlossen werden.

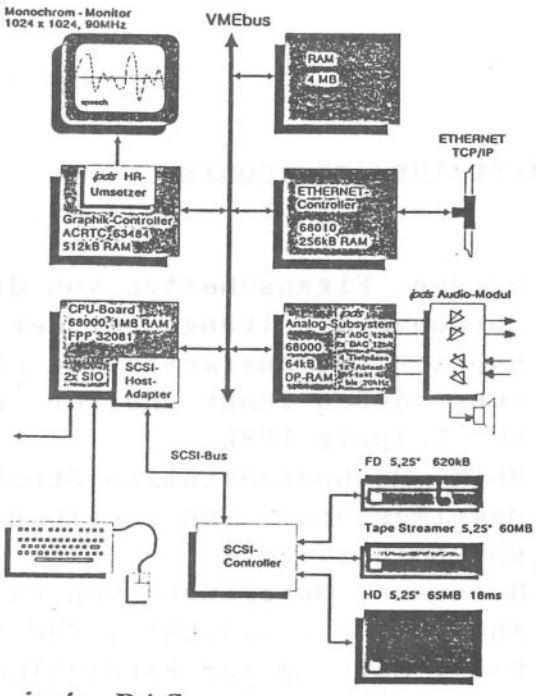

Bild 1 Struktur der Phonetischen Arbeitsstationen

Die graphische und alphanumerische Bildschirmausgabe erfolgt über eine Graphik-Controller-Karte mit einem Hitachi ACRTC 63484 und 512kB Bildspeicher. Auf ihr befindet sich ein im *ipds* entwickelter Umsetzer, der ein monochromes Videosignal mit 1024x1024 Bildpunkten und 90MHz Videotakt erzeugt. Dies wird auf einem 15"-Monitor dargestellt.

Das Analog-Subsystem besteht aus zwei Halbkarten. Der industriell gefertigte Teil enthält einen eigenen 68000-Prozessor mit 64kB 'dual ported' RAM, das auch im Adressraum der Haupt-CPU eingeblendet ist. Der im *ipds* entwickelte Teil enthält zwei 12bit-A/D-Wandler, zwei 12bit-D/A-Wandler, vier Tiefpässe mit software-steuerbarer Grenzfrequenz und einen software-steuerbaren Taktgenerator als gemeinsame Abtasttaktquelle für die vier Wandler. Die im *ipds* entwickelte Firmware des Subsystem-Prozessors erlaubt neben der Einstellung der Tiefpaßgrenzfrequenz und der Abtastrate die kontinuierliche Wandlung und den Datentransfer von bzw. zu der Festplatte in zwei Kanälen mit je max. 20kHz Abtastrate. Dies können wahlweise zwei Eingangskanäle, zwei Ausgangskanäle oder je ein Eingangs- und Ausgangskanal sein.

Dem Analog-Subsystem vorgeschaltet ist das im *ipds* entwickelte Audio-Modul, das alle weiteren Analog-Komponenten, wie Verstärker, einen Kontroll-Lautsprecher, und eine Bedienungsfrontplatte mit Kanalschalter, Lautstärkestellern, Ein/Ausgangsbuchsen sowie eine vom Analog-Subsystem gesteuerte digitale Betriebszustandsanzeige enthält.

Insgesamt bietet die Hardware und Software der PAS-Analogschnittstelle dem Benutzer die gleichen Möglichkeiten wie eine Stereo-Bandmaschine.

2.2 Netzwerk

Die sechs implementierten PAS sind an das ETHERNET-Netzwerk des *ipds* angeschlossen, das durch alle Räume des Instituts verlegt ist. Die ETHERNET-Controller der PAS verfügen über einen separaten 68010-Prozessor und 256kB Pufferspeicher. Das Netzwerk-Protokoll TCP/IP, das als Firmware im Controller vorhanden ist, ermöglicht den problemlosen Transfer von Signaldateien zwischen den Arbeitsstationen und insbesondere auch zu anderen Rechnern des Netzwerks mit verschiedenen Betriebssystemen. Neben den sechs PAS, die als Betriebssystem OS9/68k verwenden, sind zur Zeit zwei AT-Kompatible (MS-DOS) und eine Apollo DN3500 (UNIX SYSTEM V oder BSD 4.3) angeschlossen.

2.3 Ergänzungen und Erweiterungen

Ein Gerät zur Erfassung von Hörerurteilen und -reaktionszeiten bei Hörtests kann an jede der PAS angeschlossen werden. Im Hörtest wird den Hörern eine Reihe von systematisch manipulierten Sprachsignalen über die PAS-Analogschnittstelle angeboten. Die Hörer geben ihr Urteil durch Drücken einer von mehreren Tasten. Das Reaktions(zeit)meßgerät startet eine Zeitmessung gesteuert von der PAS an vom Experimentator festgelegten Zeitpunkten bezogen auf das Sprachsignal. Es überwacht die Tasten aller Hörer bis zu einer Maximalzeit; bei der ersten Reaktion jedes Hörers werden Zeitpunkt und Taste gespeichert. Am Ende des gesamten Tests werden die gespeicherten Meßdaten an die PAS zur weiteren Verarbeitung übertragen.

Es ist geplant, die PAS durch Austausch der CPU-Karten mit 68000/32081-Prozessoren durch solche mit 68040/68882-Prozessoren in ihrer Rechenleistung zu erweitern. Dies wird auch einen Ausbau des Arbeitsspeichers auf 8MB mit sich bringen und insbesondere die Geschwindigkeit der Analyse- und Synthese-Algorithmen erheblich steigern.

Die PAS dienen der Grundversorgung mit Rechenleistung für experimentelle Sprachsignalverarbeitung. Das weitverbreitete Netzwerkprotokoll TCP/IP erlaubt es, nahezu jeden für spezielle Aufgaben der Signalverarbeitung zugeschnittenen Rechner in das Netzwerk einzubinden und seine Rechenleistung den Benutzern der PAS verfügbar zu machen.

3. Software

3.1 Zielsetzung

Das Softwarepaket ASSP (advanced speech signal processor) wurde mit dem Ziel entwickelt, die unter 1. beschriebenen Möglichkeiten sowohl einem Anfänger (über eine komfortable, menügeführte Benutzer-Oberfläche) als auch einem erfahrenen Benutzer (über kommandozeilenorientierte direkte Aufrufe von ASSP-Befehlen und Unterprogrammen) mit einer einfachen Bedienung zur Verfügung zu stellen. Außerdem soll das Paket nicht nur für die Forschung, sondern auch für die Lehre benutzt werden. Das heißt, daß neben den Signalen zur interaktiven Bearbeitung auch solche Signale adäquat dargestellt werden können, die nur zu Demonstrationszwecken graphisch ausgegeben werden (z.B. Folgen von LPC-Koeffizienten usw.). Weiterhin wurde im Interesse einer leichten Wartung und Weiterentwicklung ein hohes Maß an Modularität des Pakets angestrebt. Dies ermöglicht es auch Programmierern anderer Signalverarbeitungsverfahren, in ihren Programmen durch Verwendung von ASSP-Standardroutinen zur Datenverwaltung auf existierende ASSP-Dateien zuzugreifen und neue Dateien zu erzeugen, die ohne weiteres von ASSP weiterverarbeitet werden können.

3.2 Datenstruktur

Zusammengehörige Signal- und Parameterdateien sind in besonderen Unterverzeichnissen (sogenannten AREAs) gespeichert. Jede AREA enthält eine Konfigurationsdatei in der globale Parameter (Abtastrate, LPC-Ordnung, Frame-Größe, Frame-Verschiebung, Zahl der Formanten usw.) abgespeichert sind. Diese Angaben gelten damit für alle Dateien in der AREA. Beim Erzeugen einer AREA hat der Benutzer die Möglichkeit, entweder eine aus einer Reihe von Standardkonfigurationen zu wählen oder eine speziell auf sein Problem zugeschnittene Konfigurationsdatei zu erstellen. Weiterhin enthält jede Datei einen Kopf mit Angaben über die Datenstruktur innerhalb der Datei, wodurch Routinen für Lesen, Schreiben und graphische Darstellung sehr allgemein formuliert werden konnten.

3.3 Programmstruktur

Wie Bild 2 zeigt, ist das ASSP-Paket aus drei Schichten aufgebaut: Die obere Schicht umfaßt die allgemeine ASSP-Benutzeroberfläche. Die mittlere Schicht wird vom ELK (Easy Language Kit)-Befehlsinterpreter gebildet. Die dritte Schicht ist eine Toolbox mit allgemeinen Datenverwaltungs- und Signalverarbeitungsroutinen.

Der Dialog auf der Benutzeroberfläche führt zur Erzeugung von ELK-Befehlen wie z.B. 'DISPLAY 3, (T1,T2)' ('Bilde für die ausgewählten Parameter sämtlicher Dateien, die Fenster Nr. 3 zugewiesen worden sind, den Zeitbereich T1 bis T2 ab'). Der ELK-Interpreter führt einen solchen Befehl entweder selbst mithilfe von Routinen aus der Toolbox aus oder ruft ein Unterprogramm auf. Jedes Unterprogramm verfügt über eine standardisierte, dialogorientierte Benutzerschnittstelle, kann aber auch mit einer Argumentliste aufgerufen werden. Falls die Argumentliste ausreichende Angaben für die Ausführung des Programms enthält, erfolgt kein Dialog mit dem Benutzer. Dadurch wird es möglich, zeitaufwendige Programme wie Analyse und Synthese auch als Hintergrundprozeß laufen zu lassen.

Der Benutzer kann mithilfe eines Texteditors oder über einen Makrobefehl der Oberfläche eine ELK-Befehlsdatei erzeugen und diese ausführen lassen. Z.B. kann für eine Abfolge von Parametermanipulationen mit gleicher Struktur eine Befehlsdatei erzeugt werden, die diesen Ablauf weitgehend automatisiert. Daneben können Unterprogramme direkt vom Benutzer aufgerufen werden (d.h.

ohne die ASSP-Oberfläche starten zu müssen). Dies bedeutet auch, daß neue Programme außerhalb des ASSP-Pakets entwickelt und getestet werden können. Um sie allen Benutzern zur Verfügung zu stellen, können sie entweder durch einfaches Hinzufügen eines Menüpunktes als Unterprogramm oder - nach Anpassung - als ELK-Befehl in das Paket eingebunden werden.

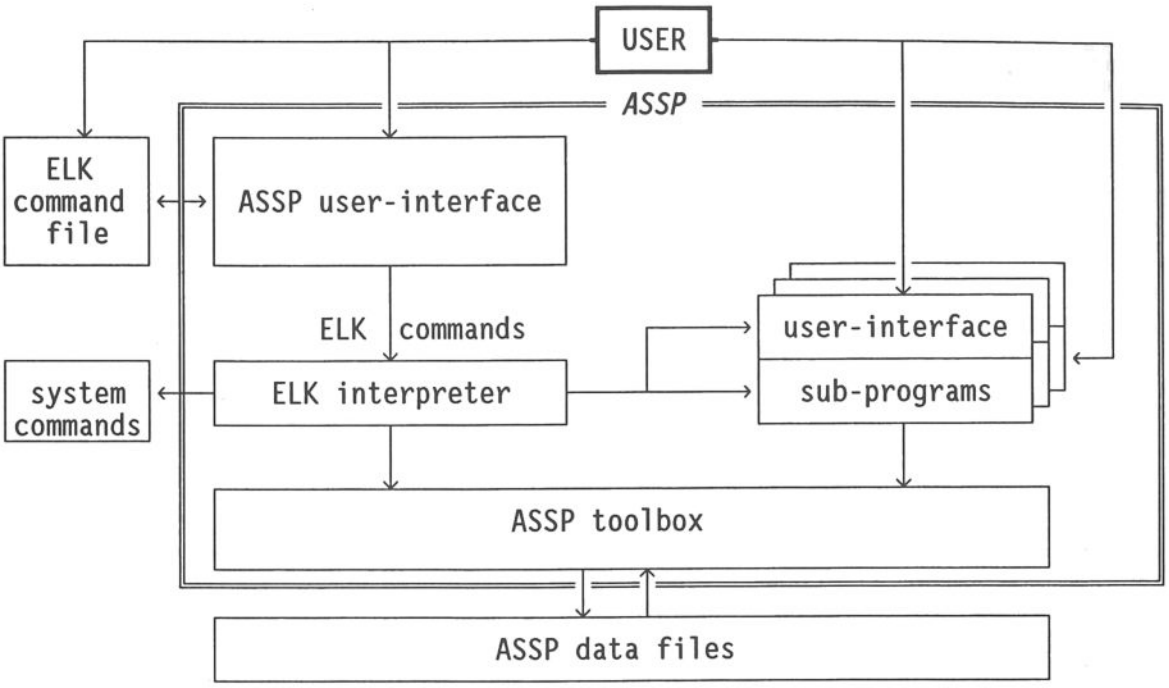

Bild 2 Struktur des Programmpakets ASSP

Aufgrund unserer Erfahrungen mit dem SSP-Paket nimmt die graphische Darstellung und interaktive Manipulation eine zentrale Stelle im Programmablauf ein (siehe Bild 3). Über das Hauptmenü können von dort aus Unterprogramme wie Aufnahme/Wiedergabe, Analyse/Synthese, nicht-interaktive Datenmanipulation und Datenverwaltung aufgerufen werden.

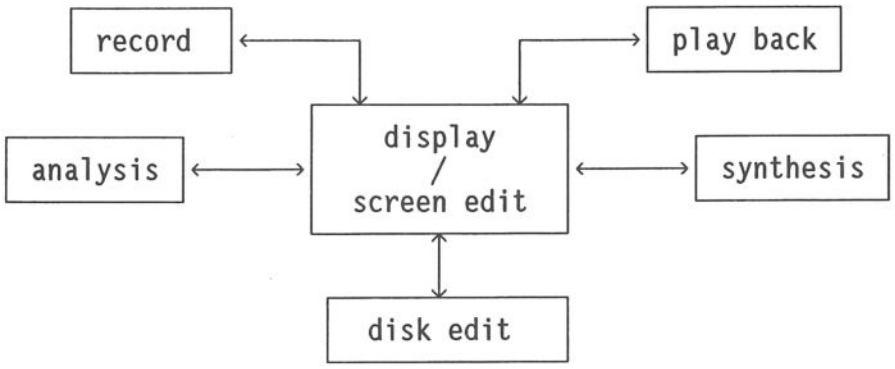

Bild 3 Programmablauf