Konzeption einer Tondatenbank

Integration von Ton, Text, Bild und Daten in S_Tools©_DB

Ralf Vollmann, Werner A. Deutsch, Anton Noll, Sylvia Moosmüller

1. Anwendungsbereiche

Im technisch-wissenschaftlichen Bereich, in der Linguistik, Musikethnologie, Volksmusikforschung, Ethnologie, Volkskunde, Musikwissenschaft, im Museumsbereich und in der Zoologie besteht ein zunehmender Bedarf an Tondatenbanken:

Technik: Straßen- und Maschinenlärm, der Lärm von Flug- und Schienenfahrzeugen, Schwingungen aller Art werden aufgezeichnet, gemessen und verglichen.

Sprache: In phonetischen, allgemein-phonologischen, dialektologischen, dialektgeographischen, ethnolinguistischen, soziolinguistischen Erhebungen, in der Aphasieforschung, in der Diskursanalyse werden Tonaufnahmen für die Analyse oder nur für die Archivierung, häufig im Feld, seltener im Studio, hergestellt.

Musik: Tonaufnahmen musikalischer Ereignisse (soziale Anlässe, ethnologische Studien), von Gesang (Tonhöhe, Melodie, Rhythmus, Singstile), von Instrumentalaufnahmen (Aufführungspraxis, Instrumentenakustik, ...) usw. werden häufig im Feld, sehr selten im Studio hergestellt.

Zoologie: Schreie und Rufe von Tieren, aber auch feinste Vibrationen (z.B. der Spinne im Spinnennetz) werden häufig im Regenwald oder in der Au, teilweise auch im Studio auf Tonträgern aufgenommen und gesammelt.

Schallarchivtechnik: Archive, Bibliotheken und Museen müssen ihre analog gespeicherten Tondokumente digitalisieren, um sie über die begrenzte Lebensdauer der Trägermedien hinaus erhalten zu können. Künftig ist es wichtig, die Daten zu erhalten und nicht Medien zu pflegen, gleichzeitig sollten historische Aufnahmen der Forschung oder den Interessenten besser zugänglich gemacht werden können.

Das Ergebnis all der vorgenannten Aktivitäten ist, physikalisch gesehen, immer eine gewisse Anzahl von vollbespielten Tonbändern mit entsprechend genauer Dokumentation und zumeist schriftlicher Protokollierung. Zunächst entsprechen die Tonaufnahmen in ihrer Chronologie der jeweiligen Aufnahmesituation und sind meistens nach der Schallquelle orientiert (Sprecher/in, Sänger/in, Instrument, Tier, Maschine, schwingendes Objekt). Je nach Applikation sind Wortlisten, Sätze, freie Rede, einzelne Lieder, Musikdarbietungen, Rufserien von Fröschen, Lärmereignisse nur sequentiell vom Tonträger abrufbar. Weder ein auditiver A/B-Vergleich noch vergleichende akustische Analysen sind ohne zeitraubende und qualitätsmindernde Umkopierung möglich. Das Umkopieren zerstört ferner den Originalkontext, womit die Möglichkeit für spätere, die akustische Umgebung umfassende Abfragen verhindert wird. Durch die Zerstörung des Originalkontexts werden die Daten zu erratischen empirischen Datenblöcken, die bloß einer einmal gestellten Fragestellung genügen und im übrigen nur allzuoft nutzlos sind oder werden, nämlich dann, wenn eine neue Fragestellung auftaucht, die mit dem zerschnittenen Material nicht mehr beantwortet werden kann. Man denke etwa an eine Wortsegmentierung, wenn man sich einige Jahre später mit der Satzintonation einer Sprache oder eines Dialekts befassen möchte, oder

an die Beurteilung eines Lärmereignisses, ohne vorher und nachher die Ruhesituation dokumentiert zu haben.

Orientiert nach der Schallquelle
↳ Nur sequentiell aufrufbar
↳ A/B-Vergleich nur durch Cutting/Umkopieren möglich
↳ Umkopieren ist
a) zeitraubend
b) qualitätsmindernd
c) destruktiv

Abb. 1: Die Nachteile der Speicherung und Archivierung auf analogen Tonträgern.

Das bedeutet, daß bei der herkömmlichen Art der Speicherung des relevanten Materials (der Aufnahmen) bereits beim Zusammenschnitt ein sehr starkes Informationsfilter angewendet wird, das die Art der zukünftigen Bearbeitung einschränkt bzw. entscheidend mitbestimmt. Erwünscht ist daher eine Art der Speicherung, bei der die Originalaufnahme nicht zerstört (also kein derartiges Informationsfilter wirksam) wird, bei der die Qualität durch Kopiervorgänge nicht leidet, die nur wenig Bearbeitungsaufwand erfordert, und bei der sowohl ein sequentieller als auch ein nichtsequentieller (Random Access) Aufruf möglich ist.

2. Anforderungen an die Tondatenbank

Die Konsequenz aus dem Anforderungskatalog der Wissenschaften für klingende Tondatenbanken ist die Zuwendung zur digitalen Speicherung von Schall, und hier wiederum mit einer Technologie, die diese Anforderungen erfüllt. Signalspeicher-, Sprachverarbeitungs-, hypermediale und Analyse-Systeme, bei denen der Vorgang der Segmentierung, des Zerschneidens der Information *vor* dem Prozeß der Digitalisierung durchgeführt werden muß oder bei denen nach der Digitalisierung „analog“ zur traditionellen Technik wiederum geschnitten wird, erfüllen die Hauptfunktionen dieses Anforderungskatalogs nicht, oder drastisch gesprochen: ob man "digits" zerschneidet oder Tonbänder, es läuft auf dasselbe hinaus: das Originalsignal ist am Schluß nicht mehr vorhanden oder es kann nicht mehr eindeutig darauf referiert werden.

Mit der Digitalisierung von Schallsignalen allein ist es also nicht getan: was nottut, ist die sinnvolle Ausnützung der Möglichkeiten der digitalen Techniken:

(1) Mehrfachkopie ohne Qualitätsminderung (im digitalen Bereich realisiert, wenn die Daten dauerhaft digital gespeichert vorliegen und folglich verlustfrei kopiert werden können).
(2) Kein digitales Cutting, sondern eine „virtuelle Segmentierung“ durch sampleweise Referenz (Labeling, Tagging) auf das Material (Adressierung statt Cutting, d.h., das Original bleibt erhalten).
(3) Programmgesteuerter Aufruf sequentiell gespeicherter Daten in beliebiger Reihenfolge bzw. programmgesteuerten Sequenzierung relevanter Tonausschnitte.
(4) Echtzeitzugriff auf große Datenmengen, realisierbar mit Harddisks, magneto-optischen Disks (MOD) bzw. über Netzwerke (LAN, MAN).
(5) Kumulative Archivierung; digitalisiertes und segmentiertes Material bleibt für kommende Untersuchungen erhalten und kann mit neuen Daten kombiniert werden.
(6) Möglichkeit einer kumulativen Bearbeitung des Materials (Labeling und Indizierung in beliebiger Form; z.B. kumulativer Grob- und Feintranskription in mehreren Durchgängen). Nachdem eine Erstsegmentierung des Materials grundsätzlich nicht alle künftigen Anwendungen zufriedenstellen wird, muß eine kumulative Bearbeitung ermöglicht werden.
(7) Reproduzierbarkeit der Analysen und Bearbeitungsdurchgänge durch Speicherung der Analyseprozeduren und nicht der Analyseergebnisse.

Bei Beachtung dieser Vorgaben wird das Informationsfilter aus der reinen Datenerfassung in die jeweilige Analyseprozedur hinein verschoben, indem nur die Abfrage, nicht aber das Material selbst, die wissenschaftliche Fragestellung beschränkt. Um nicht das Informationsfilter gewissermaßen in das System „einzubauen“, ist es ferner notwendig, ein flexibles und dynamisches Datenbankkonzept mit applikationsabhängiger Reorganisationsmöglichkeit der Feldstruktur(en) einzuführen, das Mehrfachlabeling und Mehrfachverwendung desselben Materials erlaubt.

3. Realisierung des Tondatenbankkonzepts

Eine Datenbankstruktur (STDB, „S_Tools Database“), die neben Text und Bildern Tondokumente verwaltet und bei Bedarf einem Bearbeitungs- und Analysesystem zuführt, erfüllt die wichtigsten Voraussetzungen für ein multimediales Archiv, das jederzeit zugänglich, kumulativ erweiterbar und jeder Art von Fragestellung gerecht werden kann. Um eine möglichst weitgehende Systemoffenheit zu garantieren und den Besonderheiten[1] der unterschiedlichen Informationsarten gerecht werden zu können, ist es angebracht, Ton, Text und Bild getrennt zu speichern und in der Datenbank hinlänglich relational aufeinander zu beziehen.

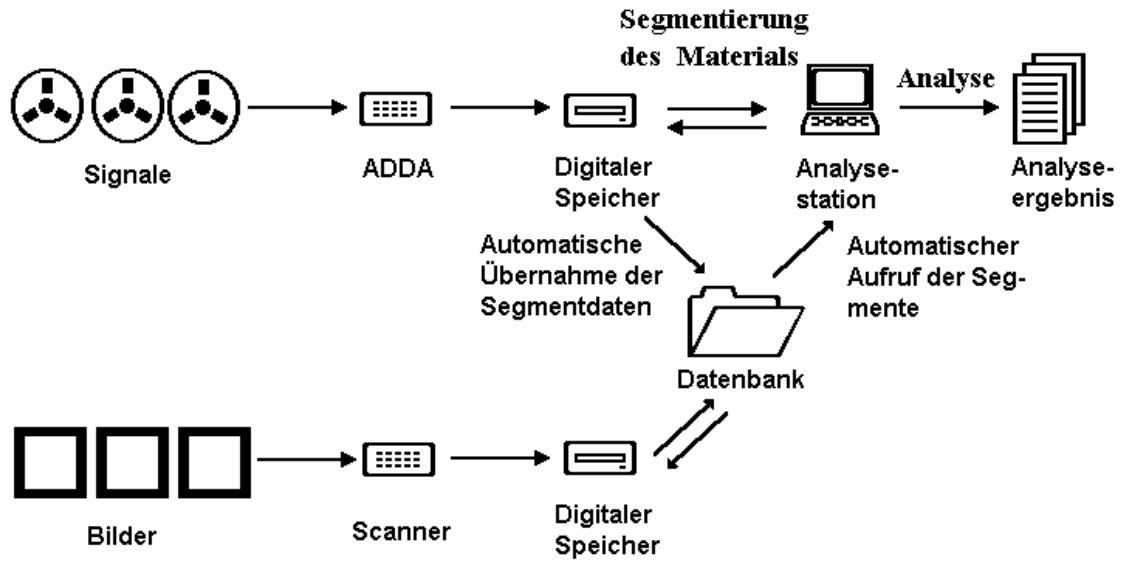

Abb. 2: Schema des multimedialen Datenbankkonzepts für Aufnahme, Wiedergabe und Analyse von akustischen Signalen jeder Art.

3.1. Tondateneingabe, Vorarchivierung und Bearbeitung mittels S_Tools[2]

S_Tools ist eine integrierte digitale Arbeitsstation zur Bearbeitung, Analyse, Speicherung und Visualisierung von akustischen Signalen aller Art; das System arbeitet kann sowohl auf lokalen Rechnern als auch im Netzwerk aufgebaut werden und ist mit einem Signalprozessor der letzten Generation (AT&T: DSP32C) ausgerüstet. Die AD/DA-Konverter (alternativ AES/EBU) erfolgt über eine professionelle Subeinheit[3]; lokal erfolgt die Speicherung auf Festplatten bis zu etlichen GB, zur Archivierung werden WORM-Laserdisks und magneto-optische Disks verwendet (MOD, 600MB/1,3GB). Bei Anschluß an ein Netzwerk stehen darüberhinaus

---

1 Unterschiedliche File-Formate, Speicherbedarf, Zugriff, etc.

2 S_Tools ist eine Entwicklung der Forschungsstelle für Schallforschung der Österreichischen Akademie der Wissenschaften (vgl. [KfS 1993]).

3 Ariel ProPort 656 bzw. AES/EBU-Schnittstelle.

mehrere GB Serverkapazität zur Verfügung. S_Tools schreibt ein eigenes Sound File-Format (*.SF), kann aber auch WINDOWS Wave-Files (*.WAV) oder einfache binäre Files erzeugen bzw. bearbeiten, und die Signalfiles können wahlweise mit Samplingraten zwischen 1,6 und 48 kHz mono/stereo aufgezeichnet oder gespeichert werden. Die Wortlänge beträgt 16 bit linear (< ~96 dB), somit ergibt sich bei einer Samplingrate von 16 kHz mono pro GB eine Speicherkapazität von 10 Stunden unkomprimiertem Signal. Bei Samplingraten von 44,1 und 48 kHz stereo ist der Speicherplatzbedarf entsprechend höher. Es wird höchst bewußt kein Datenreduktions- oder Komprimierungsverfahren (lossy coder) angewendet.

S_Tools besteht idealerweise aus einem Server und mehreren Work Stations (Server-Client-System) unterschiedlicher Ausstattung (reine Abhör- oder Analysestationen vs. vollwertige Aufnahme-, Wiedergabe und Bearbeitungseinheiten). Durch den Anschluß an Internet kann über z.B. Ethernet oder ATM (Glasfasertechnik) in Zukunft auf Ton- und Bildmaterial in Echtzeit und in voller Bandbreite über beliebige dislozierte Clients verfügbar gemacht werden.

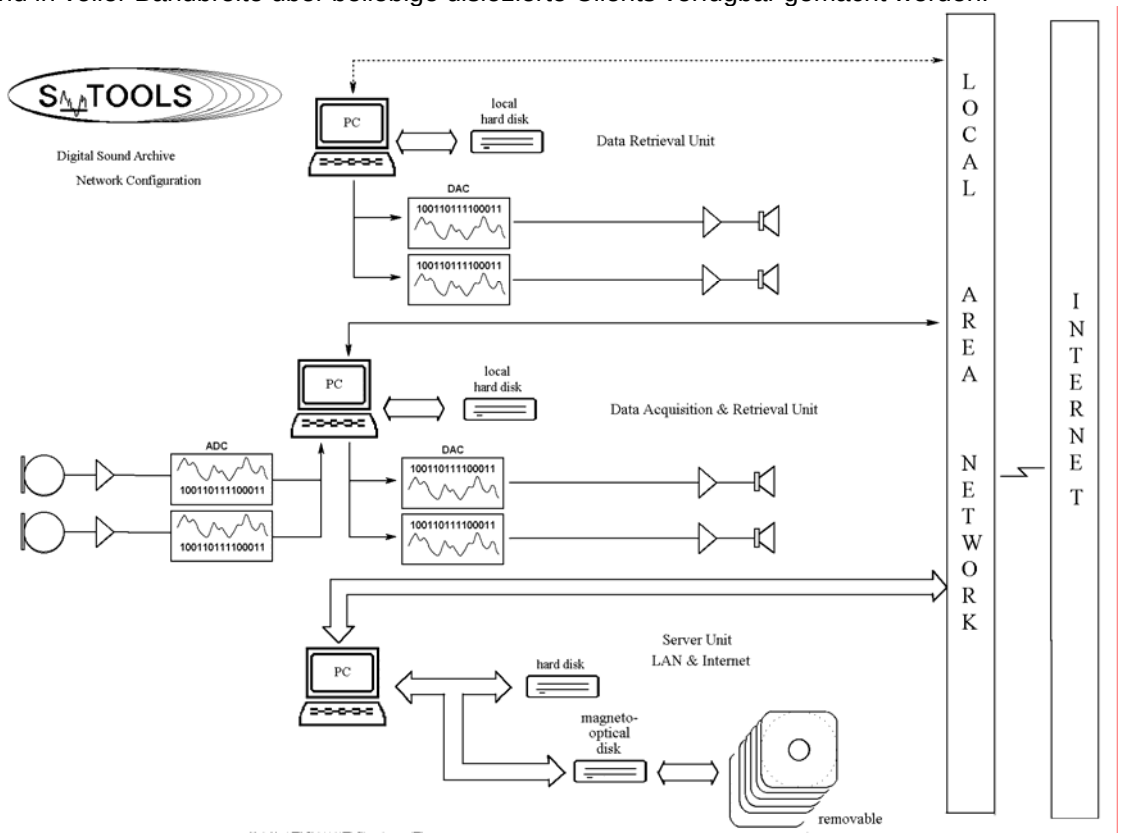

Abb. 3: *S_Tools Netzwerkkonfiguration, bestehend aus einer Servereinheit, einer Datenakquisitions- und -bearbeitungseinheit und einer Abhörstation. Der Server verfügt üblicherweise über keine Audio-Input/Output-Möglichkeiten; die Datenabfragestation hat keine Aufnahmemöglichkeit; die Datenakquisitions- und Datenbearbeitungseinheit entspricht bei Ausrüstung mit einem Signalprozessor einer Stand-alone-Version der Workstation S_Tools.*

Konzeptuell ist S_Tools© ein Harddisk-Recorder, der daher nicht memory-abhängig arbeitet; sodaß sehr große Soundfiles bis zur Größe der Plattenspeicher-Kapazität, geschrieben werden können. Dadurch werden beliebig große Segmente (Texte, Sätze, musikalische Stücke, Rufserien, usw. sowie beliebige Sequenzen im Detail und in ihrer Gesamtheit gleichzeitig bearbeitbar. Der Simultanzugriff auf mehrere geöffnete Sound Files (verschiedenen Formats) ist möglich.

Das Sound File-Format und das unter S_Tools erweiterte WAV-Format schreiben neben einem *Header* und den *Data* im File selbst auch ein *„Directory“*, das die Segmentierungen als Signaladressen verwaltet. Dieses *Sound File Directory (SFDIR)* wird standardmäßig von S_Tools erkannt. Beim Aufruf eines WAV-Files mittels einer anderen Sound Software muß die Segmentierung über ASCII-Textfiles separat verwaltet werden.

Das Sound File Directory enthält die Segmentadressen (absolut in Samples), die Länge und einen Namen für jedes Segment; Sound File Directories können aus dem Sound File exportiert und nach eventueller Bearbeitung in das Sound File reimportiert werden. Auf dieser Ebene können (a) mehrere Directories zum selben Material erstellt und (b) Directory-Informationen aus Fremdprogrammen übernommen werden. Darüberhinaus bilden die Sound File Directories die Input-Daten für Segmentverwaltung in einer Datenbank.

Die Signaladressierung kann mehrfach erfolgen: absolut in Samples, in Millisekunden, Sekunden, mittels des Segmentnamens, durch relative Adressierung (z.B. Segment1–1s bis Segment5+2s) bzw. durch Kombinationen beliebiger Adressierungsarten (z.B. Segment1+300000samples bis Segment5+3s).

Diese Art der Adressierung gestattet die Bildung von einander beliebig überlappenden Segmenten; Sequenzen und Segmenthierarchien werden im Bedarfsfall (z.B. Text–Satz–Wort–Phonem, Geräusch 1-Teilgeräusch X, usw.) erst in der Datenbank relational hergestellt.

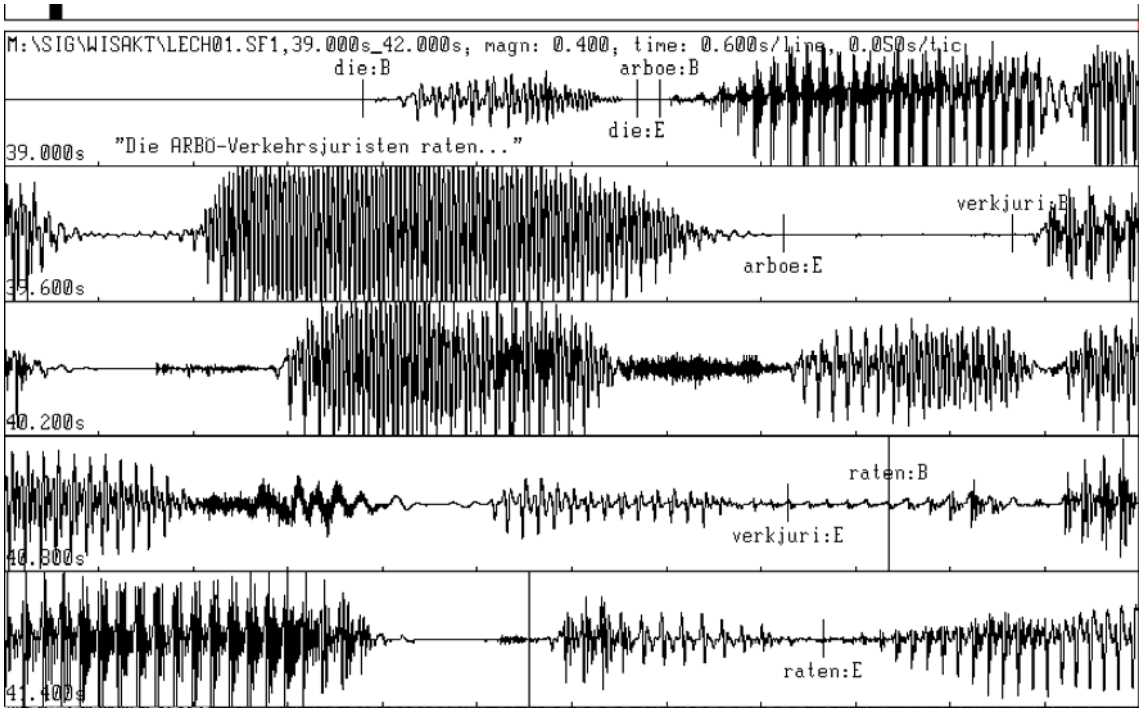

Abb. 4: Segmentierung von akustischen Signalen im Zeitbereich (S_Tools)

Das System verfügt über einen Signaleditor (mittels zweier Kursoren), der Segmentierungen sowohl im Zeitbereich als auch im Frequenzbereich (Spektrogramm) ermöglicht, Zeitlupe und Amplitudenspektrum am Ort der Kursoren sind dabei wirksame Hilfen zur raschen und exakten Segmentierung der Signale. Unter bestimmten Bedingungen ist auch eine halbautomatische Segmentierung möglich[4].

Eswird ausdrücklich darauf hingewiesen, daß S_Tools strikt das Konzept bewahrt, das Signal, die Adressierungen und gegebenenfalls Analyseprozeduren (als „Makro“ in der S_Tools-Programmiersprache) zu archivieren, nicht aber das Analyseergebnis selbst. Wo dies

[4] Eine voll ausgebaute S_Tools-Station verfügt weiters u.a. über Echtzeit-Frequenzanalyse, Echtzeit-Filter, Signalverbesserung, Spektrogramme (FFT, LPC, Wigner-Verteilung; zwei- und dreidimensional), Signalparameterextraktion (Formantextraktion, Grundfrequenz-extraktion (2 Algorithmen), RMS, RMS-Bänder).

ausdrücklich erwünscht ist, können die Analyseergebnisse mittels des *"Dataset Managers"* automatisch protokolliert und zur Weiterverarbeitung an andere Programme übergeben werden (Graphik[5], Statistik, künstliche neuronale Netzwerke).

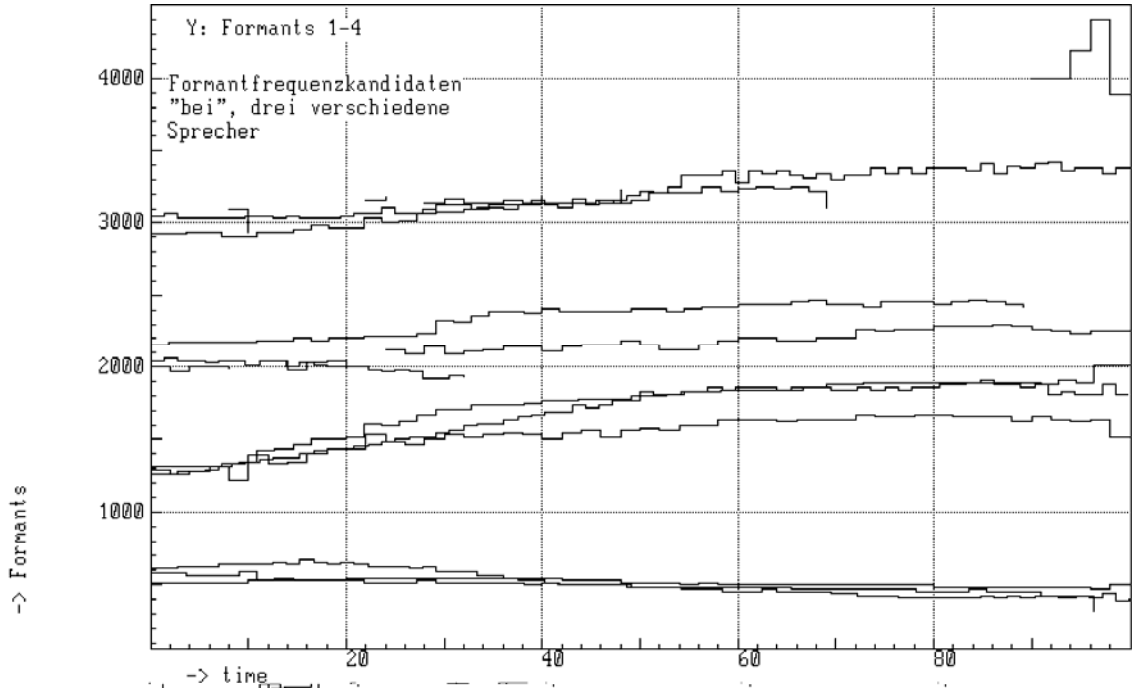

Abb. 5: Graphischer Vergleich von Signalparametern, hier: Formantfrequenzkandidaten aus verschiedenen Aufnahmen (S_Tools Dataset Manager)

Als Zwischenergebnis aus der Eingabeprozedur erhält man sauber digitalisierte, in Files und Segmenten verwaltbare Tondaten mit den dazugehörigen Directory-Informationen, die in der Folge in die Datenbankstruktur eingespeist werden.

3.2. Datenbankstruktur

Um die Vorteile der Systemarchitektur einerseits der akustischen Arbeitsstation S_Tools und andererseits der Datenbankstruktur voll nützen zu können, wurde die Trennung der verschiedenen Informationsarten (Ton, Text, Bild) strikte durchgehalten, gleichzeitig sind die Referenzierungen (Links) in der Datenbank realisierbar. Bei Verwendung entsprechender Software können die Links automatisch aus den Datenbankfeldern generiert werden.

Diesem Grundkonzept entsprach am ehesten die kommerziell erhältliche Datenbanksoftware AskSam[6] Version 5.0 (vgl. Cyffka & Bahr 1992), mit der das Retrieval bis auf weiteres durchgeführt wird (siehe weiter unten).

Im folgenden wird die Erfassung phonetischen Materials an einigen Beispielen demonstriert. Als Basis der Datenbank werden die Sound File Directories der relevanten Sound Files in einer Feldstruktur erfaßt:

```
ME[  Speichermedium, z.B. MOD Nr. ####]
DI[  Directory auf dem Speichermedium, z.B. \STUDIO]
SF[  Sound File Name, z.B. STUDIO01.SF1]
T1[  Signaladresse Beginn]
T2[  Signaladresse Ende]
```

---

5 Eine Graphikapplikation für den A/B-Vergleich von Analysedaten und einiges zur Statistik sind im Dataset Manager selbst bereits integriert.

6 North American Software.

```
TD[  Signaldauer]
TA[  Segmentname]
```

Zu diesem Grundset kommen optional folgende Informationen hinzu[7]:

```
SP[  Sprecher/innen-ID, Schallquelle]
WO[  Text, Satz, Wort, Phonem, Einheit (orthographische
     Transkription bzw. Referenz)]
PH[  phonetische Transkription bzw. genauere Referenz]
UM[  phonetische Umgebung]
IF[  (noch) nicht strukturierte Zusatzinformation]
W1[  Deskriptor(en) erster Ordnung, z.B. Wortebene, Phonemebene,
     ...[8]]
```

Wie die Adressierung in S_Tools ist die Feldstruktur von AskSam „virtuell“, d.h., die Feldnamen sind lediglich eine besondere Art von Text; eine Feldabfrage ist somit konzeptuell nur eine Sonderform der Kontextabfrage.

Wenn man zu einem Segment noch ein Bild linken möchte, kann das (a) relational geschehen (durch Verweis auf eine Bilddatenbank) oder (b) durch Spezifizierung des Bilds in der Art der Spezifizierung des Schallsignals:

```
BDI[Bild-Standort]
BTA[Bild-Name]
```

Die Relation zwischen dieser reinen Textdatenbank (die daher sehr schnell ist und die völlig unabhängig von Ton und Bild agieren kann) und dem Signal bzw. dem Bild geschieht durch Aufruf entsprechender Fremdprogramme, z.B. von S_Tools © oder eines Graphik-Programms.

Das bedeutet, daß die verschiedenen Arten von Information unabhängig voneinander verwaltet und aus der Datenbank heraus lediglich gesteuert werden. Es bedeutet, daß das Gesamtsystem austauschbare Komponenten enthält und maximal wandelbar ist; sollte es z.B. lediglich notwendig sein, Tonaufnahmen abzuhören, kann die Datenbank dazu gebracht werden, nichts weiter als die durch ein Retrieval ermittelten Sound Files mittels einer anderen Sound Software (z.B. im WAV-Format) abzuspielen; dazu ist der Standort im Archiv (MOD-Nr., Directory, Sound File) notwendig.

Bei der Analyse von Tonsignalen ist es typischerweise notwendig, vor der zeitaufwendigen Prozedur eine händische Überprüfung des automatischen Retrieval vorzunehmen; daher werden die Abfrageergebnisse als Dynaset in der Regel in ein ASCII-Text-File geschrieben und vor der Durchführung einer Analyse manuell korrigiert, worunter v.a. das Löschen von Einträgen zu verstehen ist, wenn die Abfrage zu viel Information erbracht hat oder wenn bestimmte Tonaufnahmen/-medien u.U. gerade nicht verfügbar sind; desgleichen kann auch eine Umleitung vom zeitweilig aktivierten Speichermedium MOD auf den File Server bzw. ein Jukebox-System notwendig sein, weil bestimmte Files zur Bearbeitung gerade auf dem Server vorliegen. Darüberhinaus kann ein bereits vorliegendes Dynaset durch einfache Änderung der Parameter nacheinander verschiedene Analysen in S_Tools auslösen.

---

[7] Diese Informationen ergeben sich teilweise automatisch, etwa wenn beim Import mit dem AskSam-Zusatzprogramm INSERT.EXE spezifiziert wird, daß die Extension des Segmentnamens als SP-ID verwendet wird und auch im Feld SP[(Speaker) ein zweites Mal geschrieben wird.

[8] Hier wird eine virtuelle Hierarchie eingeführt, indem z.B. ein Abfrageergebnis für ein Wort oder einen Satz die Speaker-ID ergibt, die nun in einem als DOKUMENT (W1[docs]) ausgewiesenen Datensatz gesucht wird, wodurch man sprecherrelevante Information als endgültiges Ergebnis erhält – oder umgekehrt: alle Sätze/Wörter einer Sprecherin oder eines Sprechers; oder den Satz zu einem Wort, usw. Man beachte, daß durch einen Vergleich der Beginn- und Endzeiten der Adressierung ebenfalls ermittelt werden kann, welches Segment Teil eines anderen Segments ist.

3.3. Datenbankabfrage

Die speziellen Merkmale von AskSam Version 5.0[9] bestehen v.a. darin, daß es sich (a) um eine Freitextdatenbank (b) mit der Möglichkeit der Freitext-, Kontext- und Feldabfrage handelt, die daher (c) eine maximale Offenheit des Systems gewährleistet und (d) in der Lage ist, andere Programme zu starten, die ein Abfrageergebnis übernehmen können (Aufruf von Bild und Ton). Einige Beispiele für die Möglichkeiten der Abfrage in AskSam seien hier vorgestellt:

Frage	**Bedeutung**
ana	*Zeige alle Datensätze, in denen die Zeichenfolge *ana* vorkommt*
[aber leider]	*Suche die Wortfolge „aber leider“*
SP[122 *bzw.* 122 {in} SP[	*Suche nach dem/r Sprecher/in 122*
aber {in} T1[122 bzw. 122 aber {in} T1[	*alle "aber" von Sprecher/in 122*
gemeldet {UMGEBUNG :S hat}	*Sucht Sätze, in denen "hat" und "gemeldet" vorkommt*
gemeldet {UMG[10] :4 hat}	*Sucht Sätze, in denen "hat" max. 4 Wörter vor oder nach "gemeldet" vorkommt.*
{NACH} gemeldet	*Was steht nach "gemeldet"?*
{VOR} gemeldet	*Was steht vor "gemeldet"?*
{VOR[} gemeldet	*Welche Wortgruppe steht vor "gemeldet"*
{ZEIGE :AN[11] gemeldet}	*Zeige den Absatz, der nach dem Absatz kommt, in dem "gemeldet" steht*
{SUB} SP[{FILE document DOSUB ÜBERBLICK 3}[12]	*Suche nach dem Dokument zum/r Sprecher/in im File document, Ausgabe der ersten drei Zeilen*
{lade "f:\sig\test.wav" c:\sound\play.exe}	*Spielt ein WAV-File "test.wav" vor*
{lade „K:“} DI[{RÜC} SF[{c:\sound\play.exe}	*Spielt ein durch eine Abfrage ermitteltes WAV-File vor, das sich auf K: befindet.*
{lade „bild.gif f:\sig\test.wav satz1“ stvp.exe}	*ruft das Programm STVP (S_Tools ©) auf, das ein Bild anzeigt und ein Schallsignal dazu abspielt.*
usw.	

Abb. 6: Mögliche Abfragen in AskSam©

In STDB wurde eine vernetzte (nichtlineare) Benützermenüsteuerung implementiert, von der aus Programme oder Suchvorgänge aufgerufen werden können. Genauso wie die erfaßten Datenmengen entstand die Menü-Programm-Struktur von STDB kumulativ, aufbauend auf den Erfahrungen aus dem Gebrauch der Datenbank.

---

9 Es wird hier stets nur auf die Version 5.0 referiert, weil die Version 5.1 offensichtlich einige Bugs enthält, die ein Arbeiten mit ihr zumindest sehr erschweren, und weil die Version Asksam for Windows 1.2 unverständlicherweise einiger für uns wichtiger Features ermangelt.

10 Alle Befehle können mit drei Buchstaben abgekürzt werden.

11 *:AN* bedeutet zwei Parameter des Befehls: "Absatz" und "Nach".

12 Temporäre Relationen werden als eine Art von Unterprogrammen abgearbeitet; Der Befehl ÜBERBLICK mit dem Parameter 3 bedeutet "Anzeige der ersten drei Zeilen des Datensatzes".

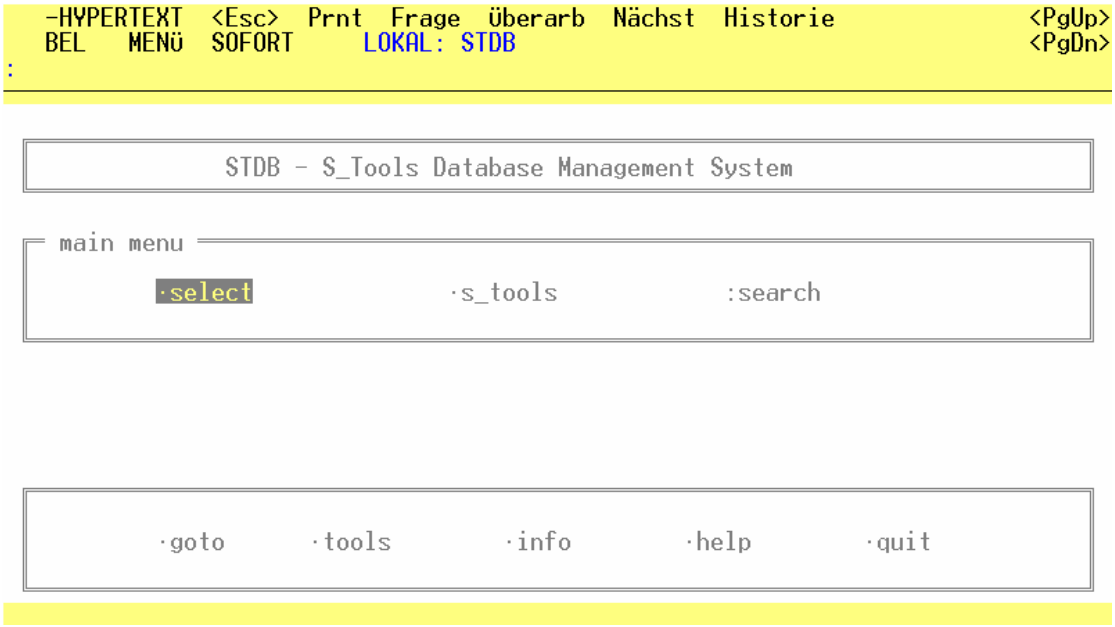

Abb. 7: S_Tools Database Management System (STDB)© – Hauptmenü

Die Systemarchitektur der Datenbank STDB („S_Tools Database Management System“)© ergibt sich nach Maßgabe der Erfordernisse für die phonetische Bearbeitung wie folgt:

Dokumentationen (orientiert nach der Schallquelle und nach der realen Archivierung in einem oder in mehreren Sound Files)	Schallarchiv (Sound Files)
Verschiedene Ebenen der Segmentierung (z.B. Satz–Wort–Phonem)	Segmentierungen innerhalb von Sound Files
Referenz auf Bilder	Bildarchiv
Referenz auf beliebige andere Datenbanken (z.B. DARPA TIMIT Datenbank[13])	Andere Zusatzinformationen / Datenbanken

Abb. 8: Systemarchitektur von STDB ©

DATENBANKFELDER:	**FUNKTION:**
WA[**&docs**]	Deskriptor: Dokumentation
DI[\VIENNA]	Subdirectory
SF[M13S05.SF]	Sound File
SP[M13c]	Schallquelle (Speaker-ID)
IF[Österreich Wien männlich; Testsätze S05; Alter: 23 Jahre (1990); Studio KfS, Aufnahmeleitung: MOOSM; SoundFile im S_Tools-Format (16 bit linear).]	Unstrukturierte Information über die Schallquelle bzw. das Material.

DATENBANKFELDER:	**FUNKTION:**
WA[**&dat**]	Deskriptor: Wortebene
DI[\VIENNA]	Subdirectory
SF[M13S05.SF]	Sound File
SP[M13c]	Schallquelle (Speaker-ID)
SNR[05]	Satznummer (Referenz zum Satz)
T1[12.35972]	Startzeit

[13] Die DARPA TIMIT Datenbank beispielsweise ist eine in sich abgeschlossene eigene relationale Datenbank, die aus den Segmentierungen und den verschiedenen Dokumentationen gewonnen wurde; sie kann aber in STDB miteingebunden werden.

T2[12.98754]	Endzeit
TD[0.78943]	Signaldauer
TA[also.004]	Segmentname
WO[also]	orthographische Transkription
PH[aE&sO]	phonetische Transkription
IF[l-voc. without a->O]	Zusätzliche Information, hier etwa über einen phonologischen Prozeß

DATENBANKFELDER:	**FUNKTION:**
WA[**&pho** &tib]	Deskriptoren: Phonemebene, Sprache Tibetisch
DI[\TIB\LHASA]	Subdirectory
SF[DORJE01.SF]	Sound File
SP[T017]	Schallquelle (Speaker-ID)
T1[12.35972]	Startzeit
T2[12.98754]	Endzeit
TD[0.78943]	Signaldauer
TA[a.012]	Segmentname
WO[a]	orthographische Transkription (Translitteration)
PH[a tone_1]	phonetische Transkription
UMG[k_#S]	phonetische Umgebung
IF[stressed]	Zusätzliche Information, hier über Betontheit

DATENBANKFELDER:	**FUNKTION:**
WA[**&txt**]	Deskriptor: Satzebene
DI[\VIENNA]	Subdirectory
SF[M13S05.SF]	Sound File
SP[M13c]	Schallquelle (Speaker-ID)
T1[12.35972]	Startzeit
T2[14.98754]	Endzeit
TD[2.78943]	Signaldauer
TA[take.004]	Segmentname
SN[05]	Satznummer (Referenz zur Einzelwortsegmentierung)
WO[Also ich weiß das auch nicht!]	orthographische Transkription

Erklärung der Feldnamen und Kürzel: DI = Directory, SF = SoundFile, SP = Speaker-ID, SNR = sentence number, T1 = start time, T2 = end time, TD = signal duration, TA = task (segment) name, WO = word, PH = phonetic transcription, UMG = phonological UMGEBUNG, IF = info (free) field; &tib = tibetan language, &pho = phone, &dat = word, &txt = utterance, whole text, &docs = document (header).

Da es nicht nötig ist, eine fixe Struktur einzuhalten, kann die Dokumentation etwa zu einem bestimmten Sprecher oder einer bestimmten Sprecherin in derselben Datenbank gesichert werden wie die übrigen Daten; durch die programmierbare "Hypertextfunktion" von AskSam können so Informationen aus der Dokumentation gesucht und eine weitere Auswahl der dieser Dokumentation zugeordneten Segmente getroffen werden; beispielsweise bei der Suche der Wörter "Wien" und "männlich" könnten die Sprecherkennungen in der Dokumentation gesucht und die diesen Dokumenten zugeordneten Realisierungen des Wortes "aber" gesucht werden: "alle vorhandenen ‘aber’ von männlichen Wiener Sprechern". Desgleichen ist es möglich, in einem ersten Durchgang ein Wort zu suchen und in einem zweiten Suchlauf über SN den dazugehörigen Satz auszugeben (relationale Abfrage innerhalb einer Datenbank).

Die Datenbank besteht also aus einem Datenbankfile, das alle Datentypen enthält: Texte/Sätze, Wörter, Silben/Biphone/Phone, und die Header-Information. Die Daten, die Programme und die Untermenüs zur Abfrage der Daten befinden sich innerhalb der Datenbank. Es liegen bislang 4 unterscheidbare Datenstrukturen vor: für phonemische Segmente (&pho), für Wörter und Wort-

gruppen (&dat), und für Sätze bzw. Texte (&txt) einerseits, sowie für die Dokumentation (&docs). Diese Trennung in verschiedene Datenformate wird erst in der Abfrage vorgenommen, d.h., man kann auch Phoneme und Wörter und Sätze gleichzeitig suchen, man kann aber auch z.B. nur die Sätze/Texte von der Suche ausschließen. Man kann durch eine relationale Abfrage auch direkt die Sätze suchen, in denen ein bestimmtes (bereits gefundenes) Wort vorkommt, usw..

3.3. Das Prinzip der Virtualität

Eine korrekte Datenbankabfrage wird von uns evtl. als Programm abgespeichert; dadurch ist das Datenset jederzeit wiederherstellbar – und wird auch kumulativ immer umfassender, da ja inzwischen neue Daten hinzukommen können. Dazu wird das Suchprogramm mit neuem Namen kopiert und anstelle der Eingabeoption eine fixe Abfrage festgelegt.

Die Segmente sind, wie bereits erwähnt, keine Segmente im physikalischen Sinn, sondern in einem durch die Segmentierung unversehrten Schallsignal werden bestimmte Bereiche angesprochen.

Die einander überlappenden Segmente können aufeinander bezugnehmen, d.h., es ist möglich, sinnvolle Kontexte zu einem Segment zu ermitteln (das Wort, in dem ein segmentiertes Phonem vorkommt, usw.).

Die extrahierten Daten können zwar als Datensets abgespeichert werden, ergeben sich aber anhand der gesetzten Parameter bei jeder gleichartigen Abfrage auch erneut in derselben Weise.

D.h., in keinem der Module von S_Tools und der Datenbank wird ein enges Informationsfilter gesetzt; damit ist die absolute Reproduzierbarkeit der Analysen von der Datenbank bis zu einer etwaigen statistischen Auswertung gewährleistet.

4. Verteilte Datenbanken / Netzwerke

Die Datenbankdaten liegen in ASCII vor bzw. sind beliebig konvertierbar; es ist daher möglich, die Steuerung der verschiedenen Medien auch von anderen Systemen aus vorzunehmen. Dabei kann daran gedacht werden, bei Bedarf in einer genormten Datenbanksprache den Zugriff auf das Material über INTERNET zuzulassen.

5. Anwendungen

Die erste Anwendung bestand darin, den Bestand an segmentiertem Material, das im Lauf der Jahre auf Laserdisks und magneto-optischen Disketten archiviert worden ist, automatisch zu erfassen; es entstand daher 1991 das Laserdisk-Directory (LDISKDIR) mit rund 28.000 Segmenten aus den Bereichen Musik und Sprache. Das Abfrageergebnis war der genaue Standort des gesuchten Segments auf einem Archivmedium.

In der Folge wurde, ausgehend von dem durch Moosmüller (1987, 1991) analysierten Material, eine spezialisierte Datenbank des österreichischen Deutsch geschaffen (STDB-G), die in der Folge durch spontansprachliches Material in unterschiedlicher Qualität ergänzt wurde; diese Datenbank wurde zuletzt durch weiteres Studiomaterial (Foltin 1994) ergänzt und umfaßt nun etwa 35.000 Segmente (Satz-, Wort-, Phon-Segmentierung) mit den entsprechenden Einträgen (orthographische und phonetische Transkription, phonologische und phonetische Bemerkungen, phonetische Umgebung des Segments). Mittels dieser Datenbank ist die Erstellung von Datenreihen für automatische Analysen realisierbar.

Zwischenzeitlich wurde die DARPA TIMIT ACOUSTIC PHONETIC DATABASE PART II mit 420 Sprecherinnen und Sprechern des Amerikanischen Englisch (4200 Satzäußerungen mit 165.000 phonemischen Segmenten) aus Praktikabilitätsgründen ins S_Tools-Format konvertiert, auf MOD archiviert und mit den im ASCII-Format beiliegenden Signaladressierungen versehen. Die Signaladressierungen und die umfangreiche Dokumentation zu den Sprecherinnen und Sprechern sowie zur Datenbank selbst wurde in einem relationalen Datenbanksystem (TIMIT) erfaßt; in derselben Weise wie STDB-G können mittels relationaler Verknüpfung der verschiedenen Informationen Datensets für die automatische Analyse erzeugt werden.

Schließlich wurden die einzelnen Datenbanken der Forschungsstelle und beispielhaft Teile der Datenbank des Österreichischen Phonogrammarchivs und der Datenbanken des Österreichischen Musiklexikons durch ein übergeordnetes System (MMDEMO) vorläufig zusammengefaßt, sodaß nun in Form eines multimedialen Systems von einer einzigen Oberfläche aus zu allen Datenbanken geswitcht werden kann. Dabei wurde mittels eines S_Tools-Zusatzprogrammes und der Einbindung eines Graphikprogrammes ermöglicht, jedes beliebige Schall-Segment sofort zu hören und, falls vorhanden, Bilder, Videos und beliebige zusätzliche Informationen abzurufen.

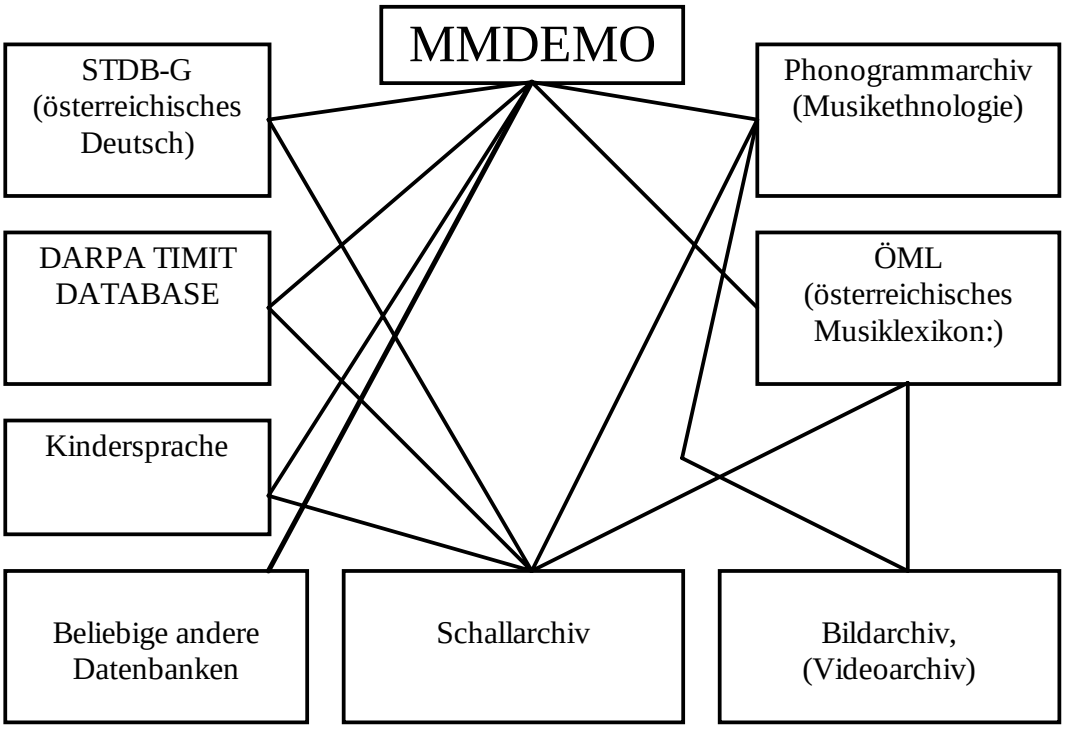

Abb.9: Das Gesamtsystem des digitalen Schallarchivs MMDEMO

In diesem Zusammenhang wird darauf hingewiesen, daß die Einbindung von in anderen Datenbanksystemen vorliegenden Informationen oder von zusätzlichen Textdatenbanken (z.B. Bibliographien, etc.) jederzeit möglich ist.

Diese Demoversion eines multimedialen Systems für wissenschaftliche Anwendungen bewährte sich auch im Betrieb über das Glasfaserkabel und erlaubt die dislozierte Abfrage von Schallsignalen bis 41 kHz Samplingrate Stereo in Echtzeit.

Der modulare Aufbau des Systems und die Unabhängigkeit der Komponenten voneinander erlauben den schrittweisen Ausbau des Systems. So wird im Rahmen eines Projekts zur Kindersprache gerade eine neue Datenbank mit den Transkriptionen der untersuchten Kinder angelegt, die je nach Stand der Bearbeitung den Aufruf der Aufnahmen ermöglicht.

Durch diese Art der Archivierung bleiben die (für einen speziellen Zweck erhobenen) Daten stets präsent und sind ohne zeitraubende Recherche jederzeit wiederverwendbar.

6. Bibliographie

Cyffka, Günther 1991: askSam: Funktionen, Befehle, Musterbeispiele. Vaterstetten: IWT-Verlag (= IWT kompakt: Datenbanken).

Cyffka, Günther & Hermann Bahr 1992: Erfolgreich arbeiten mit askSam: Vom einfachen Notizbuch bis hin zur komplexen Datenverwaltung. (mit Diskette) Vaterstetten: IWT-Verlag.

Deutsch, Werner A. & Anton Noll 1985: Halbautomatische Verfahren zur Formantfrequenzmessung. in: wlg 35-36: 95-103.

Deutsch, Werner A. & Anton Noll 1994: Datenerfassung, Speicherung und Digitale Signalverarbeitung für Akustik, Lärm, Sprache und Musik. Wien: Österreichische Akademie der Wissenschaften.

Fisher, W.M. & G.R. Doddington & K.M. Goudie-Marshall 1986: The DARPA speech recognition research database: Specifications and status. in: Proceedings: DARPA Speech Recognition Workshop, pp.93-99.

Fisher, W. & V. Zue & J. Bernstein & D. Pallett 1987: An acoustic phonetic database. in: JASA, Suppl. A 81, S92.

Klatt, D.H. 1977: Review of the ARPA Speech Understanding Project. in: JASA 62: 1345-1366.

Klimbie, J.W. & K.L. Koffeman (eds.) 1974: Data Base Management. in: Proceedings of the IFIP Working Conference on DB Management, Cargese, Corsica, April 1-5, 1974. Amsterdam, New York, Oxford.

Kuhlen, Rainer (ed.) 1979: Datenbasen – Datenbanken – Netzwerke. Praxis des Information Retrieval. Bd. 1: Aufbau von Datenbasen. München, New York, London, Paris.

Kurzidim, Michael 1993: Spürnase. Textretrieval mit AskSam für Windows. in: c't 11/93: 58.

Lamel, L.F. & R.H. Kassel & S. Seneff 1986: Speech database development: Design and analysis of the acoustic phonetic corpus. in: Proceedings DARPA Speech Recognition Workshop, pp.100-109.

Moosmüller, Sylvia 1987: Soziophonologische Variation im gegenwärtigen Wiener Deutsch. Wiesbaden: Steiner.

Moosmüller, Sylvia 1991: Hochsprache und Dialekt in Österreich. Wien: Böhlau.

Pallett, D.S. 1988: Documentation for Part I of the DARPA Acoustic-Phonetic Database. ms.

Vollman, Ralf 1993: Die „klingende“ Datenbank - Multimedia in der Forschung. in: Monitor 9/93: 58-64.