

# Ein interaktives automatisches System zur Inventarerstellung und -optimierung

Uwe Koloska

Institut für Akustik und Sprachkommunikation  
Technische Universität Dresden, D-01062 Dresden  
<mailto:kom@eakss1.et.tu-dresden.de>

**Zusammenfassung:** Um die Erstellung und Pflege der Daten, die zur Erstellung eines Syntheseinventars benötigt werden, zu erleichtern und ihre Konsistenz und Transparenz zu erhöhen, wurde ein Systemkonzept und ein Softwaresystem entwickelt. Es definiert eine Verwaltungsschicht für alle zur Inventarerzeugung nötigen Daten und bietet gleichzeitig eine Sammlung von Werkzeugen, die mit Hilfe dieser Verwaltungsinformationen einen direkten und für verschiedene Nutzer auf unterschiedlichen Rechnern transparenten Zugriff auf die gesammelten Daten erlaubt.

## Motivation

Die Erstellung von Inventaren für die verkettende Sprachsynthese ist ein aufwendiger und langwieriger Vorgang. Unabhängig von der Einheitenart sind viele verschiedene Querbeziehungen und Abhängigkeiten zu beachten. Außerdem werden die verwendeten Inventare immer komplexer – und für die unterschiedlichsten Anwendungen werden jeweils verschiedene Stimmen benötigt. Viele für ein Inventar erarbeitete Daten, könnten für eine neue Stimme wiederverwendet werden und aus den Quellen lassen sich verschiedene Inventare erzeugen. Bei der bisherigen, weitgehend manuellen Vorgehensweise geschieht die Verknüpfung und Wiederverwendung der Daten meist im Kopf des Ausführenden – und dieses Wissen geht sehr schnell verloren, wenn ein anderer Mitarbeiter dieses oder ein ähnliches Problem bearbeiten muß.

## Konzept

Grundlage ist die Einteilung der für Erstellung und Verwaltung von Inventaren nötigen Daten in Ebenen. Zweck dieser Unterteilung ist es, möglichst viele Daten für andere Projekte wiederzuverwenden. So kann die Liste der Einheiten, die ein Inventar bilden von verschiedenen Stimmen benutzt werden.

### Kontextebene

allgemeine Variablen (Klassenvariablen):

- Verwaltungsinformationen (Pfade, Protokolle, ...)
- Hilfsdaten (Phonemliste, ...)

- mögliche Beschreibungsebenen und ihre jeweilige Sprache (Phonemlabel, Periodenmarken, ...)
- Instanzen – Ausprägungen dieser Klasse (Sprecher)

Variablen für Sammlung von Kontexten (Instanzvariablen):

- Aufnahmedaten (Sprecher, Equipment, ...)
- technische Daten (Dateiformate, Speicherpfade, ...)
- Menge der Kontexte

Variablen für jeden einzelnen Kontext (Elementvariablen):

- technische Daten (Länge, Status, Dateiinformationen, ...)
- Beschreibungsebenen, notwendige (Sprachdaten, Bausteinlabel, ...) und mögliche (Silbenlabel, hierarchische Label, ...)

### Einheitenebene

Variablen für Sammlung von Einheiten (Klassenvariablen):

- Menge der Einheiten

Variablen für Sammlung von Alternativen (Instanzvariablen):

- Einheitenname
- Zuordnung aktiv / passiv

Variablen für jede einzelne Alternative (Elementvariablen):

- eindeutiger Name, Dateiname
- Beschreibung / Umgebung (phonetischer, semantischer, syntaktischer Kontext, ...)
- Verweis auf den Quellkontext

### Inventarebene

Variablen für ein Inventar (Klassenvariablen):

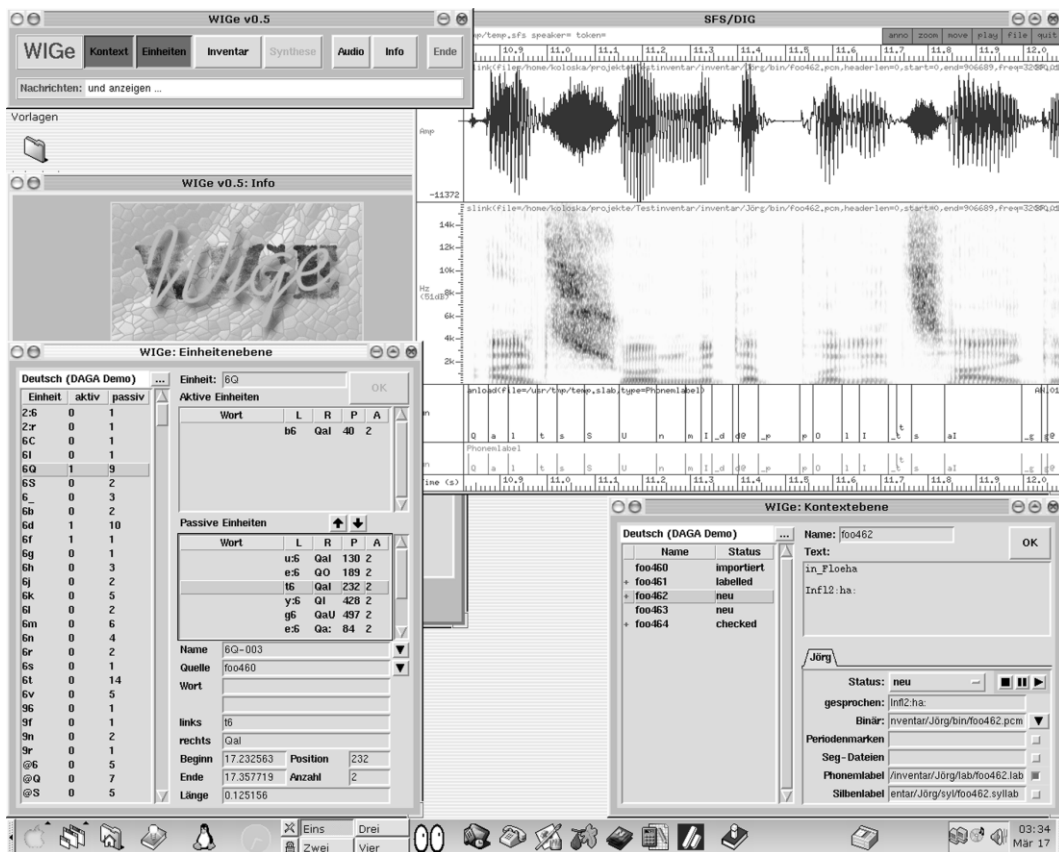
- Menge der Einheiten (Untermenge der Einheitenebene)

Variablen für spezielles Inventar (Instanzvariablen):

- Abstammung (Einheiten, Kontext)
- Speicherpfade, Verarbeitung

### Syntheseebene

Ansteuerung der Synthese und Rückkopplung aus dem Syntheseergebnis.



**Abbildung 1:** Die Oberfläche zur Verwaltung der Daten, die bei der Inventarerzeugung anfallen und verknüpft werden müssen.

Zu dieser Struktur der Daten kommt noch eine Konvention für ihre Organisation, um sicherzustellen, daß alle Daten, die für die Erzeugung eines Inventars nötig sind immer und vor allem konsistent zur Verfügung stehen. Auf der Basis dieser Struktur- und Organisationsform können nun die Daten beliebig bearbeitet werden – auch mit privaten Skripten, die aber für andere weiterhin transparent bleiben.

## Werkzeug zur Inventargenerierung

Ausgehend von den Erfahrungen mit der individuellen Skriptprogrammierung in einer heterogenen Rechnerumgebung und aufbauend auf der vorgestellten Daten- und Organisationsstruktur wurde das Werkzeug zur Inventargenerierung WIGe entwickelt.

- Es ermöglicht einen einfachen Überblick über die großen Datenmengen bei der Inventarerstellung
- Es unterstützt die Wiederverwendung der Daten
- Durch die Vereinheitlichung der Verwaltungsdaten, kann ein großer Teil der Aufgaben automatisch durchgeführt werden – insbesondere die Verwaltung der Abhängigkeiten und die automatische Etikettierung
- Es bietet eine konsistente Form der Wissensrepräsentation
- Es ist (relativ) unabhängig von der Rechnerplattform und eignet sich so hervorragend für heterogene Rechnerumgebungen

- Es macht die Abhängigkeiten und Querverbindungen zwischen den einzelnen Datenebenen und -formen für den direkten Zugriff zugänglich
- Es bietet einen intuitiven und unkomplizierten Zugang zu den Daten, so daß Arbeiten leichter delegiert werden können.

## Begriffe

- Baustein:** kleinste Einheit in den Sprachdateien, aus denen sich alle anderen Elemente zusammensetzen. (z. B. Phoneme)
- Kontext:** eine Sprachaufnahme, aus der Elemente für die Inventarerzeugung entnommen werden.
- Einheit:** Teile der Sprachaufnahmen, die sich aus Bausteinen zusammensetzen und zum Aufbau des Inventars benutzt werden. (z. B. Diphone, Silben, Wörter)
- Einheit, aktive / passive:** aus einer Menge von Kontexten lassen sich meistens mehrere Einheiten der gleichen Zusammensetzung extrahieren. Nicht alle davon sind für die Synthese geeignet. Geeignete werden als aktiv, ungeeignete als passiv markiert.
- Alternative:** gleiche Einheit aus einem anderen Kontext.
- Inventar:** geordnete Menge von Einheiten, die für das eigentliche Syntheseprogramm zusammen mit Zusatzinformationen bereitgestellt werden.