

SpeechRecorder – Mehrkanal-Sprachaufnahmen über das WWW

Christoph Draxler, Klaus Jänsch

Institut für Phonetik und Sprachliche Kommunikation, Universität München

Zusammenfassung

SpeechRecorder ist ein Softwarepaket für Sprachaufnahmen. Die Software läuft auf Clientrechnern im WWW, die Aufnahmen werden lokal oder auf einem Server gespeichert. Die Software ist plattformunabhängig in Java implementiert. Das Audiointerface kann mit vorhandenen Audiolibraries wie ASIO o. ä. implementiert werden. SpeechRecorder unterstützt Text-, Grafik-, Audio- und Videoprompts. Alle Aufnahmen werden über ein flexibles Aufnahmeskript mit präzisiertem Aufnahmeprotokoll gesteuert. Die Software ist frei verfügbar [5].

SpeechRecorder

SpeechRecorder ist eine am Institut für Phonetik (IPSK) von Grund auf neu entwickelte Applikation für Sprachaufnahmen mit den folgenden Eigenschaften:

- Plattformunabhängigkeit
- offene Schnittstelle zum Einbinden externer Audiolibraries,
- flexible Steuerung des Aufnahmeablaufs,
- Speicherung der Prompt-, Sprach- und Metadaten lokal oder auf einem Server,
- Text-, Grafik-, Audio- und Videoprompts, und
- konfigurierbare Benutzersichten.

SpeechRecorder kann in einer Client-Server Konfiguration oder als Standalone-Applikation betrieben werden.

Implementation

SpeechRecorder ist plattformunabhängig in Java implementiert und setzt Java 1.4 voraus.

Audiobibliothek

Zur Audioein- und -ausgabe werden unter den verschiedenen Betriebssystemen jeweils plattformabhängige und hardwarenahe Audiotreiber verwendet. Daneben haben sich, insbesondere für qualitativ hochwertige Audiogeräte, die plattformunabhängigen ASIO Treiber etablieren können.

Das JavaSound API ist eine umfangreiche Programmibibliothek zur Verarbeitung von Audio. Implementationen von JavaSound gibt es u. a. für Microsoft Windows, Solaris, Linux und Macintosh, allerdings unterstützt keine dieser Implementationen die ASIO Treiber.

Am IPSK wurde daher das Audio API `ipsk.audio` entwickelt. `ipsk.audio` implementiert

- Java Wrapperklassen für ASIO Treiber,
- eine vereinfachte Programmierschnittstelle für Audioaufnahmen, und
- URL-basierten Zugriff auf Ressourcen.

Damit sind qualitativ hochwertige Mehrkanal-Audioaufnahmen über das WWW möglich.

Alternative Audiobibliotheken wie z. B. QuickTime for Java oder ähnliche können in `ipsk.audio` ebenfalls eingebunden werden.

Aufnahmeskript

Das Aufnahmeskript definiert den Inhalt einer Aufnahmesitzung und die Reihenfolge, in der die Promptitems präsentiert und aufgenommen werden [2]. Ein Aufnahmeskript ist in XML formatiert und besteht aus Metadaten und Abschnitten. Ein Abschnitt besteht aus Promptitems.

Multimediaprompts

Ein Promptitem besteht aus einer Anweisung, dem eigentlichen Prompt und einem Kommentar. Anweisung und Prompt werden dem Sprecher präsentiert, der Kommentar ist nur für den Aufnahmeleiter sichtbar.

Ein Promptitem ist entweder ein Text-, Bild-, Audio- oder Videoprompt. Zur Darstellung von Textprompts kann die gesamte von Java unterstützte Unicode Zeichentabelle oder eine externe HTML-formatierte Datei verwendet werden (siehe z. B. Abb. 2). Bild-, Audio- und Videoprompts werden aus externen Dateien eingebunden. Sie sind vor allem dann sinnvoll, wenn keine konkreten Wortvorgaben gemacht werden sollen.

Aufnahmeprotokoll

In SpeechRecorder besteht eine einzelne Aufnahme aus mehreren Phasen, die zyklisch durchlaufen werden. Die wichtigsten Phasen sind

- *idle*: keine Aufnahme,
- *prerecording*: Zeit zwischen Beginn der Aufnahme und Anzeige des Prompts,
- *recording*: Aufnahme mit Anzeige des Prompts,
- *postrecording*: Nachlaufzeit der Aufnahme.

Prerecording wird üblicherweise dazu verwendet, für eine festgelegte Zeitdauer Umgebungsgeräusche ohne Sprache des Sprechers aufzunehmen. Während der Phase *recording* wird der Prompt angezeigt und die Äußerung

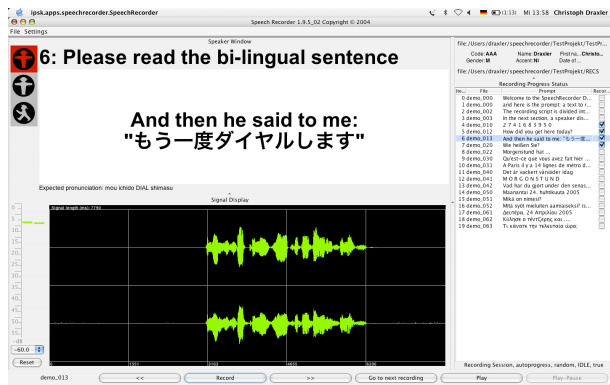


Abbildung 1: Versuchsleitersicht

des Sprechers aufgezeichnet. Diese Phase endet nach Ablauf der vorgegebenen Zeit oder durch Drücken der Stoptaste. Die Phase *postrecording* wird vorwiegend dazu benutzt, ein Abschneiden der Aufnahme durch vorzeitiges Drücken der Stoptaste zu verhindern. Der Wechsel der Phasen wird mit Zeitstempel protokolliert. Damit können Benutzerreaktionen auch zeitlich erfasst und ausgewertet werden.

Benutzersichten

SpeechRecorder bietet zwei unterschiedliche Sichten an: in der Sprechersicht sind nur der Prompt und die Ampel zur Aufnahmesteuerung sichtbar, in der Versuchsleitersicht zusätzlich noch Signaldisplay, Aussteuerungsanzeige, das Aufnahmeskript und Schaltflächen zur Steuerung des Ablaufs (Abb. 1, 2). Diese Sichten können einander überlappen oder auf getrennten Monitoren einzeln ausgegeben werden.

SpeechRecorder im Einsatz

Bereits während der Entwicklung wurde SpeechRecorder in konkreten Projekten eingesetzt. Die Erfahrungen im Einsatz der Software konnten so frühzeitig berücksichtigt werden.

Im Frühjahr 2003 hat das IPSK Sprachaufnahmen im fahrenden Auto für die Robert Bosch GmbH durch-

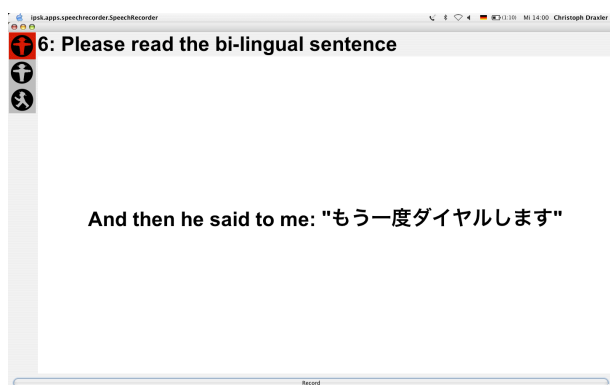


Abbildung 2: Sprechersicht

geführt. Für diese Aufnahmen wurden Audioprompts verwendet, um einerseits den Fahrer nicht durch Ablesen von Text von einem Bildschirm von der Fahrt abzulenken, und um andererseits einen typischen Dialog mit wechselnden Dialogpartnern zu simulieren.

Im Sommer 2003 wurden in St. Petersburg russische Aufnahmen gemäß dem IPA Handbuch durchgeführt [1]. Seit Herbst 2003 finden am IPSK Sprachaufnahmen für ein öffentlich verfügbares deutsches Sprachsynthese Korpus statt. Aufgenommen werden sowohl phonetisch reiche Sätze als auch Logatome mit Diphon-Kombinationen aller deutschen und der charakteristischen englischen und französischen Phoneme [3].

Seit November 2004 wird SpeechRecorder im Projekt Ph@ttSessionz eingesetzt. In Ph@ttSessionz werden im gesamten Bundesgebiet Sprachaufnahmen an Schulen durchgeführt, um eine qualitativ hochwertige Sprachdatenbank mit regional gefärbter Sprache Jugendlicher zu erstellen [4].

Ausblick

SpeechRecorder läuft stabil und hat sich im praktischen Einsatz bewährt. Die Definition eines Aufnahmeskripts allerdings muss noch manuell erfolgen, was einen XML-Editor voraussetzt. Wünschenswert ist ein grafischer Editor für Aufnahmeskripte, der einfache Konsistenzkontrollen wie die Überprüfung externer Dateien oder die Eindeutigkeit der Dateinamen der aufgenommenen Audiodateien durchführt.

In der aktuellen Version wird für jedes Promptitem eines Aufnahmeskripts eine eigene Audiodatei angelegt. Eine Alternative dazu sind kontinuierliche Aufnahmen mit externen Zeitmarken. Damit lassen sich ganze Aufnahmesitzungen komplett aufnehmen, und zugleich die einzelnen Äußerungen gezielt extrahieren.

Danksagung

Teile dieser Arbeit wurden aus Mitteln des BITS Projektes (Vorhaben 011VB01) vom Bundesministerium für Bildung und Forschung finanziert.

Literatur

- [1] O. Dioubina, H. Pfitzinger: Illustration of the Russian Language, Journal of the IPA, in Vorbereitung
- [2] Chr. Draxler, K. Jänsch: SpeechRecorder – a Universal Platform Independent Multi-Channel AudioRecording Software, Proceedings of LREC, Lissabon, 2004
- [3] T. Ellbogen, F. Schiel, A. Steffen: The BITS Speech Synthesis Corpus for German, Proceedings of LREC, Lissabon, 2004
- [4] A. Steffen, Chr. Draxler, A. Baumann, S. Schmidt: Ph@ttSessionz: Aufbau einer Datenbank mit Jugendsprache, Proceedings of DAGA 2005, München
- [5] www.phonetik.uni-muenchen.de/speechrecorder