

Prof. Dr. Marcus Spies
Christian Leibold

Übungen zur Vorlesung
Automatische Spracherkennung und integrierte Dialoganwendungen
Sommersemester 2005

Blatt 2

Zur Bearbeitung der Aufgaben können Sie in Teams zu ZWEI Personen arbeiten. Abgabe der Lösungen zu den markierten Aufgaben bis Dienstag, 14.06.2005, 12 Uhr an Christian.Leibold@stud.ifi.lmu.de. Namen und Matrikelnummer nicht vergessen!

Aufgabe 2-1 Installation CMU Sphinx:

Unter der Adresse „<http://sourceforge.net/projects/cmusphinx/>“ finden Sie die Übersicht des CMU – Sphinx Projektes. Laden Sie sich die aktuelle Version von Sphinx 4 mit den Sourcen herunter und installieren Sie die Software gemäß der Beschreibung die im Packet enthalten ist (Datei: index.html). Sie benötigen dazu eine Version von „ant“ (mindestens 1.6.0, erhältlich unter <http://ant.apache.org/>). Überprüfen Sie die Funktion Ihrer Installation durch den Aufruf der enthaltenen Demoprogramme HelloDigits.java oder ZipCity.java. Zum Test Ihrer Mikrofoneinstellungen können Sie das enthaltene Audiotool verwenden.

Aufgabe 2-2 Konfiguration von Sphinx: (Abgabe erforderlich)

Beachten Sie, dass einige Konfigurationen zu Fehlern in der Spracherkennung und im Programmverlauf führen können. Versuchen Sie dann in Ihrer Lösung auf diese Fehler einzugehen und die Ursache abzuschätzen.

- a) In den Programmordnern von Sphinx finden sie Dateien mit den Endung „*.gram“ und *.trigramm).lm. Beschreiben Sie kurz die Funktion (nicht die Funktionsweise) dieser Dateien.
- b) Für jedes Programm in Sphinx stehen mit der Datei <Programmname>.config.xml detaillierte Konfigurationsmöglichkeiten zur Verfügung. Grenzen Sie die Bedeutung der Tags <component>, <property> und <item> voneinander ab.
- c) Gehen Sie kurz auf die Funktionsweise des Interface PropertySheet ein, wie werden Variablen behandelt?

Aufgabe 2-3 Live Demo / Bewertung Frontend: (Abgabe erforderlich)

Im Verzeichnis <sphinx4>/test/live finden Sie den LiveDecoder. Diesen rufen Sie durch „ant live-ep“ auf. Weitere Informationen finden Sie in der Datei „README.html“ im selben Verzeichnis. Bitte wählen Sie nun einen Decoder für diese Aufgabe aus (am besten „weather“, „bishop“ „connected digits“ oder „alpha spelling“). Verwenden Sie für alle Teilaufgaben denselben Decoder.

Zur Lösung der Aufgabe nehmen Sie in der betreffenden *.config.xml Anpassungen vor. In der Datei <Sphinx4>/javadoc/constant-values.html können Sie weitere Informationen zu den Standardvariablen erhalten.

Vergleichen Sie die eingestellten Werte hinsichtlich Erkennungsrate (word accuracy), Echtzeitverhalten (*RT) und Geschwindigkeit der gesprochenen Sprache(subjektiv).

Aufgaben:

- a) Testen Sie wie sich das Hinzufügen und Weglassen der Items in der „pipeline“ der Komponente „Frontend“ auf die Erkennungsqualität und Laufzeit auswirken. Bei welchen Items macht das Weglassen von vorneherein keinen Sinn?
- b) Wählen Sie als `sampleRate` für das Mikrofon Werte zwischen 8000 und 44100 Hz.
- c) Vergleichen Sie unterschiedliche Werte für `windowShiftInMs` und `windowSizeInMs`.
- d) Nehmen Sie nun Feineinstellung für die Verwendung der Fourier Transformation vor. Vergleichen Sie die Ergebnisse mit den Werten 64, 128 und 512 für `numberFftPoints`.
- e) Variieren Sie Samplerate und Anzahl der eingesetzten Mel Filter. Welche Werte haben Sie ausprobiert und welche sehen sie als sinnvolle Mindestanzahl an?
- f) Reflektieren Sie kurz Ihre Ergebnisse.