

E90 - BRAVOZ - RECONHECIMENTO DE VOZ COM SUPORTE A GRANDES VOCABULÁRIOS (LVCSR) PARA O PORTUGUÊS BRASILEIRO: DESENVOLVIMENTO DE RECURSOS LINGÜÍSTICOS

Vanessa Davanzo (BIC/UCS), André Gustavo Adami - Deptº Informática/UCS - nessoila@gmail.com.br

Atualmente o processamento de voz é visto como um desafio a ser enfrentado pela computação, engenharia e lingüística no desenvolvimento de máquinas capazes de interagir de forma natural com os seres humanos. O projeto de pesquisa surgiu da importância do processamento da voz (ou fala) na computação moderna e da necessidade de se desenvolver um sistema único de reconhecimento automático de voz para o Português Brasileiro (PB), ainda mais no que diz respeito a um suporte a grandes vocabulários (acima de 30 mil palavras), tal sistema é identificado como *Large Vocabulary Continuous Speech Recognition* (LVCSR). Entre as dificuldades para o desenvolvimento de sistemas LVCSR estão a criação de um *corpus* de voz digitalizada e transcrita (para treinamento de modelo acústico) e de recursos específicos ao PB (dicionário fonético). Um dos objetivos do projeto é disponibilizar transcrições ortográficas e fonéticas de bases vocais em PB. Tais recursos poderão ser utilizados para o estudo e desenvolvimento de novos LVCSR. Nesta primeira etapa do projeto, estão sendo realizadas as transcrições fonéticas e ortográficas de bases de dados de voz como a seção das locuções em PB da *OGI 22-Language* e Spoltech (projeto desenvolvido na própria UCS). Estes recursos integrados serão utilizados, numa segunda etapa, para desenvolver um sistema de referência para o PB utilizando *toolkits open source* específicos para sistemas de reconhecimento de voz. Os recursos e conhecimentos produzidos no projeto serão compartilhados via internet para que outros pesquisadores possam utilizar em sua pesquisa e assim avançar a pesquisa em sistemas para reconhecimento de fala para o PB.

Palavras-chave; reconhecimento de voz, processamento de voz, transições fonéticas e ortográficas

Apoio: UCS