

BRAVOZ - Desenvolvimento de um Sistema de Reconhecimento de Fala para o Português Brasileiro

Cristian Keil de Abreu (PIBIC-CNPq), André Gustavo Adami (orientador) - ckabreu@ucs.br

O processamento de voz possui uma grande importância na computação moderna. Além do aspecto econômico, considerando os investimentos realizados por grandes empresas no desenvolvimento de soluções para a disseminação desta tecnologia, existe também o desafio de construir máquinas com a capacidade de interagir naturalmente com os seres humanos. Entretanto, apesar da sua reconhecida importância, ainda não existe em domínio público um único sistema de reconhecimento automático de voz para o português brasileiro (PB), sendo mais problemático ainda quando se trata de suporte para grandes vocabulários (mais de 30.000 palavras). A maioria dos trabalhos na área de processamento de voz utilizam palavras isoladas ou vocabulários reduzidos, pois nestes casos não necessitam de recursos avançados que sejam específicos à língua portuguesa. Para o desenvolvimento de um sistema de reconhecimento de fala contínua para o Português brasileiro, está sendo utilizada a ferramenta de reconhecimento de fala conhecida como HTK. Mundialmente utilizada, esta ferramenta é baseada no modelo estatístico conhecido por Modelo Oculto de Markov (HMM), a qual permite modelar qualquer série temporal. O desenvolvimento de um sistema de reconhecimento pode ser dividido em dois estágios principais. O primeiro estágio trata do treinamento do sistema, na qual os parâmetros de um conjunto de HMMs são estimados com base nos fonemas disponíveis em transcrições fonéticas (intervalos de tempo onde cada fonema é produzido) a fim de maximizar a taxa de palavras corretamente reconhecidas. O segundo estágio trata da avaliação da taxa de reconhecimento do sistema utilizando transcrições ortográficas de locuções previamente gravadas que nunca foram utilizadas durante o estágio de treinamento. Para os testes com o HTK, está sendo utilizada o corpus de fala Spoltech (coletado em diversas regiões do Brasil) com as transcrições fonéticas no padrão SAMPA. Os testes visam a avaliação do desempenho de um sistema de reconhecimento de fala treinado com base no alinhamento fonético produzido pelo HTK e no alinhamento fonético produzido por um foneticista. Resultados mostram que a diferença de desempenho não é estatisticamente significativa apesar da precisão do alinhamento manual. Uma das razões por tal resultado é que o alinhamento executado pelo HTK visa a generalização enquanto que o alinhamento manual adiciona as idiossincrasias do locutor que afetam a generalização dos modelos.

Palavras-chave: processamento de fala, reconhecimento de fala contínua, HTK.

Apoio: UCS, CNPq.