



T524

CONVERSÃO ORTOGRÁFICO-FONÉTICA AUTOMÁTICA VISANDO A CRIAÇÃO DE UMA BASE DE DADOS PÚBLICA PARA APLICAÇÃO EM PROCESSAMENTO DE FALA

Flávia Esteluti Lomba (Bolsista PIBIC/CNPq) e Prof. Dr. Fábio Violaro (Orientador), Faculdade de Engenharia Elétrica e de Computação - FEEC, UNICAMP

O projeto proposto se insere num projeto maior destinado à criação de uma base de fala pública visando pesquisas em processamento de fala em língua portuguesa falada no Brasil. Tal base é essencial para se comparar resultados obtidos por diferentes pesquisadores trabalhando em diferentes algoritmos de reconhecimento de fala com independência de locutor. O objetivo específico do presente projeto de Iniciação Científica é, a partir do software Ortofon desenvolvido no LAFAPE-IEL, adaptá-lo com as devidas alterações para efetuar a transcrição ortográfico-fônica automática de todo o material de texto empregado na criação da base de fala. Foram implementados para a nova versão do Ortofon um dicionário de abreviações, um dicionário de estrangeirismos, além de se completar o dicionário de siglas já implementado na versão mais antiga do software. Também foram implementadas as opções de coarticulação entre palavras adjacentes e de transcrição de uma lista de arquivos de texto ao invés de apenas um arquivo por vez. Adicionalmente foram aperfeiçoadas e corrigidas várias regras de transcrição implementadas anteriormente. Já foram efetuados testes de transcrição com extensos arquivos de fala, da ordem de Mbytes, demonstrando a eficácia e robustez da nova versão do software.

Conversão ortográfico-fônica - Base pública de fala - Reconhecimento de fala