



RECONHECIMENTO DE FALA PARA LEGENDAGEM DE ÁUDIO COM APLICAÇÃO DE DEEP LEARNING

LUIZ ALBERTO PINTO; KASSIUS VINICIUS SIPOLATI BEZERRA

Introdução: Sistemas de reconhecimento automático de fala, além da extração das informações fonéticas, se baseiam também na análise do modelo de linguagem. Dessa forma, para a construção de sentenças, tais sistemas consideram os aspectos semânticos, bem como as informações contextuais. A aplicação desse modelo de linguagem permite que palavras em qualquer posição da frase sejam alteradas na conclusão da sentença pelo interlocutor, o que resulta em um reconhecimento da fala mais próximo do natural. Contudo, se considerarmos uma aplicação do reconhecimento de fala para inserção de legendas tendo em vista a acessibilidade à programas de TV de pessoas com deficiência auditiva, a redução do tempo de leitura pode ser um problema para a completa acessibilidade destas pessoas ao conteúdo televisivo. **Objetivo:** Esse trabalho propõe um sistema de reconhecimento automático de voz que, enfatizando a estrutura fonética do texto reduza os erros de reconhecimento e aumente o tempo de exposição para leitura do texto legendado por pessoas portadoras de deficiências auditivas. **Metodologia:** Os modelos foram construídos com dois tipos de descritores dos sinais de fala, os coeficientes cepstrais na escala Mel (MFCC) e os coeficientes extraídos pela Short-Time Fourier Transform (STFT). Para a classificação, as MFCCs e as STFTs foram extraídas de arquivos de áudio de um conjunto de dados de teste e foram convertidas em imagens que constituíram as entradas de uma Rede Neural Convolucional. Para este trabalho foi utilizada a arquitetura de rede GoogLeNet. **Resultados:** O modelo construído com a STFT mostrou-se superior ao da MFCC com uma acurácia média de 99,5%, F1-Score igual a 99,4%, precisão de 99,4% e recall igual a 99,5%. **Conclusão:** Os resultados mostram que a combinação da STFT com a rede GoogLeNet possibilita a obtenção de sistemas de transcrição de fala para texto, de forma a melhorar as condições de acessibilidade de pessoas surdas aos programas de televisão.

Palavras-chave: **LEGENDAS OCULTAS; ACESSIBILIDADE EM PROGRAMAÇÃO DA TELEVISÃO; MFCC; STFT; GOOGLINET**