

Abstract

The range of voices is an essential aspect that a singer needs to know. This knowledge is necessary so that the singer can maximize their singing potential. Human vocal types are divided into six classes, namely alto, mezzo-soprano and soprano for women and bass, baritone and tenor for men. However, in this study the human vocal range was only classified into four classes commonly used in choirs, namely, alto (low voice for women), soprano (high voice for women), bass (low voice for men) and tenor (high voice for men). This research use Mel-frequency Cepstral Coefficient (MFCC) for its feature extraction and Convolutional Neural Network (CNN) for the classification. This study emphasized how MFCC and CNN was able to solve human vocal type classification problem. It is assisted by WavAugment for augmentation to maximize the learning process. In this study, the data used were primary so that the data were collected through surveys and experiments conducted directly by the researchers. The data used also affect the classification result, where the data need to be sparse enough to avoid the model being overfitted. The experiment is giving a good result where the training accuracy reaches 91.83% and testing accuracy is 91.14%. This model (for the extractor feature) is able to outperform the model that uses the Short-Time Fourier Transform (STFT) feature extractor, which has competitive results on several references, with a difference in training accuracy of 3.11% and test accuracy of 1.15%. In addition, the results of high precision and recall on the MFCC with an average of 91% state that the model with MFCC is able to provide accurate predictions. This study is a multi-disciplinary science that has a strong influence on music, especially in the choir. This study was conducted to improve choir music and computer technology continuity by combining music with computer science.

Keywords: MFCC, CNN, WavAugment, Vocal