Effects of measurement uncertainty on classification algorithms, as applied to vocal features for health assessment and early diagnosis

Abstract

The present manuscript deals with the analysis of vocal features applied to health assessment and early diagnosis of patients with the Parkinson's Disease and other pathological voices. In particular, the whole measuring chain has been characterized in order to identify the main uncertainty contributions of voice features and study their effect on classification algorithms. The main features analysed in this work are the pseudo-period and amplitude stability metrics, such as Jitter and shimmer, of sustained vowels recorded with a microphone in air. Additionally, the Cepstral Peak Prominence Smoothed, the Root Mean Square values and the Harmonics to Noise Ratio sequences, extracted from the sustained vowels, have been analysed in terms of descriptive statistics. To evaluate the uncertainty contributions of the stability metrics, an analytical evaluation has been carried out highlighting that the main uncertainty contributions for the period stability metrics are the time-base resolution of the Analog to Digital Converter, while for the amplitude stability metrics the Integral NonLinearity and the Gain Error were identified as important contributions along the amplitude resolution. The analytical evaluation of the features uncertainty showed to be very challenging for some stability metrics so a Monte Carlo uncertainty propagation has been carried out. The Monte Carlo propagation has been performed on the stability metrics to evaluate the effect of time and amplitude resolution on the bias and dispersion of the metrics under analysis, highlighting an important bias contribution on some stability metrics. Moreover a study on the effect of background noise has showed that the extraction algorithm represents the main uncertainty contribution for the evaluated stability metrics. In order to evaluate the uncertainty contribution of each component of the measuring chain, a vowel *re-synthesis* method has been proposed to produce artificial vowels with known pseudo-periods and amplitudes sequences. This method is based on the sampling of the original distributions of pseudo-periods and amplitudes sequences, which are used to produce reference sequences that are statistically comparable

to the original ones. Using this method a characterization of the measuring chain, composed by a microphone in air, a portable audio digital recorder and an extraction algorithm, has been carried out to evaluate the main uncertainty contributions to the voice features. This has been made possible thanks to the use of a Head and Torso Simulator, which replaces the original subject in the measuring chain and allows to perform repeatable and reproducible measurements. The analysis of the feature uncertainty highlighted that the vocal features are affected by a bias contribution and a dispersion contribution. The bias contribution can be different depending on the subject and on the length of the measuring chain, while the dispersion showed to be almost constant. This study highlighted that the extraction algorithm is the measuring chain component that mainly affects the evaluation of the voice features because the contribution of the whole measuring chain is comparable to the extraction contribution alone. The uncertainty evaluations performed in this work have been used to train binary weighted logistic regression models using a number of features in a range between 2 and 6. To train the models various strategies have been used to take advantage of the uncertainty evaluations of the measuring chain. In particular the effect of bias removal and the effect of evaluating the mixed terms of the uncertainty have been tested. The uncertainty analysis of the predicted probabilities was used to produce confidence intervals around the probabilities and thus the definition of a third class of non-classified. Thanks to this definition new classification metrics have been proposed to evaluate the classification performances and in particular the Realistic Accuracy has been defined excluding from the accuracy evaluations the non-classified subjects. These evaluations leaded to a maximum realistic accuracy of 100 % for the training and 77 % for the validation of the classification of Parkinson subject with respect to an healthy control group. For the Parkinson vs. Pathologic classification the training realistic accuracy reached values up to 96 % and 92 % for the validated one.