On the Use of Grey Zones in Automatic Voice Pathology Detection

Thomas Dubuisson[1], Thomas Drugman[1], Thierry Dutoit[1]

[1] TCTS Lab, Faculté Polytechnique, Université de Mons, Mons, Belgique
{ (thomas.dubuisson, thomas.drugman, thierry.dutoit)@umons.ac.be}

This study presents a voice pathology detection method based on features extracted from the speech signal and the glottal source. Safety mechanisms are included in the classification in order to ensure a reliable indication to the clinician about the presence of a pathology.

Features are computed from the normal and pathological sustained vowels of the MEEI database [1]. They are extracted from the speech signal and the glottal source [2], estimated by the IAIF method [3]. Considering frames centered on Glottal Closure Instant (GCIs), speech features consist in energy ratios between various perceptual spectral bands as well as the harmonic-to-noise ratio, while glottal source features are the glottal formant frequency and bandwidth, energy ratios between various perceptual spectral bands as well as the glottal discontinuity at the GCI. The most informative pair of features for the classification of normal and pathological frames is selected by computing mutual information measures. From [2], these two features are the energy ratio between a low frequency band and the whole frequency range in the speech signal and the glottal discontinuity at the GCI.

A training set is formed by randomly chosing 70% of the patients of the database. In this set, the probability distribution of the pair of features in each class is modeled by a Gaussian Mixture Model (GMM). For each possible couple value, the most likely class is the one exhibiting the highest probability. Thus, when a patient of the evaluation set (the remaining 30% of the patients of the database) has to be classified, the GCI-centered frames are extracted, the features are computed for each frame and the decision about the class of each frame is drawn based on the GMM outputs. The class assigned to the patient is finally the most represented class among its frames. The random partition of the database in a training and evaluation sets has been repeated 10 times and the average overall accuracy (the ratio between the number of correctly classified patients in the normal and pathological classes and the total number of patients classified in these two classes) of the classification is 96.7 %.

Two safety mechanisms are included in this classification. At the frame level, the first mechanism consists in considering a grey zone in the decision about the frame class. Indeed, when deciding the most likely class, the probabilities of the classes may be very close (e.g. 55% and 45%). It is therefore preferred to avoid classifying a frame as normal or pathological if the difference between the probabilities is lower than a threshold δ, leading to label the frame as unknown. At the patient level, the second mechanism is applied when the patient class is decided by merging the decision of all its frames. Let us consider a patient for whom 50%, 40% and 10% of his frames are respectively labeled as normal, pathological and unknown. This patient should be labeled as normal since this latter class is the most represented, although the percentage of pathological frames is considerable. Considering this, a patient is labeled as normal (respectively pathological) if the difference between the representation percentage of normal (respectively pathological) class and the most represented class between the two others is higher than a second threshold Δ. Otherwise, the patient is labeled as unknown. The application of the two grey zones ($\delta = \Delta = 20\%$) induces on average a decrease of the confusion between the normal and pathological classes and a decrease of the correct classification rates (for each class, the ratio between the number of correctly classified patients in the class and the total number of patients of the class in the evaluation set). However, these two effects are balanced by the fact that the overall accuracy is increased (98.7%), meaning that the application of the two grey zones allows a more reliable decision about the presence of a pathology.

References
[1] Kay Elemetrics Corporation (1994) Disodered Voice Database Model (version 1.03)
[2] Drugman T., Dubuisson T., Dutoit T. (2009) On the Mutual Information between Source and Filter Contributions for Voice Pathology Detection, *Proceedings of Interspeech 2009,* pp. 1463-1466.
[3] Alku P. (1992), Glottal Wave Analysis with Pitch Synchronous Iterative Adaptative Inverse Filtering, *Speech Communication*, vol. 11(2-3), pp. 109-177.