Pathology Detection by Speech Analysis: An Overview

Automatic speech recognition is a mature field with several commercial applications. More recently, speech analysis has been used for healthcare tasks such as pathology detection. Several pathologies have been considered so far, including physiological pathologies such as hoarseness, nodule, polyp, cyst, cancer, polypoid hyperplasia, keratosis, papilloma, and neurodegenerative brain disorders **(NBD)** such as Parkinson's disease **(PD)**, Alzheimer disease **(AD)**, and mild cognitive impairment **(MCI)**. In contrast to speech recognition, content of the speech is often not of interest but characteristics like prosody are. In this study, a literature survey on pathology detection is provided, by targeting the most studied pathologies.

For physiological pathologies, input generally consists of sustained vowels. However, some few works utilize continuous speech. Many low-level descriptors have been considered such as mel-frequency cepstral coefficients **(MFCC)**, jitter, pitch statistics, and linear predictive coding. For pathology detection, many classifiers rely on discriminative models such as support vector machine **(SVM)** and neural networks [1]. Generative models are also utilized [2]. The most used database is Kay-elemetrics-disordered-voice-database by Massachusetts-Eye-and-Ear-Infirmary (KayPentax). There are small databases for special pathologies like hoarseness and cancer. Classification accuracy between healthy and pathological speech is the standard performance metric. For KayPentax, results range from 80% to 96% for different methods and dataset configurations.

PD is the most investigated NBD. Vowel phonations are the standard input, albeit, in rare works, words and sentences are also considered [3]. Two different objectives are usually sought: speech classification as healthy or pathological, and the unified PD rating scale **(UPDRS)** estimation [4]. In addition to common features like jitter or MFCC, pitch period entropy and harmonics-to-noise-ratio have been shown to be useful. **PD classification** usually is based on discriminative models, and mostly SVM. Generative models are less used and hidden Markov-models, particularly, are never considered to the best of our knowledge. The Oxford PD Detection Dataset is the most popular dataset with 195 vowel phonations from 31 subjects, among which 23 have PD. Accuracies range from 65% to 98%. For **UPDRS estimation**, linear and non-linear regression methods are utilized. The largest dataset up-to our knowledge consists of 52 PD-patients. Samples from PD-patients only are considered. The mean absolute error is used as the performance metric and is reported to be around 2.

MCI is usually an early indicator of AD, and these two pathologies are often studied together [5]. Continuous speech, e.g. recording of personal stories is a useful type of input. Automatic spontaneous speech analysis and emotional temperature are proposed as useful descriptors. SVM is the preferred classifier for this task. No standard dataset exists, but continuous speech from MCI or AD patients and healthy controls are collected in independent works. Reported accuracies are around 79% for MCI-versus-healthy, 94% for AD-versus-healthy, 80% for MCI-versus-AD.

To conclude, early-stage pathology detection is important for patient treatment or stabilization so it may be the most important future direction of interest. Another direction is automatic extraction of features relevant to a special pathology.

This work was supported by Embassy of France in Turkey.

References

[1] Kons, Z., Satt, A. and Hoory, R., “On feature extraction for voice pathology detection from speech signals”, In: Proceedings of the 1st Annual Afeka-AVIOS Speech Processing Conference, Tel Aviv Academic College of Engineering, Tel Aviv, Israel, 2011.

[2] Little, M.; McSharry, P.; Moroz, I.; Roberts, S., "Nonlinear, Biophysically-Informed Speech Pathology Detection," Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on , vol.2, no., pp.II,II, 14-19 May 2006.

[3] Sakar, B.E.; Isenkul, M.E.; Sakar, C.O.; Sertbas, A.; Gurgen, F.; Delil, S.; Apaydin, H.; Kursun, O., "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings," in Biomedical and Health Informatics, IEEE Journal of , vol.17, no.4, pp.828-834, July 2013.

[4] Tsanas, Athanasios; Little, Max A.; McSharry, Patrick E.; Ramig, Lorraine O., “Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity”, Systems Analysis, Modelling and Prediction (SAMP) group, Mathematical Institute and Department of Engineering Science, University of Oxford, Oxford, UK, Journal of The Royal Society Interface, 11/2010; 8(59):842-55.

[5] Alexandra König, Aharon Satt, Alexander Sorin, Ron Hoory, Orith Toledo-Ronen, Alexandre Derreumaux, Valeria Manera, Frans Verhey, Pauline Aalten, Phillipe H. Robert, Renaud David, Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, Volume 1, Issue 1, March 2015, Pages 112-124, ISSN 2352-8729.