Abstract

The present thesis addresses the automatic analysis of speech disorders resulting from Parkinson's disease and hearing loss. For Parkinson's disease, the progression of speech symptoms are evaluated considering speech recordings captured in the short-term (4 months) and long-term (5 years). Machine learning methods are used to perform three tasks: (1) automatic classification of patients vs. healthy speakers, (2) regression analysis to predict the dysarthria level and neurological state, and (3) speaker embeddings to analyze the progression of the speech symptoms over time. For hearing loss, automatic acoustic analysis is performed to evaluate whether the duration and onset of deafness (before or after speech acquisition) influences the speech production of cochlear implant users. Additionally, articulation, prosody, and phonemic analyses are performed to show that cochlear implant users present altered speech production even after hearing rehabilitation.

Automatic acoustic analysis is performed considering phonation, articulation, prosody, and phonemic features. Phoneme precision is characterized using the posterior probabilities obtained from recurrent neural networks trained in German and Spanish. The phonemic analysis considers three main dimensions: manner of articulation, place of articulation, and voicing. This thesis also proposes a methodology for automatically detecting voice onset time in voiceless stop consonants.

Furthermore, this thesis studies the acoustic cues that reflect changes in elderly people due to the aging process. Regression analysis is performed to estimate a person's age using the phonation, articulation, prosody, and phonemic features. Additionally, the use of smartphones for health care applications is considered here.

Zusammenfassung

Die vorliegende Dissertation befasst sich mit der automatischen Analyse von Sprachstörungen infolge von Parkinson und Hörverlust. Bei der Parkinson-Krankheit wird der Verlauf der Sprachsymptome anhand von Sprachaufzeichnungen bewertet, die kurzzeitig (4 Monate) und langfristig (5 Jahre) aufgenommen wurden. Methoden des maschinellen Lernens werden verwendet, um drei Aufgaben zu erfüllen: (1) automatische Klassifikation von Patienten vs. gesunde Sprecher, (2) Regressionsanalyse zur Vorhersage des Dysarthrie-Levels und des neurologischen Zustands und (3) Sprechereinbettungen zur Analyse des Verlaufs der Sprachsymptome im Laufe der Zeit. Bei den Patienten mit Hörverlust wird eine automatische akustische Sprachanalyse durchgeführt, um zu beurteilen, ob die Dauer und das Einsetzen der Taubheit (vor oder nach dem Spracherwerb) die Sprachproduktion von Cochlea-Implantat-Trägern beeinflusst. Darüber hinaus werden Artikulations-, Prosodie- und Phonemanalysen durchgeführt, um zu zeigen, dass Träger von Cochlea-Implantaten auch nach einer Hörrehabilitation eine veränderte Sprachproduktion unterschiedlichen Ausmasses aufweisen.

Für automatischen akustischen Analysen werden wird Phonation, Artikulation, Prosodie und phonemischen Merkmalen berücksichtigt. Die Phonempräzision wird durch die Posterior-Wahrscheinlichkeiten charakterisiert, die aus rekurrenten neuronalen Netzen gewonnen werden, die auf Deutsch und Spanisch trainiert wurden. Die phonemische Analyse fokussiert auf drei Hauptdimensionen: Artikulationsart, Artikulationsort und Stimmgebung. Diese Arbeit schlägt auch eine Methodik zur automatischen Erkennung der Stimmeinsatzes nach stimmlosen Stoppkonsonanten vor.

Darüber hinaus untersucht diese Arbeit die akustischen sprachlichen Charakteristika, die Veränderungen bei älteren Menschen aufgrund des Alterungsprozesses widerspiegeln. Eine Regressionsanalyse wird durchgeführt, um das Alter einer Person unter Verwendung der Phonation, Artikulation, Prosodie und phonemischen Merkmale zu schätzen. Darüber hinaus wird hier der Einsatz von Smartphones für Anwendungen im Gesundheitswesen betrachtet.

Resumen

La presente tesis aborda el análisis automático de los trastornos del habla derivados de la enfermedad de Parkinson y la pérdida auditiva. En el caso de la enfermedad de Parkinson, el progreso de los síntomas del habla se evalúa considerando las grabaciones capturadas a corto (4 meses) y largo plazo (5 años). Métodos de aprendizaje automático son utilizados para realizar tres tareas: (1) clasificación automática de pacientes contra a hablantes sanos, (2) análisis de regresión para predecir el nivel de disartria y el estado neurológico, y (3) modelos de hablante para análisis longitudinal del progreso de los desórdenes en la voz. En el caso de la pérdida auditiva, se realiza un análisis acústico automático para evaluar si la duración y el inicio de la sordera (antes o después de la adquisición del habla) influye en la producción del habla de los usuarios de implantes cocleares. Además, se realizan análisis de articulación, prosodia y fonémicos para demostrar que los usuarios de implantes cocleares presentan una producción del habla alterada incluso después de la rehabilitación auditiva.

El análisis acústico automático se realiza considerando fonación, articulación, prosodia y características fonémicas. La precisión de la producción de fonemas se caracteriza mediante el cálculo de las probabilidades obtenidas de redes neuronales recurrentes entrenadas en Alemán y Español. El análisis fonémico considera tres dimensiones principales: forma de articulación, lugar de articulación y sonorización. Esta tesis también propone una metodología para la detección automática del tiempo de inicio de la voz en consonantes oclusivas sordas.

Además, en este trabajo se analiza la influencia de la edad en el análisis acústico. El análisis de regresión se realiza para estimar la edad de una persona utilizando las características de fonación, articulación, prosodia y fonética. También, en esta tesis se considera el uso de smartphones para aplicaciones en el sector médico.

Chapter 1

Introduction

1.1 Motivation

Oral communication of adults and children can be affected by developmental or acquired speech disorders resulting from motor/neurological impairments (e.g., brain injuries, Parkinson's disease) or sensory/perceptual disorders (e.g., hearing loss)[1]. On the one hand, neurological diseases such as Parkinson's disease (PD) affect certain regions in the brain and the muscles involved in the speech production process, leading to different motor speech-based impairments such as imprecise articulation, slower speaking rate, monotonous speech, hoarse quality of voice, among others (Ho et al., 1999; Trail et al., 2005). On the other hand, perceptual disorders such as sensorineural hearing loss cause decreased speech intelligibility, changes in terms of phoneme articulation, abnormal nasalization, slower speaking rate, and decreased variability in fundamental frequency (Hudgins and Numbers, 1942; Langereis et al., 1997; Leder et al., 1987). One of the aims of pathological speech processing is the development of technology to support the diagnosis and monitoring of different medical conditions through speech (Gupta et al., 2016). This thesis focuses on the automatic acoustic analysis of speech signals captured from PD patients and people with hearing loss. Furthermore, as the speech of elderly people

[1]www.asha.org/Practice-Portal/Clinical-Topics/Articulation-and-Phonology

changes due to the aging process, a clinical condition, or both, the description of acoustic cues in the speech that reflect such differences is a topic that deserves special attention.

PD is a neurodegenerative disease characterized by the progressive loss of dopaminergic neurons in the substantia nigra of the midbrain (Hornykiewicz, 1998). The primary motor symptoms of PD include tremor, slowness, rigidity of the limbs and trunk, postural instability, swallowing disorders, and speech impairments. Many of the symptoms are controlled with medication, however, there is no clear evidence indicating positive effects of those treatments on the speech impairments (Skodda et al., 2010), but there is evidence showing that speech therapy combined with the pharmacological treatment improves the communication ability of PD patients (Schultz and Grant, 2000). The evaluation of PD requires the patient to be present at the clinic, which is time-consuming and expensive for both, the patient and healthcare system (Yang et al., 2020), however, the continuous monitoring of PD patients could help to make timely decisions regarding their medication and therapy.

In the case of hearing loss, there are different treatments available for different types and degrees of deafness. A Cochlear implant (CI) is the most suitable device for severe and profound deafness when hearing aids do not improve sufficiently speech perception. A CI uses a sound processor to capture audio signals and send them to a receiver implanted under the skin behind the ear. The receiver transforms the signal into electrical impulses which are sent to electrodes implanted in the cochlea. However, CI users often present altered speech production and limited understanding even after hearing rehabilitation. Thus, if the deficits of speech would be better known the rehabilitation might be properly addressed (Pomaville and Kladopoulos, 2013). CI users require assistance before, during, and after surgery from audiologists, medical specialists in Otorhinolaryngology, and speech-language pathologists [2]; however, speech production quality is seldom assessed in outcome evaluations, thus including speech technology could lead to a reliable outcome evaluation

[2] www.asha.org/Practice-Portal/Professional-Issues/Cochlear-Implants/

contributing to the rehabilitation success.

This thesis addresses the automatic evaluation of speech production from PD patients and CI users by combining signal processing techniques with machine learning methods. Such methods are also considered to analyze the effect of age as another possible source of changes in speech production. Additionally, since the use of smartphones for health care has become more frequent, some of the speech processing techniques addressed in this thesis are implemented in Android-based applications.

1.2 Speech disorders in selected populations

1.2.1 Parkinson's disease

Clinical diagnosis

Parkinson's disease is characterized by a combination of some symptoms regarding motor control. Moreover, next to motor control, other symptoms such as mood changes, cognitive decline, and sleep disorders might occur (Poewe, 2008). There is no standard method to diagnose PD. Doctors rely on the clinical history and physical examination to assess the patients. Additionally, the severity of the disease is evaluated by neurologist experts using different scales such as the Movement Disorder Society–Unified Parkinson Disease Rating Scale (MDS-UPDRS) (Goetz et al., 2008). This is a perceptual scale used to assess motor and non-motor abilities of the patients with 65 items distributed in four sections:

- Section 1 (MDS-UPDRS-I, 13 items) concerns the non-motor experiences of daily living such as cognitive impairment, depressed mood, and fatigue.
- Section 2 (MDS-UPDRS-II, 13 items) considers motor experiences of daily living such as eating, dressing, handwriting, and tremor.