Aalto University

CS-E4070, Special Course in Machine Learning and Data Science: Signal processing and machine learning methods for speech-based biomarking of human health (5 credits)

Prof. Paavo Alku
Department of Signal Processing and Acoustics
School of Electrical Engineering
Aalto University

27.1.2021

Contents

1. Course practices in spring ‘21
2. Introduction to the topic
3. Schedule

1. Course practices in spring ‘21

- Topic of this semester: Signal processing and machine learning methods for speech-based biomarking of human health.
- Pre-requisites: In order to register to the course in 2021, the student must have **basic knowledge in speech processing**. This means that the student must have taken either ELEC-E5500 Speech Processing or a similar course in another university. In the latter case, the student must contact the teacher before registering to the course.
- The course is held as a seminar (in zoom) where the teacher first gives an introduction lecture on the topic.

- After the introduction lecture, the course continues in weekly seminars where **students present selected articles** on the topic. The articles will be selected by the teacher and assigned to each student. Each student gives 1-2 presentations. The student presentation includes:

(1) Preparation of slides (about 10-20 pages) on the selected article.

(2) Presenting the slides to the audience in zoom meetings.

(3) Preparation of 2-3 simple assignments on the presented article.

(4) The student returns the model solution/answer to the teacher

- In order to pass the course, the student must (a) give his/her own presentation(s), (b) must have answered correctly to 80% of the assignments given by the other presenters and (c) must be present in at least 80% of the presentations by other students.

- In order to get a deeper understanding of the topic **using real speech data**, the students who have passed the course are encouraged to continue by the following means:

  (1) MSc students:

  (1.1): You have a possibility to take SPA-EV, a course with varying contents (1-10 credits), see:

  www.aalto.fi/en/services/course-codes-at-school-of-electrical-engineering

  (1.2) Discuss with the teacher and your own professor about possibilities to start a MSc thesis on the topic.

  (2) PhD students: select a topic of your own interest (which is related to speech-based biomarking of human health) and continue studying the topic by taking course ELEC029Z-LZ (1-10 credits), see:

  www.aalto.fi/en/services/course-codes-at-school-of-electrical-engineering

2. Introduction to the topic

- The main role of speech is to enable communication between people by transferring linguistic information between speakers.
- In addition to its linguistic content, the speech signal, however, includes plenty of other information. This information includes **paralinguistic** issues such as vocal emotions (e.g. angry/sad/happy speech) and **speaker states and traits** (e.g. gender, age, height, state of health etc. of the speaker).
- A research question of increasing interest belonging to the latter category is **how to predict the speaker's state of health using his/her speech signal.**
- This research question calls for signal processing and machine learning knowledge and will be the topic of the course in 2021.

- The topic will be studied in the course by discussing a group of recent articles investigating the speech-based biomarking of human health. The articles have been selected in order to address the following issues of the topic.

2.1 The benefits of speech-based technologies

- Speech-based biomarking of the state of health **does not** replace the clinical diagnosis.
- However, the topic benefits from the following issues.

(1) The input signal to the biomarking system, the speech microphone signal, can be recorded **non-invasively** in a **comfortable** manner using a **cost-effective** devise (e.g. mobile phone).

(2) The speech-based biomarking can be **conducted outside hospital** using a system that is easy to administer and can be used by the patient at home, thereby avoiding frequent and often inconvenient visits to the clinic.

- The issues above are particularly useful, for example, in efforts to tackle neurodegenerative diseases (such as Parkinson’s disease and Alzheimer’s disease) which have become serious global health problems due to aging of populations.
- Particularly in neurodegenerative diseases, the speech-based biomarking can be used in **the early detection of the disease from telephone speech recordings**.
- Even though the speech-based biomarking does not replace clinical examinations, it can be used in **preventive healthcare** technology to detect diseases at **an early stage** and to track physiological changes caused by the disease.

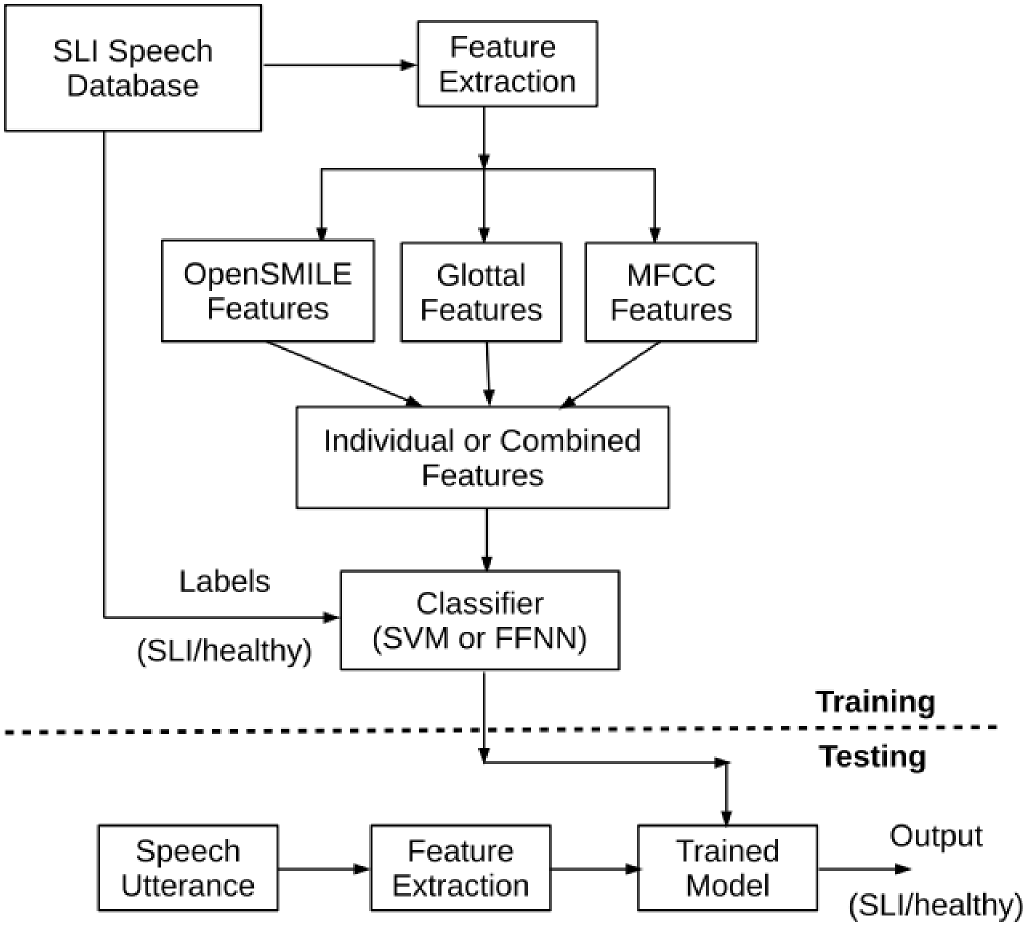
2.2 Machine learning tasks studied in the topic

- The topic has been mainly studies as **a binary classification problem** by investigating, for example, the speech-based detection of patients with Parkinson’s disease from their healthy controls (Orozco-Arroyave et al., 2016).
- Some of the studies have addressed the biomarking topic from the **multiclass classification’s** point of view by studying, for example, 4-class classification where patients suffering from three known voice production disorders are classified from healthy talkers (Chui et al., 2020).
- The **assessment of the severity** of the underlying disorder has also been studies related, for example, to Parkinson’s disease (Bayestehtashk et al., 2015).
- Some studies have addressed progression of the underlying disorder using **longitudinal analysis** (e.g. Arias-Vergara et al., 2018).

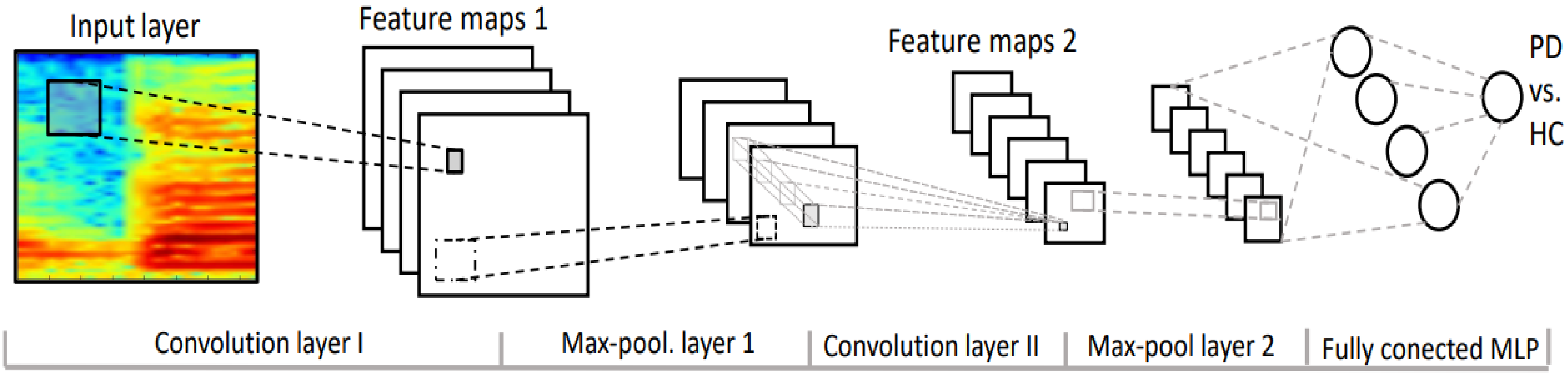
2.3 Technology

- One of the main goals of the seminar is to familiarise the student with the signal processing and machine learning methods which have been used in the study area.
- A conventional method is to use methods consisting of two parts: feature extraction and classifier.
- In the former, a wide range of acoustical methods have been used including classical, low-dimensional feature extraction methods such as mel-frequency cepstral coefficients (MFCCs) (Davis and Mermelstein, 1980) but also more high-dimensional features (such as openSMILE, Eyben et al. 2013) consisting of tens of different individual parameters.
- In the latter, conventional ML methods, particularly support vector machines (SVMs) (Cortes and Vapnik, 1995) have been taken advantage of.

- More recently, these **classical pipeline systems** have been increasingly replaced by **end-to-end systems** where the speech signal, either the raw time-domain signal (e.g. Millet and Zeghidour, 2019) or the spectrogram (e.g. Vasquez-Correa et al. 2017) is processed directly by deep learning methods (typically CNN) to solve the underlying task.
- Some studies have also combined classical, **hand-crafted features** and deep-learned features (e.g. He and Cao, 2018).



An example of a classification system based on the traditional pipeline approach in the detection of special language impairment of children (Reddy et al., 2020).



An example of a classification system based on the end-to-end approach in the detection of Parkinson’s disease (Vasquez-Correa et al., 2017).

2.2 Disorders

- Neurodegenerative diseases, particularly Parkinson’s disease and Alzheimer’s disease, are becoming prevalent globally due to aging of the populations. Parkinson’s disease has particularly been studied in the area of speech-based biomarking of human health (e.g. Arias-Vergara et al. 2018; Bayestehtashk et al., 2015; Vasquez-Correa et al., 2017.
- In addition, neurodegenerative diseases such as Alzheimer’s disease (e.g. Warnita et al., 2018) and ALS (e.g. Norel et al., 2018) have been investigated in the study area.
- Other examples of disorders investigated are depression (e.g. Jiang et al. 2017), voice production disorders (e.g. Gomez-Garcia et al., 2019) and sleep apnea (e.g. Botelho et al., 2019).

2.3 Speech databases and speaking tasks

- Studying the topic involves using data-driven approaches where network parameters are trained using real speech produced by speakers affected by the underlying health problem.
- Publicly available databases exist for some diseases such as dysarthria (the TORGO database (Rudzicz et al., 2012) and the UA Speech database (Kim et al., 2008)), voice production disorders (the Saarbrucken voice database, SVD), and special language impairment (Grill and Tučková, 2016).
- Some of the open databases are, however, fairly small (e.g. with 10-20 speakers each producing a few utterances) which might limit the use of modern data-hungry ML networks.

- Speaking tasks are various, including (a) simple repetitions of isolated vowel sounds and words, (b) text reading and (c) spontaneous speech.
- Some of the speaking tasks have been tailored to be more challenging to b produced by the underlying patient population. An example is the diadochokinetic (DDK) task where the speaker is asked to repeat three-syllable units (i.e. /pa/-/ta/-/ka/). The DDK task is widely used in studying Parkinson’s disease (Rusz et al., 2011).

References:

- Arias-Vergara et al. Speaker models for monitoring Parkinson’s disease progression considering different communication channels and acoustic conditions. Speech Communication, 101: 11-25, 2018.
- Bayestehtashk et al. Fully automated assessment of the severity of Parkinson’s disease from speech. Computer Speech and Language, 29: 172-185, 2015.
- Botelho et al. Speech as a biomarker for obstructive sleep apnea detection. Proc. ICASSP, pp. 5851-5855, 2019.
- Chui et al. Combined generative adversarial network and fuzzy c-means clustering for multi-class voice disorder detection with an imbalanced dataset. Applied Sciences 10: 4571, 2020.
- Cortes, Vapnik. Support-vector networks. Machine Learning 20: 273-279, 1995.

- Davis, Mermelstein. *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences.* IEEE Transactions on Acoustics, Speech, and Signal Processing, 28*:* 357-366, 1980*.*
- Eyben et al. Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In: Proc. ACM International Conference on Multimedia, pp. 835–838, 2013.
- Gomez-Garcia et al. On the design of automatic voice condition analysis systems Part II: Review of speaker recognition techniques and study on the effects of different variability factors. Biomedical Signal Processing and Control, 48: 128-143, 2019.
- Grill and Tučková. Speech databases of typical children and children with SLI. PLoS ONE, 11: Art. no. e0150365, 2016.
- He and Cao. Automated depression analysis using convolutional neural networks from speech. Journal of Biomedical Informatics, 83: 103-111, 2018.

- Jiang et al. Investigation of different speech types and emotions for detecting depression using different classifiers. Speech Communication, 90: 39-46, 2017.
- Kim et al. Dysarthric speech database for universal access research. Proc. Interspeech, pp. 1741-1744, 2008.
- Millet, Zeghidour. Learning to detect dysarthria from raw speech. Proc. ICASSP, pp. 5831-5835, 2019.
- Norel et al. Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis. Proc. Interspeech, pp. 377-381, 2018.
- Orozco-Arroyave et al. Automatic detection of Parkinson’s disease in running speech spoken in three different languages. Journal of the Acoustical Society of America, 139: 481-500, 2016.
- Reddy et al. Detection of specific language impairment in children using glottal source features. IEEE Access, 8: 15273-15279, 2020.

- Rudzicz et al. The TORGO database of acoustic and articulatory speech from speakers with dysarthria. Lang. Resour. Eval. 46: 523–541, 2012.
- Rusz et al. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease. Journal of the Acoustical Society of America, 129: 350-367, 2011.
- Saarbrucken voice database. http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4
- Vasquez-Correa et al. Convolutional neural network to model articulation impairments in patients with Parkinson’s disease. Proc. Interspeech, pp. 314-318, 2017.
- Warnita et al., Detecting Alzheimer’s disease using gated convolutional neural network from audio data. Proc. Interspeech, pp. 1706-1710, 2018.

3. Schedule

- Information about the seminar dates, presenters and papers to be discussed in the seminar will be given in the MyCourses
- A preliminary list of dates/presenters/papers will be mailed to the students by ??
- Is Wednesday at 13:15 – ca. 14:30 ok for all??