Automatic Detection and Severity Estimation for Oral Cancer Speech

Janay S.C. Monen[1], Bence M. Halpern[3,4,5], Teja Rebernik[1,4], Thomas Tienkamp[1], Rob J.J.H. van Son[3,4], Vass Verkhodanova[1], Max J.H. Witjes[1,2] and Martijn Wieling[1]
[1]University of Groningen, [2]University Medical Center Groningen, [3]University of Amsterdam, [4]Netherlands Cancer Institute, [5]Delft University of Technology
janay.monen@mpi.nl

University of Groningen

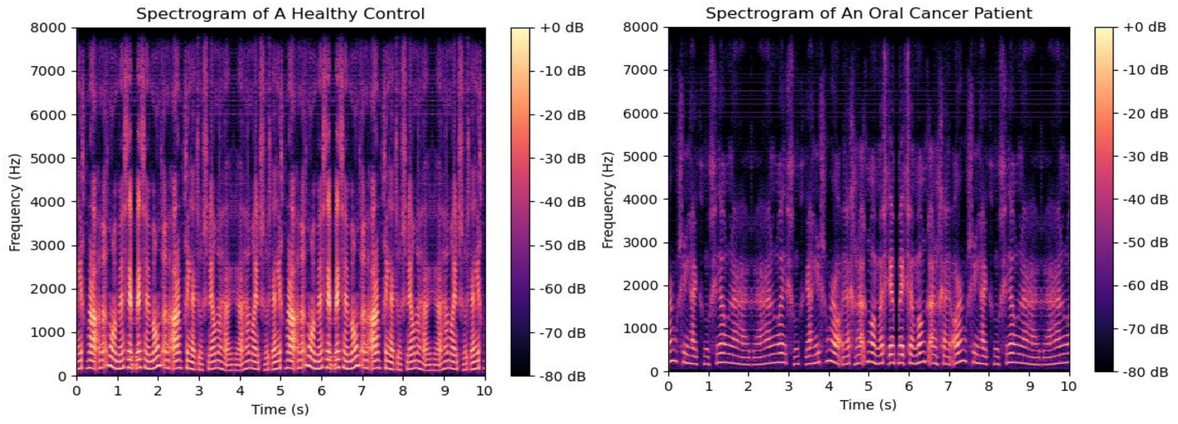
Can classification-based **machine learning** methods **detect oral cancer speech** AND/OR **estimate** the **severity** of it?

1. Motivation

- There is very limited research on oral cancer (OC) speech severity estimation using machine learning (ML) methods [1–3], and even less on the automatic detection of OC speech [4].
- Given the social and functional impact of impaired speech, severity estimation is a crucial element of the OC post-treatment phase, i.e., it allows for speech monitoring.
- Detection can help identify OC speech characteristics that can be used for severity estimation.
- Implementing ML methods generates more robust and unbiased outcomes.

Sentence – Dutch
Met zijn mand aan de arm ging hij toen snel naar huis.

Translation
With his basket on his arm, he then quickly went home.



2. Oral cancer

Definition
A type of abnormal and excessive tissue growth on the lip or oral cavity.

Main causes – etiological factors
Tobacco consumption and alcohol abuse [5].

Common speech impairments
- Reduced speech intelligibility
- Unstable phonation

Treatments
- Glossectomy (partial or full tongue removal)
- Surgical reconstruction
- Jaw surgery
- Radio-/Chemotherapy

3. Methods

Stimuli
- 227 Dutch sentences – read speech
  - Newspaper articles and stories
  - ca. 330 minutes of speech in total
    - ca. 30 minutes/participant

Feature extraction
- Long-term average spectrum (LTAS)
- Mel frequency cepstral coefficient (MFCC)

Building 4 simple classification models
- Logistic Regression (LR)
- Support Vector Machines (SVM)
- Multilayer Perceptrons (MLP)
- 1D Convolutional Neural Networks (1D-CNN)

Scan this QR code to see the model architectures

SHI: Speech Handicap Index – Dutch Adaptation

Translated example questions*
P1 *The speed with which I speak has changed.*
F5 *I avoid conversations with my family, friends, neighbors.*
E4 *People seem irritated by my speech impairments.*
* P = physical impact; F = functional impact; E = emotional impact.

Scoring
- Self-reported
- *'never'*, *'almost never'* *'sometimes'*, *'almost always'*, *'always'*
- 0–60 points

Assigning severity
- Score 0–20 => level 1
- Score 21–31 => level 2

Patient	Score	Severity
PT1	24	2
PT2	18	1
PT3	29	2
PT4	6	1
PT5	13	1
PT6	31	2

Participant information

OC patients	3 male, 3 female, >1 year post-surgery
Controls	2 male, 3 female
Age	44–77 years
Language	Dutch

Test conditions

OC Patients
Had tongue sensors attached to document articulatory trajectories.

Tasks: Detection and severity estimation

Detection	- OC and healthy speech (5 vs 5 speakers) - 5 * 5-fold cross-validation - 25 different speaker combinations - Leave-two-speaker-out
Severity estimation	- OC speech and SHI scores - 3-fold cross-validation - Leave-two-speaker-out
Metrics	- Accuracy - Area under curve (AUC) - Sensitivity and specificity
Chance level	- 50% for both tasks

4. Results

Model accuracy (%)

Detection

Model	MFCC	LTAS
LR	36.11	59.00
SVM	49.62	57.22
MLP	37.31	64.72
1D-CNN	49.31	**67.41**

Severity estimation

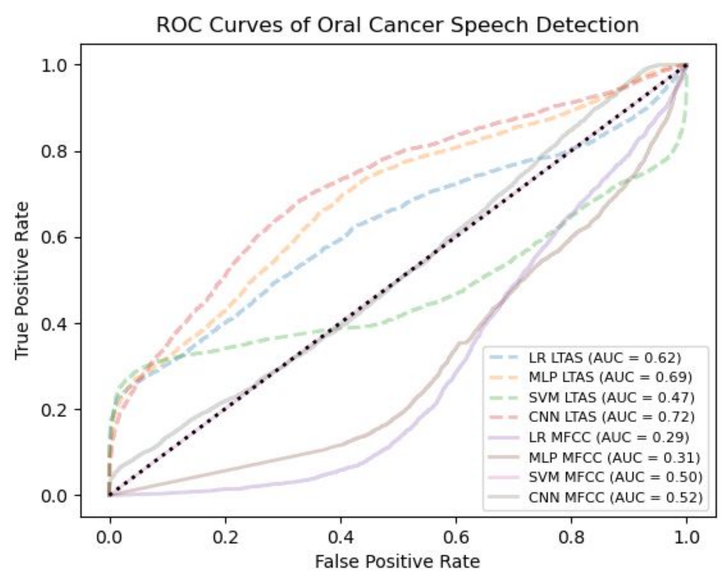
Model	MFCC	LTAS
LR	57.84	35.64
SVM	**68.73**	32.10
MLP	47.52	49.09
1D-CNN	46.37	48.43

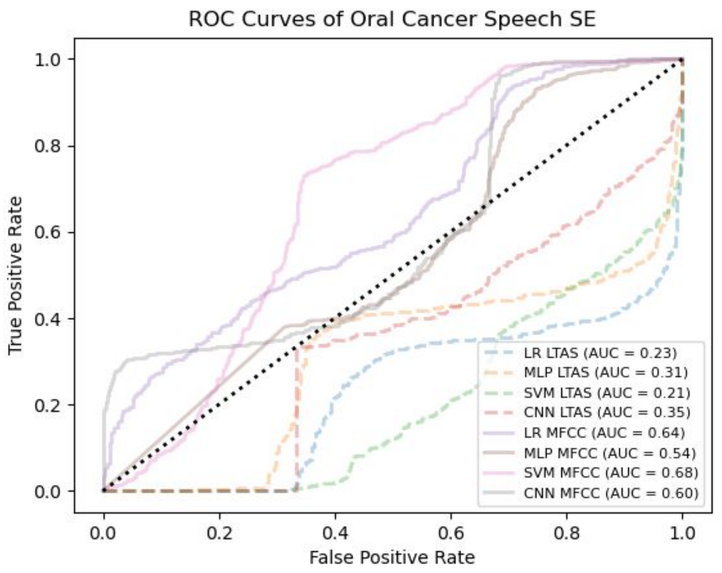
Main findings
It is possible to detect and estimate the severity of OC speech using ML methods.
- Detection: **1D-CNN** trained on **LTAS**,
  - **but** all models trained on MFCCs performed poor.
- Severity estimation: **SVM** trained on **MFCC**,
  - **but** mainly the models trained on LTASs performed poor.

Sensitivity and specificity

Best sensitivity score
Detection: **LTAS MLP**, **67.82%**
Severity estimation: **MFCC SVM**, **71.29%**

Best specificity score
Detection: **LTAS SVM**, **99.01%**
Severity estimation: **MFCC 1D-CNN**, **95.54%**





5. Discussion

Implications
Depending on the task, **1D-CNNs**, **SVMs** and even **MLPs** could be useful assistive screening tools to detect OC speech characteristics that can help determine a patient's speech severity.

Some limitations
- The tongue sensors could have increased the speech severity.
- The controls may have displayed OC speech features.
- No tuning was done to avoid overfitting.
- A dataset with a wider variety of severity scores could improve model performance.
- Uneven female-to-male ratios could have affected model performance.

SCAN ME
Scan this QR code to get the references

Scan this QR code to read the full paper

rijksuniversiteit groningen / campus fryslân

YFRSW 2022
Incheon – 인천