Neural Technologies
@
BME TMIT SmartLabs

Géza Németh

Speech Communications and Smart Interactions Labs

Department of Telecommunications and Artificial Intelligence

Budapest University of Technology and Economics

About us

BME, Budapest University of Technology and Economics, Hungary

Since 1782, 8 faculties, cca. 21.000 students

Department of Telecommunications and Artificial Intelligence

Since 1949, ~100 employees

Head of department: Dr. Pál Varga

https://www.tmit.bme.hu/

SmartLabs (3 contributing groups)

Since 1978, ~15-20 employees, 3 DSc, 9 PhD, 8 PhD (international) students and 30-40 (international) MSc/BSc students

Head of Laboratory: Dr. Géza Németh

http://smartlab.tmit.bme.hu/

SmartLab
Intelligent Interactions

MŰEGYETEM 1782

Definition(s) of Artificial Intelligence

Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. (source: www.britannica.com)

My preferred (own) definition:

Artificial implementation or extension of human cognitive functions

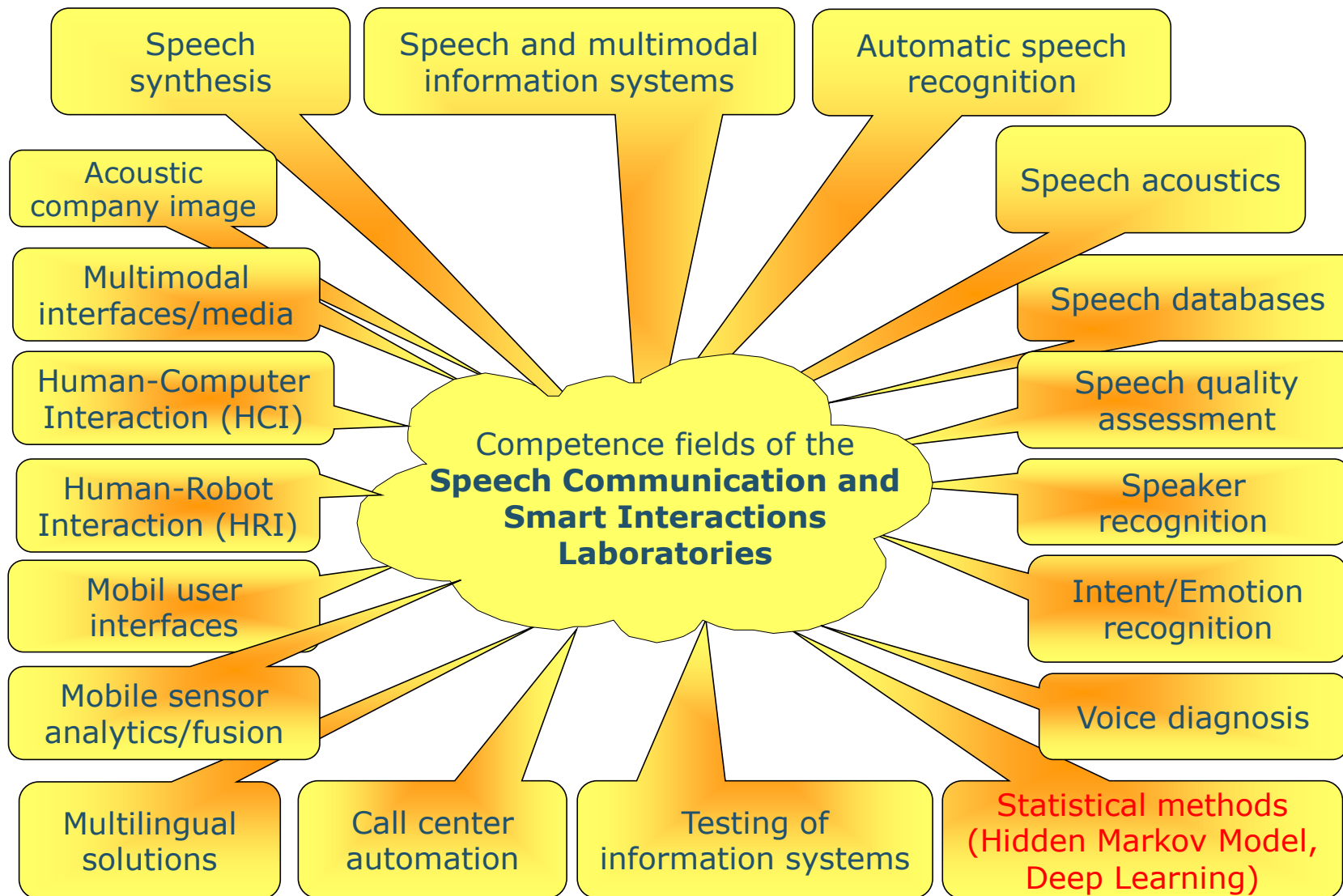*Our focus: Speech & Smart Interactions*

Potential application areas:

• Intelligent Assistants, Service Automation

• Artificial Intelligence Analytics & Informatics

• Robotics

• Health and Medical Technologies

http://smartlab.tmit.bme.hu

SmartLab
Intelligent Interactions

M Ű E G Y E T E M 1 7 8 2

Competence fields of the **Speech Communication and Smart Interactions Laboratories**

- Speech synthesis
- Speech and multimodal information systems
- Automatic speech recognition
- Acoustic company image
- Speech acoustics
- Multimodal interfaces/media
- Speech databases
- Human-Computer Interaction (HCI)
- Speech quality assessment
- Human-Robot Interaction (HRI)
- Speaker recognition
- Mobil user interfaces
- Intent/Emotion recognition
- Mobile sensor analytics/fusion
- Voice diagnosis
- Multilingual solutions
- Call center automation
- Testing of information systems
- Statistical methods (Hidden Markov Model, Deep Learning)

http://smartlab.tmit.bme.hu

SmartLab Intelligent Interactions

MŰEGYETEM 1782

RESEARCH

1. **Deep Learning fundamentals**

2. **Time series prediction and classification**
   - multiple timeframes
   - weak coherence
   - attention mechanism

3. **Deep Reinforcement Learning**
   - Physical architecture search
   - Neural architecture search
   - Automated Driving

4. **Natural Language Processing**

5. **Speech technologies**
   - Text-To-Speech Synthesis
   - Automatic Speech Recognition
   - Medical / rehabilitation applications
   - Emotion / intent detection
   - Intelligent Assistants
   - Human-Robot Interaction

6. **Sensor data analysis**

7. **Embedded devices**

   - Model compression
   - Mixed precision
   - Smartphone/UAV implementations

http://smartlab.tmit.bme.hu

SmartLab
Intelligent Interactions

Speech Synthesis
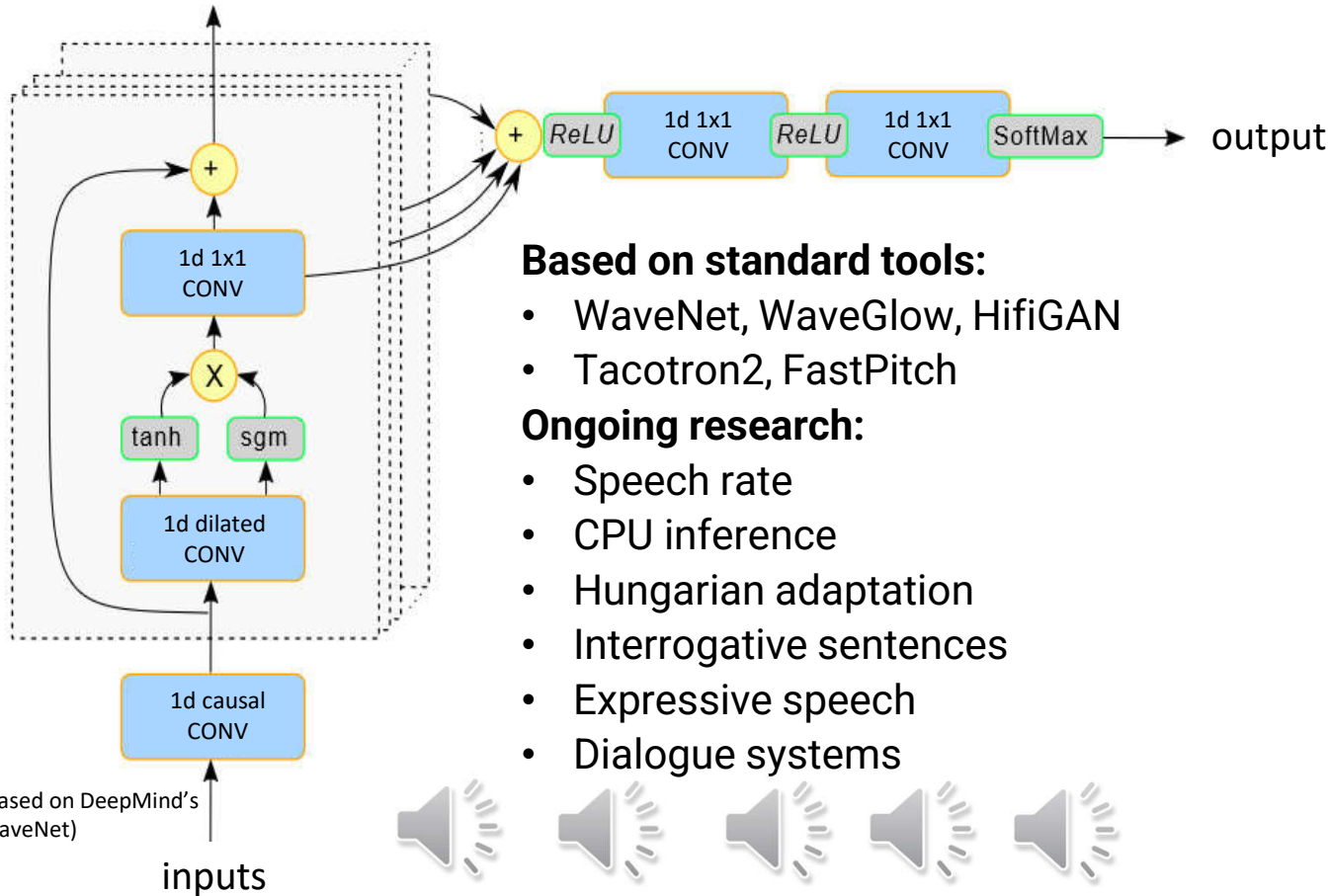
Based on standard tools:
- WaveNet, WaveGlow, HifiGAN
- Tacotron2, FastPitch

Ongoing research:
- Speech rate
- CPU inference
- Hungarian adaptation
- Interrogative sentences
- Expressive speech
- Dialogue systems

(Based on DeepMind's WaveNet)

http://smartlab.tmit.bme.hu

AI driven railway announcement system **running at more than 100 railway stations in Hungary!**

Automated call center for Hungarian rail timetable

Intelligent talking assistant for government services

**Automatic film narration for the visually impaired**

Silent Speech Interfaces (SSI)

- biosignals
  - Electromagnetic Articulography (EMA)
  - surface Electromyography (sEMG)
  - Lip video
  - Ultrasound Tongue Imaging (UTI)
  - Permanent Magnetic Articulography (PMA)
  - Magnetic Resonance Imaging (MRI)
  - EEG
  - Multimodal

Speech Production

http://smartlab.tmit.bme.hu

MŰEGYETEM 1782
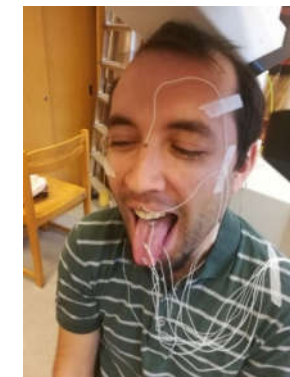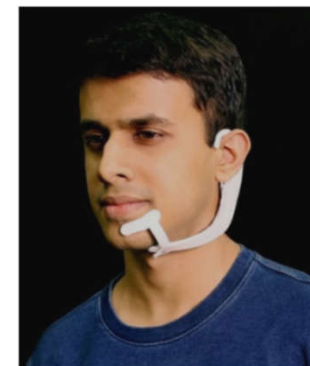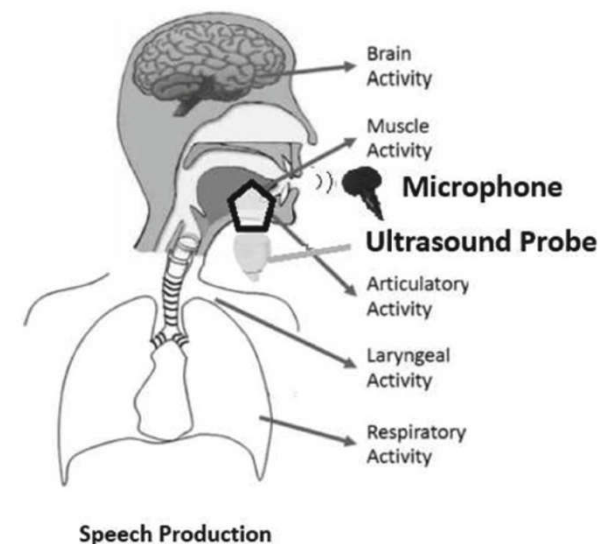
Speech Recognition
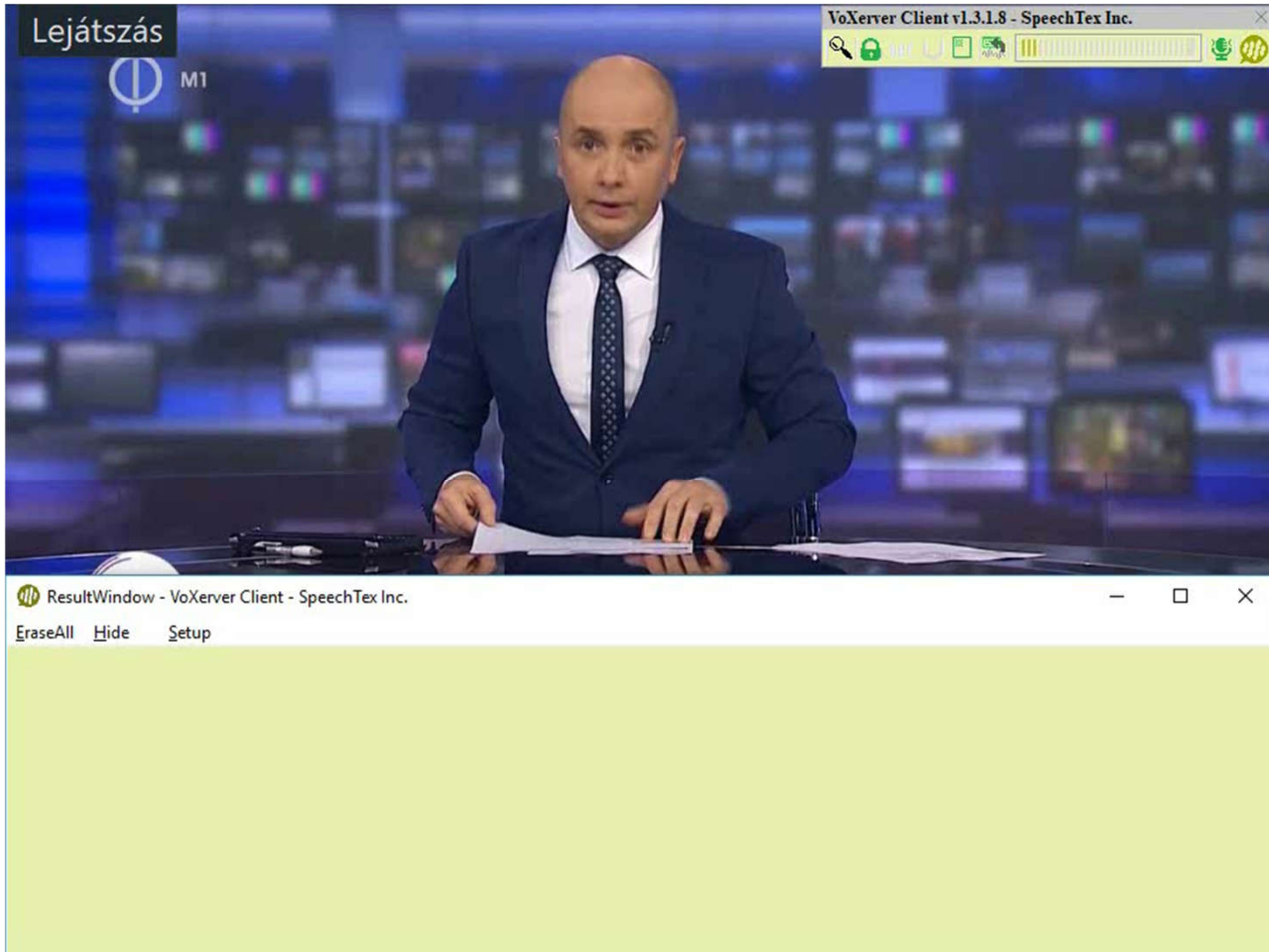
Real-time
deep learning powered
automatic speech
recognition system

Deployed to Hungarian National Television
(real-time subtitling),
European Commission
(translation support),
Banks, government services
(call center automation,
intelligent assistant), etc.

http://smartlab.tmit.bme.hu

SmartLab
Intelligent Interactions

MŰEGYETEM 1782

NONVERBAL SPEECH PROCESSING

mental disorder detection

pathological voice analysis

multimodal Parkinson's disease detection

speaker recognition

forensic voice comparison

speech therapy

http://smartlab.tmit.bme.hu
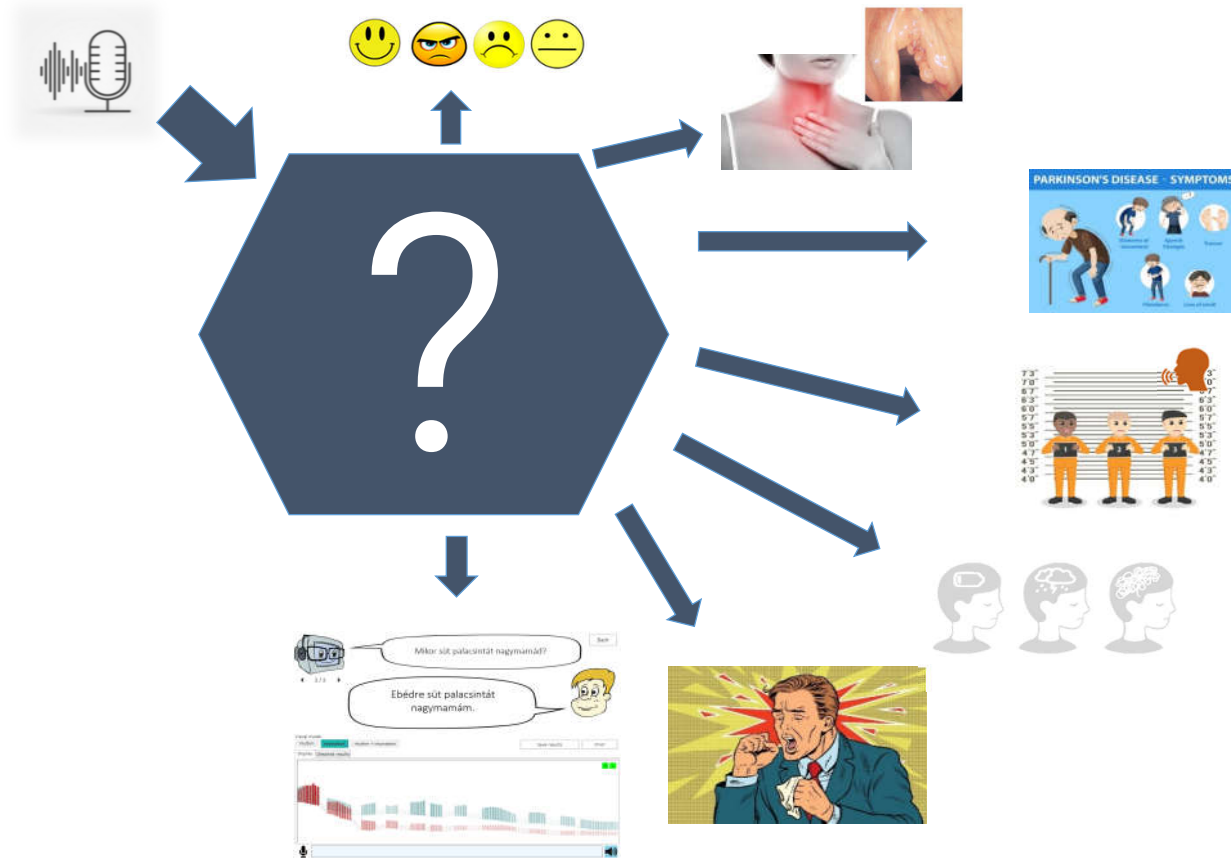
Development of voice diagnostic applications

automatic diagnostic possibilities

The physiological, psychological and neurological
state of the speaker affects the physical parameters of produced speech

- laryngological diagnostic support tool

  - by examining the effect of pathological changes in the vocal organs on speech

- depression detection

  - examining the effect of psychological changes on speech

- Parkinson's disease detection

  - examining the effect of neurological changes on speech

- measuring the acoustic effect of cleft palate surgery

  - objective acoustic measurement of the improvement effect of the surgery for the treatment of the disease

http://smartlab.tmit.bme.hu

**SmartLab**
Intelligent Interactions

MŰEGYETEM 1782
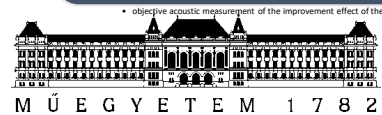
Language independency of speaker verification
## deep learning models for forensic voice comparison

Language differences matter in voice comparison? A model pre-trained on a language different from the target samples can be used for forensic voice comparison? YES! The lowest EER was obtained with the model pre-trained on the VoxCeleb dataset and evaluating it on the Hungarian ForVoice120+ corpus. Deep learning speaker embedding models can be used cross-lingual.

Sample length analysis shows that the longer the duration of the sample in question the better the performance is, as one would naturally expect.
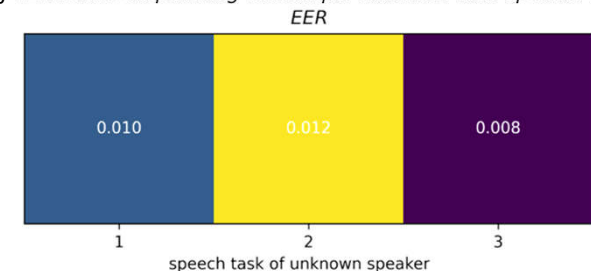
There is no real difference if various speaking styles are applied, and it doesn't really matter.
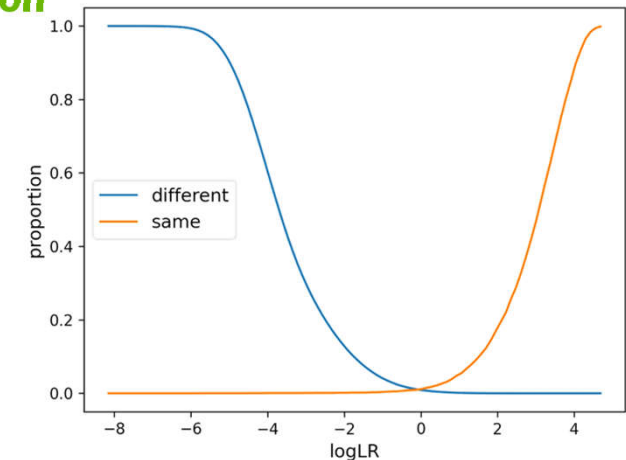
Tippet plot of same and different speaker verifiaction in a cross-lingual experiment (ECAPA-TDNN pretrained on VoxCeleb and LR calibration on ForVoice120+)
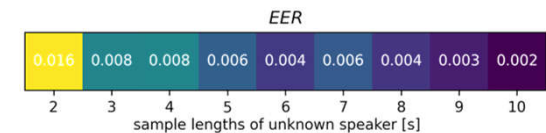
Heatmap of EER values depending on sample duration with speaker enrollment

Heatmap of $Cllr_{min}$ and EER values depending on speech task with speaker enrollment
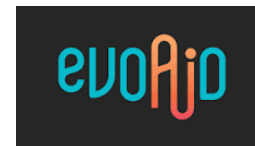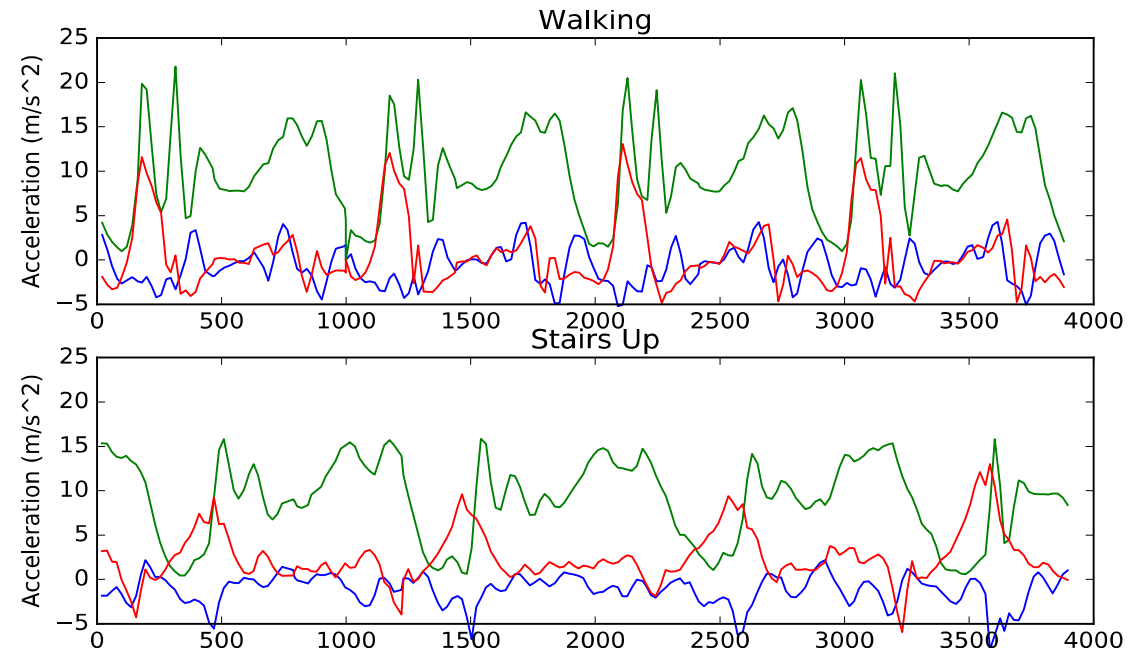
Sensor Data Analysis

Sensor data

- Accelerometer
- Orientation
- GPS, WiFi, etc.

Deep Learning based on device realtime classification

- Behaviour
- Activity
- Motions, gestures

Applications

- Emergency alerts
- Behaviour analysis
- User and user group identification

http://smartlab.tmit.bme.hu

Economic indicators from economic news

using deep learning-based natural language processing methods●

Traditional economic sentiment indicator
- Inconvenient data collection (questionnaire)
- Low frequency time series
- Results for the past

»

Algorithm-based economic sentiment indicator
- Automated data collection
- High frequency time series
- Results for the present
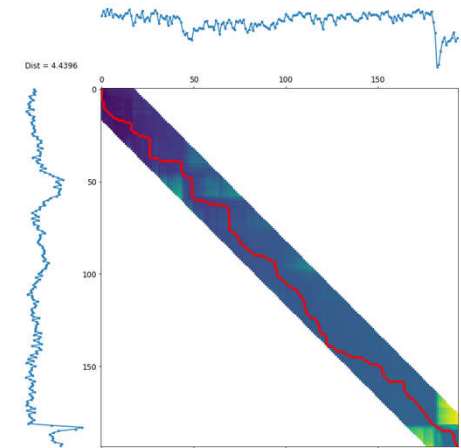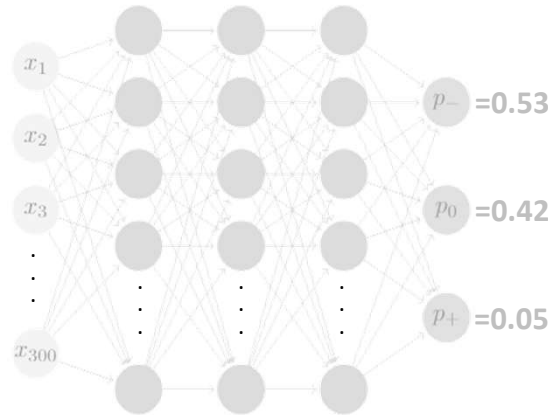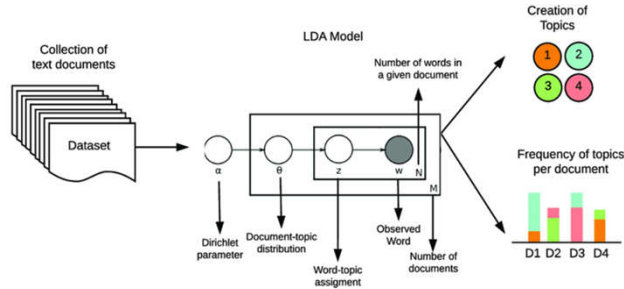
Topic modeling – LDA
Cca. 800.000 economic news

Sentiment prediction - DNN
Own dataset with consensus labeling

Time series matching – DTW
Economic vs predicted indices

Forrás: Buenaño-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*, *8*, 35318-35330.

Partner and industrial exploitation: Central Bank of Hungary
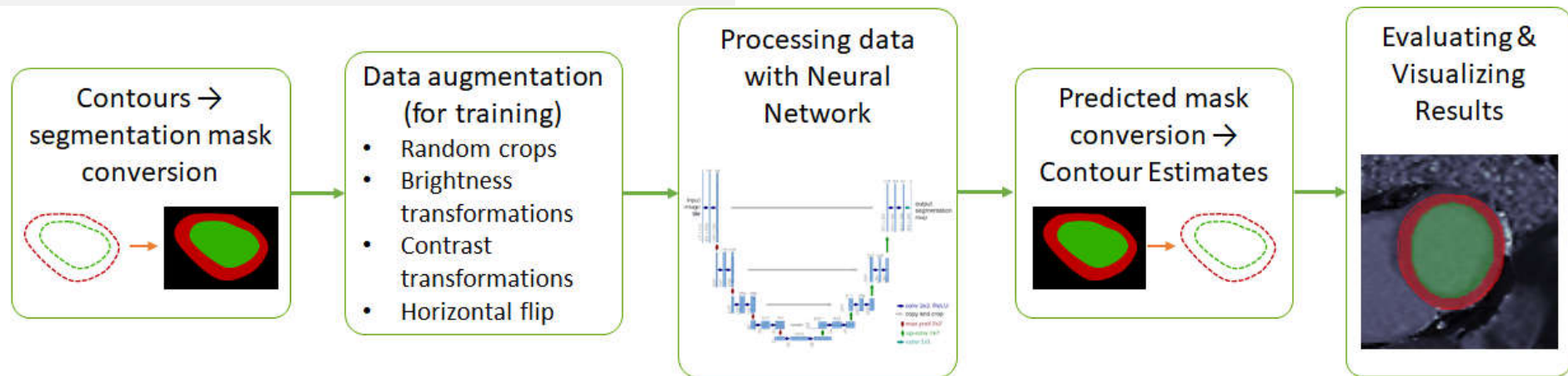
MŰEGYETEM 1782

Understanding the factors and effects of premature birth.
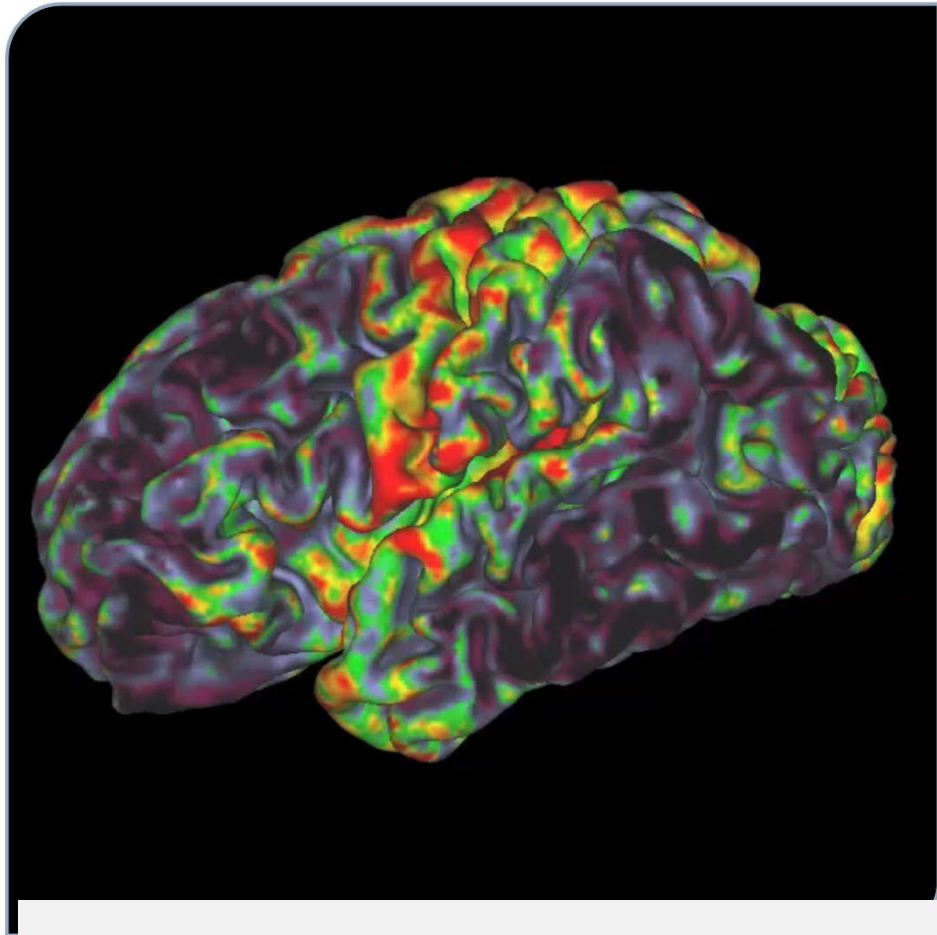
- ➢ dHCP, Developing Human Connectome Project (EU)
- ➢ database of brain MRI scans: 783 newborns
- ➢ 3D representations
- ➢ Goals
  1. Deep learning on a curved surface (cerebral cortex).
  2. Identification of areas indicative of normal development.
  3. Identification of areas of abnormal development in premature birth.

Cardiac T1&T2 mapping images-based semantic segmentation with deep learning

- Labeled scans of more than 250 patients.
- Consensus labeling from 3 medical experts
- Unlabeled scans of more than 1100 patients
- Cca. 300.000 DICOM files altogether
- General architecture for segmentation tasks (UNet with ResNet encoder)
- Hausdorff distance-based evaluation

Contours → segmentation mask conversion

Data augmentation (for training)
- Random crops
- Brightness transformations
- Contrast transformations
- Horizontal flip

Processing data with Neural Network

Predicted mask conversion → Contour Estimates

Evaluating & Visualizing Results

Partners and industrial exploitation: Semmelweis University, Siemens Healthineers GmbH

SIEMENS
Healthineers

Recent AI Research Projects

- AI4EU (www.ai4europe.eu), Hungarian NCP: Géza Németh (BME TMIT)
- ENFIELD: European Lighthouse to Manifest Trustworthy and Green AI (Horizon Europe, https://www.enfield-project.eu/ 2023-)
- CherISH: Cochlear implants and spatial hearing: Enabling access to the next dimension of hearing (Marie Curie Doctoral Networks, 2023-)
- MILab (Hungarian AI NoE member in NLP and AI Fundamentals)
- InfoLab (Hungarian NoE in Infocommunications)
- APH-ALARM H2021-AAL https://www.aph-alarm-project.com/
- FIKP (Hungarian Higher Education)
- EFOP (Hungarian Higher Education)
- NKFIH OTKA - FK, PD -, Bolyai, ÚNKP-4 (Hungarian Academic Excellence)
- NVidia GPU Research Grants

http://smartlab.tmit.bme.hu

SmartLab
Intelligent Interactions

M Ű E G Y E T E M 1782

COLLABORATIONS

Industrial Partners and Spinoffs
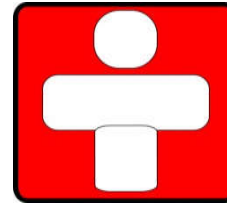
IT.DOT Kft

Latest spinoff: EvoAid
(http://www.evoaid.com/en/home/)

Catalyst of the smart city on a global scale.

evoAid

Guarding Beats
- AI based touchless safety app
- gesture control and fall recognition
- pictogram-based text message compilation for people with disabilities
- only a smartphone is required
- white label business model
  - References: B2B, B2G

Tracking Vibe
- automatic and sustainable road quality detection and CO_2 logging system
- Real-Time edge computing
- server-side app
- white label business model
  - Reference: B2G

UTILITY HOLE
POTHOLE
UTILITY HOLE
POTHOLE

Accounting Swish
- automatic and sustainable invoice processing
- payroll optimizations
- server-side app
- white label business model
  - Reference: B2B

EvoAid is considered among the **most innovative Artificial Intelligence companies** in Hungary.

BEST STARTUP.eu | DATA | FUTUROLOGY | NVIDIA INCEPTION PROGRAM | newchip | AAL PROGRAMME

MŰEGYETEM 1782

http://smartlab.tmit.bme.hu

**SmartLab**
Intelligent Interactions

Professional Intelligence for Automotive
(BME TMIT and IIT + Continental)

➢ Duckietown, **AI Driving Olympics**

➢ Organized at Neural Information Processing Systems (**NeurIPS**) Conference since 2018

➢ Dominated the compeion for the past **2 years**

➢ 2020 AIDO5 **– 1st place** in 2 competitions out of 3

➢ 2021 AIDO6 **– 1st place** in 2 competitions out of 3

➢ **+100 competitors** from top universities and research centers around the world (MIT, Berkley, ETZH Zurich, Jetbrains Russia, University du Montreal, etc)

http://smartlab.tmit.bme.hu

SmartLab
Intelligent Interactions

MŰEGYETEM 1782
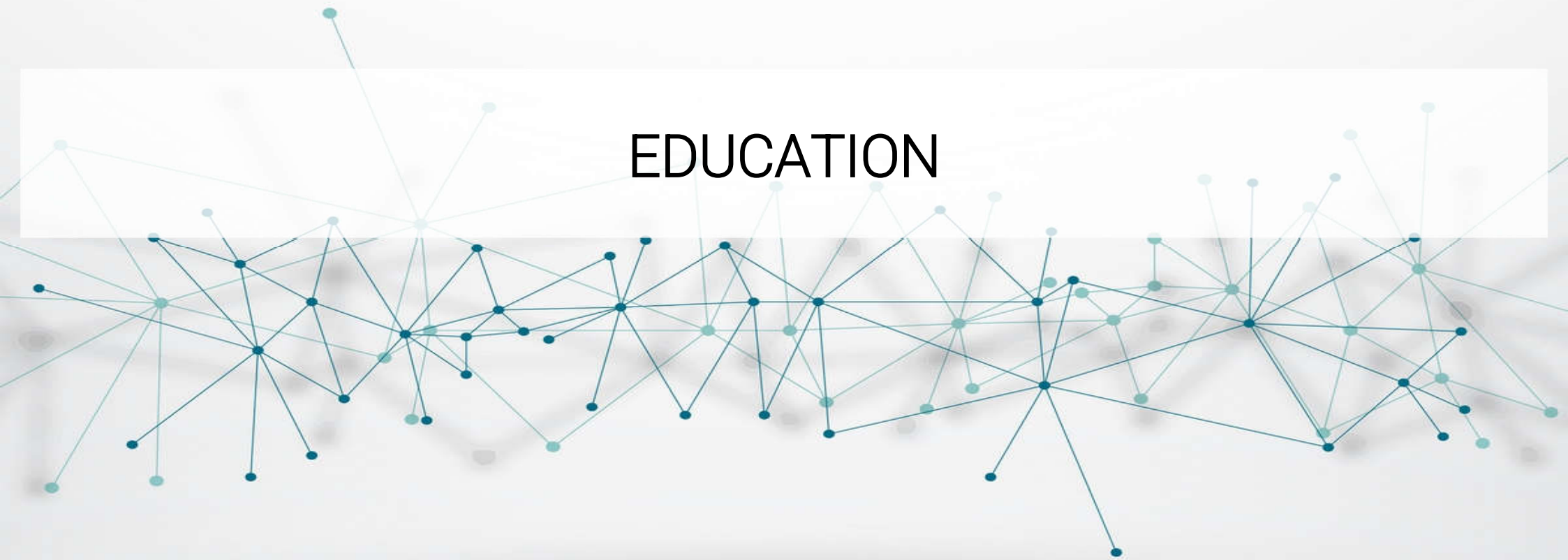
EDUCATION

DEEP LEARNING EDUCATION

Most important courses

- Speech Information Systems (BSc, forced elective, app. 120, 20+ years)
- Deep Learning in Practice (BSc, elective, project-based, app. 100, 2016-)
- Data Science and Artificial Intelligence (MSc major, app. 80, 2023-)
- UX and Interactions (MSc minor, project-based teamwork, app. 40, 2023-)

Training individual students

- Preparation for project laboratory
- Project Laboratory
- BSc and MSc thesis work
- PhD students

Working on AI related topics

- 10 PhD students
- 30-40 BSc and MSc students

http://smartlab.tmit.bme.hu

THANK YOU

nemeth@tmit.bme.hu