Sequence Learning

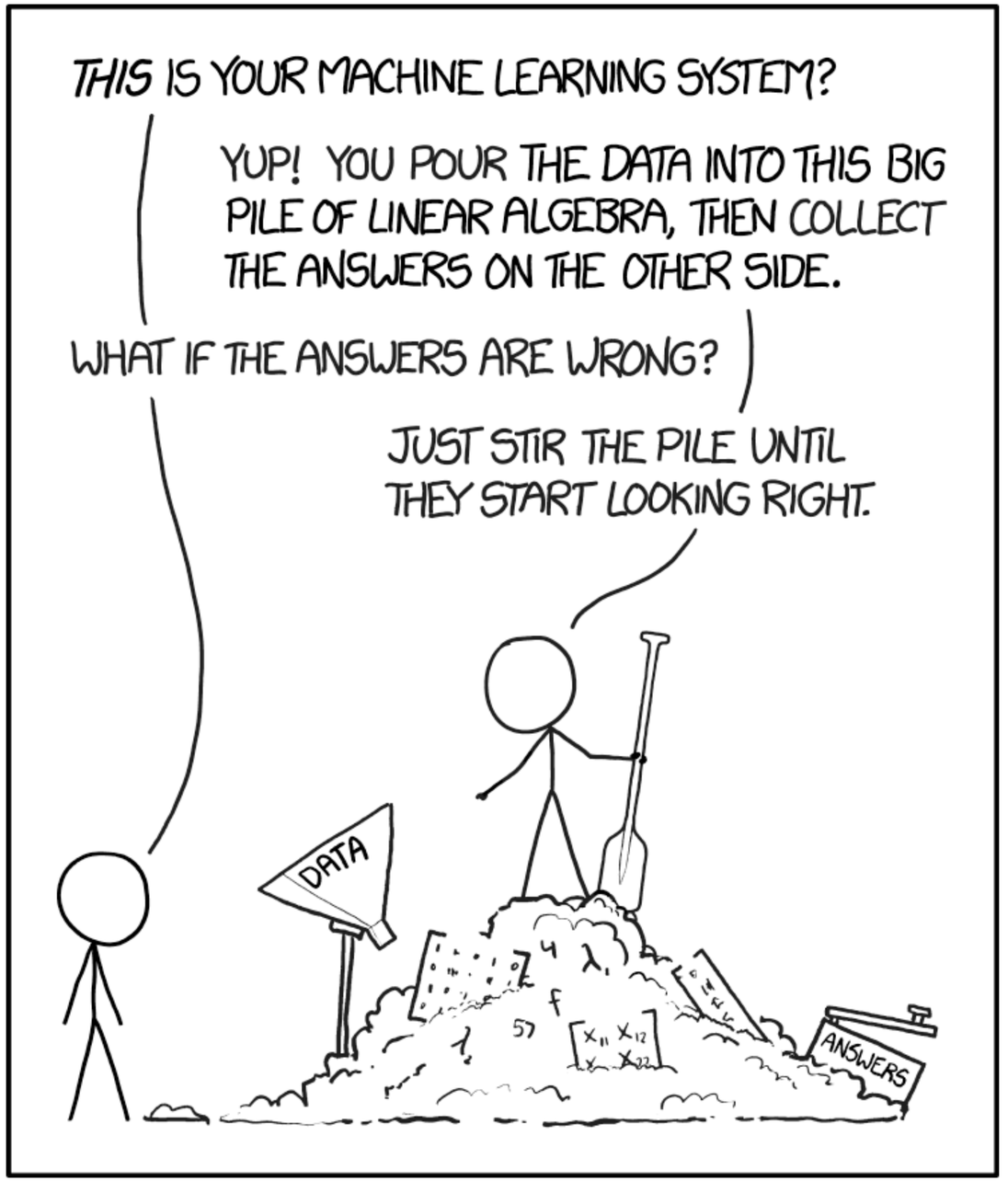
Introduction

Korbinian Riedhammer

TECHNISCHE HOCHSCHULE NÜRNBERG
GEORG SIMON OHM

Today

- Logistics
- Why I teach this class
- Why you should take this class
- Motivation
- Syllabus
- What you should bring to this class



https://xkcd.com/1838/

Logistics

- Tuesdays at 11.30a HQ.406, discussion on Teams (Code: `5kxlgot`)
- Materials: https://seqlrn.github.io (continuously updated…)
- Exam:
  - mandatory assignments in python (pair-programming ok)
  - 20’ oral exam in the last week of lecture period (Week 28)

Why I teach this class

- Industry background in speech recognition/indexing (mod9.io)
- Research focus
  - Speech processing for medical applications (eg. stuttering, dementia)
  - Speech recognition for indexing/search
  - Sequence learning for industrial applications (mostly anomaly detection)

Why you should take this class

- Machine learning is the future*
- Many applications are to sequences, not single observations
- Understand the foundations of sequence classification

*or at least a very well paid part of it

Flashback: Verbmobil

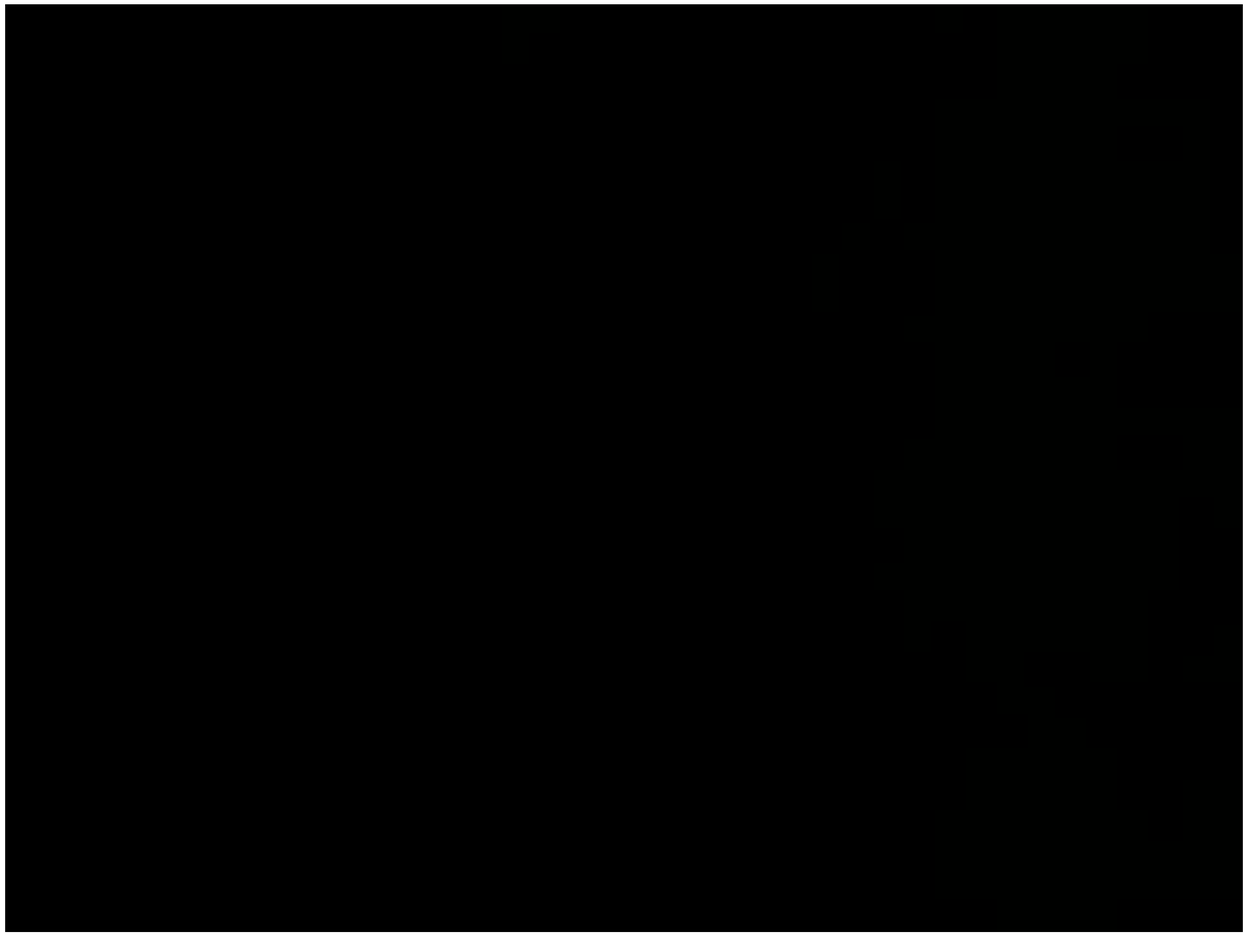
Research project 1993-2000 (!)

https://www.youtube.com/watch?v=DcG9-KWx0Fg

Motivation: AI in Fiction

Star Trek & Star Trek TNG

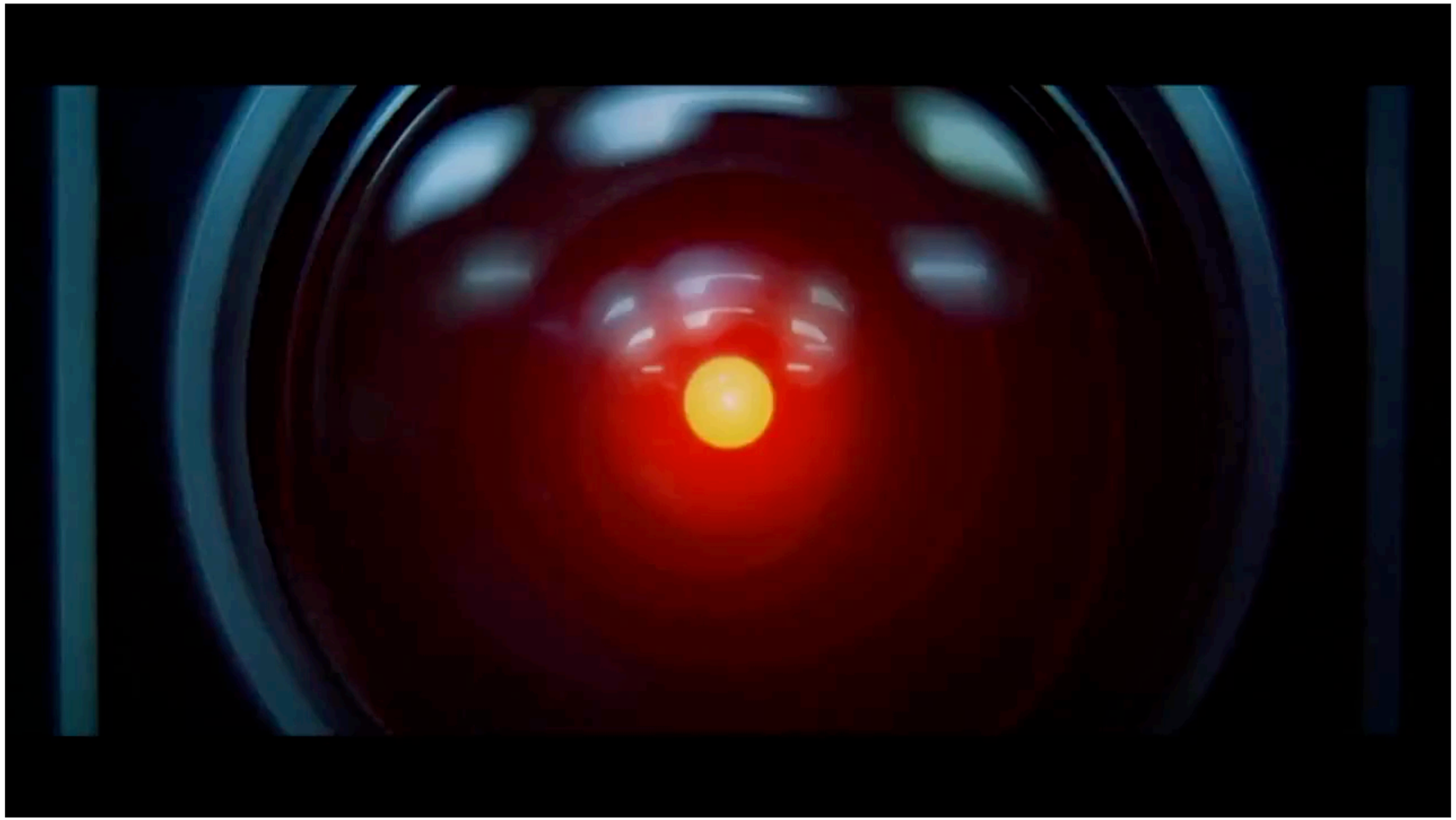
1966, 1987



https://www.youtube.com/watch?v=tpKx7Oi0oeM

2001: Space Odyssee

1968



https://www.youtube.com/watch?v=ARJ8cAGm6JE

Star Wars: A New Hope

1977

https://www.youtube.com/watch?v=EIZfE1AVDPQ

Knight Rider

1982

https://www.youtube.com/watch?v=OY_nGtN4ARA

Charlie’s Angels

2000

https://www.youtube.com/watch?v=rRHRa80wVq8

24 (S02E23)

2003

"Cypress Recording" wurde durch Sound Engine mit angelernter Sprachsynthese erstellt.

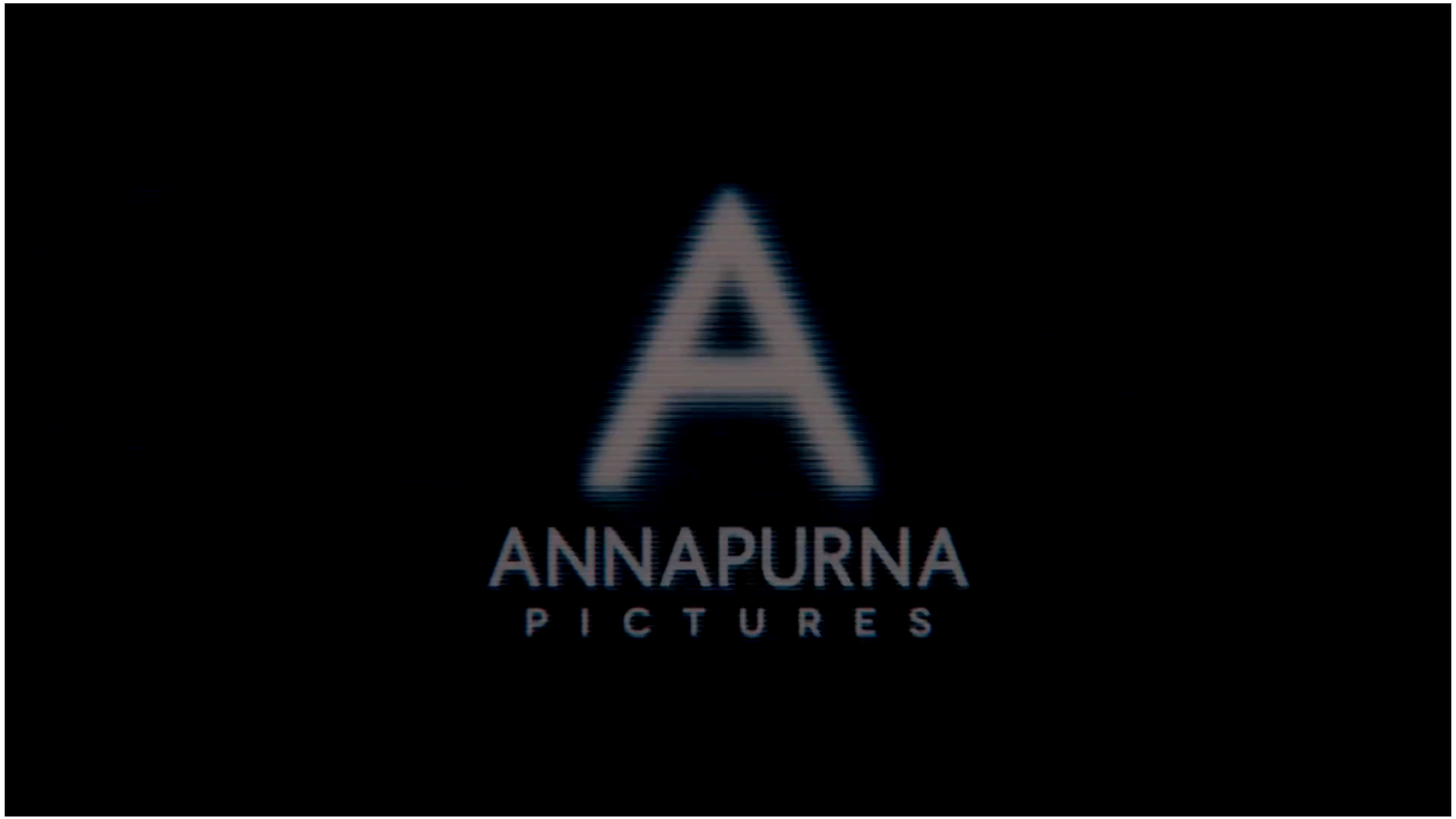
Wild Hogs

2007

https://www.youtube.com/watch?v=qVZtE3rL_sQ

Her

2013



https://www.youtube.com/watch?v=ne6p6MfLBxc

Blade Runner 2049

2018

https://www.youtube.com/watch?v=RL0gX1_NWTk

Motivation: AI in Products

Radio Rex

1920

“Classic” signal processing: triggers on 500Hz (“reks”)

Worlds of Wonder’s Julie Doll

1987

https://www.youtube.com/watch?v=UkU9Sblictc

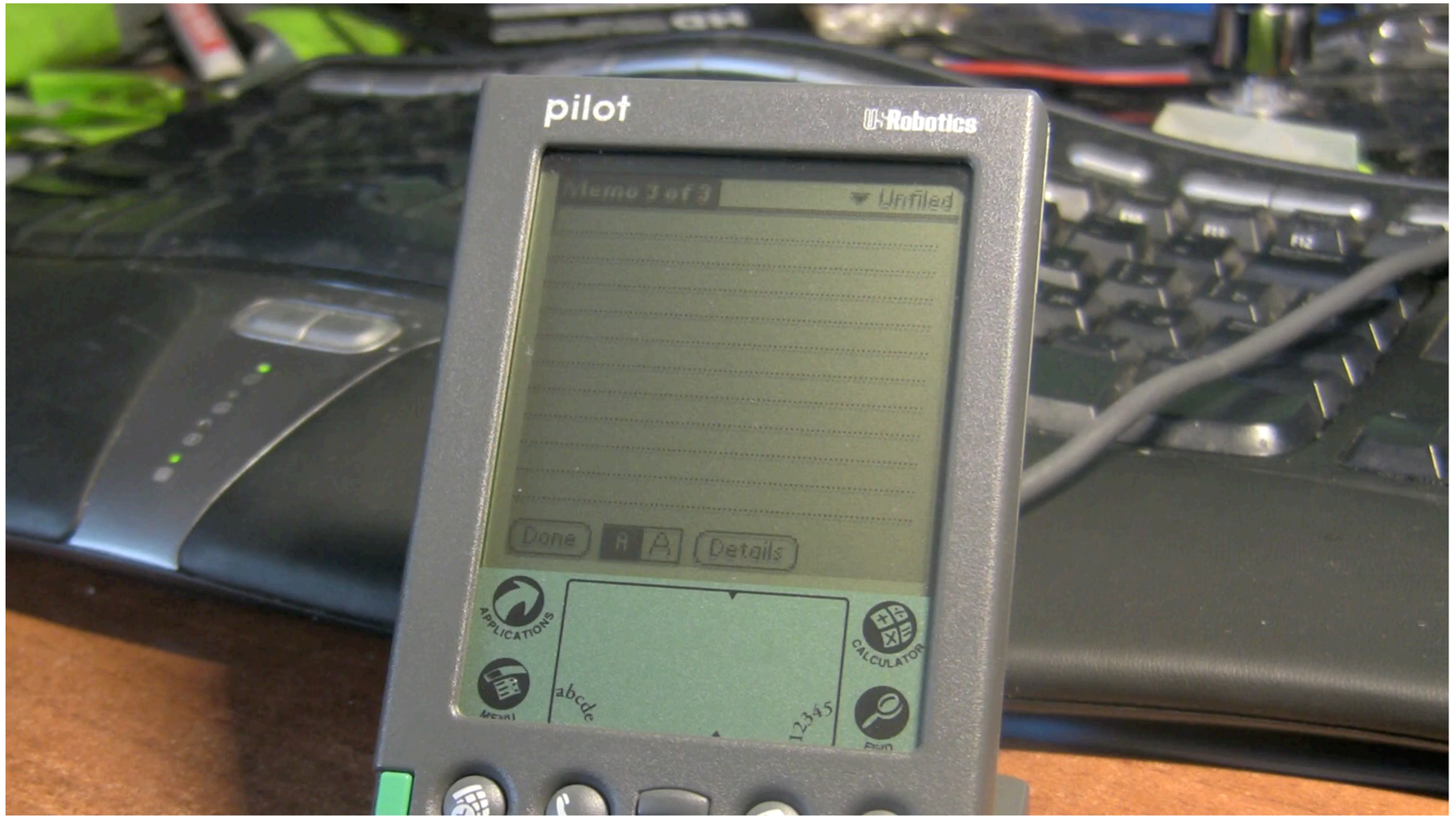
PenPoint OS

1991

https://www.youtube.com/watch?v=x0XE08BjQDQ

Graffiti (Palm OS)

1997



https://www.youtube.com/watch?v=iL0YLuClysY

Microsoft Speech Recognition

2008

https://www.youtube.com/watch?v=-0kDcUEDfmY

BMW Voice Control

2009

https://www.youtube.com/watch?v=xJo9pK42VRs

Apple Siri

2011

https://www.youtube.com/watch?v=agzItTz35QQ

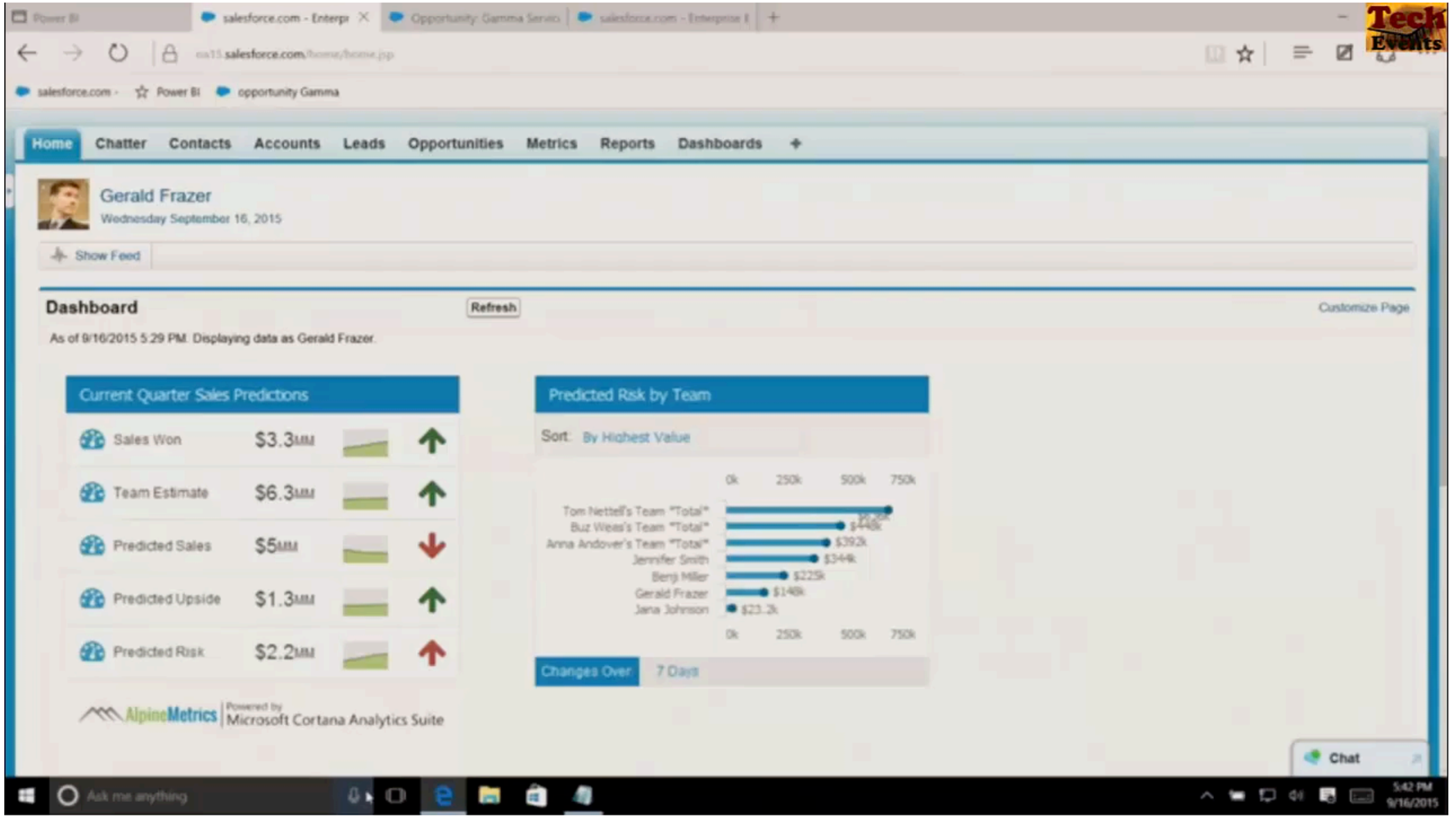
Amazon Alexa

2014

https://www.youtube.com/watch?v=YvT_gqs5ETk

Microsoft Cortana

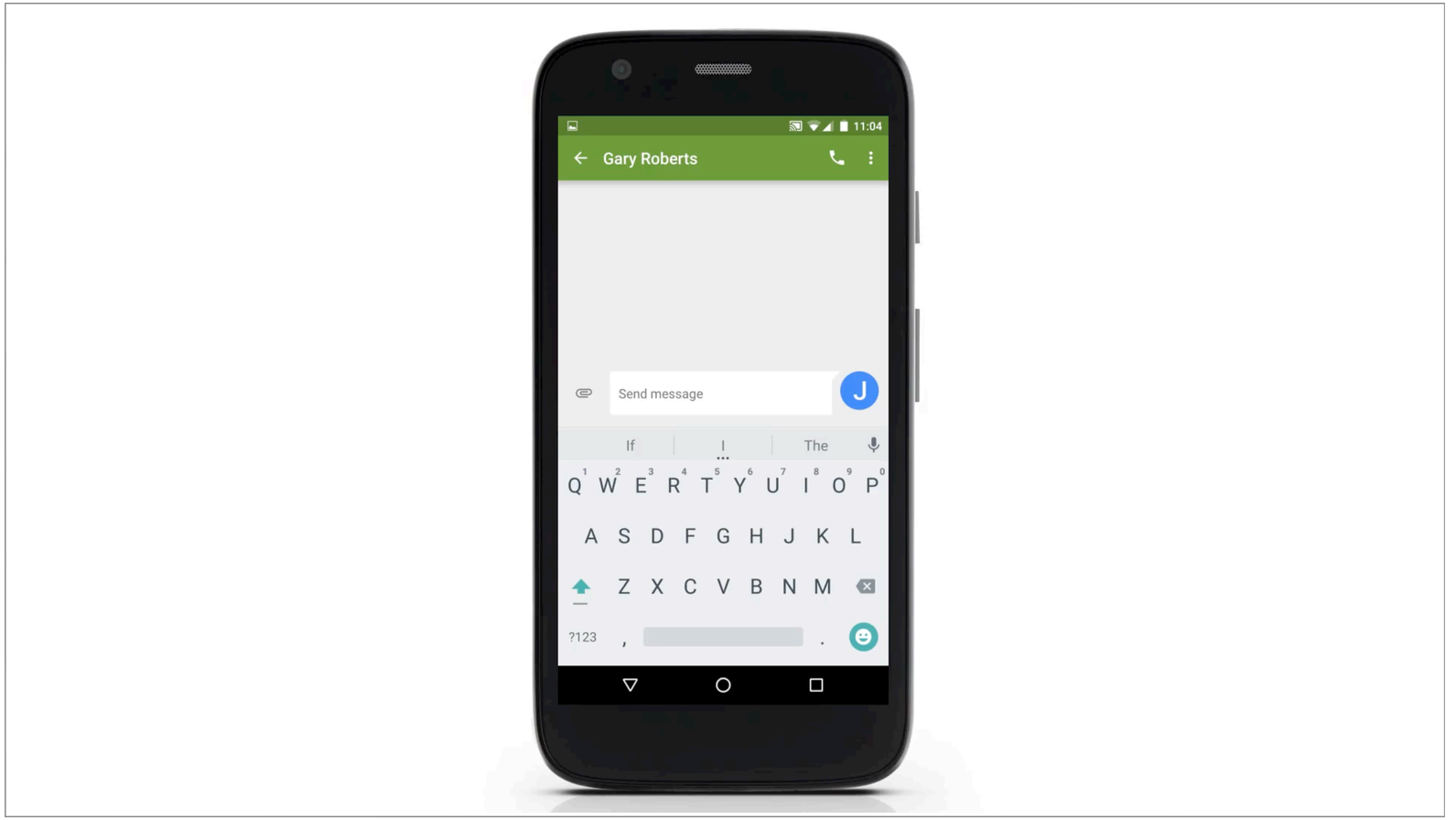
2015



https://www.youtube.com/watch?v=DDqrfCmIPxI

Google GBoard

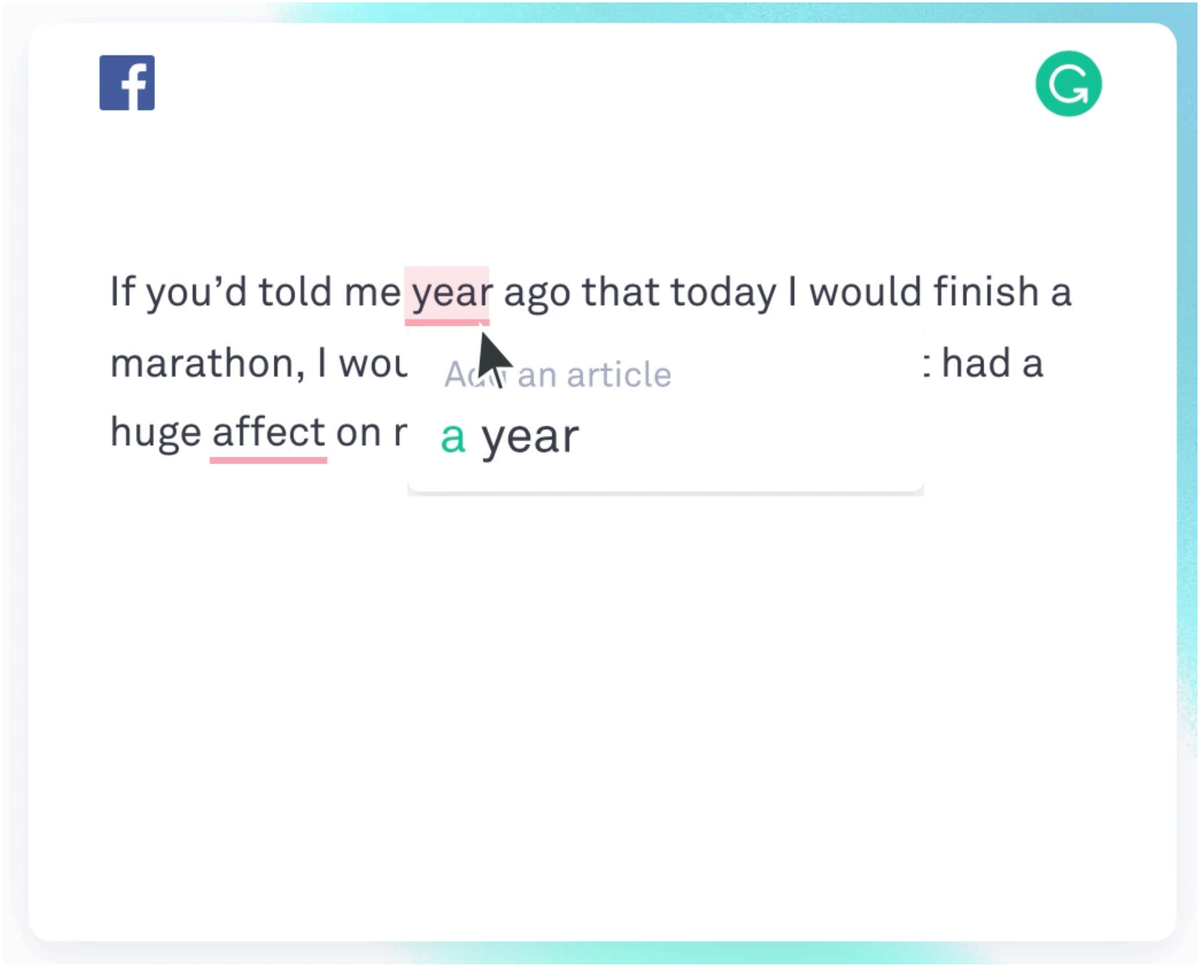
Predictive Text, Auto-Correct, Glide Typing



https://www.youtube.com/watch?v=5DSfFDdybzg

Grammarly

Spelling and grammar Correction



Screen capture of https://www.grammarly.com on 2021-03-21

- Machine translation
- Automatic summarization
- Stock market prediction
- Anomaly detection
- Controller automation (eg. MarI/O)
- Music composition
- Human-machine co-creation

Drop Jump Classification

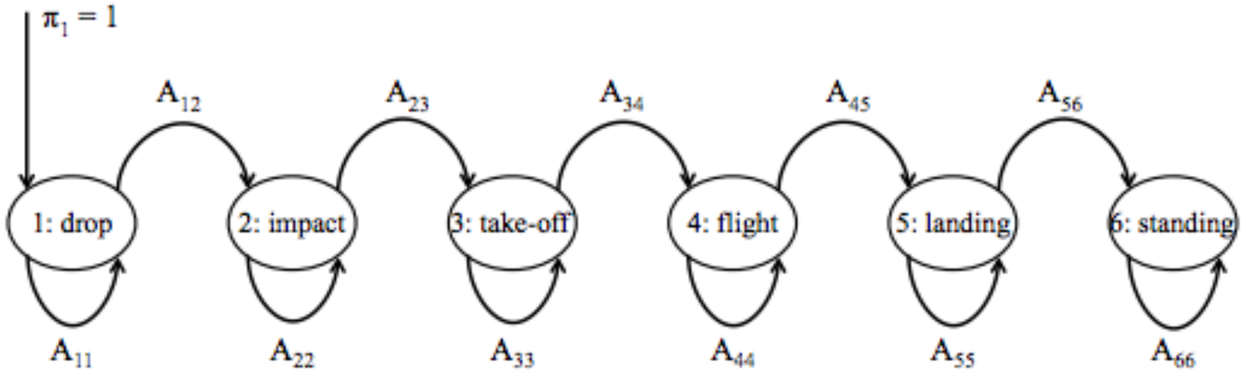


Figure 1. Left-right state transitions of the HMM that was used to analyze a drop jump sequence.

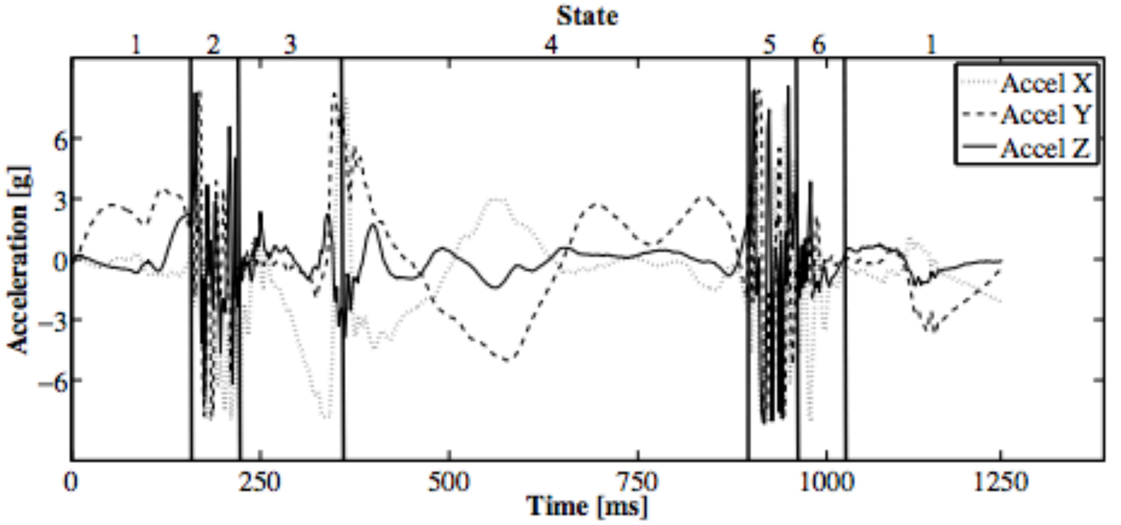
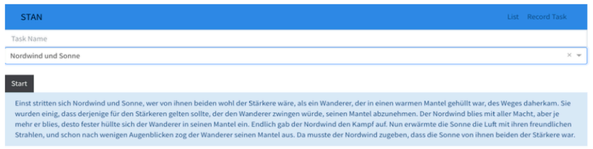
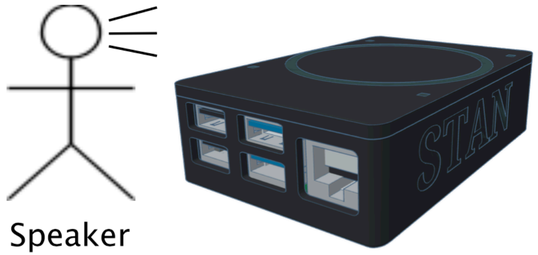
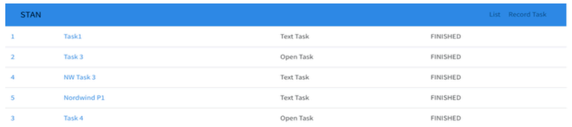
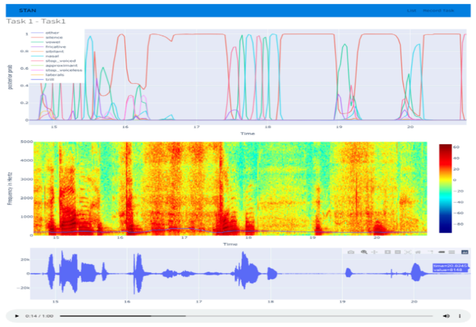
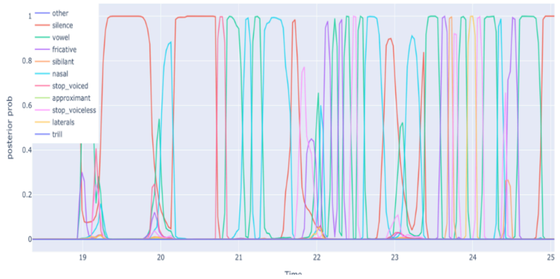
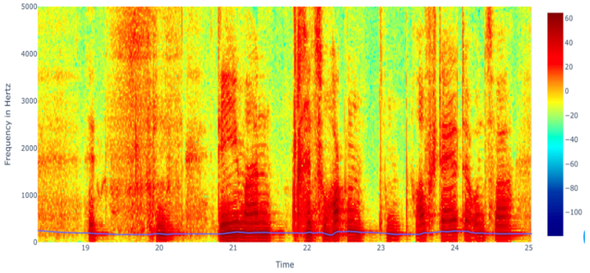


Figure 2. Left foot accelerometer signal of one drop jump and corresponding state segmentation.

Stuttering Detection

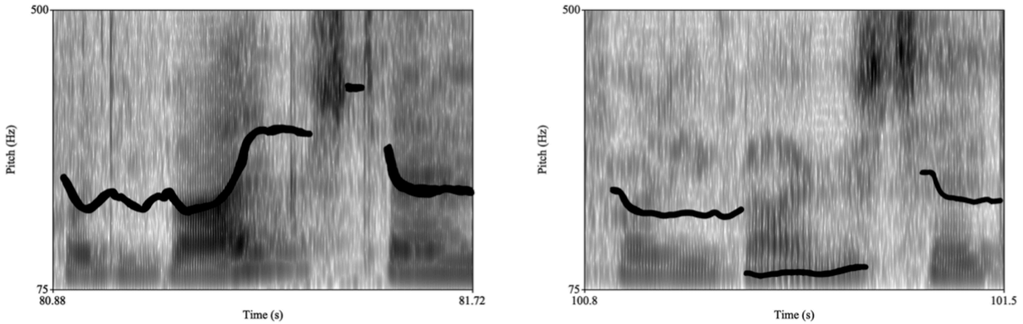
Ongoing research project (w/ Kasseler Stottertherapie)

<table>
<tr><td rowspan="2">Data Collection</td><td>Starting and Stopping of tasks</td><td>Record audio</td><td>Task List</td></tr>
<tr><td> 


</td><td> 
</td></tr>
<tr><td>Analysis</td><td>Full details with audio player</td><td>Posteriogram detail</td><td>Spectrogram details</td></tr>
<tr><td></td><td> 
</td><td> 
</td><td> 
</td></tr>
</table>

Emotional Carriers

Ongoing research project (w/ U Trento)



(a) emotion carrier (b) non emotion carrier

Figure 2: *Spectrograms with f0-contour (a) showing the bold part of the phrase: "Dann haben sie mich sozusagen* ***vor die Wahl gestellt****, ja, entweder du kannst studieren gehen oder du hast ein Pferd, weil beides kann man nicht finanzieren."*[1]*, which was marked as an EC. (b) was taken from the same recording session, but was not marked as EC.*

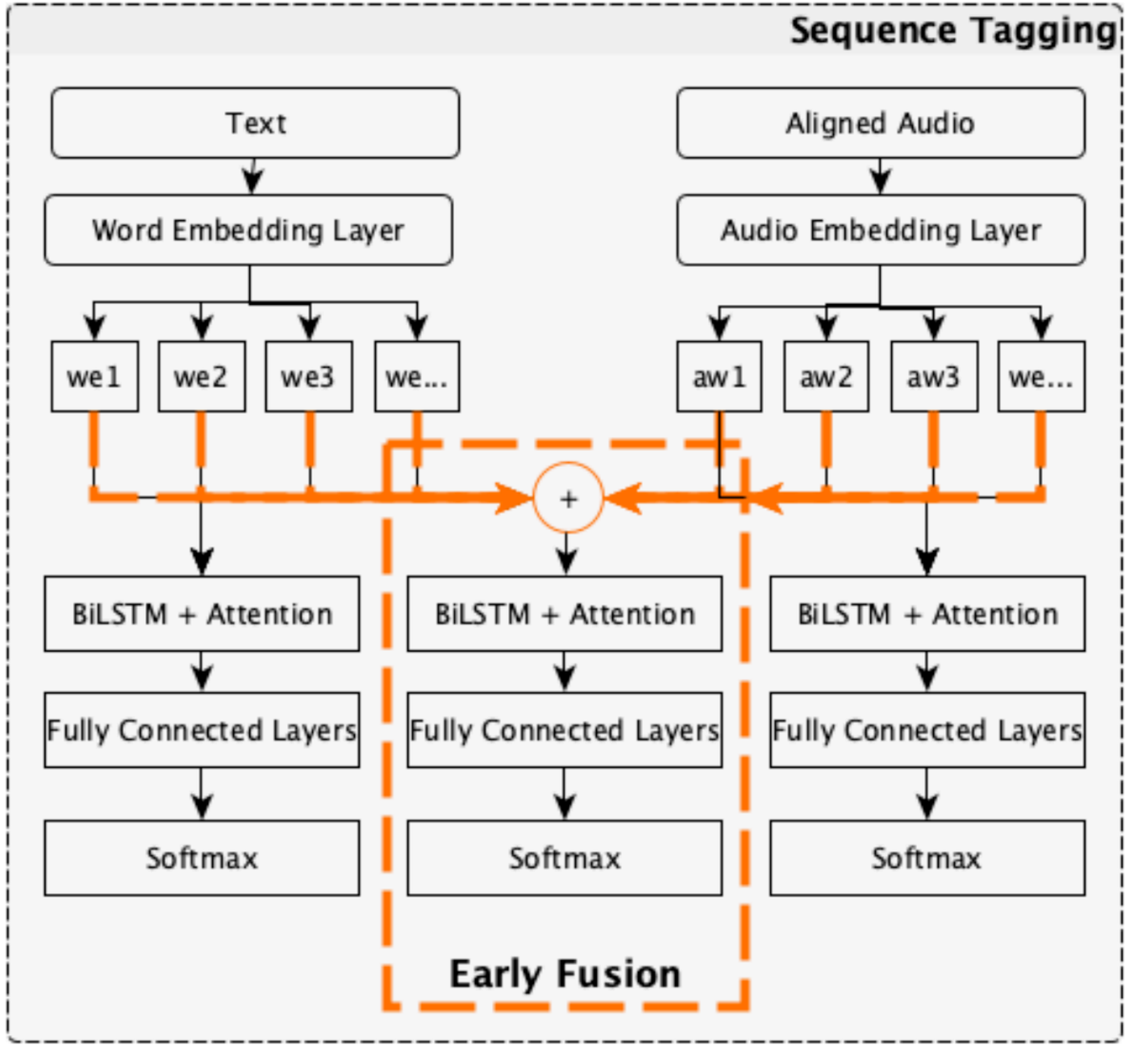
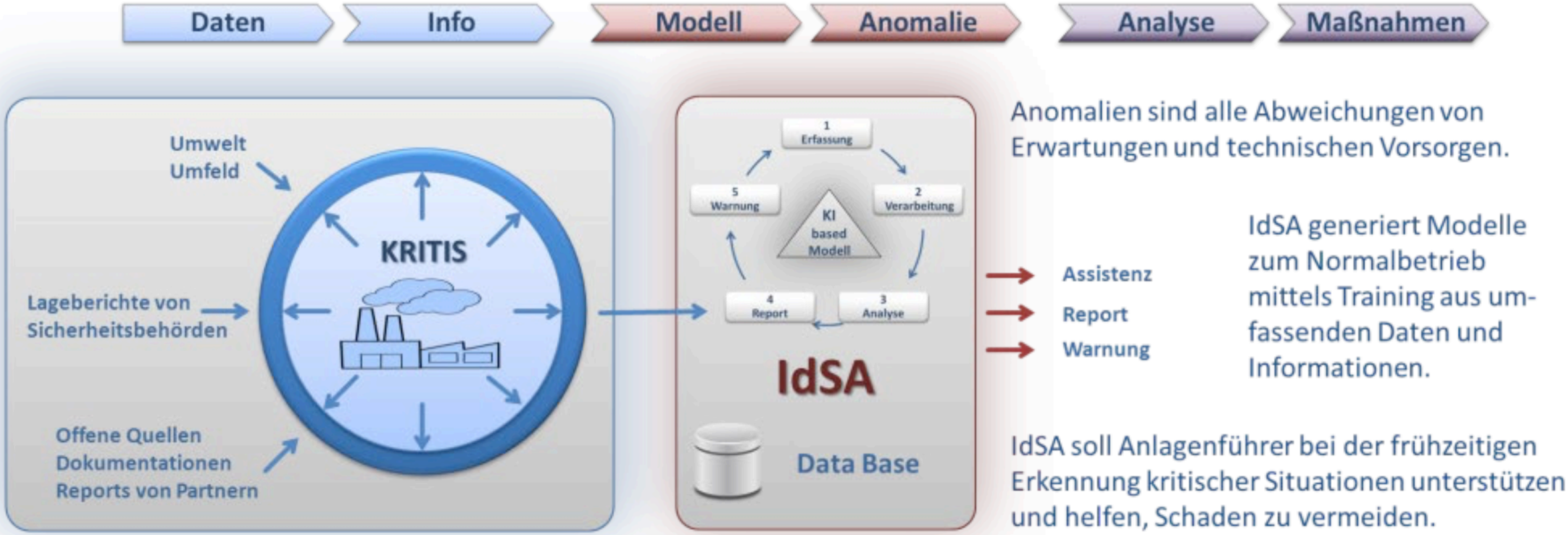


Figure 3: *Sequence tagging (ST) architecture that can operate using either Word Embeddings (WE), Audio Word Embeddings (AWE) or a combination of WE and AWE in an early fusion (EF) approach. The embedding layer of the ResNet classifier is used to extract AWE for each word.*

Anomaly Detection

Ongoing research project (w/ ESTW)



Instrumentation and Control Engineering

Ongoing research project (w/ AC)

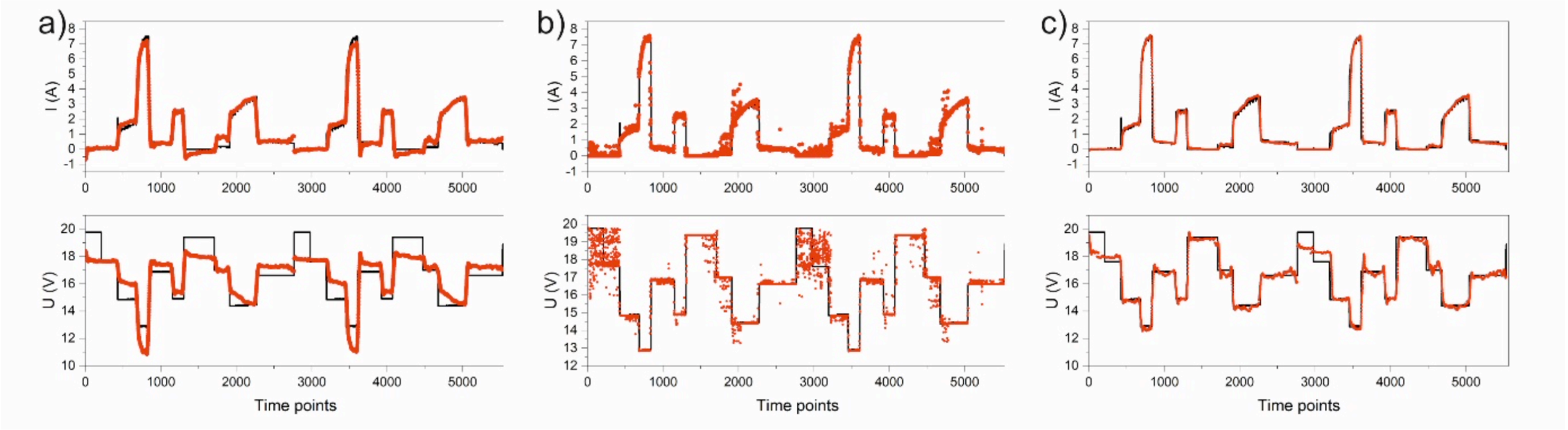
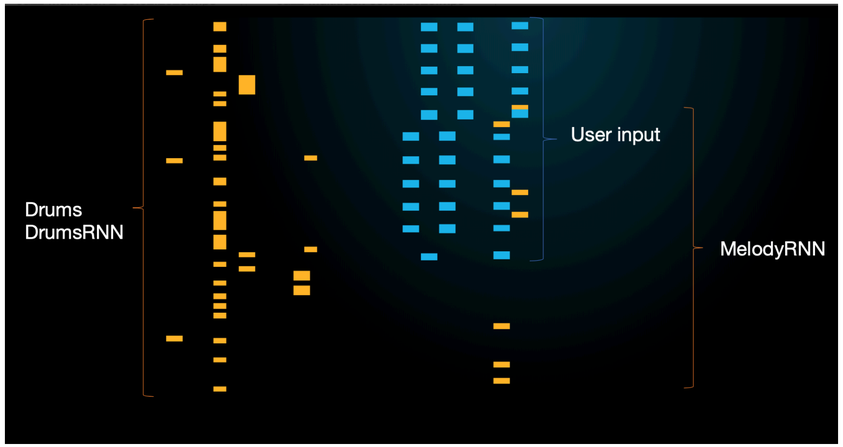
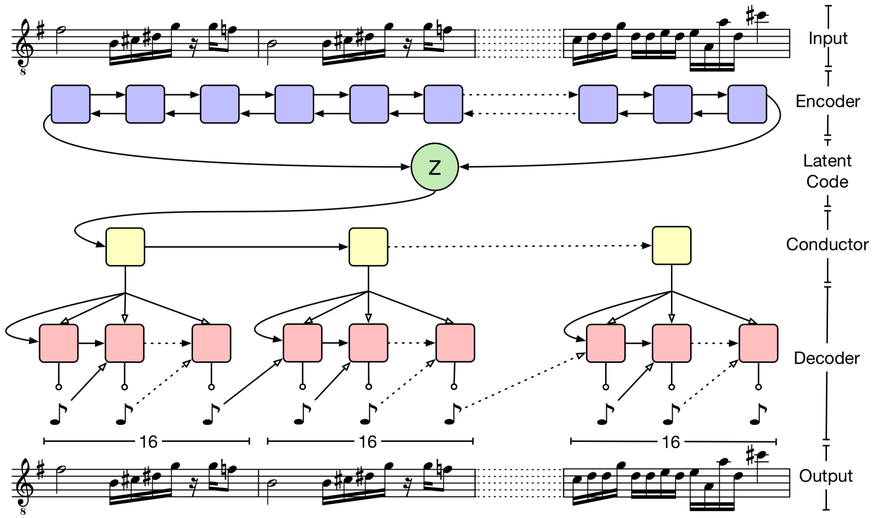


Figure 2. Calculated (red points) and measured (black lines) output parameters stack current I and voltage U of a 100 W fuel cell stack. The calculated values were obtained by training and testing MLR (a), RFR (b) and DNN (c) models with static performance profiles as shown in Figure 1a) using stack temperature and H_2 flow as input parameters. In a statistical error analysis root mean square errors of 0.837 (MLR), 0.616 (RFR) and 0.537 (DNN) were obtained for the calculated output parameters.

Spirio Sessions: Human-Machine Co-Creation

Ongoing research project (w/ HfM Nürnberg)

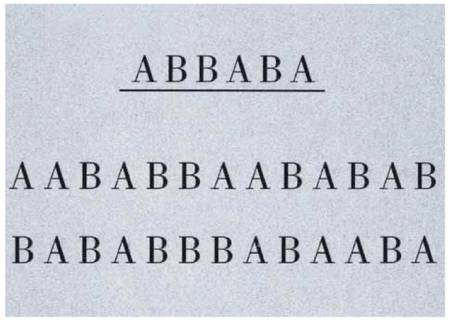




https://magenta.tensorflow.org/music-vae

Memory Clinic

Cooperation with Klinikum Nürnberg



- Can we use speech processing to automate dementia tests?

Table 1: *Automated SKT scoring on manual transcriptions (Trans.) and automatic speech recognition with (ASR-5) and without (ASR-1) the top five word alternatives. Column Top-21 refers to top 21 speakers and ASR-5.*

ID	Test/Task	Trans.	ASR-1	ASR-5	Top-21
1	naming objects	0.89	0.70	0.81	0.89
2	reproducing objects	1.00	0.58	0.71	0.83
3	reading numbers	0.94	0.85	0.86	0.94
6	counting symbols	0.90	0.59	0.58	0.54
7	interference test	0.99	0.97	0.98	0.99
8	naming after distraction	1.00	0.75	0.90	0.97
9	recognizing objects	0.89	0.50	0.55	0.68
-	attention score	0.92	0.84	0.82	0.85
-	memory score	0.98	0.62	0.78	0.93
-	total score	0.97	0.81	0.89	0.94

Table 2: *Automated CERAD scoring on manual transcriptions (Trans.) and automatic speech recognition with (ASR-5) and without (ASR-1) the top five word alternatives. Column Top-21 refers to top 21 speakers and ASR-5.*

ID	Test	Trans.	ASR-1	ASR-5	Top-21
1	verbal fluency test	0.98	0.82	0.85	0.91
2	Boston Naming Test	0.70	0.14	0.24	0.47
3	MMSE	0.71	0.07	0.35	0.52
4	word list learning	0.94	0.62	0.70	0.75
6	word list recall	0.99	0.68	0.81	0.78
-	total	0.71	0.37	0.49	0.61

Visualizing Pathology in Voice

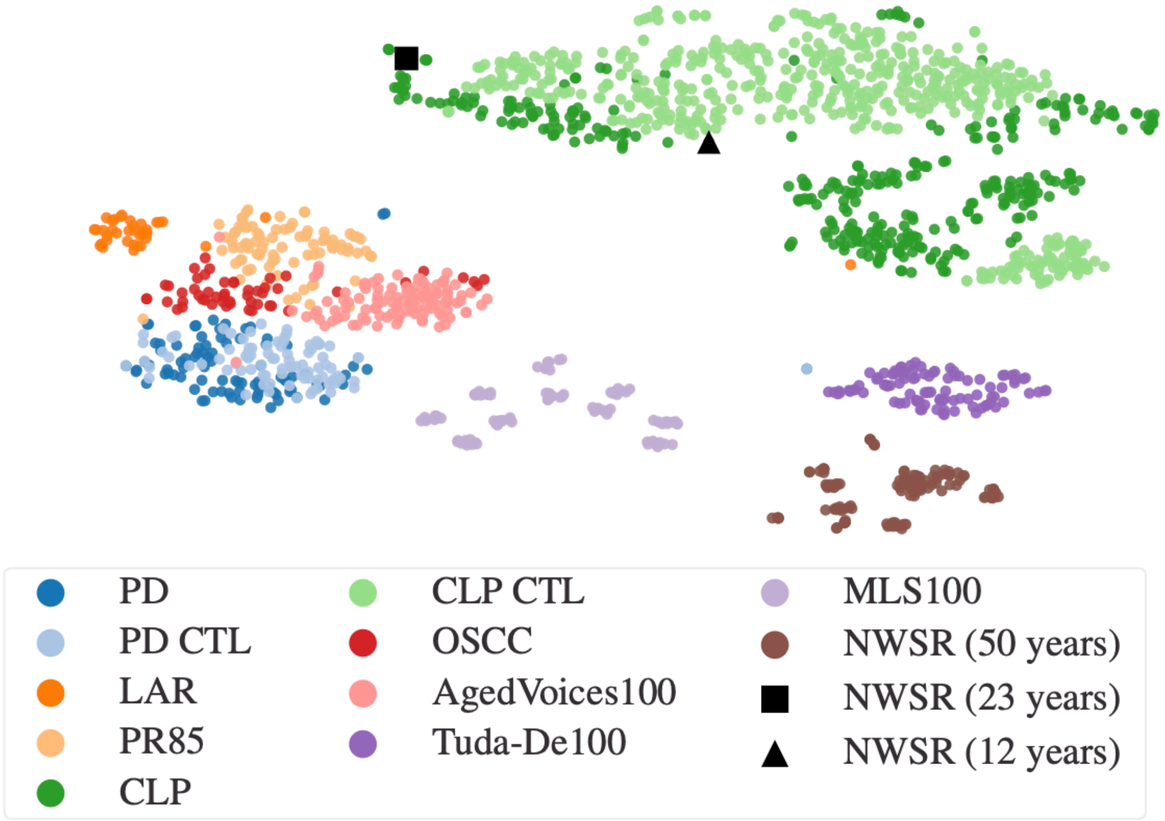


Fig. 1. 2-dimensional t-SNE projection of 768-dimensional latent features extracted at W2V2 layer 4 for all datasets used in this study ($perplexity = 30$).

Data Sources

Analog signals (discretized)

- Microphones
- Vibrations
- Conductivity
- Ambient: pressure, temperature, humidity, …
- Positional: GPS, gyro, distances
- User input: key-press, gestures, pressure, swipe, …

Data Sources

Digital or “Big Data” signals

- Log streams
- Network traffic
- Events (IoT, MQTT, …)
- User-generated content (Twitter, blogs, …)

Toolkits

- Python3 (some parts: Java)
- Unittest, code cells (Visual Studio Code)
- numpy/scipy
- PyTorch (basics)

Syllabus

- Basic algorithms
  - Matching and comparing (discrete) sequences
  - Dynamic programming
- Statistical modeling
  - Markov chains, hidden Markov models
  - Maximum likelihood, expectation maximisation
- Neural networks
  - Feed-forward and recurrent networks
  - Attention and transformers
  - Transfer learning

Assignments

- Assignments due…
  - April 17: Dynamic Programming
  - May 1: Markov Chains
  - May 8: Hidden Markov Models
  - May 22: RNN
  - June 12: Attention
  - June 19: Transformers

What you should bring to this class

- A little bit of probability theory
- A little bit of optimization theory
- Algorithms and programming
- Curiosity and perseverance: understanding is hard, implementing sometimes even harder…