USI/SUPSI Istituto Dalle Molle di studi sull'intelligenza artificiale IDSIA

Biosignal Processing

Machine Learning 2024

Michael Wand

Biosignals

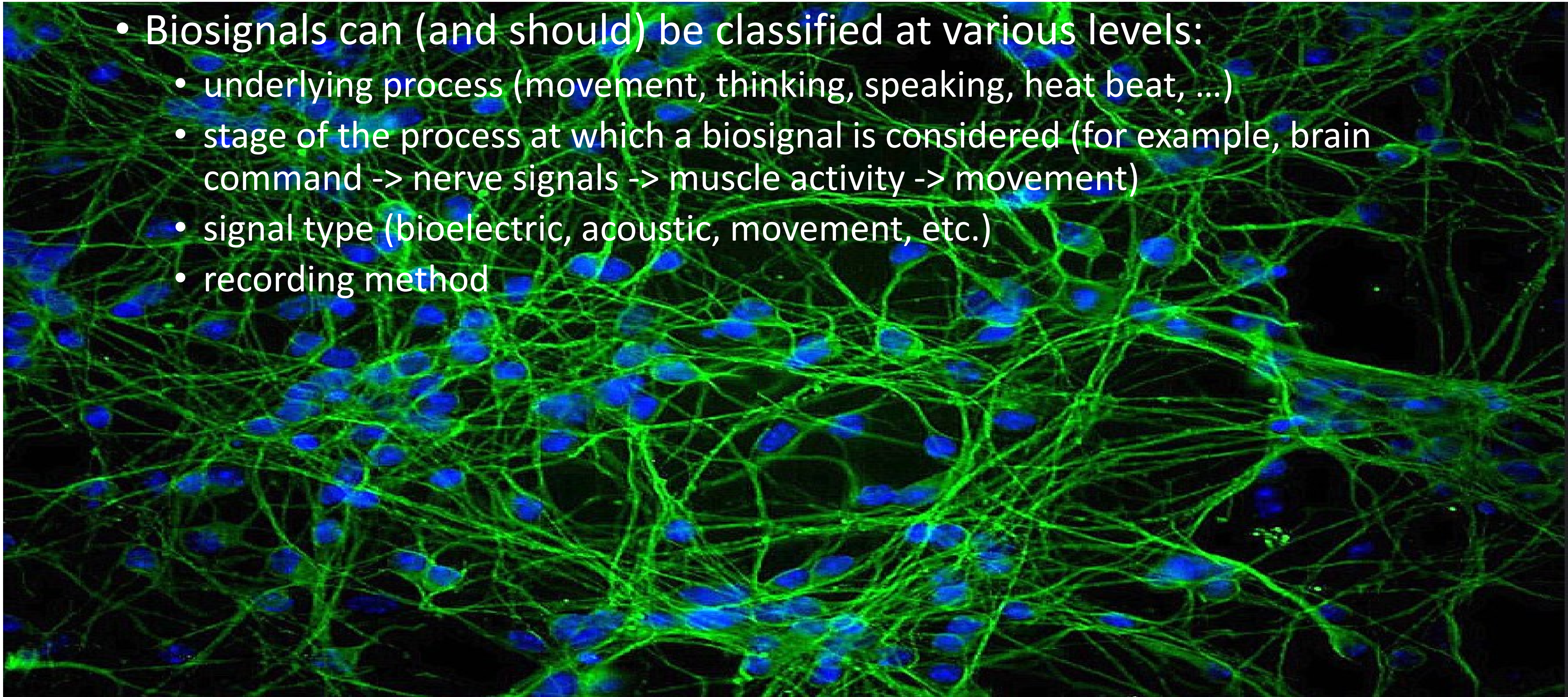
- “autonomous signals produced by human activities measured in physical quantities using different sensor technologies” [1]
- large variety: *electrical* signals, *mechanical* signals, *visual* signals, *acoustic* signals, …
- reflect a variety of biophysiological processes: brain activity, muscle control & movement, inner processes…
- integral part of control chain in our bodies
- highly useful in medicine, prosthetics, monitoring, user interfaces, etc.
- Biosignals are *fun!*

[1] Schultz, Wand, et al: *Biosignal-based Spoken Communication: A Survey*. IEEE Trans. Acoustics, Speech, and Language Processing 25(12), 2017, pp. 2257 - 2271

Elements of a Taxonomy

Istituto Dalle Molle di studi sull'intelligenza artificiale
USI/SUPSI
IDSIA

- Biosignals can (and should) be classified at various levels:
  - underlying process (movement, thinking, speaking, heat beat, …)
  - stage of the process at which a biosignal is considered (for example, brain command -> nerve signals -> muscle activity -> movement)
  - signal type (bioelectric, acoustic, movement, etc.)
  - recording method



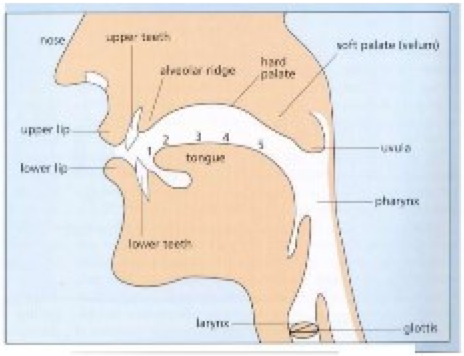
Measuring Biosignals

- Classifying biosignals by signal type / capturing method:
  - electrical biosignals: surface or indwelling electrodes
  - mechanical biosignals: accelerometer, inertial sensors, body-attached devices (angle, strain, …), magnetic coils
  - acoustic signals: microphone (normal/body-conducting), stethoscope, ...
  - visual biosignals: camera (normal / IR), ultrasound, …

Taxonomy Example

USI/SUPSI
Istituto
Dalle Molle
di studi
sull'intelligenza
artificiale
IDSIA

- Take speech as an example:
  - measure speech traces in the brain using ECoG [2]
  - measure speech-related muscle activity with electrodes (we will talk about this) or articulation activity (articulogram [3])
  - measure the speech itself with a microphone.



[2] Herff, et al: *Interpretation of convolutional neural networks for speech spectrogram regression from intracranial recordings.*Neurocomputing 342, 2019, pp. 145 – 151
[3] Gonzalez et al: *Direct Speech Reconstruction From Articulatory Sensor Data by Machine Learning.* IEEE Trans. Acoustics, Speech, and Language Processing 25(12), 2017, pp. 2362 - 2374

Challenges of ML with Biosignals

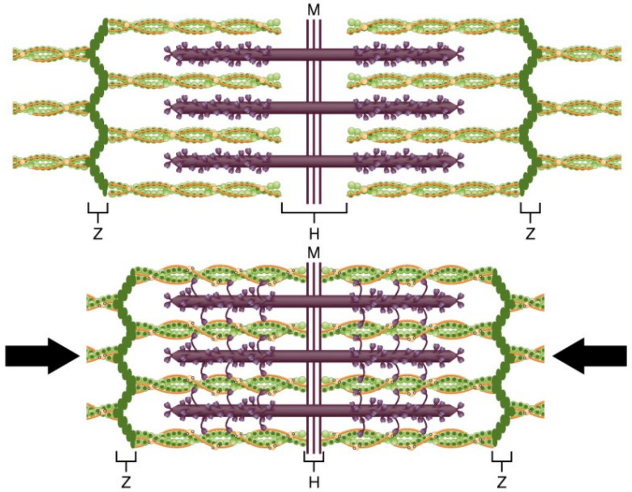
Depending on the concrete application, one observes:

- Signal noise / artifacts
- Subject dependence, occasionally session dependence
  - requires adaptation / disentangling of subject-specific signal structure
- Often small data corpora
- Inexact labels (for supervised training)
- Difference between laboratory-style data recording and everyday usage
- Possibly multimodality issues (but that can also be an advantage!)

The Electromyographic Signal

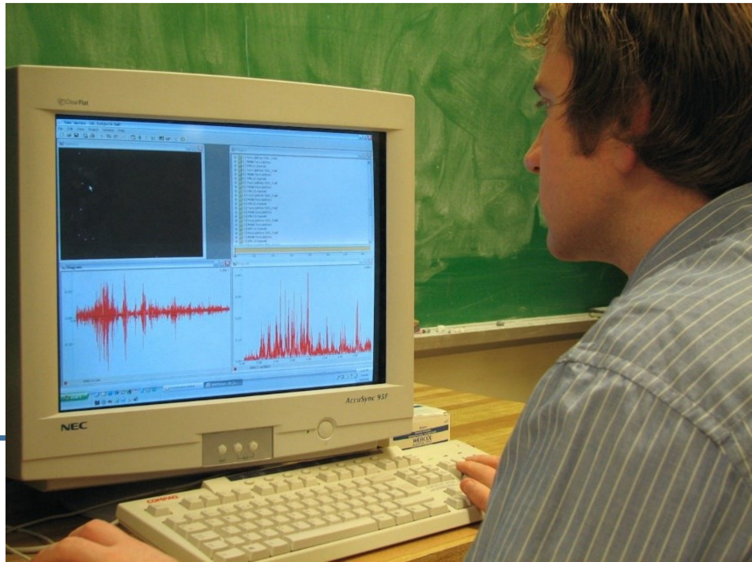
Electromyography (EMG)

- Muscle contraction causes a series of electrical "spikes" which stem from the neural innervation and directly cause muscle contraction.
- With each spike, the violet *Myosin* filaments slide along the green *Actin* filaments
  - (imagine a rowing boat)
- We measure the electrical activity by means of needle or surface electrodes.



fine structure of a muscle fiber

EMG recording



Electromyography (EMG)

- Image: The EMG signal acquisition chain as a whole.
- The motoneuron *innervates* a set of muscle fibers which then contract in unison (*Motor Unit*).
- Force control:
  - by varying the frequency (but not the amplitude) of the neural spikes
  - by recruiting multiple subsets of muscle fibers

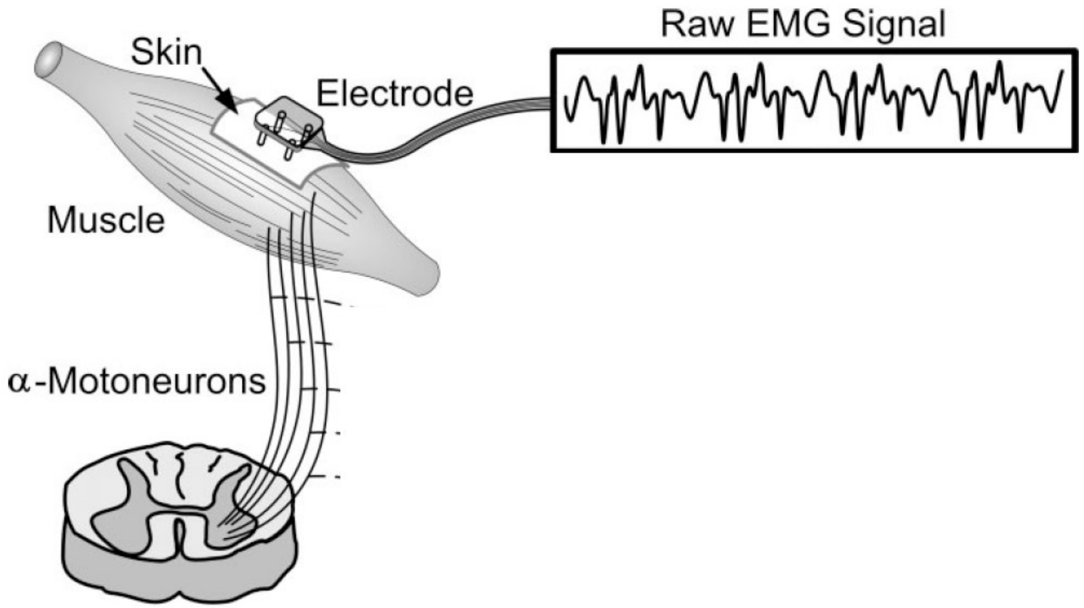


Image adapted from De Luca et al., *Decomposition of Surface EMG Signals.* Journal of Neurophysiology 96, 2006.

Electromyography (EMG)

- Many motor units are involved in a muscle contraction.
- When using surface electrodes, measured signal (sEMG) is a *superposition* of the activity of many such Motor Units, plus measurement noise.
- Effectively, one obtains a stochastic signal whose energy reflects the local muscle activity.
  - Interpretation can be quite complex…
- Indwelling EMG (iEMG) gives a much clearer signal, but at a higher price…

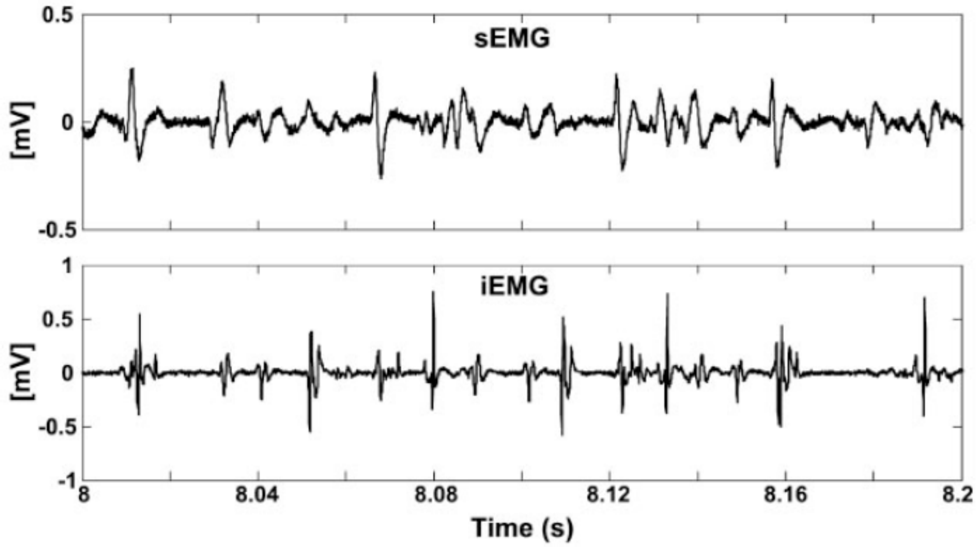


Figure from De Luca et al., *Decomposition of Surface EMG Signals.* Journal of Neurophysiology 96, 2006.

Electromyography (EMG)

Istituto Dalle Molle di studi sull'intelligenza artificiale
USI/SUPSI
IDSIA

- The EMG signal is used for a variety of purposes:
  - medical diagnosis
  - training and rehabilitation
  - *prosthesis control*
  - *EMG-based speech recognition*
- Different purposes might require different means to process the signal.

Electromyography (EMG)

- Idea: *Decompose* the signal into its constituent motor units.
  - Might give detailed insight into muscle fiber recruitment patterns, and into the shape of the motor unit action potentials.
  - Useful for medical purposes, and for biophysiological research.
  - Usually done by recording multiple EMG *channels* and comparing.
- No reliable fully automatic solution available as of now!
  - Usually semi-automatic, requires surveillance by professional.

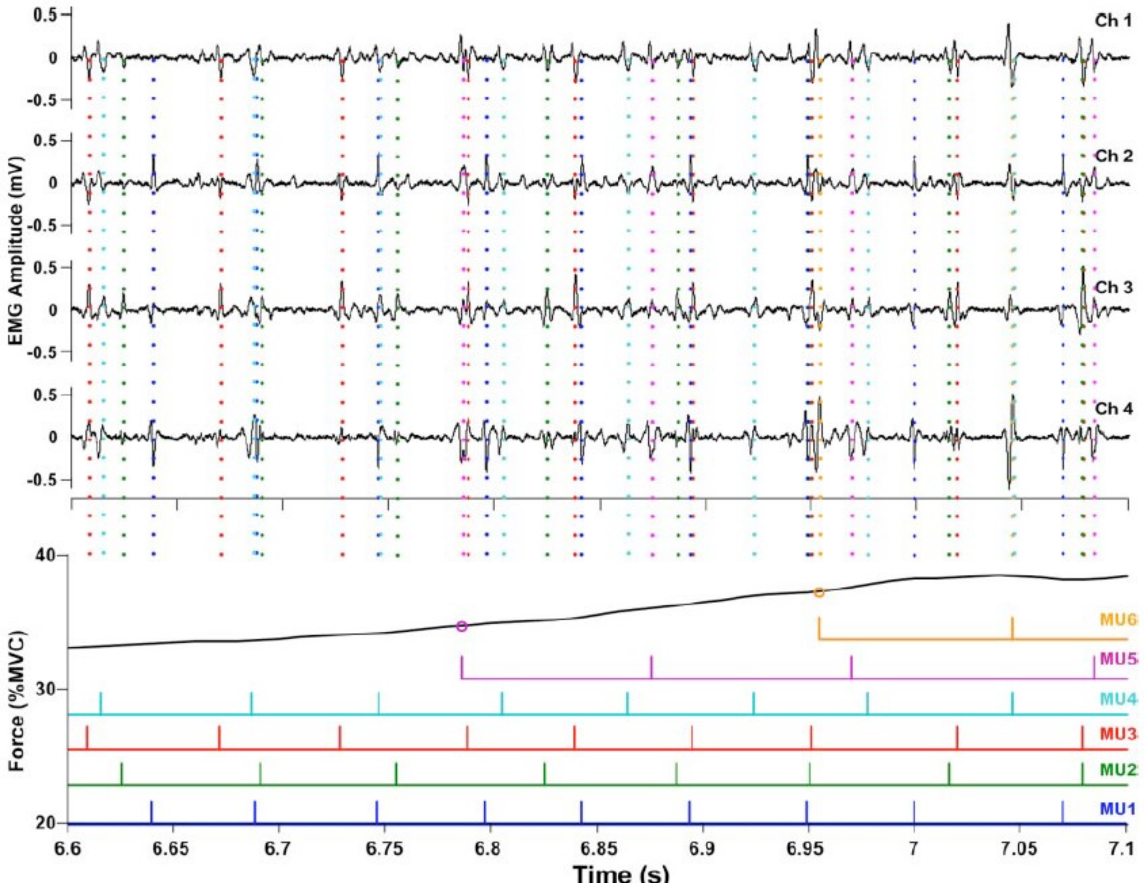


Figure from De Luca et al., *Decomposition of Surface EMG Signals.*

Electromyography (EMG)

- Idea: Take the signal as it is and train a Machine Learning system to understand patterns.
  - somewhat a black box approach
  - but the only feasible one for large-scale applications.
- In this lecture: Two applications!
  - EMG-based prosthesis control
  - EMG-based speech recognition.

Case 1: EMG-based Prosthesis Control

Prosthesis Control

- Goal: control of a multifunctional hand prosthesis with EMG signals using advanced machine learning techniques.
- Commercially available EMG-controlled prostheses are less dexterous than laboratory results suggest
  - usually, no way to control several degrees of freedom simultaneously
  - hardcoded rules and assumptions.
- We want the control (training and testing) to be as intuitive and natural as possible.
- Machine learning is the central method – the system learns how to perform control, avoid hardcoding assumptions
- EU-funded project (INPUT, http://input-h2020.eu), 2016 – 2020.

Data Recording

USI/SUPSI
Istituto Dalle Molle di studi sull'intelligenza artificiale
IDSIA

- Project partners designed state-of-the-art recording system
- Getting rid of artifacts was a major piece of work
- Note that a lot of tests were performed on able-bodied subjects.

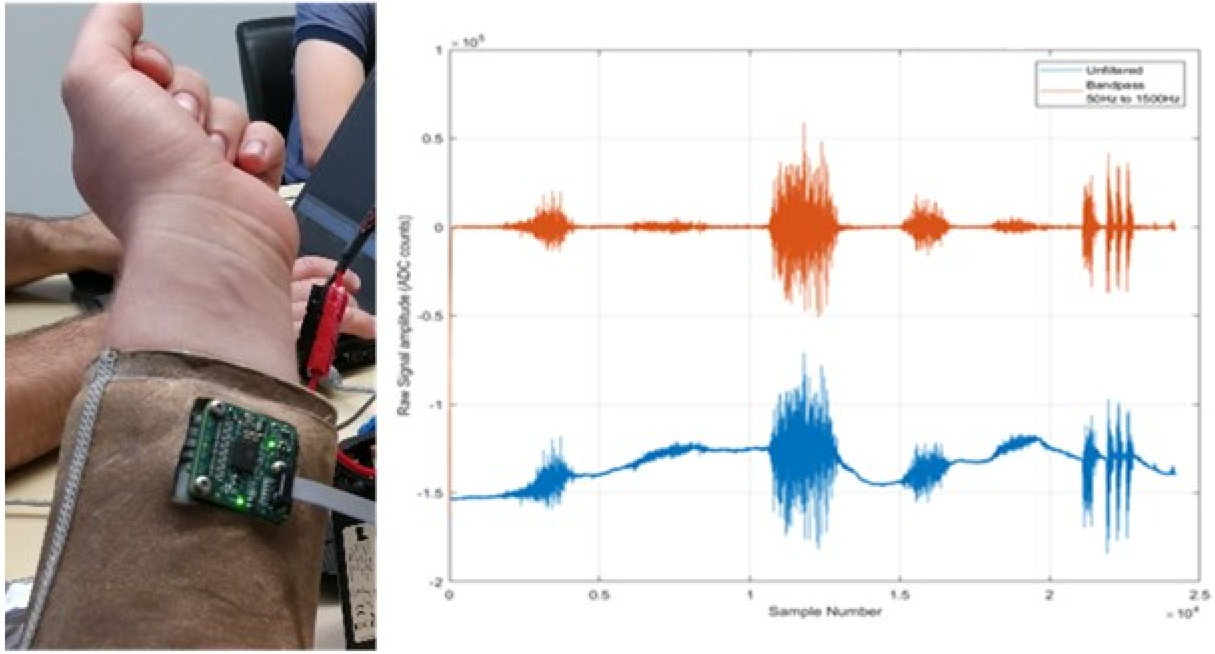
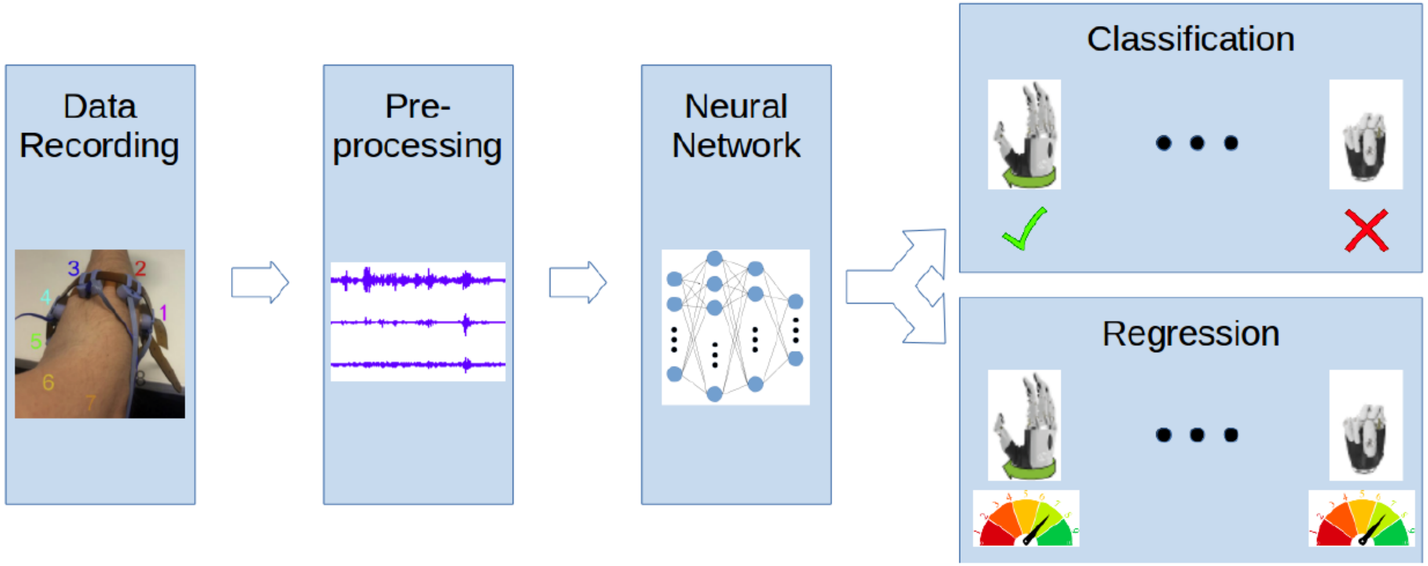


Figure 9: Left: ADS1298 PCB embedded into the electrode liner. Right: EMG signal collected by the new ADS1298 PCB design (orange –filtered EMG, blue- unfiltered EMG)

Offline System

- We ran a variety of experiments on recorded “offline” data:
  - linear regression
  - classification + regression
  - neural networks, diverse architectures
  - standard EMG features vs raw data
  - sequence modeling
- In the end, standard feedforward networks had the best efficiency – performance ratio!



Offline System

- Insights into data processing: Compare linear regression and NN regression on the prosthesis task!
- Ellipses show two first principal components of the data
- Can you see why linear regression is “linear”?

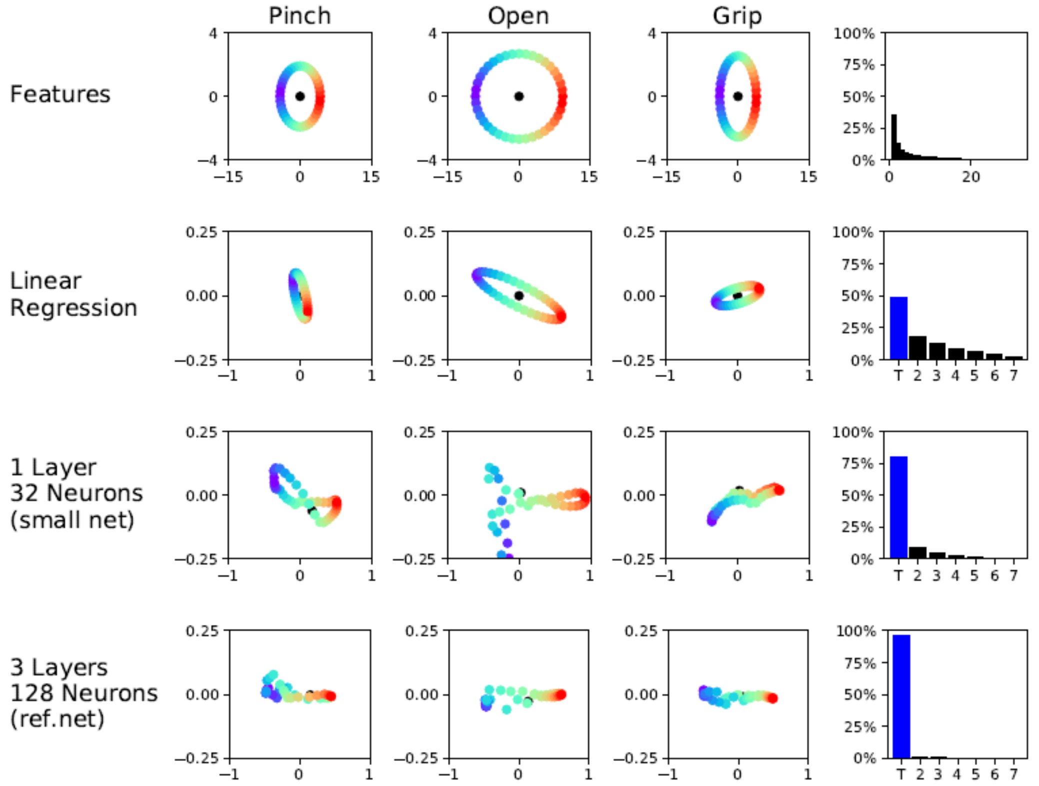
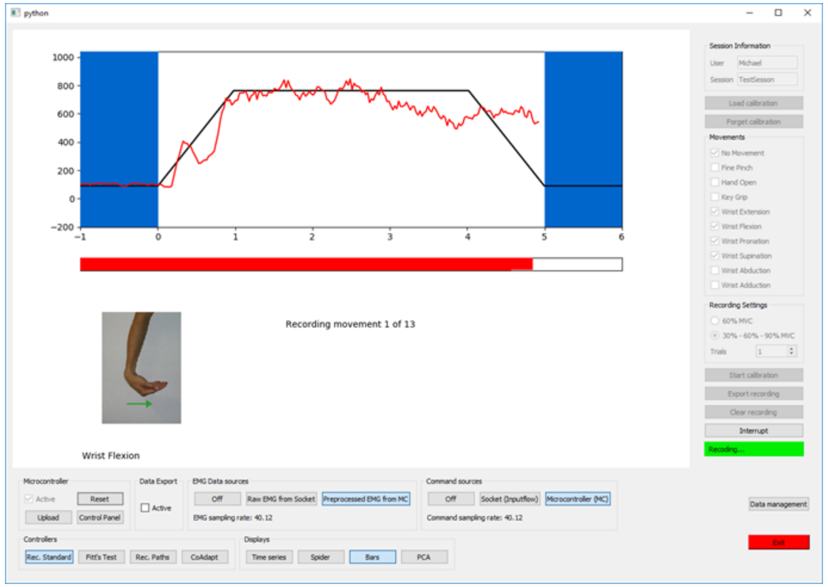


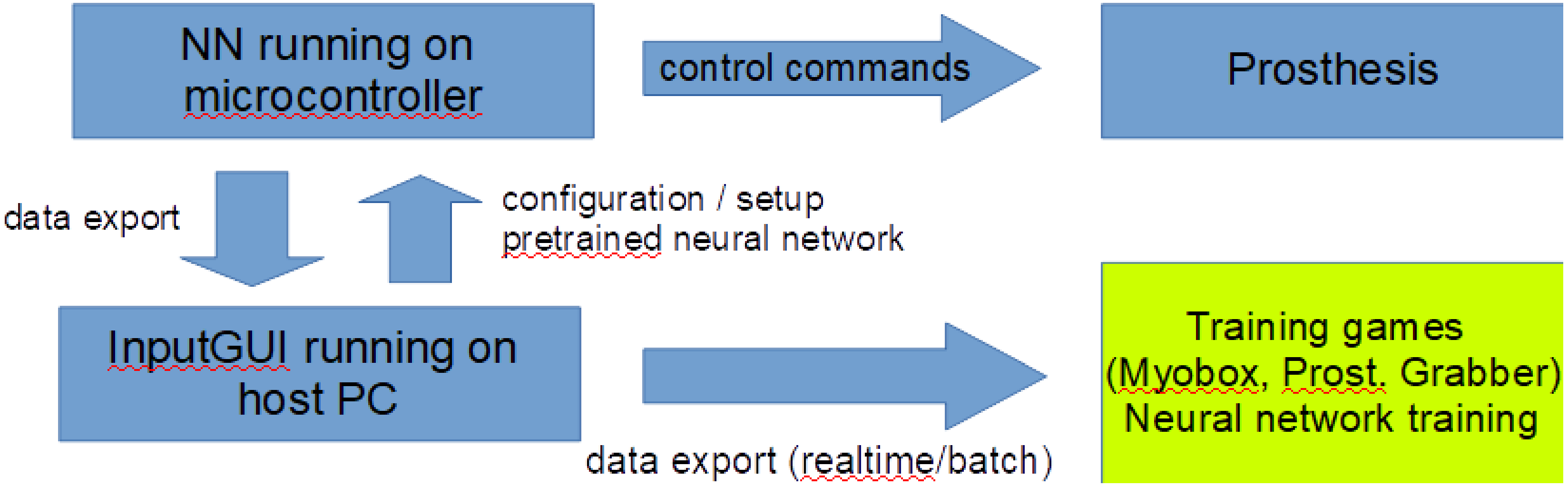
Fig. 6: Mapping of diverse movements of subject 7 in PCA space, for several regressors. Right column shows explained variances averaged over all movements.

Online System

USI/SUPSI
Istituto Dalle Molle di studi sull'intelligenza artificiale
IDSIA

- Idea: Pretrain the system using a computer (might be at rehabilitation facility)
  - Afterwards, system needs to run on integrated microcontroller.
- User training (also done by project partners) with *Serious Games.*

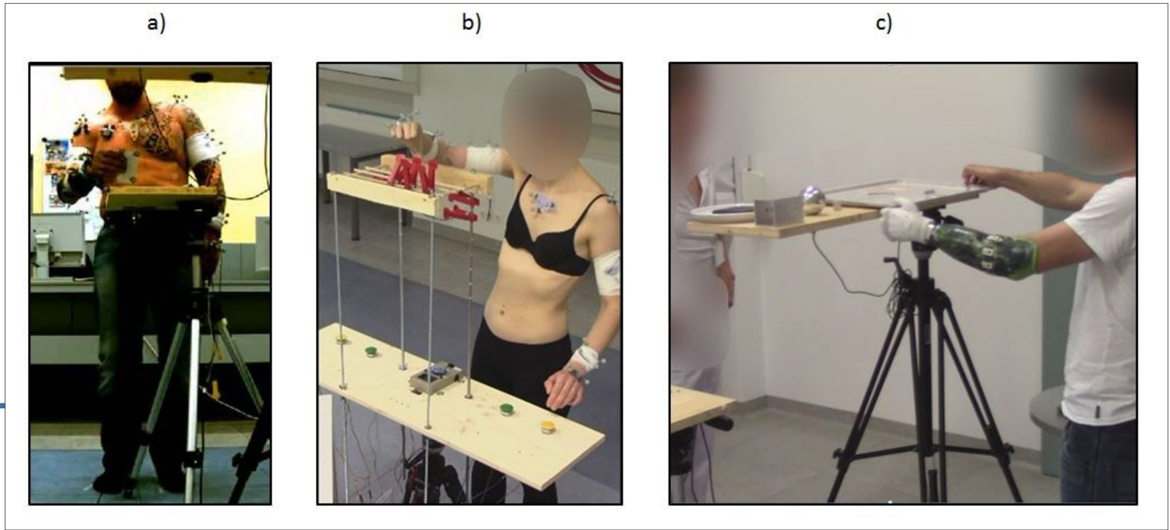




Kristoffersen et al: *Serious gaming to generate separated and consistent EMG patterns in pattern-recognition prosthesis control.* Biomedical Signal Processing and Control, 2020

Application

- Patient tests showed efficiency of the system
- Difference between controlled training data and practical use is evident
- In practical applications, *user training* can make the difference between success and failure!

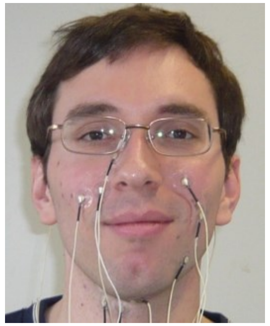


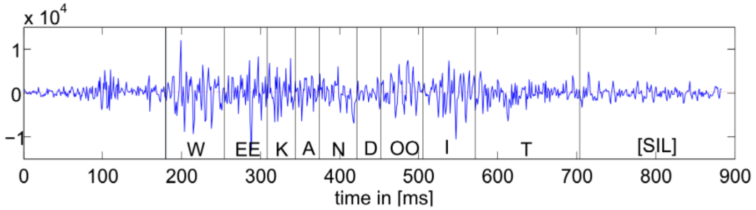
Case 2: Silent Speech Interfaces and EMG

EMG-based Silent Speech Interface

USI/SUPSI
Istituto
Dalle Molle
di studi
sull'intelligenza
artificiale
IDSIA

- Silent Speech: Soundless Articulation
- Silent Speech Interface (SSI): processes speech even when no sound can be measured
- EMG-based speech recognition or synthesis: an instance of an SSI





Electromyography (EMG)

Istituto Dalle Molle di studi sull'intelligenza artificiale
USI/SUPSI
IDSIA

Speech is

- the most natural means of human communication
- an increasingly important way to interact with computers

… but it must be clearly audible

Silent Speech makes speech-based communication/interaction available

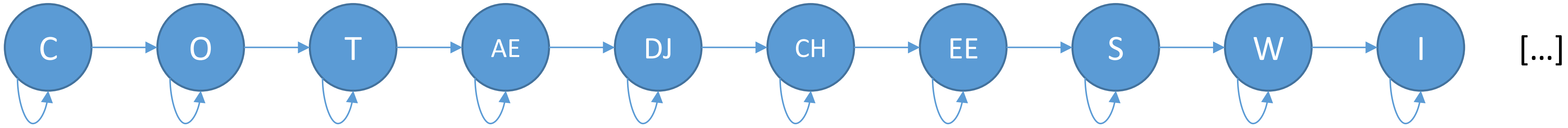
- while preserving confidentiality of communication
- while avoiding disturbances for bystanders
- for people without a voice (e.g. laryngectomees)

hot current topic with lots of exciting cross-disciplinary research to be done!

Speech Recognition Overview

USI/SUPSI Istituto Dalle Molle di studi sull'intelligenza artificiale IDSIA

- Let's start with simple phone-based models



"Cottage cheese with chives is delicious"

- Input: frame-based features, usually based on fundamental properties of the EMG signal (energy etc.)
  - stationarity assumption
  - context features, dimensionality reduction
- Until ~10-15 years ago, modelling by GMMs.
- For training, ideally one should have exact frame alignments, particularly with small amounts of data.

Hudgins et al: *A new strategy for Multifunction Myoelectric Control.* IEEE Transactions on Biomedical Engineering, 1993
Jou et al: *Towards continuous Speech Recognition using Surface Electromyography.* Proc. Interspeech, 2006

Speech Recognition Overview

- Further components:
  - Creation of adapted phonetic models (decision tree style)
  - Dictionary and language model to determine possible/probable words
    - modeling of unknown words, foreign words, names, etc……………
  - efficient search (Beam Search)
  - specific training and adaptation methods.

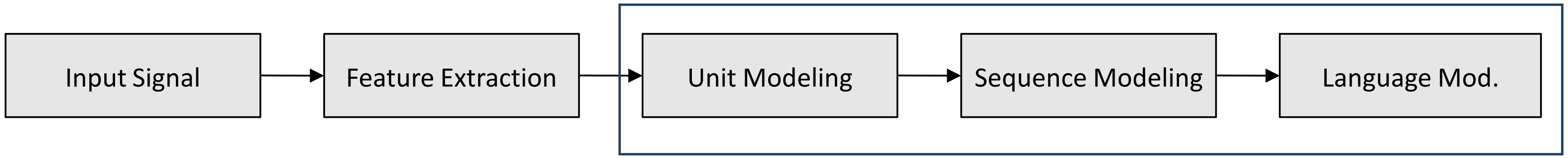
State-of-the-art NN based Speech Recognition

- Great achievement of modern methods (sequence-to-sequence with attention, transformers, ...): no more necessary to explicitly create these complex models!
- Instead, one can train a system to directly output (e.g. English characters) from speech input (can even be the raw speech signal).
- We have presented first NN end-to-end system for EMG-based speech recognition (2024)

Hou et al., *emg2vec: Self-supervised Pretraining in Electromyography-based Silent Speech interfaces.* EMBC 2024.

The EMG-based Speech Recognizer

- Have a look at the (classical) speech recognition processing chain

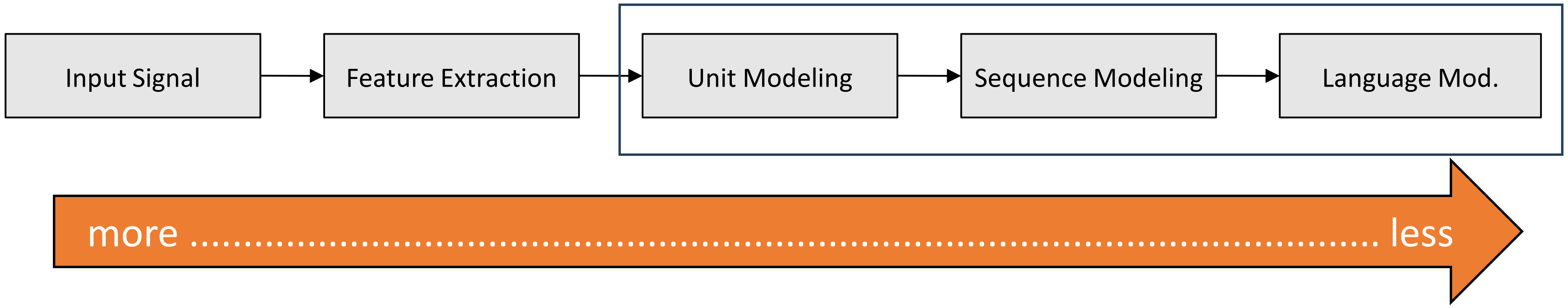


This is somewhat one single block in end to end NN systems

- Which parts do you think need to be changed in order to apply speech recognition to EMG signals?

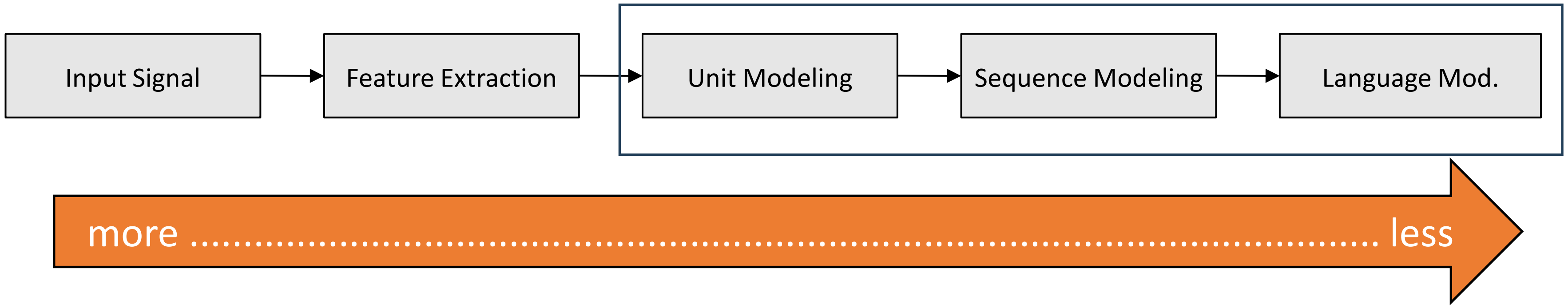
The EMG-based Speech Recognizer

- Have a look at the (classical) speech recognition processing chain



The EMG-based Speech Recognizer

- Have a look at the (classical) speech recognition processing chain



- The closer we are to the signal input, the more we need to adapt the system.
- Latter parts can remain structurally unchanged.

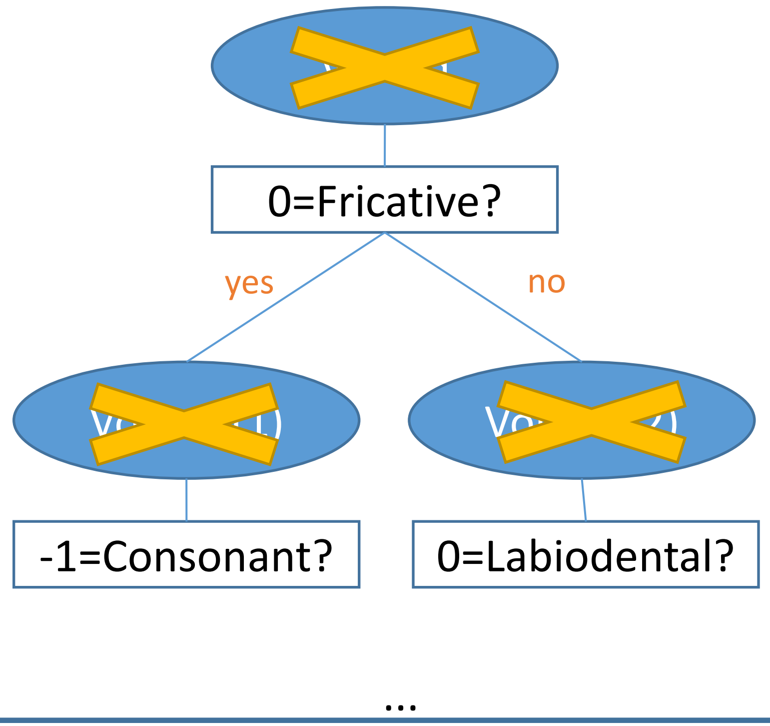
EMG Signal Features

- EMG data recording:
  - 6 channels, 600Hz sampling rate
  - often just a few minutes of data per recording *session*
  - *Broadcast News* style sentences
- EMG features:
  - time-domain features (cp. Hudgins 1983): frame-based power, zero crossing rate, ...
  - articulatory gestures have very long temporal context -> currently modeled by stacking adjacent frames to form a context vector
  - dimensionality: 500 – 1000 features
  - notably, no dimensionality reduction
    required for the neural network frontend

Wand et al: *The EMG-UKA Corpus for Electromyographic Speech Processing*. Interspeech 2014

Modeling of EMG Units

- *Bundled Phonetic Features*
- We perform a clustering (as we saw before, in context dependent speech recognition)
  - with binary phonetic features as starting points
  - the clustering algorithm is based on the GMM models (not covered here)
  - context can additionally be used



Schultz/Wand: *Modeling Coarticulation in EMG-based Speech Recognition* (Speech Comm. 2009)

Modeling of EMG Units

- One phonetic feature "stream" is not enough
- Merge several streams, having different phonetic features as starting points, into one observation probability
- works well for both GMM and DNN frontends (but for GMMs it is even more important)

- A powerful method for
  - a) binning training data in the most efficient way for distinct models
  - b) *reusing training data!*

Unit model frontend

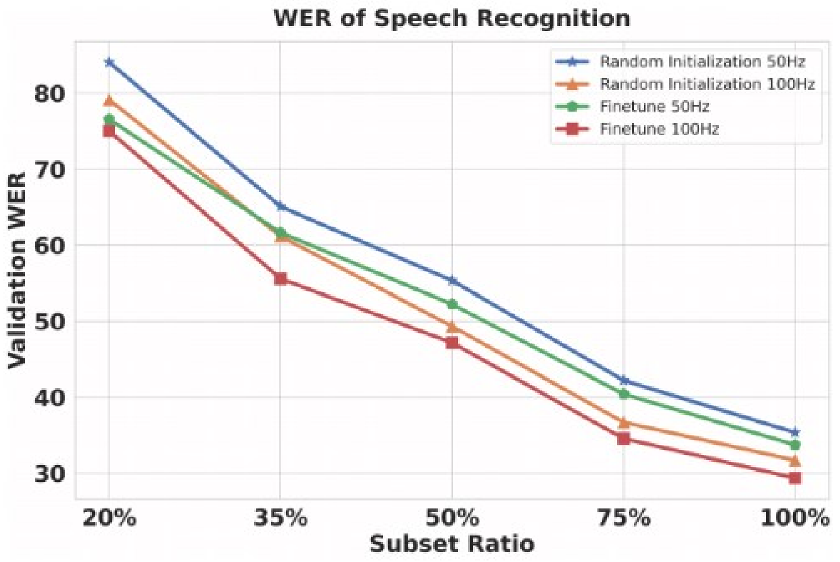
- Two *frontends* for the unit models (which map input vectors to Bundled Phonetic Features) have been explored:
  - Gaussian Mixture Models
  - Feedforward (deep) neural networks
- Neural networks work substantially better (30% error reduction)
  - but currently still depend on the EMG units which are derived from the Gaussian Mixture models.

Sequence Modeling

- Beam search over possible paths
- Dictionary: 100 words ... 2000 words
- Language model: trigram statistical language model (legacy)

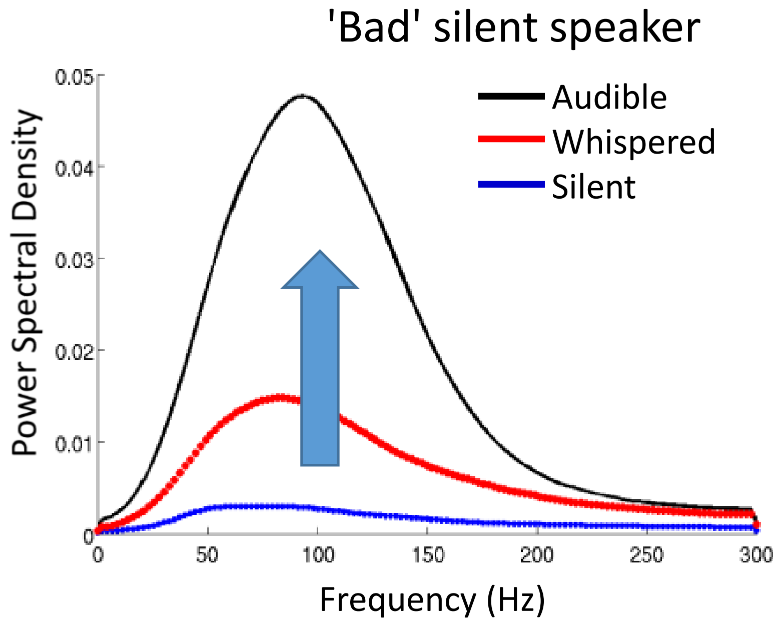
End-to-end NN system

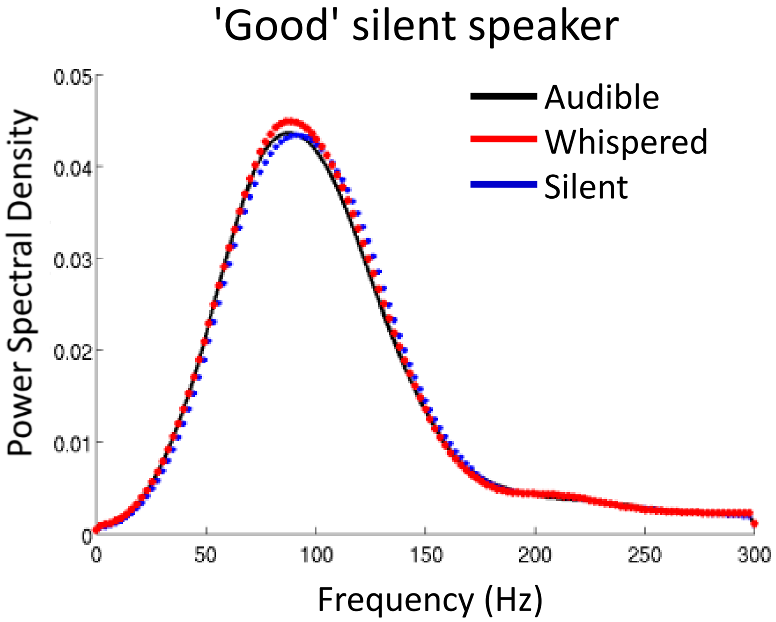
- Most recent work: train a neural network for EMG-based speech recognition end-to-end
- Architecture of choice: transformer encoder + CTC
- Output units?
  - Characters, phones, maybe even the good old bundled phonetic features?
- Features?
  - CNNs on raw EMG work surprisingly well
  - Self-supervised pretraining brings significant improvement



Audible and Silent Speech

- Silent and audible articulation differs
  - most easy to observe: different signal energy
  - experimented with different kinds of signal-based mapping algorithms
  - limited success





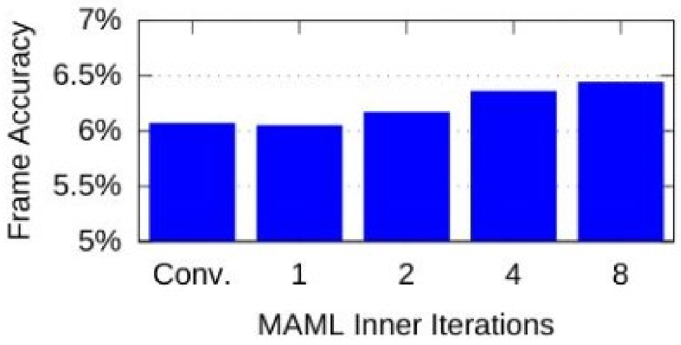
[1] Wand et al: *Tackling Speaking Mode Varieties in EMG-based Speech Recognition.* IEEE Trans. Biom. Eng., 2014

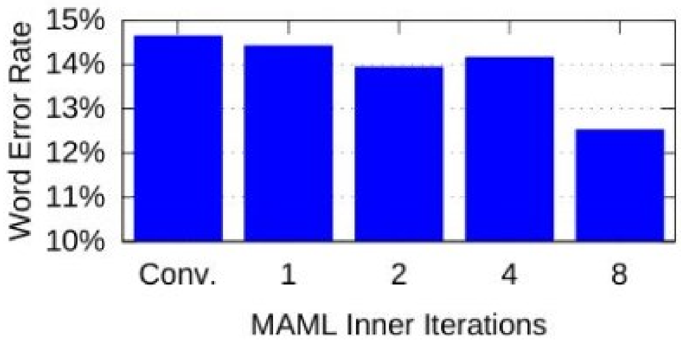
Adaptation to Unseen Sessions

- A *session* is an uninterrupted recording during which electrodes are not removed or reattached
- Between sessions, the EMG signal differs slightly (environmental factors, skin properties, repositioning)
  - huge issue for practical application!
- Best realistic solution: *pretrain* a system and *adapt* it to a new session, with
  - as little data as possible
  - possibly with limited supervision

Adaptation to Unseen Sessions

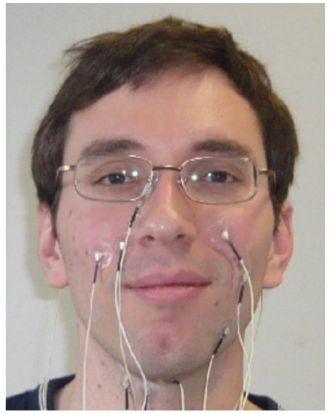
- Several successful GMM-based experiments [1,2]
- State of the art: *Metalearning* [3] for adaptation of the DNN-based recognizer [4]
- ~45 minutes of pretraining, ~1.5 minutes of adaptation





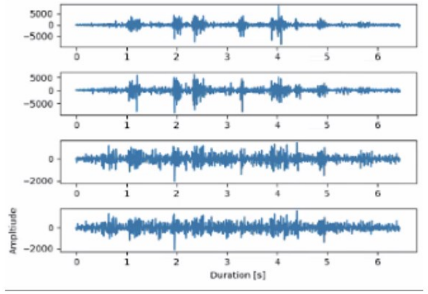
[1] Wand et al: *Session-independent EMG-based Speech Recognition*. Proc. Biosignals, 2011
[2] Wand et al: *Towards Real-life Application of EMG-based Speech Recognition by using Unsupervised Adaptation.* Interspeech 2014
[3] Finn et al: *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.* Proc. ICML, 2017,
[4] Prorokovic et al: *Adaptation of an EMG-Based Speech Recognizer via Meta-Learning.* Proc. Globalsip, 2019

EMG-to-Speech: direct synthesis



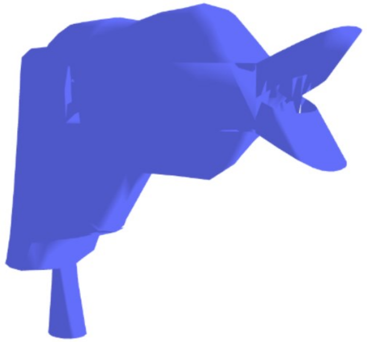
EMG electrodes

Signal acquisition



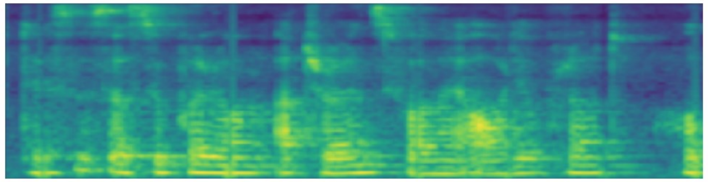
Multichannel EMG recordings

Physiologically-motivated CNN encoder



Physiology-informed articulatory latent space

MeshCNN decoder

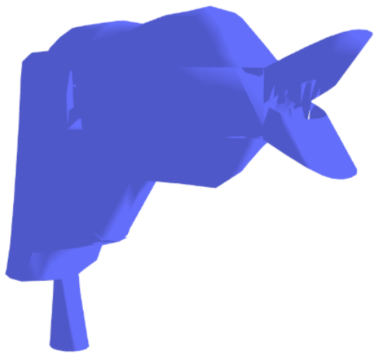


MFCC features

Vocoder

Speech at 16kHz

EMG-to-Speech: direct synthesis



Physiology-informed articulatory latent space

Deep supervision:
- parameters that control vocal tract shape known for different phonemes (Birkholz et al., Story et al.)
- increase robustness?
- speed up training?
- increase data efficiency?

Decrease number of trainable parameters:
- no need to connect every EMG channel to every parameter that controls vocal tract shape

Visual feedback:
- can a visual feedback in a GUI speed up user co-adaptation?
- increase model interpretability?
- facilitate debugging?

EMG-based SSI: Conclusion

Istituto Dalle Molle di studi sull'intelligenza artificiale
USI/SUPSI
IDSIA

- The EMG-based SSI is under active research at IDSIA (MyVoice project, funded by SNF and DFG, 2023 – 2026)
- If you are interested to collaborate, please let us know!
  - data recording
  - research assistance
  - theses

- Send an email to michael.wand@idsia.ch!