Towards Exploiting ASR Error for Generating Synthetic Clinical Speech Data

Hali Lindsay[1], Johannes Tröger [2], Mario Mina[2], Nicklas Linz[2], Philipp Müller[1], Jan Alexandersson[1], Inez Ramakers[3]

[1] *Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)*
[2] *ki:elements*
[3] *Maastricht University Medical Center (MUMC+)*

Clinical Speech Research Pipeline

Clinical Speech

Clinicians record pathological speech (e.g. SVF)

Transcription

Manually or with ASR

Features

Extracted from speech and transcripts

Models

Test diagnostic Value (ML classification)

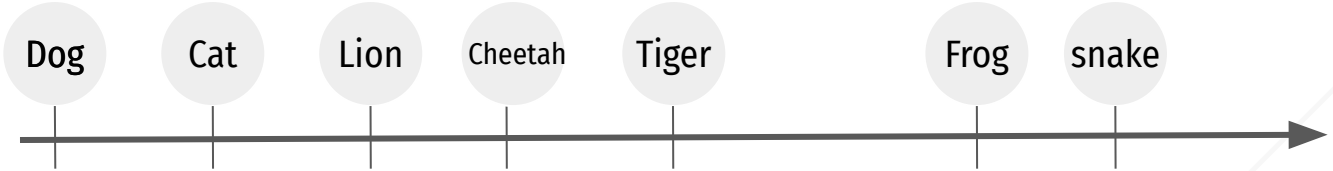
Example| Can we diagnose cognitive impairment from Speech?

“*Name as many animal as you can in sixty seconds*”

Clinical Speech

Transcription

"*Name as many animal as you can in sixty seconds*"



Clinical Speech

Transcription

Features

Cognitive Processes in the Semantic Verbal Fluency.

Semantic Memory

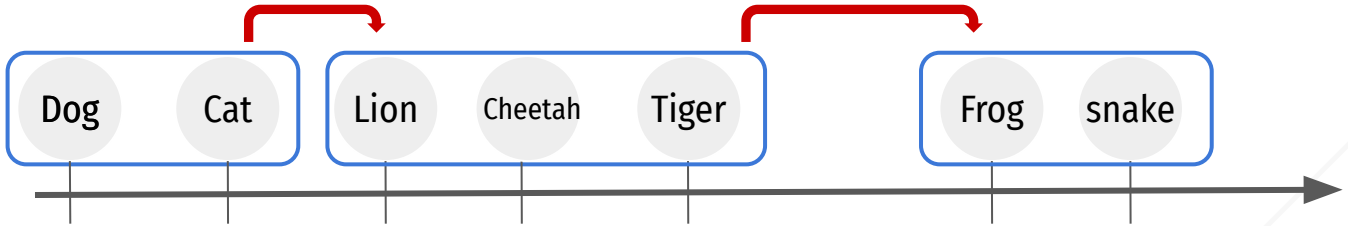
Long-term memory store of knowledge

Cluster of semantically related words

Executive Functions

Search through the semantic memory

Switch between topics to exploit



Clinical Speech

Transcription

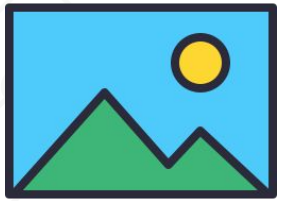
Features

Models

Challenge| Clinical Data is expensive and scarce.

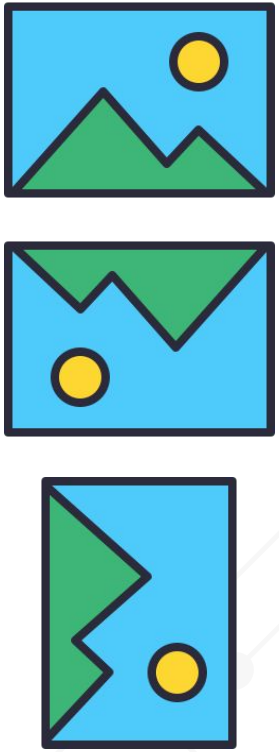
Borrowing Augmentation Techniques from Computer Vision

Original

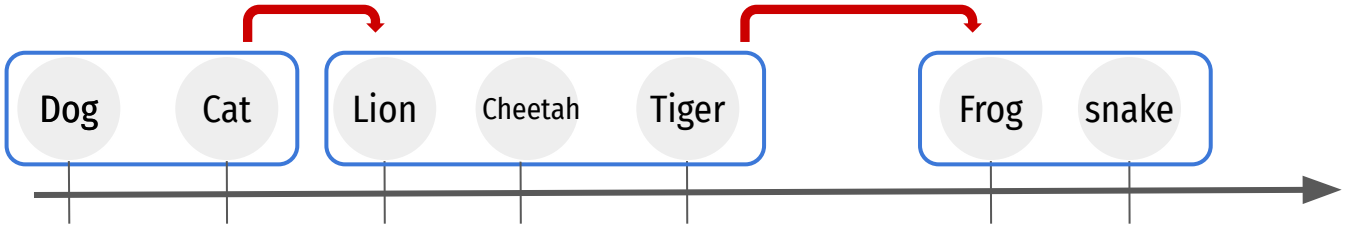


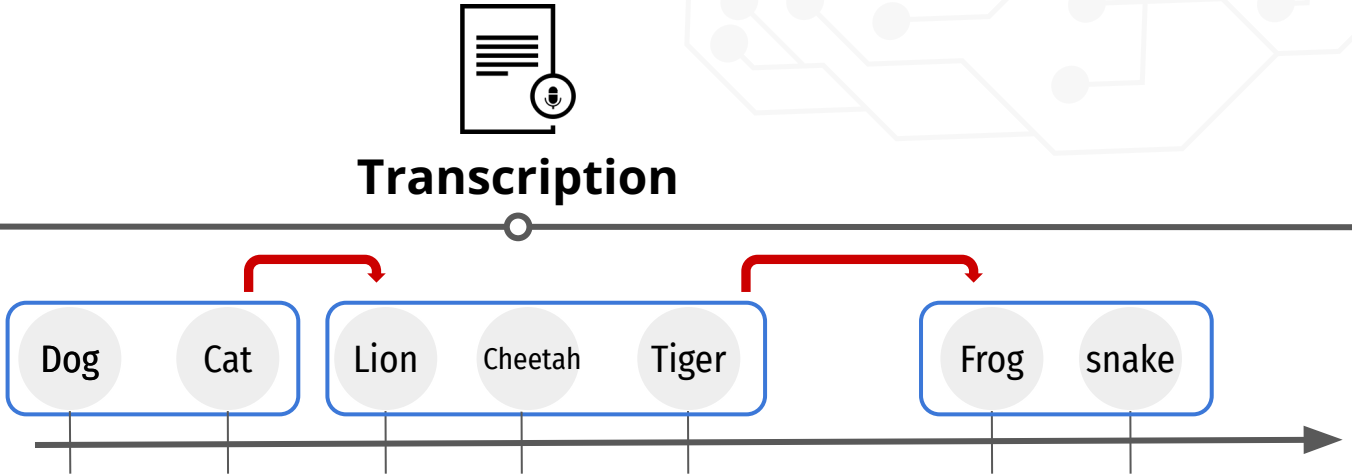
Random Erasing

Flipping/Rotating

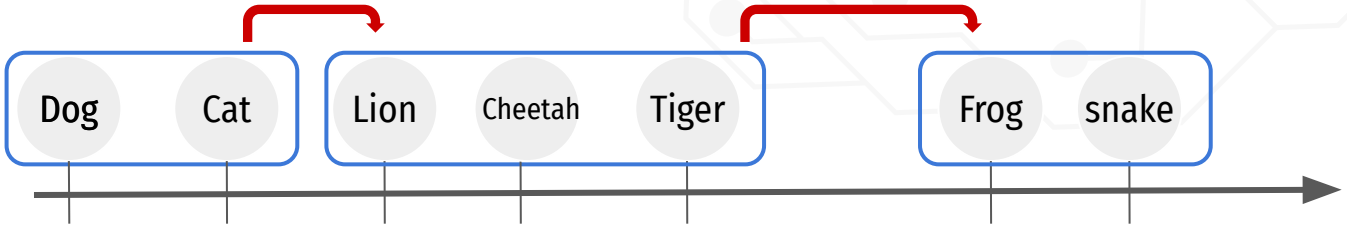


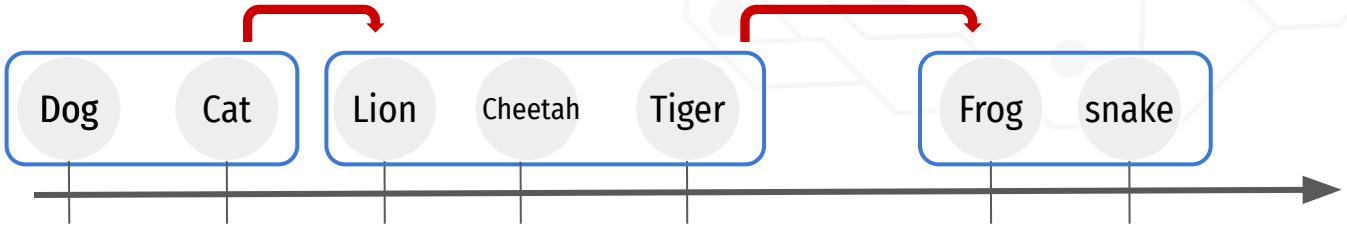
Transcription



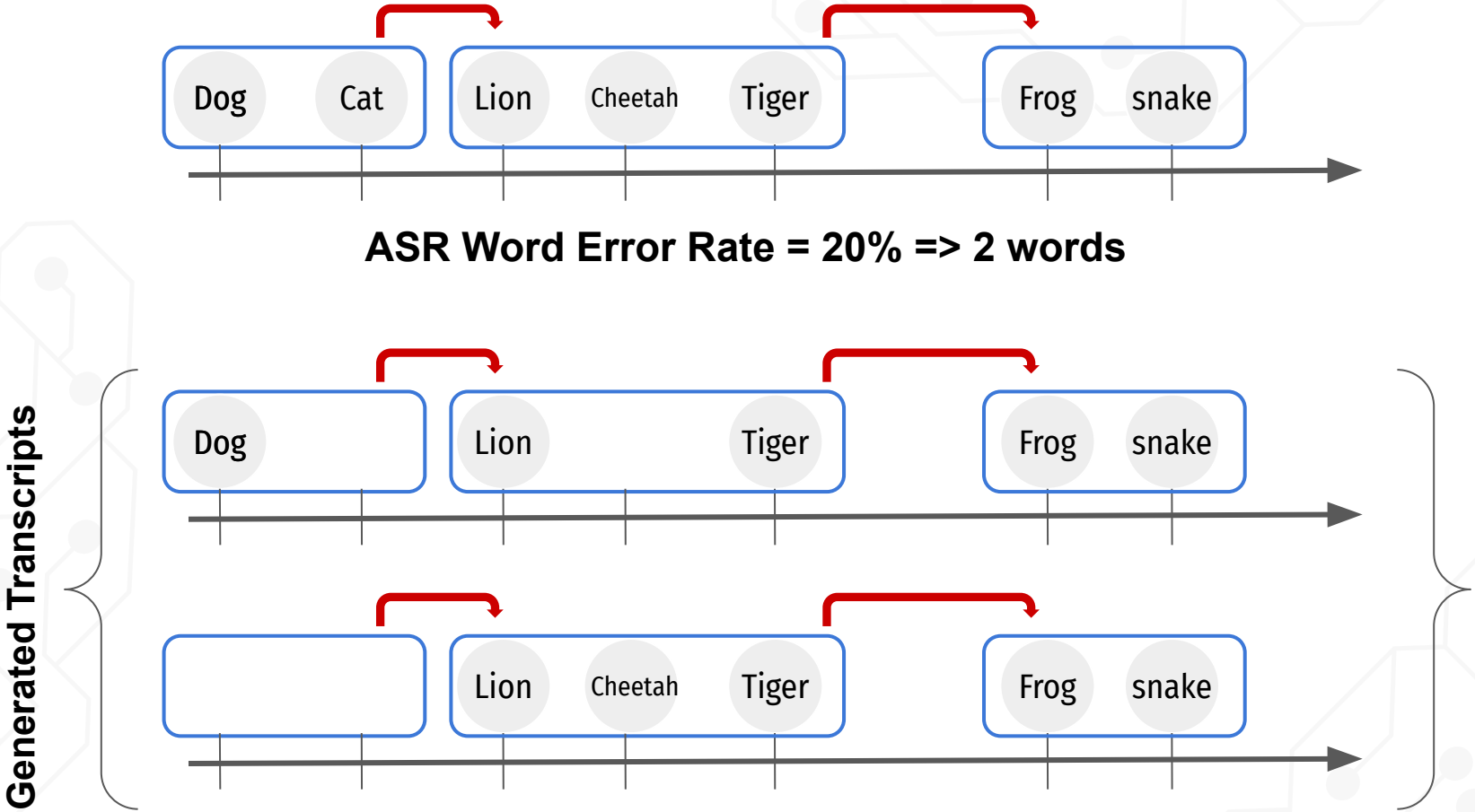


- Evaluate ASR quality by Word Error Rate (WER)
- 3 types of Errors; insertion, substitution, **deletion**
- We do not see performance difference between ASR and manual transcript (Konig et al., 2018)


Dog
Cat
Lion
Cheetah
Tiger
Frog
snake



ASR Word Error Rate = 20% => 2 words


Dog
Cat
Lion
Cheetah
Tiger
Frog
snake
ASR Word Error Rate = 20% => 2 words
Generated Transcripts
Dog
Lion
Tiger
Frog
snake
Lion
Cheetah
Tiger
Frog
snake

Clinical Speech

SVF from 100
dutch speakers
(50 HC, 50 MCI)

Clinical Speech

SVF from 100 dutch speakers (50 HC, 50 MCI)

Transcription

Manually & ASR

Compute WER

Clinical Speech

SVF from 100 dutch speakers (50 HC, 50 MCI)

Transcription

Manually & ASR

Compute WER

Data Augmentation

Generate synthetic transcripts by randomly deleting from Manual transcript

Clinical Speech

SVF from 100 dutch speakers (50 HC, 50 MCI)

Transcription

Manually & ASR

Compute WER

Features

Extract features

Data Augmentation

Generate synthetic transcripts by randomly deleting from Manual transcript

Clinical Speech

SVF from 100 dutch speakers (50 HC, 50 MCI)

Transcription

Manually & ASR

Compute WER

Features

Extract features

Models

Train on synthetic
Test on ASR data

Data Augmentation

Generate synthetic transcripts by randomly deleting from Manual transcript

Diagnosis	N	Gender	Age	MMSE	WER
HC	50	18/32	70.66(8.96)	28.68(1.27)	20.29
MCI	50	19/31	65.94(7.80)	26.92(2.07)	23.13

Generate 10 synthetic files per person based on personal WER

Model Specifications:

Logistic Regression
Leave One Out Cross Validdation
Train on 990 synthetic; Test on 1 ASR transcript
Univariate Feature Selection

Results

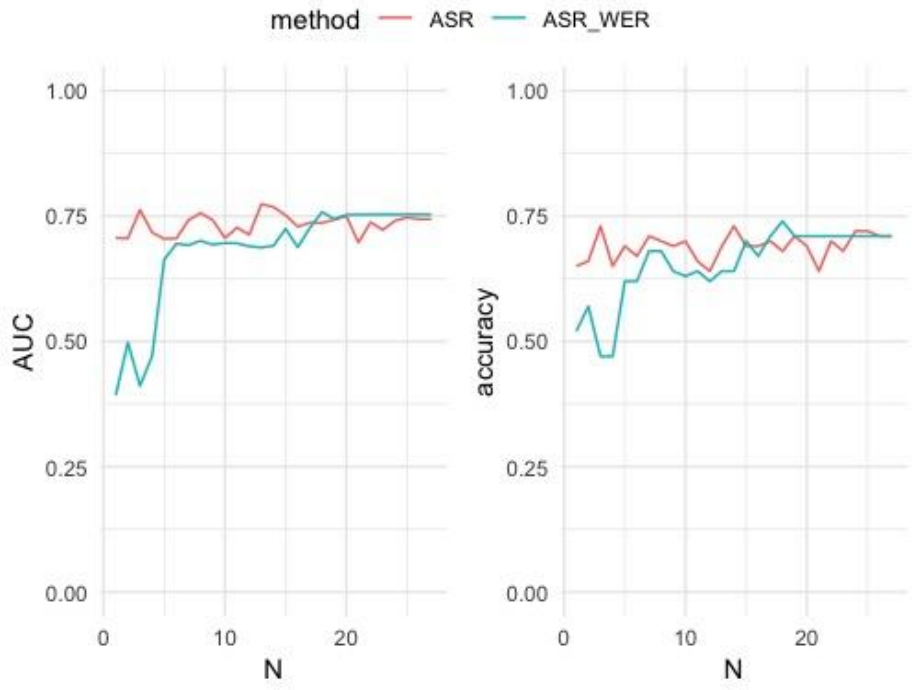
Synthetic data performs similarly to real ASR data

Increasing amount of data performs better with more features

Future Work

Try deep learning architectures to see if classification can be improved

Try in other datasets (additional languages, diagnoses)



Number of Features

Extra Slides

Thanks!

hali.lindsay@dfki.de

This research was funded by MEPHESTO projectQ10 (BMBF Grant Number 01IS20075).

Clinical Speech Research Pipeline

Clinical Speech

Clinicians record pathological speech (e.g. SVF)

Transcription

Manually or with ASR

Features

Extracted from speech and transcripts

Models

Test diagnostic Value
(ML classification)

Cognitive Processes in the Semantic Verbal Fluency.

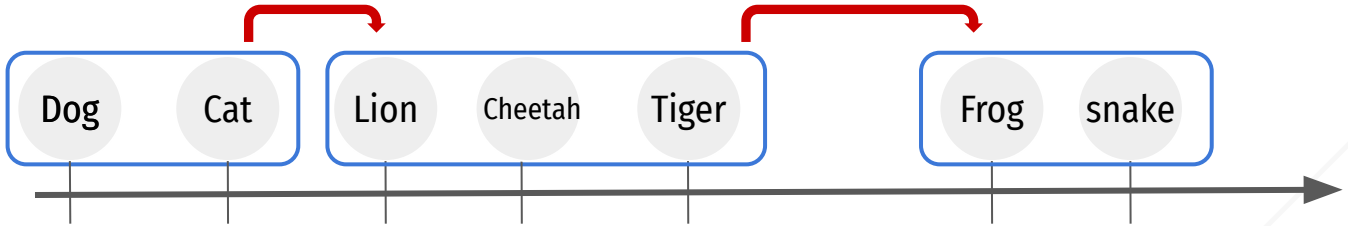
*VF performance hinges on intact **semantic memory** stores as well as the ability to access and search these memory stores, via **executive function.** To investigate strategy, **clustering** and **switching** have been proposed.*

Semantic Memory

Long-term memory store of knowledge
Cluster of semantically related words

Executive Functions

Search through the semantic memory
Switch between topics to exploit



Clinical Speech

Clinicians record pathological speech (e.g. SVF)

Transcription

Manually or with ASR

Features

Extracted from speech and transcripts

Models

Test diagnostic Value
(ML classification)