Friedrich-Alexander-Universität
Technische Fakultät

Pattern Recognition Lab

FAU

Analysis of Pathological Speech

—

Pitfalls along the Way

Elmar Nöth
and many current and former members of the
FAU Speech Group

I would like to thank many current and former colleagues who helped along the way with the analysis of pathological speech and this talk (I am responsible for any faults in the slides)

In alphabetical order:

T. Arias, A. Batliner, I. Baumann, S. Bayerl, T. Bocklet, F. Braun, T. Haderlein, F. Hönig, P. Klumpp, A. Maier, R. Orozco, P. Pérez, K. Riedhammer, M. Schuster, S. Steidl, C. Vasquez, D. Wagner, T. Weise

The FAU Team


T. Arias


A. Batliner


I. Baumann


S. Bayerl


T. Bocklet


F. Braun


T. Haderlein


F. Hönig


P. Klumpp


A. Maier

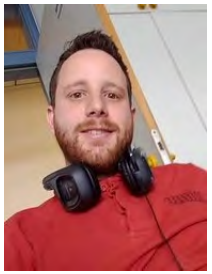

R. Orozco

P. Pérez


K. Ried-hammer


M. Schuster


S. Steidl


C. Vasquez


D. Wagner


T. Weise

Outline

Pattern Recognition Lab FAU

- Motivation
- Examples for pathologies
- Goals of (automatic) analysis
- Some use cases
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- Summary and Take-Aways

Motivation

Pattern Recognition Lab FAU

- Analysis of pathological speech: strongly increasing interest
- Speech is a “cheap bio-signal”
- Can reveal early signs of change
- But: Many different pathologies across different age groups, no recipe for analysis

In this talk:

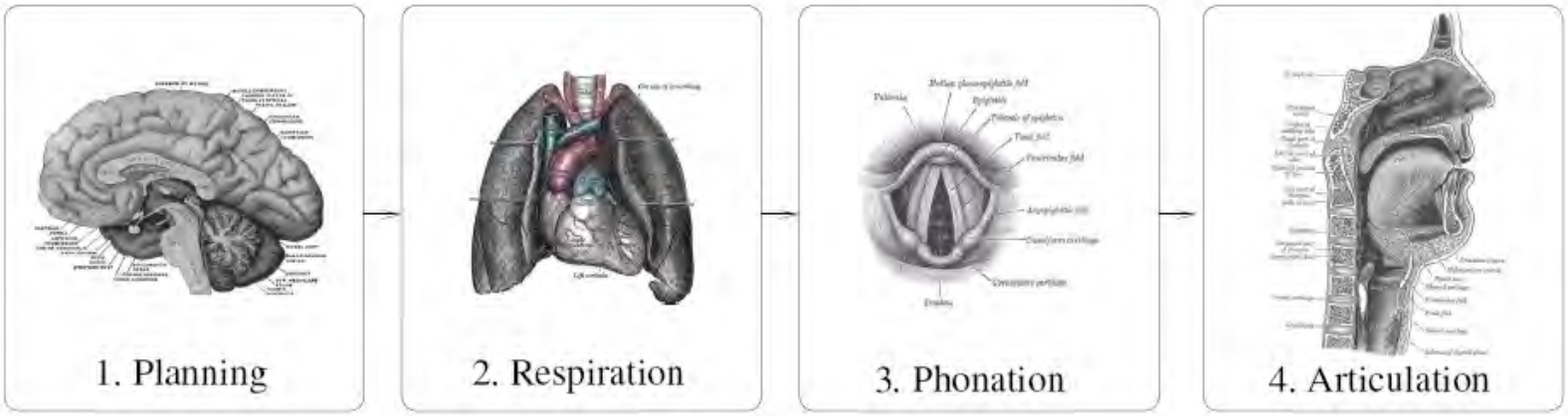
- Examples for analysis of speech pathologies and problems along the way to routine application
- May seem evident – but are often neglected
- Most examples from our work group

Outline

Pattern Recognition Lab FAU

- Motivation
- **Examples for pathologies**
- Goals of (automatic) analysis
- Some use cases
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- Summary and Take-Aways

Disorders



- Dysfunction of excitation → **voice** disorders
- Dysfunction in articulation → **speech** disorders
- Dysfunction of the neurological processing → **language** disorders

Examples for Disorders – Stuttering

- Repetition type
  - Sequences of fast contractions of the vocal cords
  - Repetitions of phonemes, syllables, words, phrases
  - Example: *t-t-t-t-table*

- Blocking type
  - Long-lasting contractions of the vocal cords
  - Pauses occurring before or within words
  - After the spasmic plosion rising tone pitch
  - Example: *t---able*

- Often combined

Nöth E., Niemann H., Haderlein T., Decher M., Eysholdt U., Rosanowski F., Wittenberg T.:
Automatic Stuttering Recognition using Hidden Markov Models, *Interspeech 2000*

Examples for Disorders – Stuttering

- Repetition type
  - Sequences of fast contractions of the vocal cords
  - Repetitions of phonemes, syllables, words, phrases
  - Example: *t-t-t-t-table*

- Blocking type
  - Long-lasting contractions of the vocal cords
  - Pauses occurring before or within words
  - After the spasmic plosion rising tone pitch
  - Example: *t---able*

- Often combined

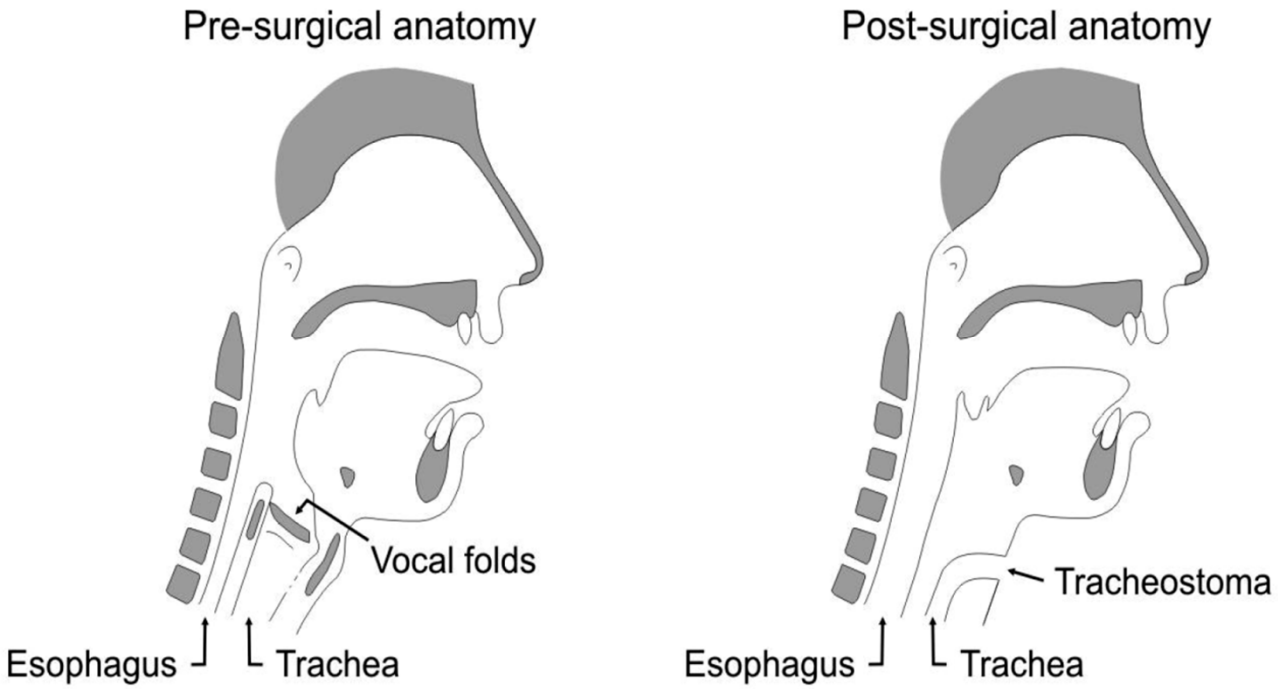
Einst stritten sich Nordwind und …

The North Wind and the Sun were

Nöth E., Niemann H., Haderlein T., Decher M., Eysholdt U., Rosanowski F., Wittenberg T.:
Automatic Stuttering Recognition using Hidden Markov Models, *Interspeech 2000*

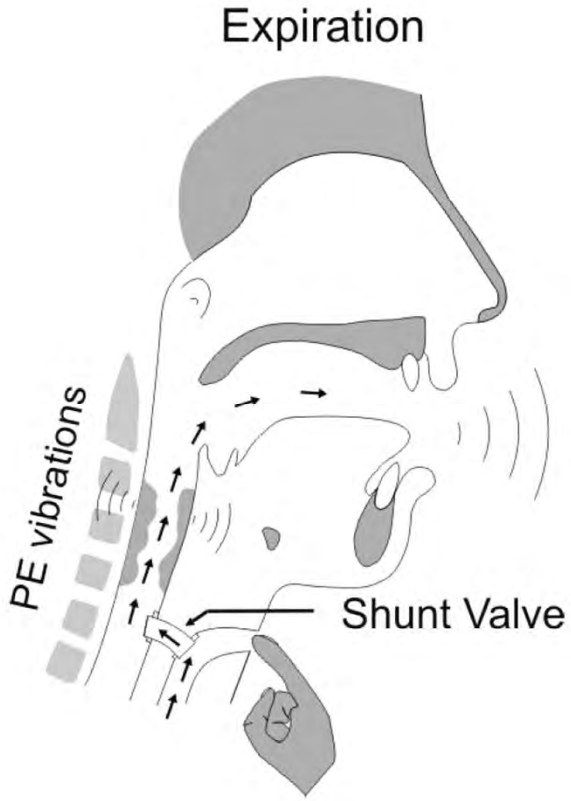
Examples for Disorders – Alaryngeal (Substitute) Voice

- Removal of the larynx due to cancer
- Breathing is detoured through a tracheostoma



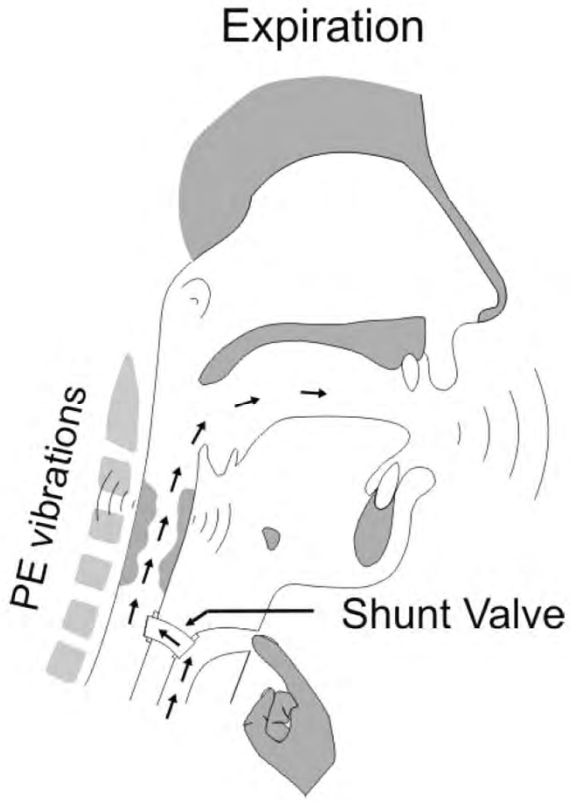
Examples for Disorders – Alaryngeal (Substitute) Voice

- Removal of the larynx due to cancer
- Breathing is detoured through a tracheostoma
- Speaking is enabled by a substitute voice



Examples for Disorders – Alaryngeal (Substitute) Voice

- Removal of the larynx due to cancer
- Breathing is detoured through a tracheostoma
- Speaking is enabled by a substitute voice



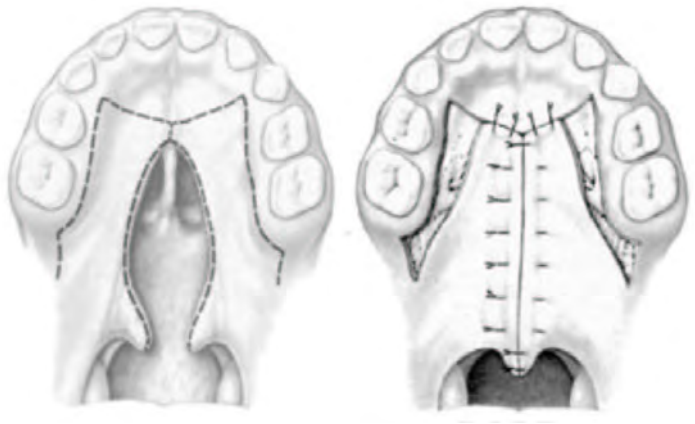
Einst stritten sich Nordwind und …

The North Wind and the Sun were

Examples for Disorders – Cleft Lip and Palate (CLP)

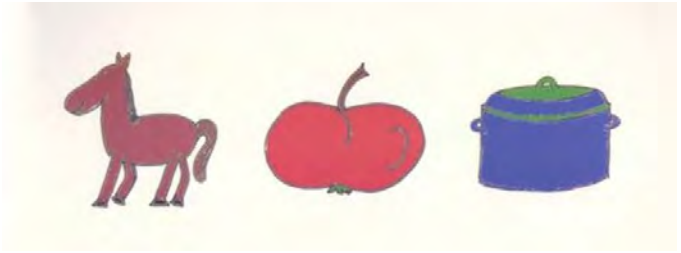
- Structural malformations of nose, throat. mouth, jaw
- Negative effects on respiration, nutrition, hearing, speaking, psychosocial competence





Examples for Disorders – Cleft Lip and Palate (CLP)

- Structural malformations of nose, throat. mouth, jaw
- Negative effects on respiration, nutrition, hearing, speaking, psychosocial competence



Es Pferd, Topf, Apfel

The horse, pot, apple

Examples for Disorders – Parkinson's Disease (PD)

- Degenerative disorder of central nervous system
- Motor and non-motor deficits
  - Motor symptoms: shaking, rigidity, problems with movement, **speech**
  - Non-motor symptoms: sleep, emotion

- Articulation and phonation affected in 70-90% of speakers with PD
- Dysarthric speech: reduced loudness, monotonic speech, breathy voice, imprecise articulation, accelerated or slowed

- Speech is early indicator of PD

Examples for Disorders – Parkinson's Disease (PD)

Examples for Disorders – Parkinson's Disease (PD)

Examples for Disorders – Parkinson's Disease (PD)

Examples for Disorders – Parkinson's Disease (PD)

Pattern Recognition Lab FAU

Clinical Evaluation

- Standard evaluation for PD:
  Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)
- 65 items across 4 sections:

(1) non-motor experiences of daily living	0 … 52
(2) motor experiences of daily living	0 … 52
(3) motor examination	0 … 132
(4) motor complications	0 … 24

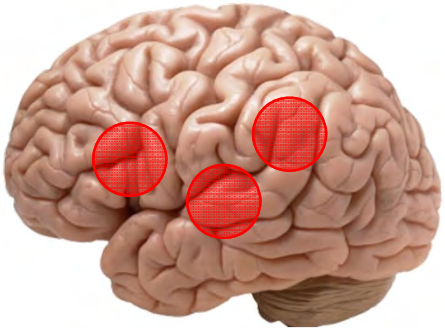
- Only 1 item in sec. (3) concerns speech 0 ... 4
- Evaluators are neurologists

Examples for Disorders – Dementia

Dementia:
- loss in memory or mental function
- interferes with daily life
- caused by physical changes in the brain

Alzheimer's disease (AD):
- progressive neurodegenerative dementia
- declines in
  - cognitive ability (e.g., memory, reasoning)
  - social ability (e.g., linguistic abilities)
  - functional capacity (e.g., executive power)

Brain regions associated with *language*
among the first affected



Examples for Disorders – Dementia

Many established tests to measure cognitive impairment (CI):

Mini-Mental State Examination (MMSE)

Patient's Name: _____ Date: _____

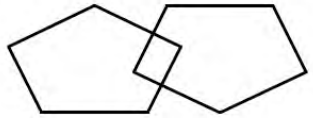
Instructions: Score one point for each correct response within each question or activity.

Maximum Score	Patient's Score	Questions
5		"What is the year? Season? Date? Day? Month?"
5		"Where are we now? State? County? Town/city? Hospital? Floor?"
3		The examiner names three unrelated objects clearly and slowly, then the instructor asks the patient to name all three of them. The patient's response is used for scoring. The examiner repeats them until patient learns all of them, if possible.
5		"I would like you to count backward from 100 by sevens." (93, 86, 79, 72, 65, …) Alternative: "Spell WORLD backwards." (D-L-R-O-W)

Examples for Disorders – Dementia

Many established tests to measure cognitive impairment (CI):

- 24 – 30 normal cog.
- 19 – 23 mild CI
- 10 – 18 moderate CI
- ≤9 severe CI

3		"Earlier I told you the names of three things. Can you tell me what those were?"
2		Show the patient two simple objects, such as a wristwatch and a pencil, and ask the patient to name them.
1		"Repeat the phrase: 'No ifs, ands, or buts.'"
3		"Take the paper in your right hand, fold it in half, and put it on the floor." (The examiner gives the patient a piece of blank paper.)
1		"Please read this and do what it says." (Written instruction is "Close your eyes.")
1		"Make up and write a sentence about anything." (This sentence must contain a noun and a verb.)
1		"Please copy this picture." (The examiner gives the patient a blank piece of paper and asks him/her to draw the symbol below. All 10 angles must be present and two must intersect.) 
30		TOTAL

Examples for Disorders and Affected Level

Pattern Recognition Lab FAU

	Voice	Articulation	Morphology Syntax	Semantics Lexicon	Pragmatics
Stuttering					
Alaryngeal voice					

Affected Can be affected * Presbyphonia

Examples for Disorders and Affected Level

Pattern Recognition Lab FAU

	Voice	Articulation	Morphology Syntax	Semantics Lexicon	Pragmatics
Sigmatism					
Cleft Lip and Palate					
Stuttering					
Alaryngeal voice					
Cancer of the Oral Cavity	*				

Affected Can be affected * Presbyphonia

Examples for Disorders and Affected Level

Pattern Recognition Lab FAU

	Voice	Articulation	Morphology Syntax	Semantics Lexicon	Pragmatics
Developmental Speech Disorders					
Sigmatism					
Cleft Lip and Palate					
Stuttering					
Alaryngeal voice					
Cancer of the Oral Cavity	*				

Affected Can be affected * Presbyphonia

Examples for Disorders and Affected Level

Pattern Recognition Lab FAU

	Voice	Articulation	Morphology Syntax	Semantics Lexicon	Pragmatics
Developmental Speech Disorders					
Sigmatism					
Cleft Lip and Palate					
Stuttering					
Alaryngeal voice					
Cancer of the Oral Cavity	*				
Dysarthria (Parkinson, Stroke, ALS)	*				

Affected Can be affected * Presbyphonia

Examples for Disorders and Affected Level

Pattern Recognition Lab FAU

	Voice	Articulation	Morphology Syntax	Semantics Lexicon	Pragmatics
Developmental Speech Disorders					
Sigmatism					
Cleft Lip and Palate					
Stuttering					
Alaryngeal voice					
Cancer of the Oral Cavity	*				
Dysarthria (Parkinson, Stroke, ALS)	*				
Aphasia & Dementia	*				

Affected Can be affected * Presbyphonia

Outline

Pattern Recognition Lab FAU

- Motivation
- Examples for pathologies
- **Goals of (automatic) analysis**
- Some use cases
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- Summary and Take-Aways

Use of Speech Technology for

Pattern Recognition Lab FAU

- Screening
  - *Does the patient show **early signs** of Alzheimer?*
- Diagnosis
  - *How **intelligible** is the patient?* (holistic impression)
  - *How strongly does the patient **nasalize**?* (distinct aspect)
- Monitoring
  - *Is there a **change in a patient´s situation**?*
- Therapy control
  - *Has the situation of the patient **improved during therapy**?*
- Comparison of therapy methods
  - ***Which therapy method** leads to the best results for a group of patients?*
- Computer-assisted therapy
  - *Did the patient perform the **exercise correctly**?*

Outline

Pattern Recognition Lab FAU

- Motivation
- Examples for pathologies
- Goals of (automatic) analysis
- **Some use cases**
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- Summary and Take-Aways

Example Use Case
Developmental Speech Disorders

- **Screening**
  - ***Does the child show developmental disorders***
- Diagnosis
  - *How intelligible is the patient?* (holistic impression)
  - *How strongly does the patient nasalize?* (distinct aspect)
- Monitoring
  - *Is there a change in a patients situation?*
- Therapy control
  - *Has the situation of the patient improved during therapy?*
- Comparison of therapy methods
  - *Which therapy method leads to the best results for a group of patients?*
- **Computer-assisted therapy**
  - ***Child-appropriate exercise App***

Developmental Speech Disorders

- Typically: Children are screened before entering elementary school
- Nursery/preschool often not mandatory
- Not all teachers & parents are conscious about Developmental Speech Disorders
- Can affect all linguistic levels
- “Standardized” “manual” tests

- Fully automated tests not feasible
- Automate existing tests
- Automatic analysis could be 2nd (objective) opinion

- If tested 1 year before elementary school, children can get speech therapy or access to special therapy apps

Developmental Speech Disorders
Phonemic Awareness

Test, e.g., syllables/words (minimal pairs) w.r.t. place of articulation or fortis/lenis production

2 different forms:

1) Child has to say “same” or “different”

- *Kragen (collar)* *tragen (to carry)*
- *Kämmen (to comb)* *kennen (to know)*

place of articulation

2) Child has to repeat the two words or syllables

Developmental Speech Disorders
Phonemic Awareness

Test, e.g., syllables/words (minimal pairs) w.r.t. place of articulation or fortis/lenis production

2 different forms:

1) Child has to say “same” or “different”

- *Kragen (collar)* *tragen (to carry)*
- *Kämmen (to comb)* *kennen (to know)*

2) Child has to repeat the two words or syllables

- *Seide (silk)* *Seite (side)*

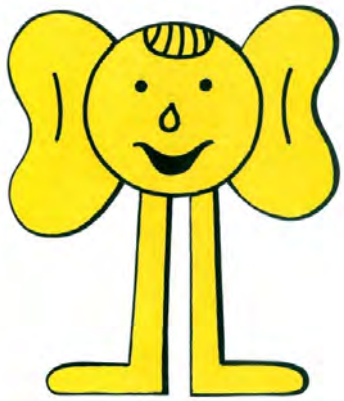
place of articulation

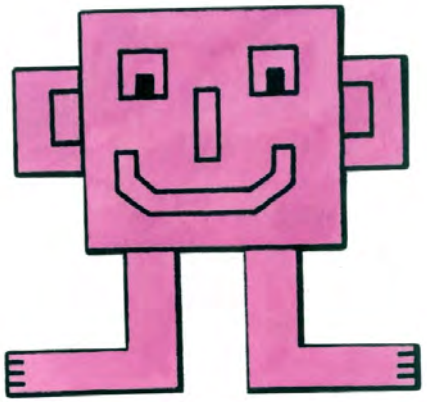
Fortis/lenis

Child repeats correctly

Developmental Speech Disorders
Phonemic Awareness

- Repeat nonsense words and sentences
- Ability to use known phones to create nonsense words
- How accurate are phones represented in the memory?
- Examples
  - *Ronterklabe*
  - *Seregropist*
  - *Glösterkeit*

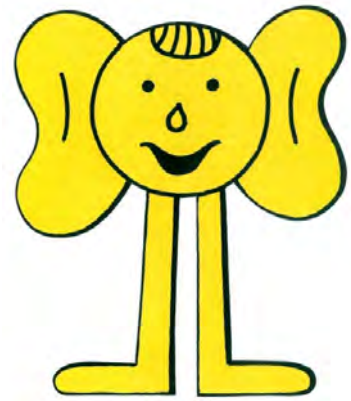


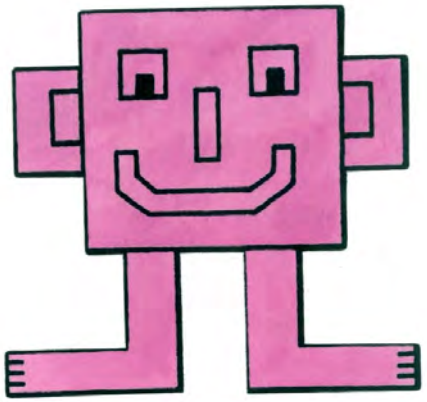


Developmental Speech Disorders
Phonemic Awareness

- Repeat nonsense words and sentences
- Ability to use known phones to create nonsense words
- How accurate are phones represented in the memory?
- Examples
  - *Ronterklabe*
  - *Seregropist*
  - *Glösterkeit*

  Glösterheit

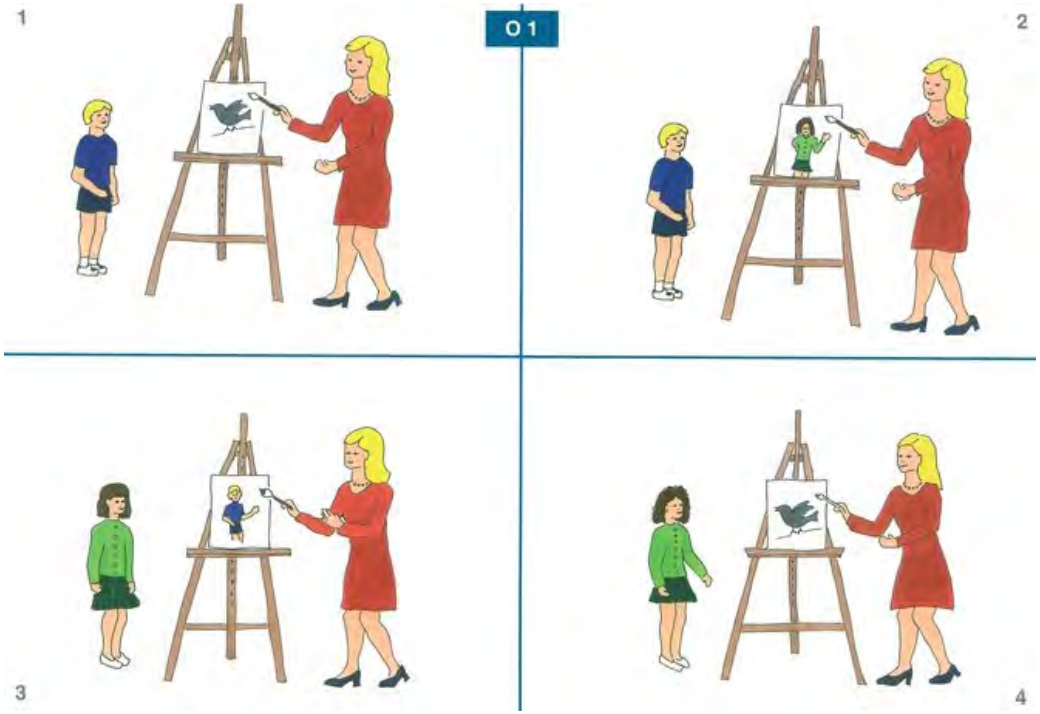




Automatic Evaluation of a Sentence Memory Test for Preschool Children → Special Session SS8, Thursday 13:30–15:30

Developmental Speech Disorders
Grammatical Understanding

Die Frau malt dem Jungen das Mädchen
(The woman is painting the girl for the boy)



Developmental Speech Disorders
Grammatical Understanding

- Child has to tell a continuous story
- Is the grammatical usage appropriate for the age?

Developmental Speech Disorders Pitfalls

Pattern Recognition Lab FAU

Automation of these tests is straightforward, but

- Equipment in preschools
- Skeptical personnel (“*My performance is being controlled*”)
- Representative reference data to train the system
  Parents (*“I don’t want my child’s voice to be recorded”*)
- Recording from other groups are not available, and even if
  “*All children in this study were native speakers of …”*
  → little knowledge about bilingual children (increasing problem)

- Training app: Automatic speech analysis is NOT the problem
  Good graphics, good feedback, good rewards, … → Need good game developer
  → Success does not depend on going from 91% recognition rate to 93% but on constant interaction with motivated preschool teachers, therapists, and game developers

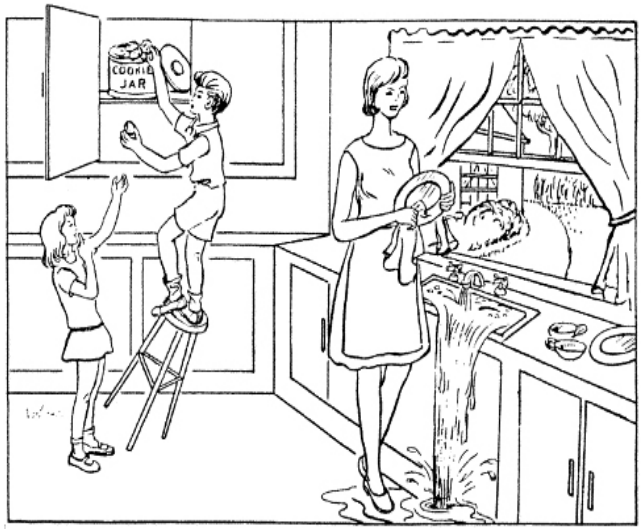
- Large amount of subjective evaluation – can that have an influence?

Example Use Case
Test for Dementia

Pattern Recognition Lab FAU

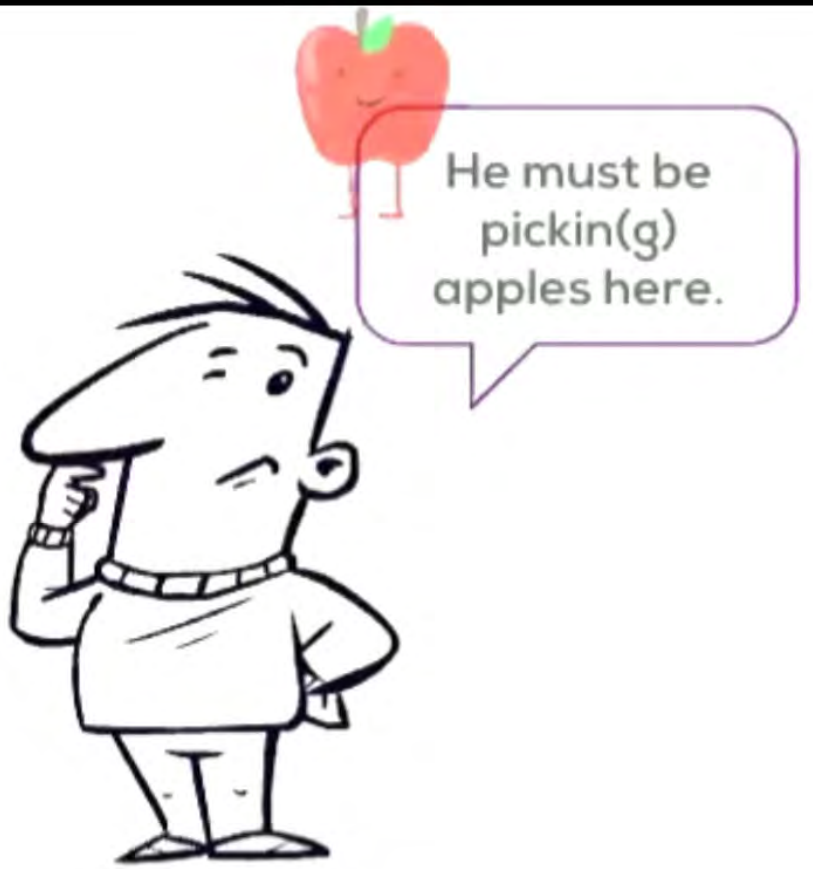
- Screening
  - *Does the child show developmental disorders*
- **Diagnosis**
  - ***Does the patient have dementia?***
- Monitoring
  - *Is there a change in a patients situation?*
- Therapy control
  - *Has the situation of the patient improved during therapy?*
- Comparison of therapy methods
  - *Which therapy method leads to the best results for a group of patients?*
- **Computer-assisted therapy**
  - ***Child-appropriate exercise App***

Diagnosis of Alzheimer Disease



- Many standard tests available, typically in interview form
- Widely used corpus: **Pitt Corpus**
- Subset of Pitt Corpus used in ADReSSo Challenge at Interspeech 2021
- 87 speakers with Alzheimer's Disease (AD) and 79 age- and sex-matched Healthy Control speakers (HC) for training
- **Cookie Theft picture** from the Boston Diagnostic Aphasia Exam; Data: speech signal, transcription, time stamps for interviewer and participant, Mini Mental State Exam (MMSE) → **degree of cognitive impairment**


COOKIE
JAR


He must be pickin(g) apples here.

Diagnosis of Alzheimer Disease

Interviewer: *Just look at the picture and tell me everything that you see.*
Participant: *Somebody's getting cookies out of the cookie jar, standing on a stool. The stool's gonna tip over. And the girl's saying. Shh don't let somebody hear. And the mother's drying dishes. The water's running out into the floor. Cups and plates are sitting on the counter. And you can look out the window and see the shrubbery. &um the lid's sideways from the cookie jar. The door is open on the cabinet. Mother apparently doesn't really notice well enough. Curtains are hanging on the window. The window is open.*
Interviewer: *That's fine. Thank you.*

Diagnosis of Alzheimer Disease

Interviewer: *Okay and there's the picture. Tell me all the action.*
Participant: *Okay he's fallin(g) off a chair.*
Interviewer: *mhm. Anything else ?*
Participant: *She's uh running the water over.*
Interviewer: *Okay.*
Participant: *Can't see anything else.*
Interviewer: *Okay anymore action ?*
Participant: *no.*
Interviewer: *Okay.*
Participant: Okay.
Interviewer: *Pardon me?*
Participant: *She's she's step in the water.*
Interviewer: *Okay. Anything else?*
Participant: No.
Interviewer: *Okay. Okay*

Diagnosis of Alzheimer Disease

Interviewer: *Just look at the picture and tell me everything that you see.*
Participant: *Somebody's getting cookies out of the cookie jar, standing on a stool. The stool's gonna tip over. And the girl's saying. Shh don't let somebody hear. And the mother's drying dishes. The water's running out into the floor. Cups and plates are sitting on the counter. And you can look out the window and see the shrubbery. &um the lid's sideways from the cookie jar. The door is open on the cabinet. Mother apparently doesn't really notice well enough. Curtains are hanging on the window. The window is open.*
Interviewer: *That's fine. Thank you.*

Interviewer: *Okay and there's the picture. Tell me all the action.*
Participant: *Okay he's fallin(g) off a chair.*
Interviewer: *mhm. Anything else ?*
Participant: *She's uh running the water over.*
Interviewer: *Okay.*
Participant: *Can't see anything else.*
Interviewer: *Okay anymore action ?*
Participant: *no.*
Interviewer: *Okay.*
Participant: Okay.
Interviewer: *Pardon me?*
Participant: *She's she's step in the water.*
Interviewer: *Okay. Anything else?*
Participant: No.
Interviewer: *Okay. Okay*

Diagnosis of Alzheimer Disease

Pattern Recognition Lab FAU

Interviewer: *Just look at the picture and tell me everything that you see.*
Participant: *Somebody's getting cookies out of the cookie jar, standing on a stool. The stool's gonna tip over. And the girl's saying. Shh don't let somebody hear. And the mother's drying dishes. The water's running out into the floor. Cups and plat[illegible] are sitting on the counter. And w[illegible] look out the window and [illegible] shrubbery. &um[illegible] [illegible]s from the cooki[illegible] [illegible] open on the c[illegible] [illegible]pparently doesn't really [illegible] enough. Curtains are hanging [illegible]he window. The window is open.*
Interviewer: *That's fine. Thank you.*

Interviewer: *Okay and there's the pictur[illegible] me all the action.*
Participant: *Okay he's fa[illegible]*
Interviewer: *mhm[illegible]*
Participa[illegible] *[illegible]ing the water over.*
In[illegible]
[illegible]: *Can't see anything else.*
Interviewer: *Okay anymore action ?*
Participant: *no.*
Interviewer: *Okay.*
Participant: Okay.
Interviewer: *Pardon me?*
Participant: *She's she's step in the water.*
Interviewer: *Okay. Anything else?*
Participant: No.
Interviewer: *Okay. Okay*

Which one is the AD participant and which one the Healthy Control?

Diagnosis of Alzheimer Disease

Pattern Recognition Lab FAU

Paper

Speaker	Source	Classification	MMSE Prediction RMSE	MMSE Prediction ρ
Participant Only	Acoustic	75 %	6.40	0.42
	Linguistic	78 %	5.14	0.68
	Fusion	83 %	4.87	0.70

5-fold CV on training data. Classification: accuracy. MMSE prediction: Root Mean Squared Error (RMSE). Spearman's Correlation (ρ) with MMSE

Diagnosis of Alzheimer Disease

Pattern Recognition Lab FAU

Speaker	Source	Classification	MMSE Prediction	
			RMSE	ρ
Complete Recording	Acoustic	**81 %**	6.27	0.51
	Linguistic	**81 %**	5.16	0.68
	Fusion	**86 %**	**4.86**	**0.72**
Participant Only	Acoustic	75 %	6.40	0.42
	Linguistic	78 %	5.14	0.68
	Fusion	83 %	4.87	0.70

Paper

5-fold CV on training data. Classification: accuracy. MMSE prediction: Root Mean Squared Error (RMSE). Spearman's Correlation (ρ) with MMSE

Diagnosis of Alzheimer Disease

Pattern Recognition Lab FAU

Speaker	Source	Classification	MMSE Prediction	
			RMSE	ρ
Complete Recording	Acoustic	**81 %**	6.27	0.51
	Linguistic	**81 %**	5.16	0.68
	Fusion	**86 %**	**4.86**	**0.72**
Participant Only	Acoustic	75 %	6.40	0.42
	Linguistic	78 %	5.14	0.68
	Fusion	83 %	4.87	0.70
Interviewer Only	Acoustic	78 %	6.02	0.51
	Linguistic	71 %	10.89	0.37
	Fusion	77 %	5.65	0.53

Paper

5-fold CV on training data. Classification: accuracy. MMSE prediction: Root Mean Squared Error (RMSE). Spearman's Correlation (ρ) with MMSE

Diagnosis of Alzheimer Disease Pitfalls

Pattern Recognition Lab FAU

- Many standard tests are in interview form (necessary)
- Interviewer‘s behaviour and dialogue contributions
  - Big influence on tested person
  - Sometimes already reveal result
- Often identity of interviewer is not documented
- Many other external factors, e.g., microphone, technical problems, ...
- ADReSSo data contain recordings with mild, medium and severe AD
  → relevant for diagnosis but not so much for screening
- The amount of training data is very small – especially with the rise of deep learning
  Is there open source data?

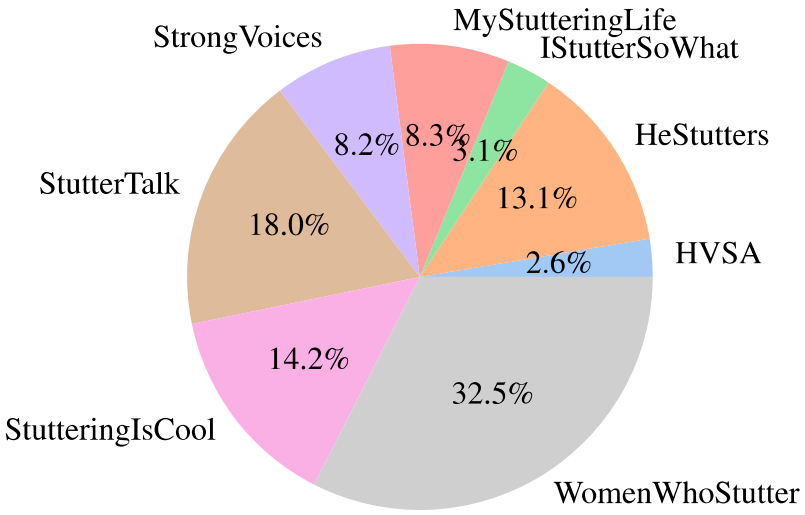
Example Use Case
Test for Stuttering

Pattern Recognition Lab FAU

- Screening
  - *Does the patient show early signs of depression?*
- **Diagnosis**
  - ***Does the participant stutter and what kind of stuttering is it?***
- Therapy control
  - *Has the situation of the patient improved during therapy?*
- Comparison of therapy methods
  - *Which therapy method leads to the best results for a group of patients?*
- Monitoring
  - *Is there a change in a patients situation?*
- Computer-assisted therapy
  - *Did the patient perform the exercise correctly?*

Open Source Stutter Data SEP-28K

- ~28000 3-sec long clips
- Labeled with five types of stuttering and some meta labels
- Biggest resource of stuttered data (~ 23h)
- Podcasts data from 385 unique episodes of 8 stuttering related podcasts



SUPER!

Data in the wild

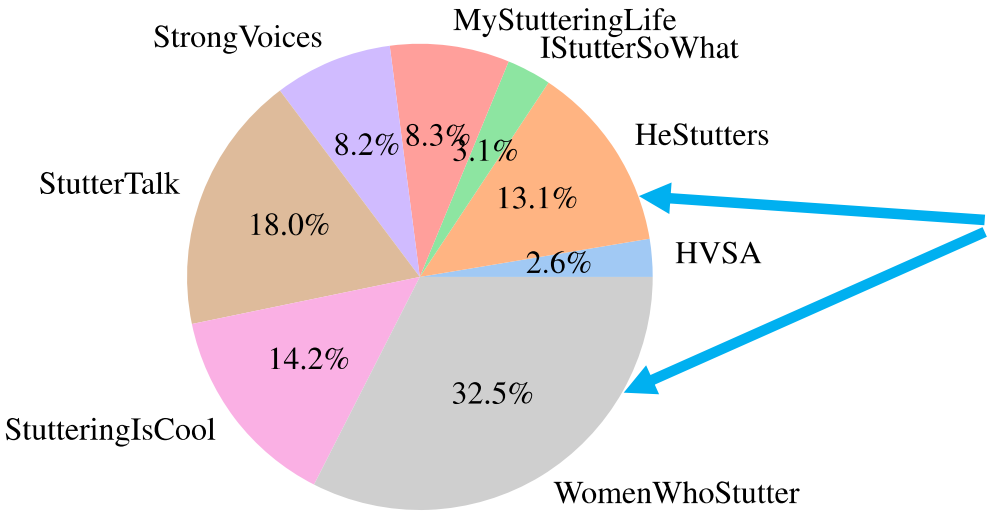
High prior of stuttering due to selection of broadcasts (~50% of the clips have stuttering)

No “standard” partition into Train, Dev, Test provided in Lea, C. et al. *SEP-28k: A Dataset for Stuttering Event Detection from Podcasts with People Who Stutter*. In *ICASSP 2021*

→ Closer look at the corpus

Open Source Stutter Data SEP-28K

- ~28000 3-sec long clips
- Labeled with five types of stuttering and some meta labels
- Biggest resource of stuttered data (~ 23h)
- Podcasts data from 385 unique episodes of stuttering related podcasts



- Sessions from 2 podcasts > 45% of the data
- No speaker labels
- Unbalanced stutter-types across podcasts

Open Source Stutter Data SEP-28K

Pattern Recognition Lab FAU

- Meta data per clip: podcast name and episode number
  → manual lookup of host name, guest names, gender, # of guests
- "Speaker" labels using automatic clustering → identification of host (dominating speaker)
- 4 hosts make up for 59% of all 28k clips!
- Rest from ~ 500 speakers

→ Created 2 own splits, and made them open source

→ Much better balance

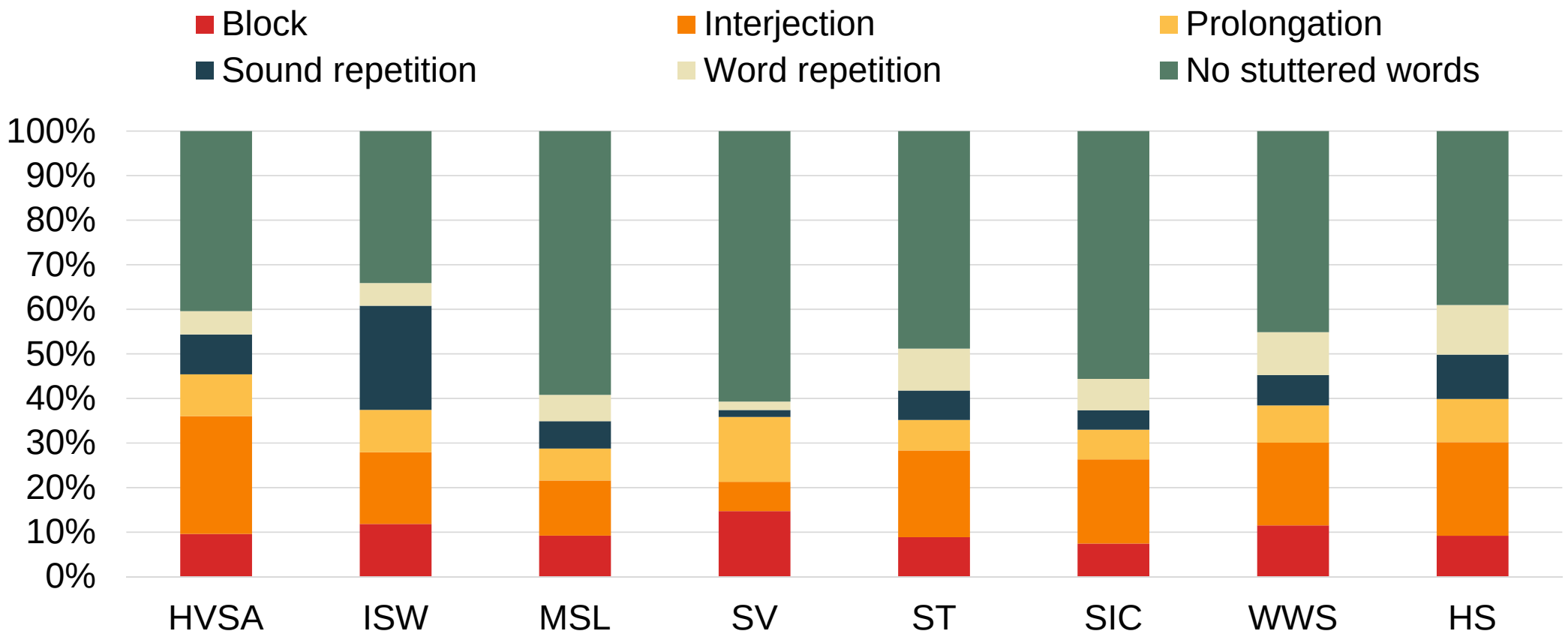
→ Modulo clustering errors no speaker overlap in train, development, test & 5-fold CV

Paper

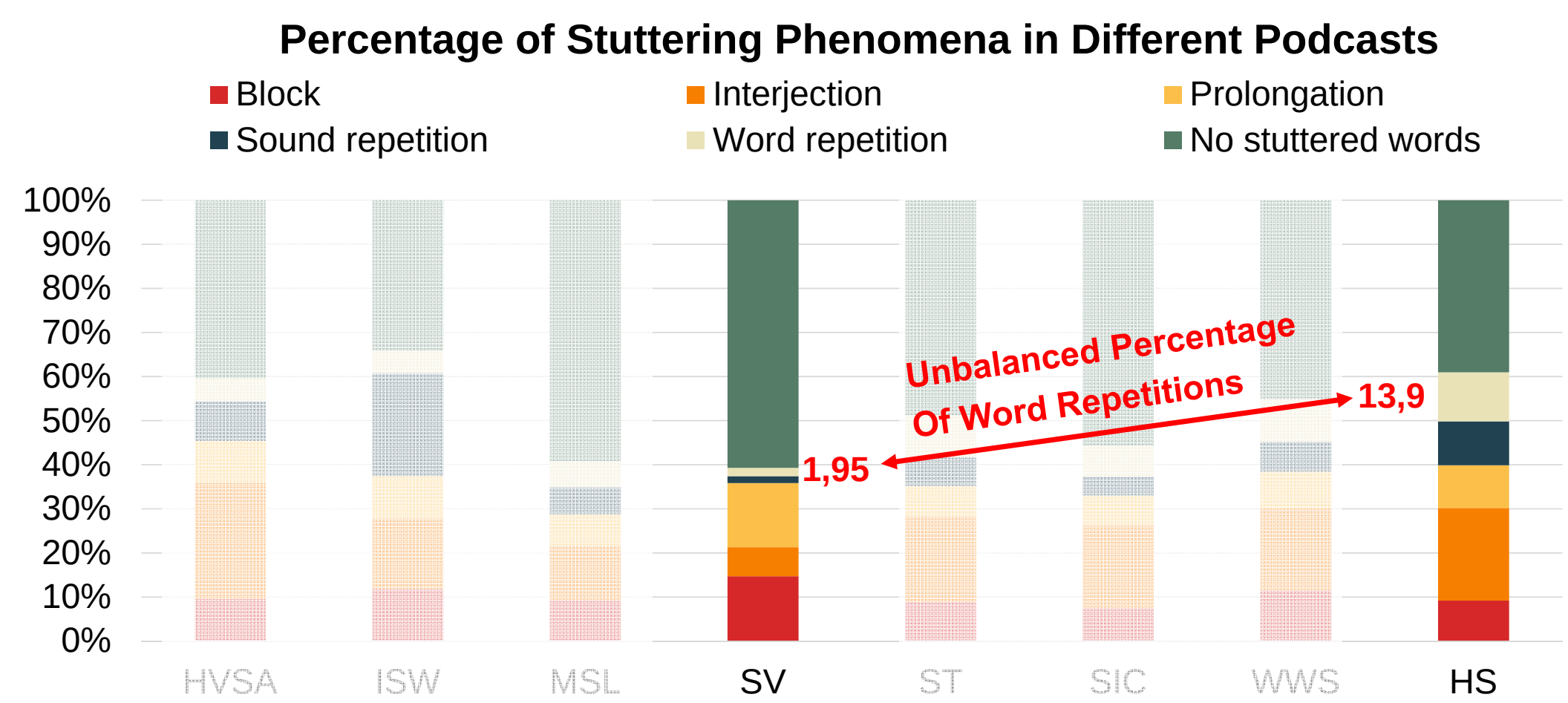
Open Source Stutter Data SEP-28K

Pattern Recognition Lab FAU

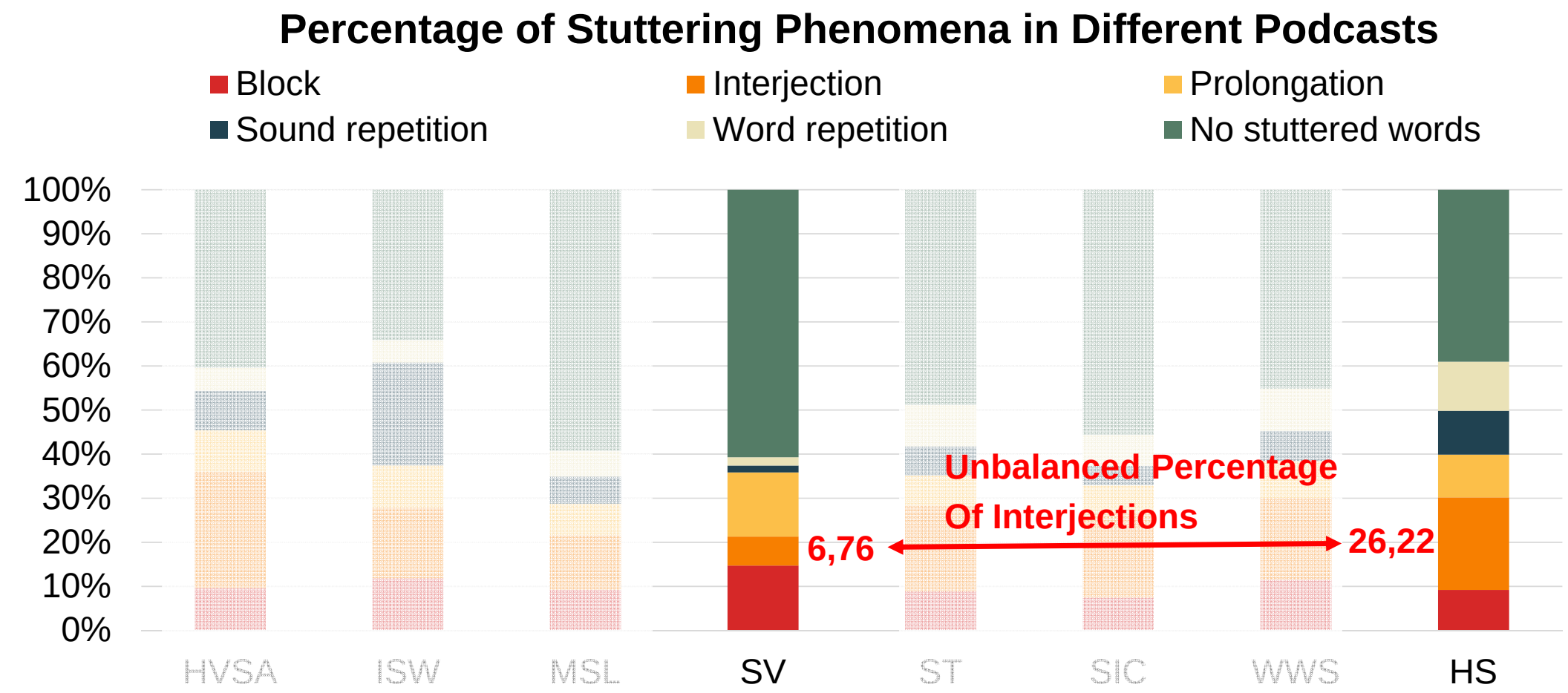
Percentage of Stuttering Phenomena in Different Podcasts



Open Source Stutter Data SEP-28K



Open Source Stutter Data SEP-28K



Open Source Stutter Data SEP-28K

Pattern Recognition Lab FAU

	SEP-28k-E train	SEP-28k-E dev	SEP-28k-E test	SEP-12k
Block	11.57 %	12.84 %	12.01 %	12.48 %
Interjection	22.94 %	18.79 %	19.51 %	19.10 %
Prolongation	9.87 %	10.07 %	10.13 %	10.15 %
Sound repetition	8.13 %	10.40 %	6.69 %	8.57 %
Word repetitions	9.98 %	8.79 %	10.48 %	9.67 %
No stuttered words	56.92 %	55.78 %	58.15 %	56.81 %
Total #	15213	6402	6562	12804

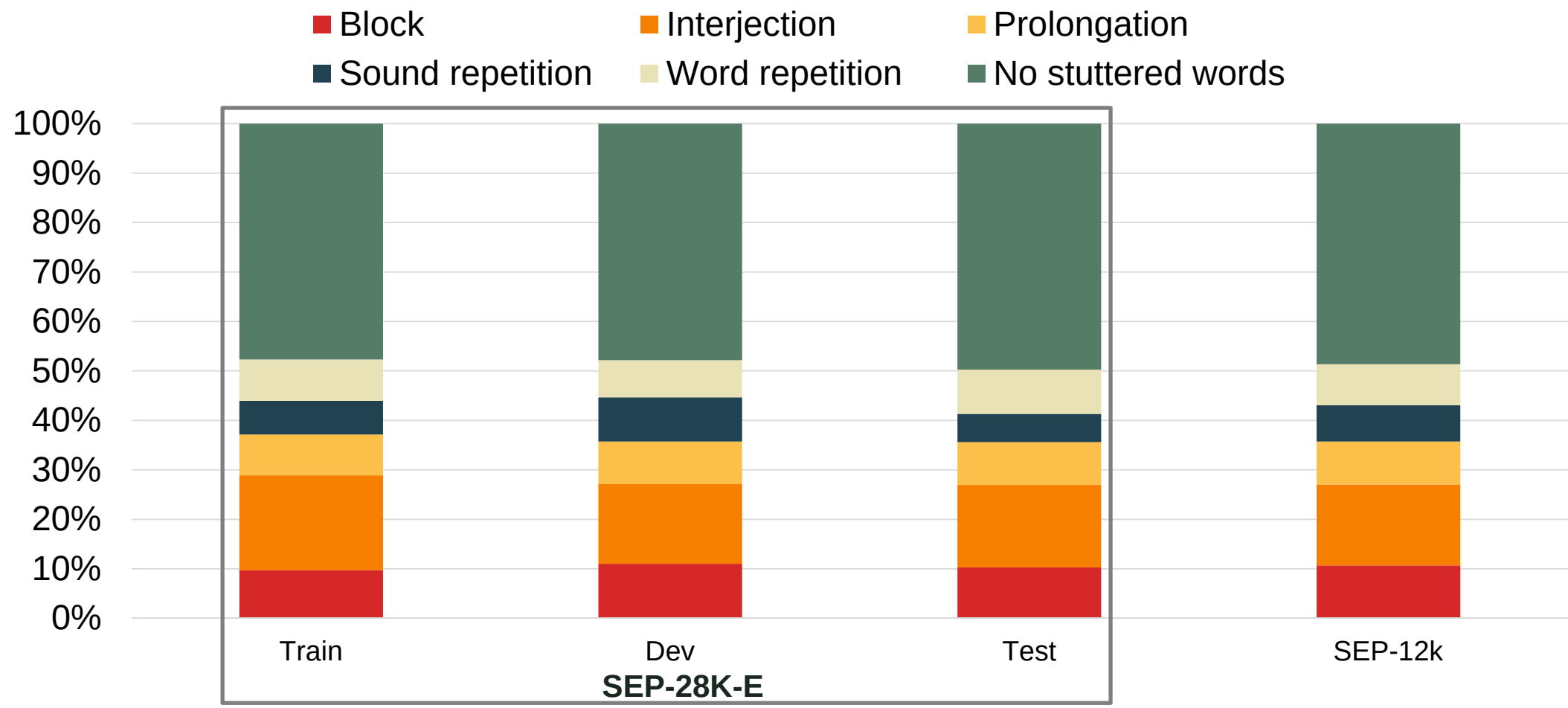
4 most frequent speakers

Rest of the speakers

All but the 4 most frequent speakers 5-fold CV

- No overlap of speakers in train, dev, test & and across the 5 folds
- Better balance between the stutter types

Open Source Stutter Data
SEP-28K



Open Source Stutter Data
SEP-28K - Pitfalls

- Open source datasets might come with many undocumented unknowns
  → “Listen” to your data!
- “In the wild” data can contain many new challenges, e.g., missing speaker ID, strongly dominating speakers
- Sometimes, no standard division into training / development / test is available
  → Can have large influence on results
  → makes direct comparisons difficult

- Can we take advantage of the huge corpora on the internet?

Example Use Case
Speech Analysis of PD Patients

- Screening
  - *Does the patient show early signs of depression?*
- **Diagnosis**
  - ***How strong is the dysarthria of a patient***
- **Therapy control**
  - ***Has the situation of the patient improved during therapy?***
- Comparison of therapy methods
  - *Which therapy method leads to the best results for a group of patients?*
- Monitoring
  - *Is there a change in a patients situation?*
- Computer-assisted therapy
  - *Did the patient perform the exercise correctly?*

Example Use Case
Speech Analysis of PD Patients

Pattern Recognition Lab FAU

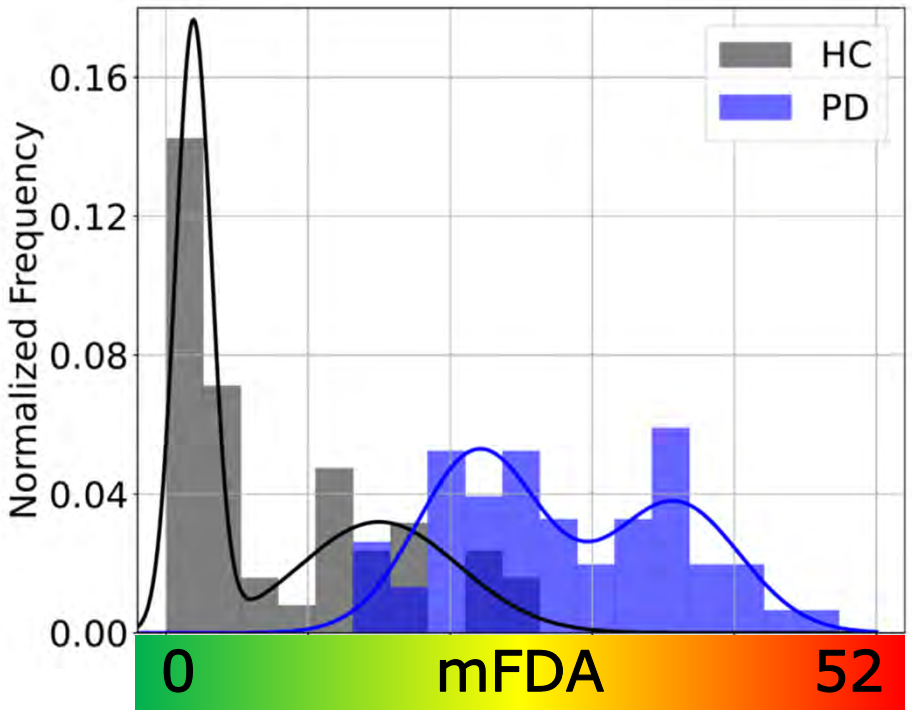
- We want to analyze PD speech
- Standard UPDRS evaluation only contains 1 speech item
  → modified Frenchay Dysarthria Assessment (mFDA, based on speech signal only)
- mFDA contains 6 subtests that test impediment of:

Respiration	0 … 8
Lips	0 … 8
Palate	0 … 8
Larynx	0 … 16
Tongue	0 … 8
Intelligibility	0 … 4

  → 0 (absolutely normal) … 52 (extremely impaired)
  → overall and distinct aspects

Example Use Case
Phonetic Analysis of PD Patients

- 58 PD & 86 HC
- Each participant‘s speech was evaluated w.r.t. mFDA
- Different tasks (diadokinetic, read & spontaneous speech)



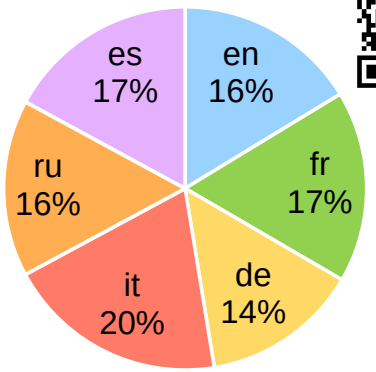
Distribution of m-FDA scores across PD and HC

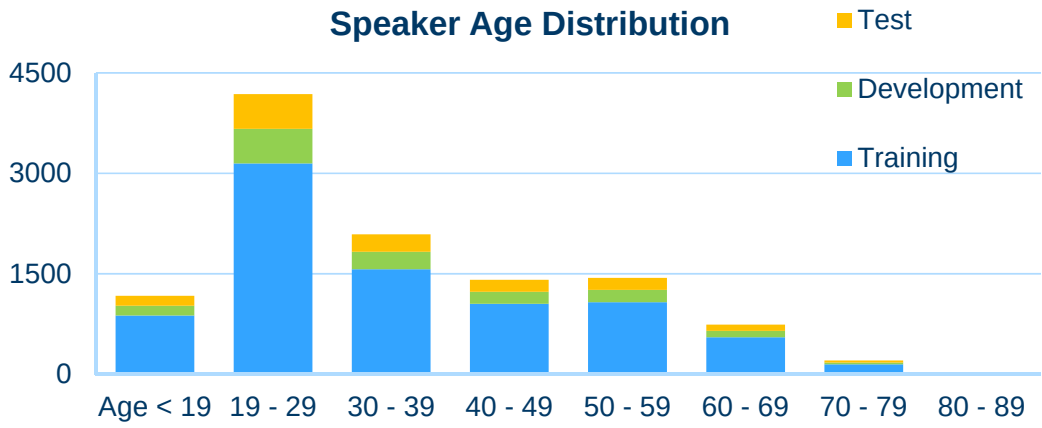
Example Use Case
Speech Analysis of PD Patients

- Created Common Phone (Subset of Common Voice):
  - Open source
  - 6 languages
  - Eliminates imbalances
  - Enriched annotation (only logged-in users)
  - Gender-, age- and language-balanced
- Provide reference dataset for:
  - Robust acoustic modelling
  - Testing in real-world environment
  - > 11k speakers, > 116 h

- Fine tuned Wav2Vec 2.0 Base & Large (94 M & 315 M parameters)
- 101 multilingual phones
- Used phone recognizer on PD & HC (has not seen PD)

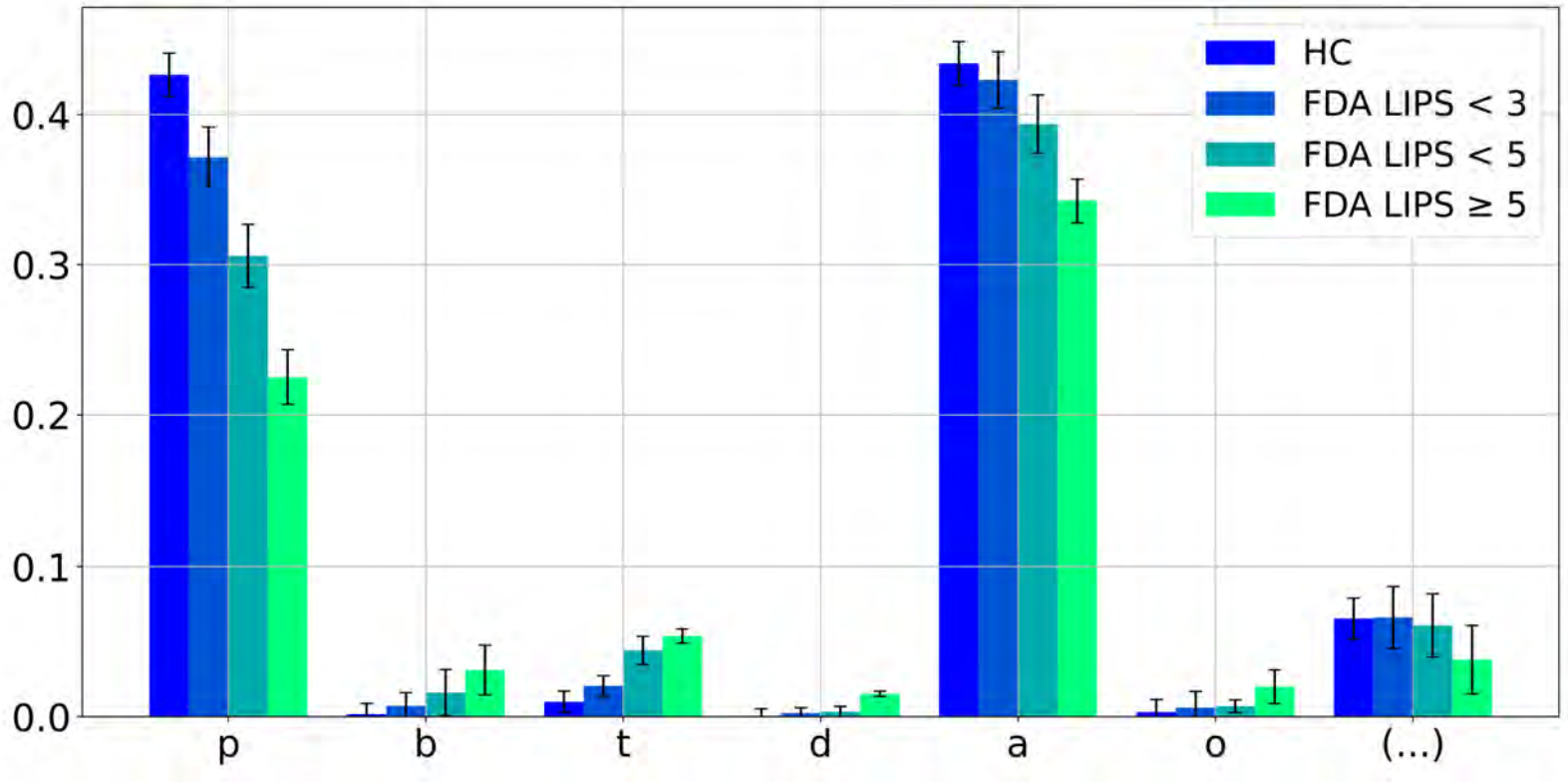
Paper





Phonetic Footprints

Rapid repetition of syllable “*pa*”



*Phonetic footprint for speech exercise “**pa**”*

Example Use Case Pitfalls

Pattern Recognition Lab FAU

- Open-source data can contain imbalances
  - English is often over-represented
  - In some “smaller” languages: dominating speaker
- Multi-lingual phone recognizer better for analyzing “non-standard” speech
- Phoneme posteriors are easier to explain than changes in 315 M parameters

Outline

- Motivation
- Examples for pathologies
- Goals of (automatic) analysis
- Some use cases
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- Summary and Take-Aways

Pitfalls – Ethical and Legal Issues

Pattern Recognition Lab FAU

- Speech is a bio-signal that can reveal the **identity and traits** of the speaker
  → agreement to record, store, & share
- The pathologies can concern children or elderly placed under **guardianship**
  → agreement might be complicated; **legal restrictions**
- Patients are no laboratory animals
  → cannot easily test, e.g., **influence of medication** on speech with Parkinson
  → might twist the truth (*When did you take your medication? – Right before I came)*
  → have to be respected when they are too **tired** or overwhelmed by the task
- Research results might be too late for patient → **motivation**
- Patients might be ashamed of their performance → relevant for **longitudinal data**
- Data might contain very **private information**, especially spontaneous speech
  → sharing is problematic

Pitfalls – Labelers

- Labeling a pathology can be difficult
  → misdiagnoses are frequent especially in early stages
- Even more difficult: degree of pathology
  → necessary for therapy control & monitoring

Remember for PD:

- The medical expert labels the degree of the pathology, e.g. PD, not of the speech impediment
- Evaluator is a neurologist not a speech therapist
- Evaluator might never have seen the patient or months ago
- Standard evaluation for PD: UPDRS, where only 1 item concerns speech

- Dysarthria assessment often not available
  → we introduced the mFDA together with experienced speech therapists

Pitfalls – Labelers

- Labeling a pathology can be difficult
  → misdiagnoses are frequent especially in early stages
- Even more difficult: degree of pathology
  → necessary for therapy control & monitoring
- Detailed labeling is very expensive → not available for most data
- Strong inter-labeler disagreement possible
- Therapy control and monitoring: We need **longitudinal** data, but medical evaluators often change, last evaluation several month ago → documentation based on clinical record
- “professional who knows the patient” vs. “professional” vs. “naïve” labelers?
- How to combine several labelers?

Pitfalls – Metadata

- Speech pathologies are rare → Take all speakers with pathology and collect matched controls
- Whatever metadata are not documented are lost for interpretation, e.g.,
  - level of education
  - local dialect
  - medication intake
  - depression
  - smoker y/n
  - microphone
  - …
- Makes interpretation and comparison difficult

Pitfalls – Intra- & Inter-Corpora Inconsistencies

- Many corpora consist of several tasks to cover different communication situations
  read sentences, read paragraphs, picture descriptions, naming tasks, diadokinetic tasks, …
- Some exercises of some patients can be missing
  - Patient might have gotten tired
  - Surgeon: *“If the patient says, he forgot his glasses, then he is probably an analphabetic”*
  - Technical problems
- Some patients were in On-state (under medication), some in Off-state (no medication in the last 6 hours), not documented
- The HC were recorded in a different room than the patients, not documented
- Patients were instructed differently:
  *“Say pa-**ta**-ka as fast and as long as you can”* vs. *“Repeat **pa**-ta-ka for about 3 seconds”*
- Sub-tasks might be missing (different protocols at various sites)

- Difficult to combine corpora

Pitfalls – Open Source Data

Pattern Recognition Lab FAU

Some open source data lack necessary documentation on

- Intra-corpus inconsistencies
- Speaker ID
- Train/Development/Test partition
- background noise
- …

- **Remember**: Corpora are collected with an application in mind, e.g., diagnosis vs. early detection vs. pre-clinical

Pitfalls – Explainability

Pattern Recognition Lab FAU

- A simple classification *“The patient has improved”* is not sufficient
- The therapist wants a medical/linguistic explanation, e.g.,
  “loudness & pitch variation has increased”
- The patient wants encouragement and details of what improved
- Current state-of-the-art classifiers have millions of parameters and lack the ability to explain their decision

Outline

Pattern Recognition Lab FAU

- Motivation
- Examples for pathologies
- Goals of (automatic) analysis
- Some use cases
- Pitfalls
  - Ethical issues
  - The labelers
  - The patients
  - Open source software and data
  - Interpretability of the data
- **Summary and Take-Aways**

Summary and Take-Aways for the Community

- Speech is a biosignal and carries important information about speaker properties including pathologies
- Pathological data are hard to get (typical: pathologic speakers and then matched controls)
- Whatever is not matched might make interpretation faulty
- Pathological data can be unrealistic for real applications
  (*we excluded patients with ...*) or (*all patients were ...*)
- Pathological data are hard to combine
- We need standardized metadata-forms for pathologic corpora
- Longitudinal data are very hard to get → Need for a multi-site collection effort with standardized protocols for the different applications

Summary and Take-Aways for You as a Researcher

Pattern Recognition Lab FAU

- Open source data are sometimes missing detailed documentation
  → Listen to some examples, make (automatic) consistency checks, ...
- Have a good look at outliers
  (strongly affected performing well and weakly affected performing bad)
- Inter- / intra-corpus inconsistencies can influence choice of algorithm, e.g., early fusion
- Look for interpretability, e.g., use interpretable features like
  prosodic features if the medical textbook says monopitch is a marker, phonological features, ...
- Can you explain your results to a neurologist or speech therapist?
- Can you explain your results to a patient?

Friedrich-Alexander-Universität
Technische Fakultät

FAU

Thank you for your attention

Questions?


INTERSPEECH 2024