VOCAL AND SPEECH BIOMARKERS OF SLEEPINESS AND PSYCHIATRIC DISORDERS

Vincent P. MARTIN

October 27, 2022

HELLO!

I would be happy to explain more over coffee (= I have extra slides)

I am Vincent P. MARTIN

- **Ph.D. (2022)** « Voice biomarkers of sleepiness »,
  Université de Bordeaux
  J.L. Rouas (LaBRI) & P. Philip (SANPSY/CHU)
- **Eng. Degree (2018)**
  Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA)
- **DIU Philosophy of psychiatry (2021)**
  Université de Bordeaux

vincent.martin@labri.fr
@V_P_Martin
Vincent-P-Martin

SLEEPINESS AND PSYCHIATRIC DISORDERS
PUBLIC HEALTH PROBLEMS

Clinicians' needs :

- High prevalence of sleepiness and Ψ disorders
- Inter-consultations follow-up
- Symptoms expression outside the hospital env.

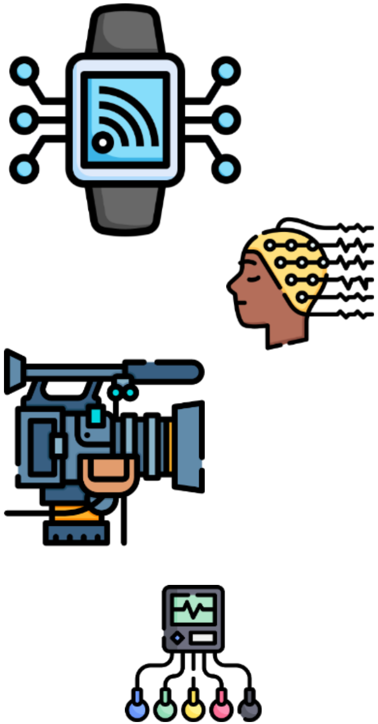
→ **Ecological* Momentary Assessment (EMA)**

- **Regular** and **ecological** measurement of symptoms

KANOPÉE

**ecological* = in the usual living conditions of the patients

SPEECH
A PROMISING MEASUREMENT TOOL



- "Physiological" measurement
- Non invasive / passive
- Few calibration / computational ressources
- Already implemented in smartphones
  - **80%** de la pop. mondiale

→ Usefull for EMA

1.
State of the art

What is the community focused on ?

STATE OF THE ART

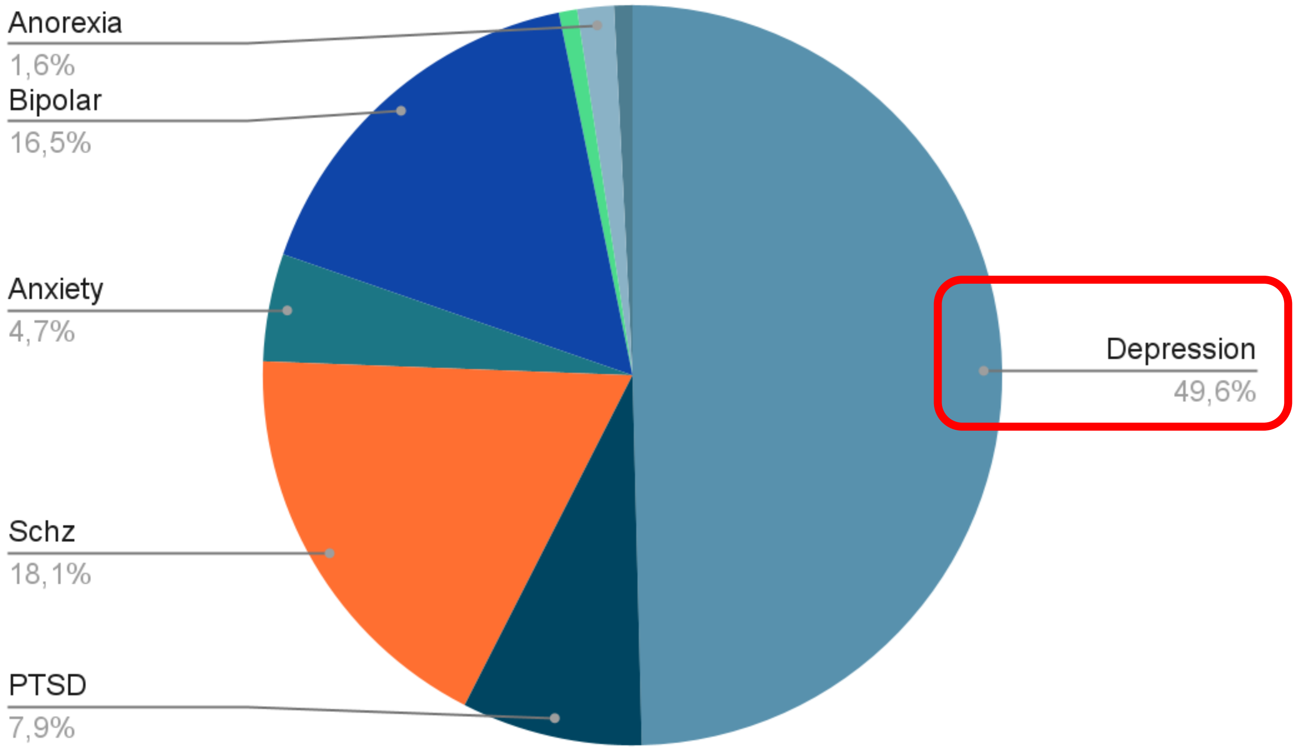
- Low et al. 2020,
  « Automated assessment of psychiatric disorders using speech: A systematic review »,
  Laryngoscope Investigative Otolaryngology

STATE OF THE ART Low et al. 2020
METHOD

- Google Scholar
- 2009-2019
- **127** **studies**

Supplementary data available online!

STATE OF THE ART Low et al. 2020
RESULTS



STATE OF THE ART Low et al. 2020
RESULTS

Label

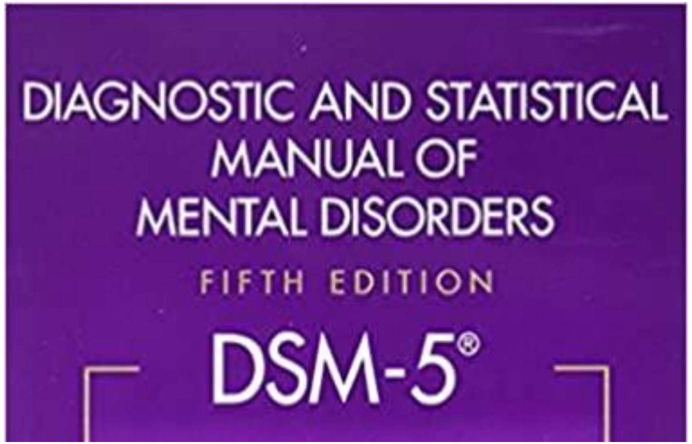
- Questionnaires (ex. PHQ9)

		Not at all	Several days	More than half the days	Nearly every day
1.	Little interest or pleasure in doing things	0	1	2	3
2.	Feeling down, depressed, or hopeless	0	1	2	3
3.	Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4.	Feeling tired or having little energy	0	1	2	3
5.	Poor appetite or overeating	0	1	2	3
6.	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7.	Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8.	Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9.	Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

STATE OF THE ART Low et al. 2020
RESULTS

Label

- Questionnaires (ex. PHQ9)
- Classification (e.g., DSM or ICD)



Major Depressive Disorder

Diagnostic Criteria

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

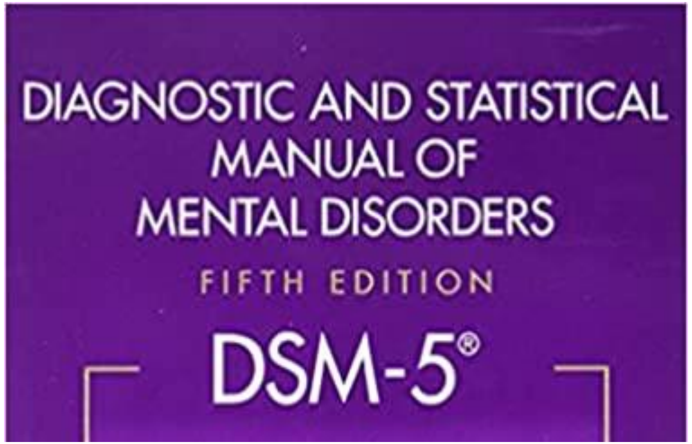
Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

STATE OF THE ART (Low et al. 2020)
RESULTS

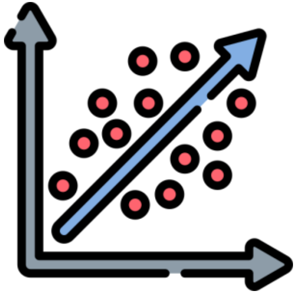
Label

- Questionnaires (ex. PHQ9)
- Classification (e.g., DSM or ICD)



Tasks

- **diagnostic**: binary classification
- **severity estimation**: regression with score



What do clinicians and patients need?

WHAT DO CLINICIANS AND PATIENTS NEED?
ACCORDING TO **SPEECH/ML ENGINEERS**

"There is an **urgency** to **objectively diagnose**, monitor over time, and provide evidence-based interventions for individuals with mental illnesses"
Low et al. 2020

"Gold-standard diagnostic and assessment tools for depression and suicidality remain rooted, almost exclusively, on the **opinion of individual clinicians** risking a range of **subjective biases**. […] Currently there is no **objective measure**, with **clinical utility**, for either depression or suicidality"
Cummins et al. 2015

No.

WHAT DO CLINICIANS AND PATIENTS NEED? ACCORDING TO **SPEECH/ML ENGINEERS**

How Does Comparison With Artificial Intelligence Shed Light on the Way Clinicians Reason? A Cross-Talk Perspective

Vincent P. Martin[1,2], Jean-Luc Rouas[1], Pierre Philip[2,3], Pierre Fourneret[4], Jean-Arthur Micoulaud-Franchi[2,3] and Christophe Gauld[4,5]*

WHAT DO CLINICIANS AND PATIENTS NEED?
ACCORDING TO **CLINICIANS**

"the main aim of the psychiatric science **is not classification** as an end in itself but rather **identification of causes** and **interventions**"
Keneth Kendler, 2012

« […] one of its most important goal is to **facilitate communication among clinicians, researchers, administrators and patients** […] by establishing a common language."
Derek Bolton, 2012

"[…] **classification in itself is** less important than often supposed to be, and **less important than other tasks**."
Derek Bolton, 2012

PITFALLS OF DIAGNOSTIC CRITERIA

Diagnosis	Symptoms
Time dependent *e.g. DSM IV, DSM 5, …*	Stable through time
Cultural dependent *e.g. Hikikomori*	Independent from culture
Heterogeneous	Homogeneous
Symptoms → Syndromes → Diagnostic	
-	Mechanistic explanation
-	Necessary for diff. diag and prog.

What do we do now?

We estimate ***symptoms***

Symptoms → Syndromes → Diagnostic

SYMPTOMS

SYMPTOMS vs. DIAGNOSIS

Diagnosis	Symptoms
Time dependent *e.g. DSM IV, DSM 5, …*	**Stable** through time
Cultural dependent *e.g. Hikikomori*	**Independent** from culture
Heterogeneous	Homogeneous

Example :

Major Depressive Disorder

Diagnostic Criteria

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

Example :
sleepiness

*Is it possible to use **voice/speech** as a measuring tool of **excessive sleepiness** for the follow-up of sleep disorders **patients**?*

WHAT DOES 'BEING SLEEPY' MEAN?
AND HOW TO MEASURE IT

- Sleepiness =
  - ✗ Fatigue?
  - ✗ Performances?

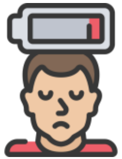


Table 1
Examples of some words used to describe fatigue, sleepiness, or both

Fatigued	Sleepy	Either or Both
Beat	Crashing	Exhausted
Languor	Drowsy	Burned out
Lassitude	Fading	Bushed
Lethargic	Groggy	Gassed
Listless	Narcotized	Pooped
Knackered	Heavy-headed	Played-out
Sluggish	Punchy	Tired
Weariness	Gorked	Tuckered-out
Whipped	Yawny	Wiped
Zoned	Slap happy	Zonked

Hirshkowitz 2013

- **Subjective** sleepiness
  - ✓ Long-term, e.g. measured by the **Epworth Sleepiness Scale**

ESS
Epworth Sleepiness Scale

TABLE 1. *The Epworth sleepiness scale*

THE EPWORTH SLEEPINESS SCALE

Name: ____________________

Today's date: ____________ Your age (years): ________

Your sex (male = M; female = F): ____________

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently try to work out how they would have affected you. Use the following scale to choose the *most appropriate number* for each situation:

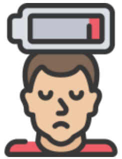
0 = would *never* doze
1 = *slight* chance of dozing
2 = *moderate* change of dozing
3 = *high* chance of dozing

Situation	Chance of dozing
Sitting and reading	____
Watching TV	____
Sitting, inactive in a public place (e.g. a theater or a meeting)	____
As a passenger in a car for an hour without a break	____
Lying down to rest in the afternoon when circumstances permit	____
Sitting and talking to someone	____
Sitting quietly after a lunch without alcohol	____
In a car, while stopped for a few minutes in the traffic	____

Thank you for your cooperation

WHAT DOES 'BEING SLEEPY' MEAN?
AND HOW TO MEASURE IT

- Sleepiness =
  - Fatigue?
  - Performances?



- **Subjective** sleepiness
  - Long-term, e.g. measured by the **Epworth Sleepiness Scale**
  - Short-term, e.g. measured by the **Karolinska Sleepiness Scale**

Table 1
Examples of some words used to describe fatigue, sleepiness, or both

Fatigued	Sleepy	Either or Both
Beat	Crashing	Exhausted
Languor	Drowsy	Burned out
Lassitude	Fading	Bushed
Lethargic	Groggy	Gassed
Listless	Narcotized	Pooped
Knackered	Heavy-headed	Played-out
Sluggish	Punchy	Tired
Weariness	Gorked	Tuckered-out
Whipped	Yawny	Wiped
Zoned	Slap happy	Zonked

Hirshkowitz 2013

KSS

Karolinska Sleepiness Scale

	Français	Anglais
1	Parfaitement éveillé(e)	Extremely alert
2	Très éveillé(e)	Very alter
3	Éveillé(e)	Alert
4	Assez éveillé(e)	Rather alert
5	Ni éveillé(e) ni somnolent(e)	Neither alert nor sleepy
6	Un peu somnolent(e)	Some signs of sleepiness
7	Somnolent(e), mais sans effort pour rester éveillé(e)	Sleepy, but no effort to keep awake
8	Somnolent(e), mais avec des efforts pour rester éveillé(e)	Sleepy, but great effort to keep awake, fighting sleep
9	Très somnolent(e), luttant contre le sommeil	Extremely sleepy, can't keep awake
10	Extrêmement somnolent, ne peut rester éveillé	Extremely sleepy, can't keep awake

WHAT DOES 'BEING SLEEPY' MEAN?
AND HOW TO MEASURE IT

- Sleepiness =
  - ✗ Fatigue?
  - ✗ Performances?

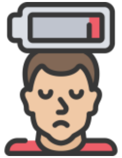


Table 1
Examples of some words used to describe fatigue, sleepiness, or both

Fatigued	Sleepy	Either or Both
Beat	Crashing	Exhausted
Languor	Drowsy	Burned out
Lassitude	Fading	Bushed
Lethargic	Groggy	Gassed
Listless	Narcotized	Pooped
Knackered	Heavy-headed	Played-out
Sluggish	Punchy	Tired
Weariness	Gorked	Tuckered-out
Whipped	Yawny	Wiped
Zoned	Slap happy	Zonked

Hirshkowitz 2013

- **Subjective** sleepiness
  - ✓ Long-term, e.g. measured by the **Epworth Sleepiness Scale**
  - ✓ Short-term, e.g. measured by the **Karolinska Sleepiness Scale**

- **Objective** sleepiness
  - ✓ EEG (Multiple Sleep Latency Test)

STATE OF THE ART
CORPORA

[1][Schuller et al. 2011] [2][Schuller et al. 2019]
[3][Huang et al. 2014] [4][Gosztolya et al. 2019]

Sleepy Language Corpus (SLC)[1]	SLEEP[2]	
State of the art[3] : UAR = 71.7% ✓	State of the art[4] : ρ = 0.387 ✓	→ Good performances
German + English	✗	→ French speakers
General population	✗	→ Patients
8.2s (sd: 15.3s) ✗	3.9s (sd: 0.6s) / 5s max ✗	→ minimum = 20 s.
Avg. of three KSS (instantaneous subjective sleepiness)	✗	→ No medical validity

UAR : *Unweighted Average Recall*

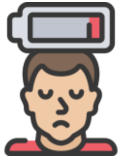
WHAT DOES 'BEING SLEEPY' MEAN?
AND HOW TO MEASURE IT

Table 1
Examples of some words used to describe fatigue, sleepiness, or both

Fatigued	Sleepy	Either or Both
Beat	Crashing	Exhausted
Languor	Drowsy	Burned out
Lassitude	Fading	Bushed
Lethargic	Groggy	Gassed
Listless	Narcotized	Pooped
Knackered	Heavy-headed	Played-out
Sluggish	Punchy	Tired
Weariness	Gorked	Tuckered-out
Whipped	Yawny	Wiped
Zoned	Slap happy	Zonked

Hirshkowitz 2013

- Sleepiness =
  - ✗ Fatigue?
  - ✗ Performances?



- **Subjective** sleepiness
  - ✓ Long-term, e.g. measured by the **Epworth Sleepiness Scale**
  - ✓ Short-term, e.g. measured by the **Karolinska Sleepiness Scale**

- **Objective** sleepiness
  - ✓ EEG (Multiple Sleep Latency Test)

MSLT
METHOD

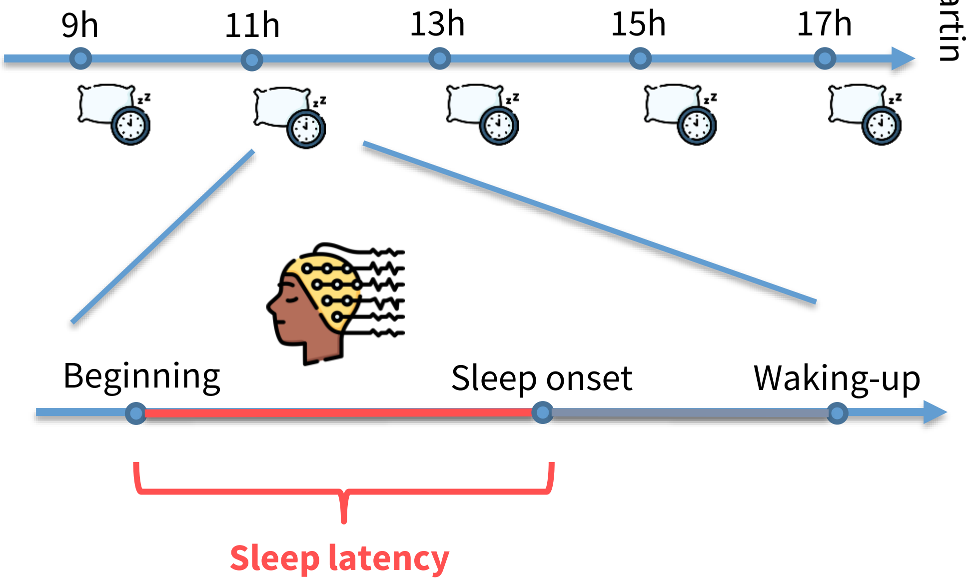
What is the MSLT ?

Multiple Sleep Latency Test

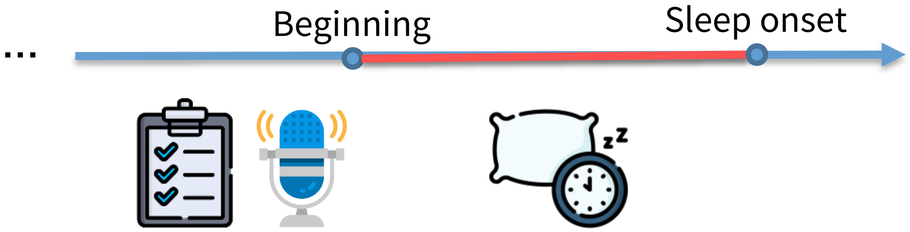
- 5 nap opportunity
- **Polysomnographic** recordings (PSG = EEG + EKG + EMG)
- Sleep Latency

  0 min. → 20 min.

  → Main label of the MSLTc
- Pathological threshold : **avg. Sleep latency ≤ 8min.**



MSLT CORPUS
METHOD



Voice recordings

- Sleep Clinic of Bordeaux
- Few interferences with MSLT
- **Reading** texts from *Le Petit Prince* (250 words / 1min 30s)
- 106 subjects, 5 samples/subjects ≈ 11h 30min
- Inclusion/Exclusion criteria based on reading level

Label and metadata

- **Sleep latency (Objective sleepiness)**
- Age, Sex, BMI, Neck circomference, Edu.
- Fatigue, Anxiety, Depression, …
- **Short-** and **long-**term subj. sleepiness

3.
Vocal and speech features

Hypothesis, definition and validation

VOCAL AND SPEECH FEATURES
CONSTRAINTS & METHOD

Explainability

- State of the art : openSMILE IS11 (#4368)
- “4th coefficient of the linear prediction of the derivative of the 25th coefficient RASTA”

Explainibility to clinicians

→ Interdisciplinary translation

Psychophysio. mecanisms

→ Integrative model

VOCAL AND SPEECH FEATURES

ACOUSTIC FEATURES

Acoustic quality of voice

A

READING ERRORS

Mistakes during the reading of a text out loud

E

P

Number, Duration and Location of pauses during reading out loud

READING PAUSES

A

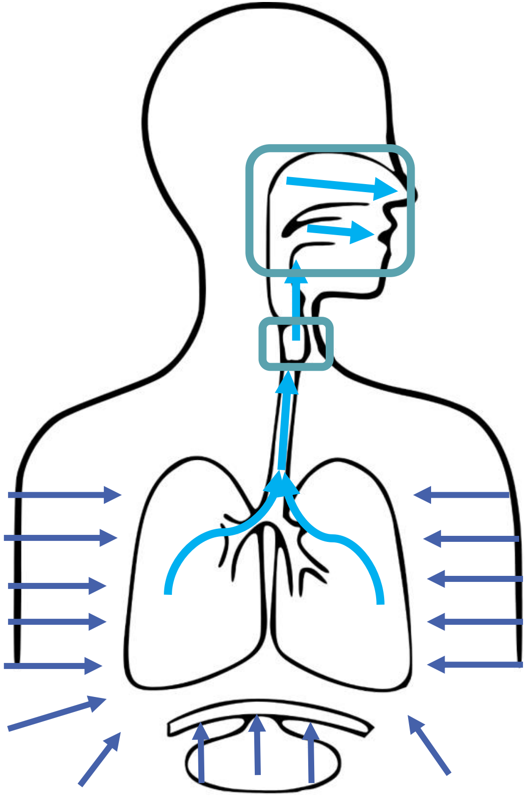
Errors made by an Automatic Speech Recognition System

ASR ERRORS

ACOUSTIC FEATURES
HYPOTHESIS AND METHOD

MSLT 1 min

MSLT 20 min



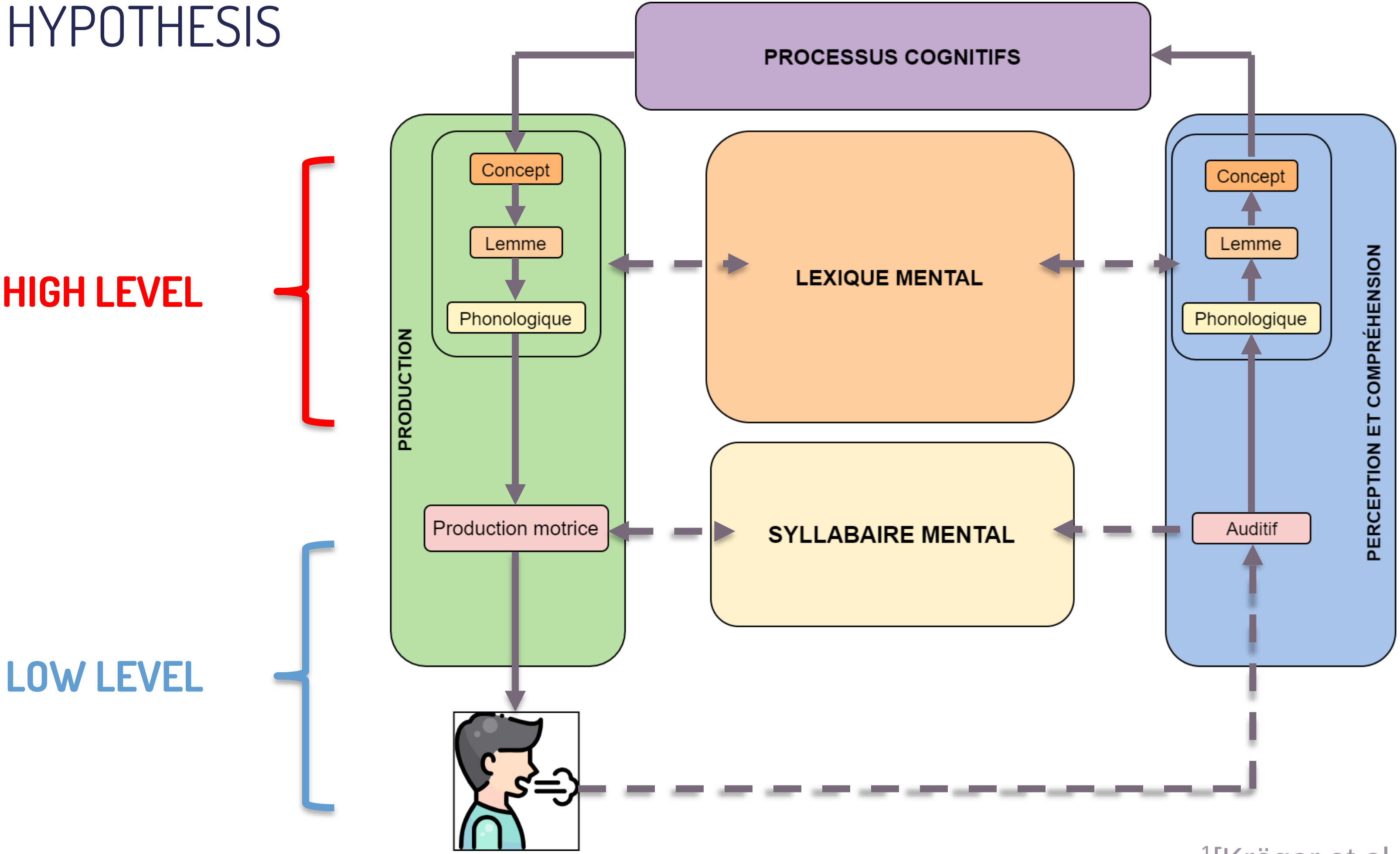
Is it possible to estimate **sleep latency** using **acoustic quality descriptors**?

Acoustic features (voiced parts)

- F0/NRJ mean, std, max, min, bdw, slope
- Harmonics: H1, H2, H4
- Formants: (amplitude, bandwidth, amplitude)
- diff. Harmonics/Formants
- HNR
- CPP

→ 44 acoustic features

HYPOTHESIS



[1][Kröger et al. 2020]

VOCAL AND SPEECH FEATURES

ACOUSTIC FEATURES

Acoustic quality of voice

A

READING ERRORS

Mistakes during the reading of a text out loud

E

P

Number, Duration and Location of pauses during reading out loud

READING PAUSES

A

Errors made by an Automatic Speech Recognition System

ASR ERRORS

READING MISTAKES
HYPOTHESIS

Is it possible to estimate **sleep latency** using **reading mistakes**?

MSLT
18.6 min.

KSS
3

Avg. MSLT.
8.3 min

Quand le mystère est trop impressionnant, on n’ose pas
« il »
désobéir. Aussi absurde que cela me semblât à mille milles
« semblais »
de tous les endroits habités et en danger de mort, je sortis
<ach>
de ma poche une feuille de papier et un stylographe.

READING MISTAKES
METHOD

Manual annotation of **530** samples of the MSLTc

- **Stumblings :** « hesitation, breaks in the speech rythm »
  Dictionnaire d'orthophonie, Brin (2018)
- **Deletions**
- **Additions**
- **Paralexia :** « identification error of written words consisting in the production of a word instead of another»
  Dictionnaire d'orthophonie, Brin (2018)
- **Words inversion**

VOCAL AND SPEECH FEATURES

ACOUSTIC FEATURES

A

Acoustic quality of voice

READING ERRORS

E

Mistakes during the reading of a text out loud

P

Number, Duration and Location of pauses during reading out loud

READING PAUSES

A

Errors made by an Automatic Speech Recognition System

ASR ERRORS

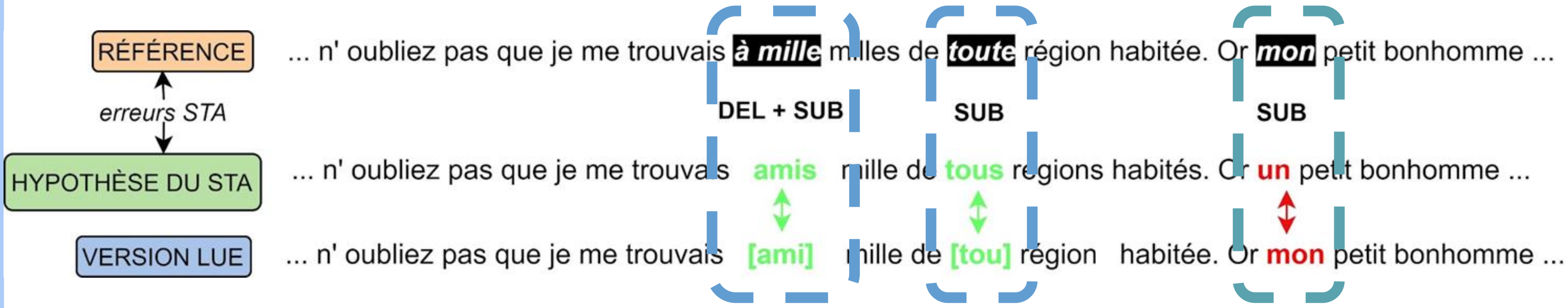
ASR ERRORS
HYPOTHESIS

Is it possible to automize reading mistakes annotations ?

MSLT
18.6 min.

KSS
3

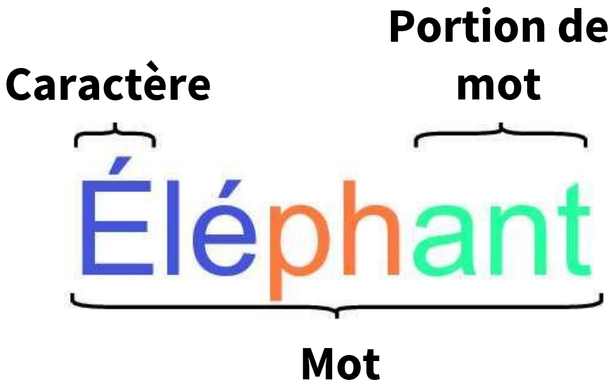
Avg. MSLT
8.3 min



ASR ERRORS
METHOD

- End-to-end (PhD Thesis of F. Boyer)[1]
- 3 different units (word, char, BPE)
- 7 configurations
- 4 errors : insertions, deletions, substitutions, nb of correct
- Word or char errors, nb or %

→ **112 features**



[1][Boyer 2021]

VOCAL AND SPEECH FEATURES

ACOUSTIC FEATURES

A

Acoustic quality of voice

READING ERRORS

E

Mistakes during the reading of a text out loud

P

Number, Duration and Location of pauses during reading out loud

READING PAUSES

A

Errors made by an Automatic Speech Recognition System

ASR ERRORS

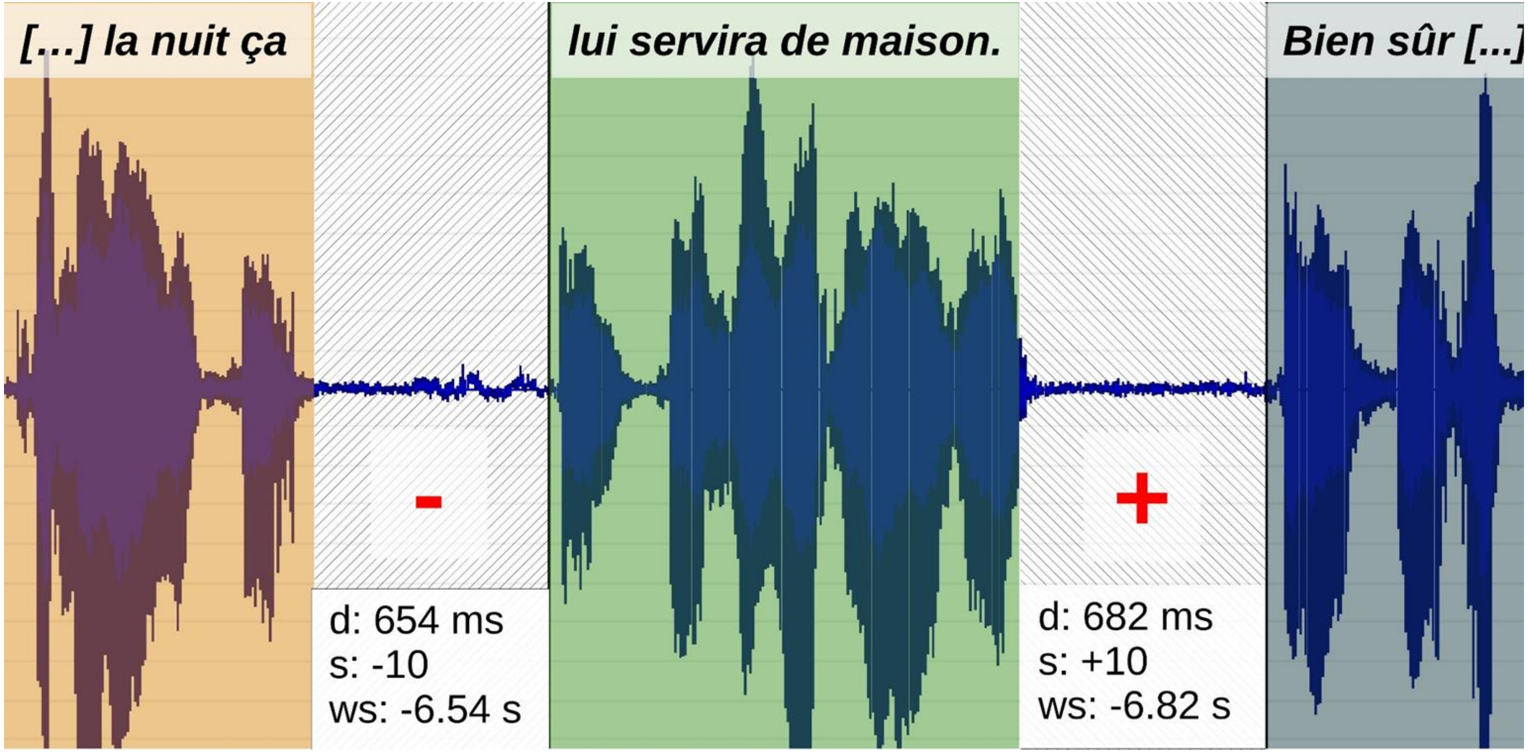
READING PAUSES
HYPOTHESIS

Are **reading pause locations** linked to **sleep propensity** ?

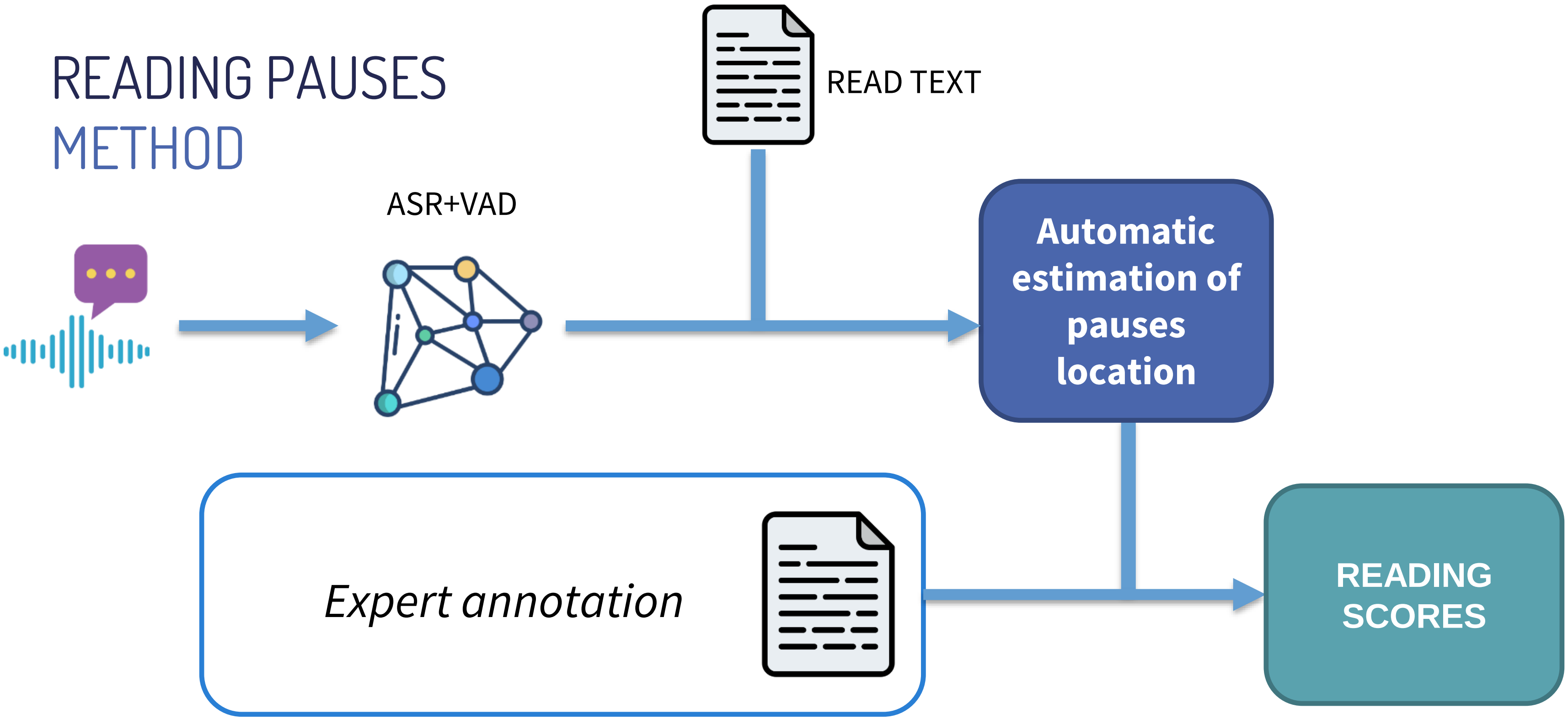
MSLT
14.5

KSS
7

Avg. MSLT
17.5



READING PAUSES
METHOD



Annotated text

J' -10 ai -10 beaucoup -10 vécu -8 chez -10 les -10 grandes -10 personnes. +10

Intraclass correlation coefficient = 0,97

Features: conclusion

ACOUSTIC FEATURES

Acoustic quality of voice

A

RRORS

reading of a

out loud

E

P

Number, Duration and Location of pauses during reading out loud

READING PAUSES

A

Errors made by an Automatic Speech Recognition System

ASR ERRORS

4.
Classification & interpretation

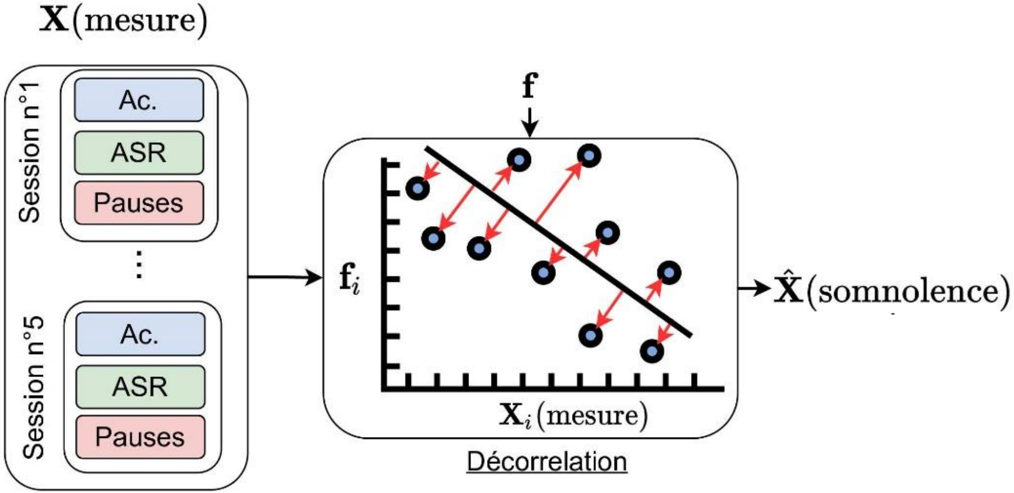
AUTOMATIC ESTIMATION
CONSTRAINTS

How to detect **sleep propensity** using the previous features?

Explainibility
Ability to explain the decision to clinicians

Biomarkers =
Sensitivity **AND** *Specificity*

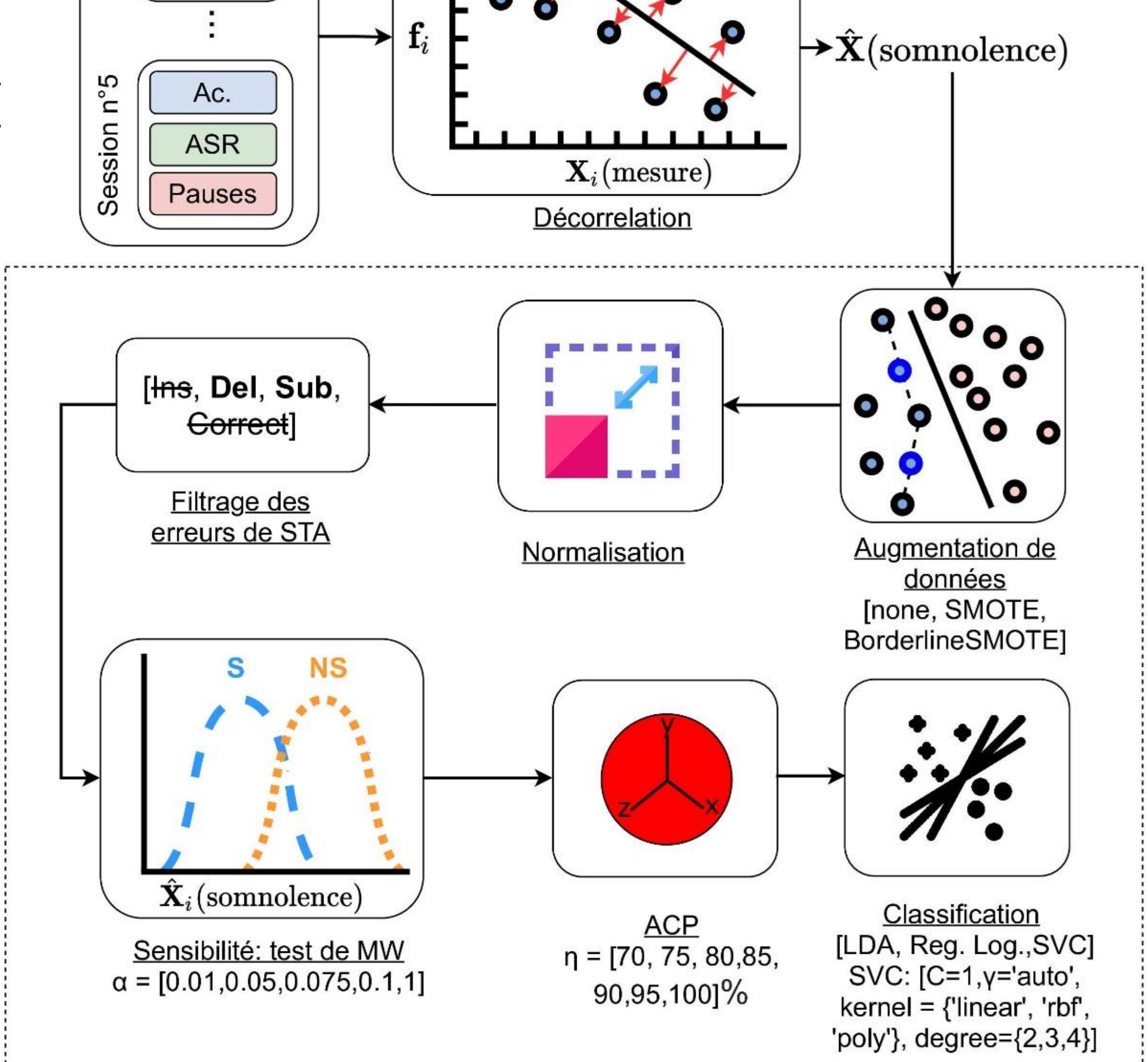
DÉTECTION DE LA PROPENSION À L'ENDORMISSEMENT
MÉTHODE



- Âge
- Sexe
- IMC
- Cou
- Édu.
- Anx.
- Dep.

Spécificité

AUTOMATIC E METHOD



Sensitivity

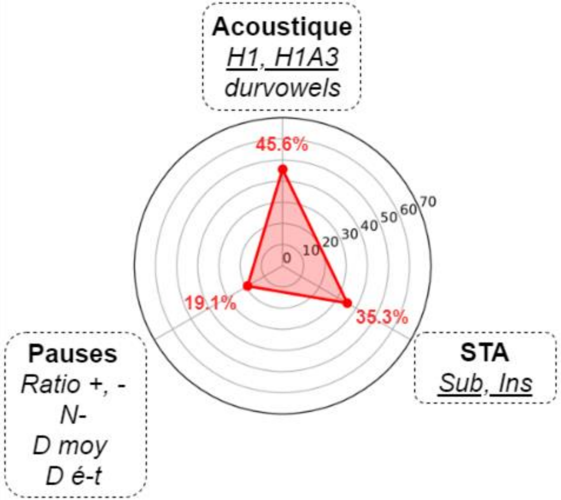
Specificity

AUTOMATIC CLASSIFICATION
RESULTS

Obj.Sleepiness

Avg. MSLT≤ 8min

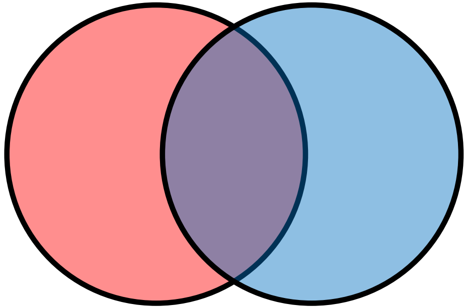
UAR = 84,6%



AUTOMATIC CLASSIFICATION
OBJECTIVES

Is it possible to detect **other symptoms?**

Pathological sleep propensity
Avg. MSLT ≤ 8min.
Objective evaluation
21 Subjects



Excessive Daytime Sleepiness
ESS > 15
Subj. evaluation (1 execution)
39 Subjects

ESS
Epworth Sleepiness Scale

TABLE 1. *The Epworth sleepiness scale*

THE EPWORTH SLEEPINESS SCALE

Name: ______________________________

Today's date: ______________ Your age (years): ________

Your sex (male = M; female = F): ________________

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently try to work out how they would have affected you. Use the following scale to choose the *most appropriate number* for each situation:

0 = would *never* doze
1 = *slight* chance of dozing
2 = *moderate* change of dozing
3 = *high* chance of dozing

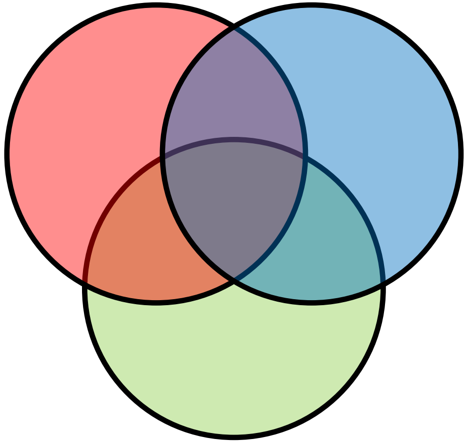
Situation	Chance of dozing
Sitting and reading	______
Watching TV	______
Sitting, inactive in a public place (e.g. a theater or a meeting)	______
As a passenger in a car for an hour without a break	______
Lying down to rest in the afternoon when circumstances permit	______
Sitting and talking to someone	______
Sitting quietly after a lunch without alcohol	______
In a car, while stopped for a few minutes in the traffic	______

Thank you for your cooperation

AUTOMATIC CLASSIFICATION
OBJECTIVES

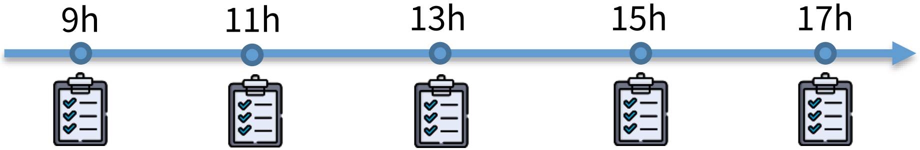
Is it possible to detect **other symptoms?**

Pathological sleep propensity
Avg. MSLT ≤ 8min.
Objective evaluation
21 Subjects



Excessive Daytime Sleepiness
ESS > 15
Subj. evaluation (1 execution)
39 Subjects

Average daytime sleepiness
Avg. Of 5 KSS > 5
Subj. evaluation (5 executions)
27 Subjects



KSS

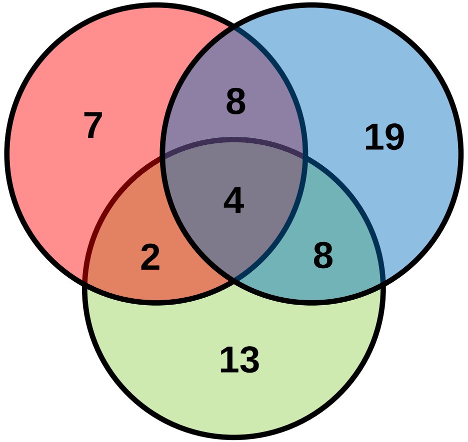
Karolinska Sleepiness Scale

	Français	Anglais
1	Parfaitement éveillé(e)	Extremely alert
2	Très éveillé(e)	Very alter
3	Éveillé(e)	Alert
4	Assez éveillé(e)	Rather alert
5	Ni éveillé(e) ni somnolent(e)	Neither alert nor sleepy
6	Un peu somnolent(e)	Some signs of sleepiness
7	Somnolent(e), mais sans effort pour rester éveillé(e)	Sleepy, but no effort to keep awake
8	Somnolent(e), mais avec des efforts pour rester éveillé(e)	Sleepy, but great effort to keep awake, fighting sleep
9	Très somnolent(e), luttant contre le sommeil	Extremely sleepy, can't keep awake
10	Extrêmement somnolent, ne peut rester éveillé	Extremely sleepy, can't keep awake

AUTOMATIC CLASSIFICATION
OBJECTIVES

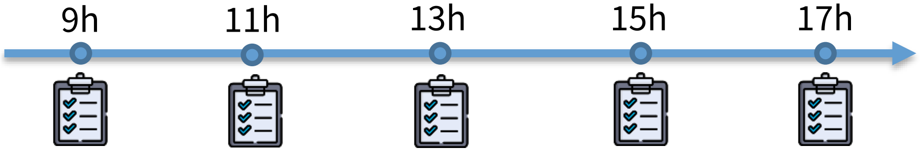
Is it possible to detect **other symptoms?**

Pathological sleep propensity
Avg. MSLT ≤ 8min.
Objective evaluation
21 Subjects



Excessive Daytime Sleepiness
ESS > 15
Subj. evaluation (1 execution)
39 Subjects

Average daytime sleepiness
Avg. Of 5 KSS > 5
Subj. evaluation (5 executions)
27 Subjects

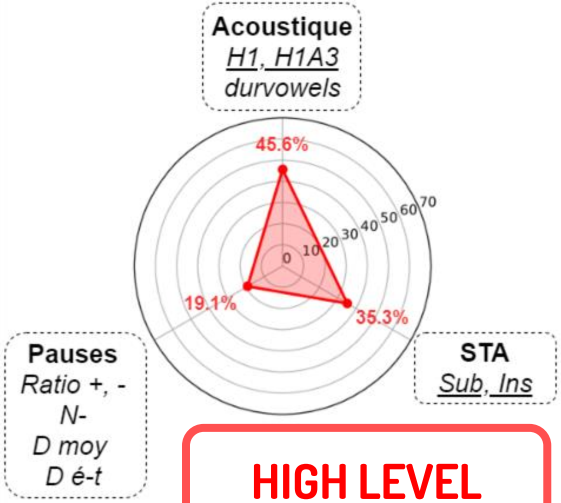


AUTOMATIC CLASSIFICATION
RESULTS

Obj.Sleepiness

Avg. MSLT≤ 8min

UAR = 84,6%



HIGH LEVEL

LOW LEVEL

Classification: conclusion

- **Simple** pipeline (explainability)
- **Objective** sleepiness → **High-level** features
- **Subjective** sleepiness → **Low-level** features

Perspectives

New databases & Symptom networks

PERSPECTIVES
NEW DATABASES

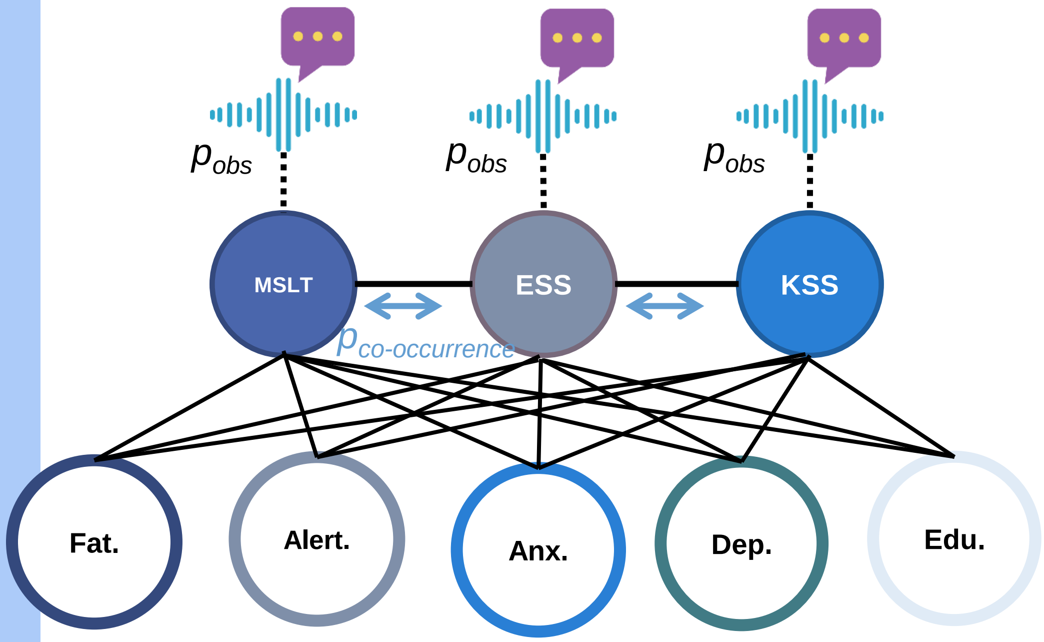
SOMVOICE

- 32 healthy subjects
- MSLT after Total Sleep Deprivation / after normal night
- Under recording

MEDISPEECH

- Colleen Baumard
- Spontaneous speech / Smartphone interaction
- Clinical MSLT / MWT
- Sleepiness/Fatigue/Depression

PERSPECTIVES
SYMPTOM NETWORKS



Symptom Networks

- **Bayesian networks**

→ **Data processing perspectives**

- Joint information
  - *Belief propagation*
  - What graph?
  - Transitions?

→ **Clinical perspectives**

- Interaction between symptoms
- Prognostic / therapeutic targeting
- Inaccessible symptoms
- Multimodality?

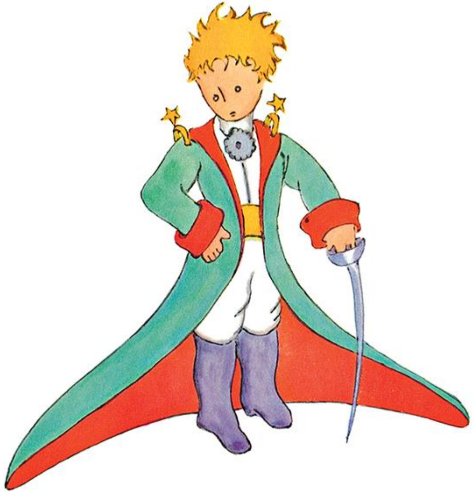
Conclusion

Doggy bag

DOGGY BAG

- **Symptoms** instead of diagnosis
- Databases with **obj. and subj.** sleepiness
- Simple explainable (**to clinicians**) features and pipeline
- **Biomarkers** = sensibility + **specificity**

Thank you for your attention!



QUESTIONS?

vincent.martin@labri.fr

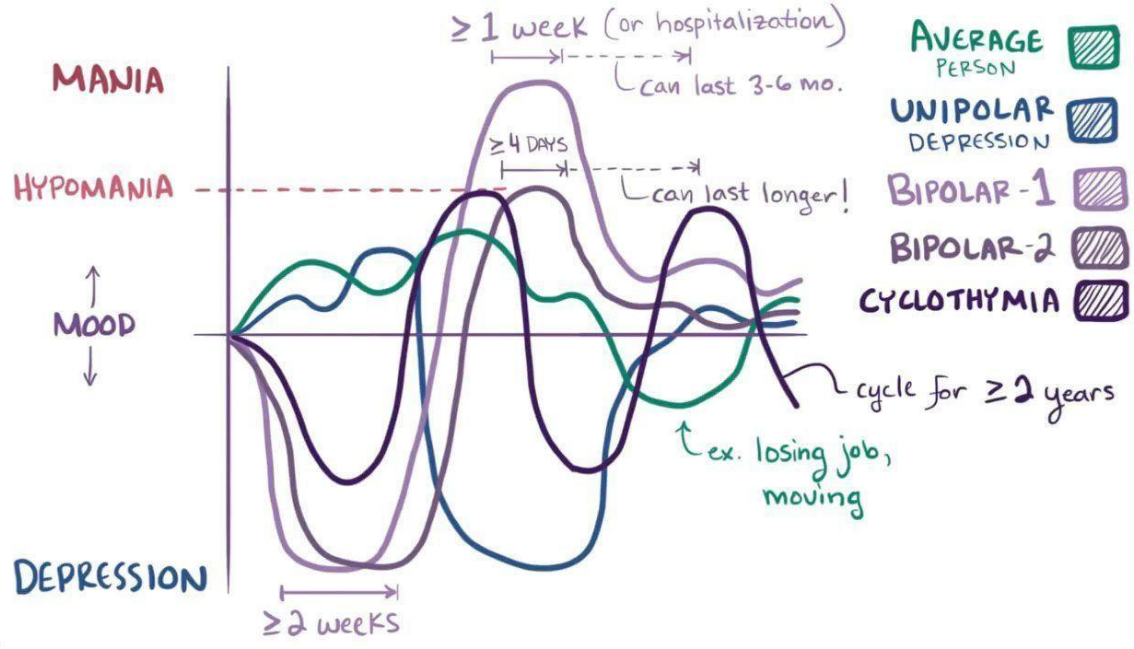
@V_P_Martin

Vincent-P-Martin

WHAT DO WE DETECT ?
EXAMPLE 1: BIPOLAR DISORDER

Bipolar disorderS

- Diag. = based on variations and duration
- How to detect BD with only **1 recording ?**
- **State** vs. **Trait** (same for all Ψ disorders)



https://www.osmosis.org/learn/Bipolar_disorder

WHAT DO WE DETECT?
EXAMPLE2: DEPRESSION

Depression

- Number of semiological profiles
- $n = \binom{2}{1} \times \left(\binom{8}{4} + \binom{8}{5} + \cdots + \binom{8}{8}\right)$
- = **326 unique profiles**
- **Eiko Fried**: STAR*D (2015) : 1030 profiles on 3703 “depressive” patients
- Rutowski et al. 2022 (IS22): (1) **test sizes** below **1K** samples gave noisy results, even for larger training set sizes; (2) **training set sizes** of at least **2K** were needed for stable results;

Major Depressive Disorder

Diagnostic Criteria

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

WHAT DO WE DETECT?
EXAMPLE2: DEPRESSION

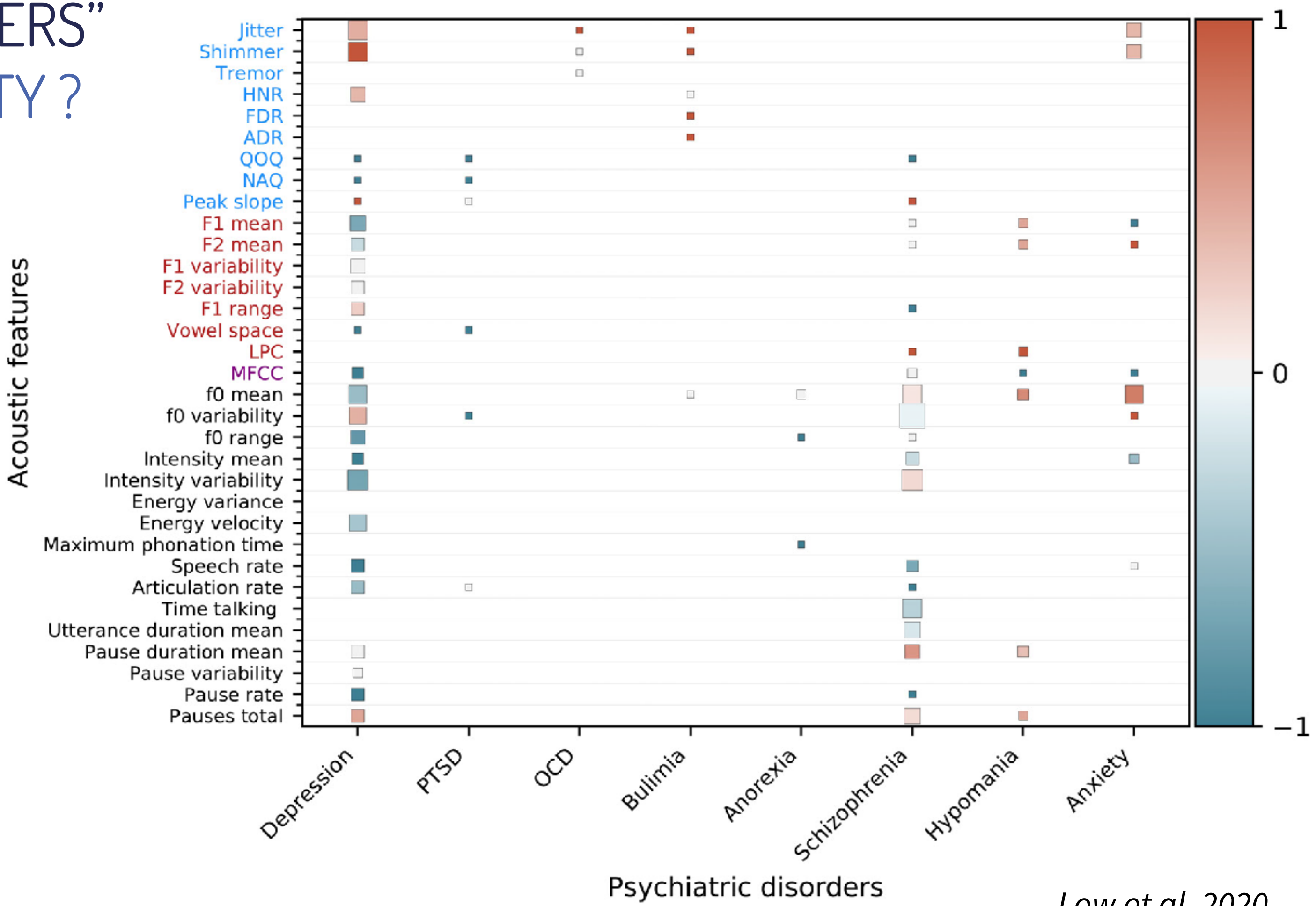
What does a ML classifier learn ?

- **Difference between groups**
  - Sub-group?
  - Symptom?
  - Other bias?
  - ?
- **+ /!\ Temporality /!**

Depressive vs. HC with bad mood

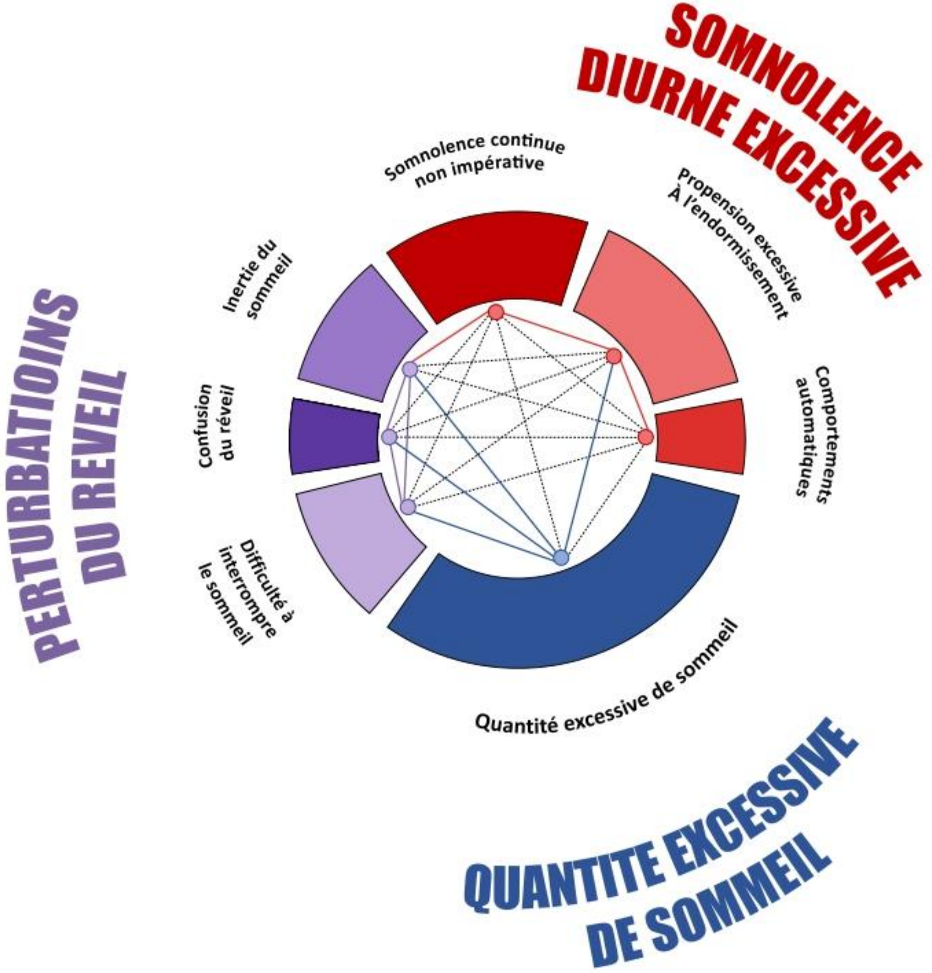
- **NOT “depression disorder”**

“BIOMARKERS”
SPECIFICITY ?



Low et al. 2020

SLEEPINESS



Lopez et al. 2022

DEEP LEARNING ?

1. IS19 challenge: winner = Fischers vectors + SVR
   Recent DL models : perf < IS19
2. C. Rudin 2019 « Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead », *Nature Machine Intelligence*
3. Meta analyses : Christodoulou « A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models» *J. Clinic. Epidemio.*

➔Did you put as much efforts in logistic regression than in tunning a deep learning model?

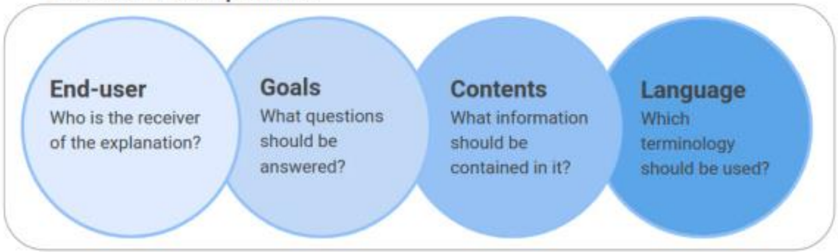
IS EXPLAINABILITY ENOUGH?

[Vilone et al. 2021 “Notions of explainability and evaluation approaches for explainable artificial intelligence”, Information Fusion]

Algorithmic transparency	The degree of confidence of a learning algorithm to behave ‘sensibly’ in general [2], [26]
Causality	The capacity of a method for explainability to clarify the relationship between input and output [8], [21], [22], [23], [24], [25], [29]
Comprehensibility	The quality of the language used by a method for explainability [9], [31], [32], [33], [34], [35], [36], [37], [38]
Effectiveness	The capacity of a method for explainability to support good user decision-making [40], [41], [42], [43]
Efficiency	The capacity of a method for explainability to support faster user decision-making [20], [41], [42]
Explicability	The degree of association between the expected behaviour of a robot to achieve assigned tasks or goals and its actual observed actions [44]
Explicitness	The capacity of a method to provide immediate and understandable explanations [45]
Faithfulness	The capacity of a method for explainability to select truly relevant features [45]

Interestingness	The capacity of a method for explainability to facilitate the discovery of novel knowledge and to engage user’s attention [33], [34], [36], [53], [54]
Interpretability	The capacity to provide or bring out the meaning of an abstract concept [9], [18], [33], [35], [55], [56], [57], [58], [59], [60], [61], [62], [63]
Informativeness	The capacity of a method for explainability to provide useful information to end-users [21]
Justifiability	The capacity of an expert to assess if a model is in line with the domain knowledge [1], [33], [40], [55], [64], [65]
Mental Fit	The ability for a human to grasp and evaluate a model [33], [66]
Persuasiveness	The capacity of a method for explainability to convince users perform certain actions [20], [41], [42]
Selection/ simplicity	The ability of a method for explainability to select only the causes that are necessary and sufficient to explain the prediction of an underlying model [25]
Soundness	The extent to which each component of an explanation’s content is truthful in describing an underlying system [27], [28]
Transparency	The capacity of a method to explain how the system works even when it behaves unexpectedly [9], [10], [11], [12], [20], [26], [40], [41], [47], [58], [59], [63], [64], [76], [77], [78]

Structure of an explanation



IS EXPLAINABILITY NEEDED?

Bourla et al. 2018: "Psychiatrists' Attitudes Toward Disruptive New Technologies: Mixed-Methods Study", JMIR Mental Health

- Sleep specialists vs. EEG
- **TRUST** does not reduce to explainability
- **Bourla et al.:**
  - 515 psychiatrists
  - 3 scenarios: biosensors comprising a connected wristband-based digital phenotype, ML-based blood test, ML-based magnetic resonance imaging (MRI).
  - 4 acceptability domains usefulness, usability, reliability, and risk

  - Overall acceptability=moderate.
  - All systems = risky(410/515, 79.6%).
  - Acceptability = strongly influenced by socioepidemiological variables (professional culture), such as gender, age, and theoretical approach.
  - Worries = therapeutic relationship, data security, data storage, and privacy risk

'OBJECTIVITY' ?