Big Data Approaches to Personalised Hearing Care

Dr. Marta Campi
Marie Curie Fellow,
Deep Hearing Lab, University of Zurich
CERIAH, Institut Pasteur

6th French-Brazilian Symposium ,
Belo Horizonte 2026

From Statistics to Hearing Care

Paris

Institut Pasteur

Postdoc

- Big Data for audiological care
- Computational Modelling for ANSD

Zurich

University Hospital Zurich

Marie Curie Fellowship

- Outcome prediction in Cochlear Implant
- Computational Modelling

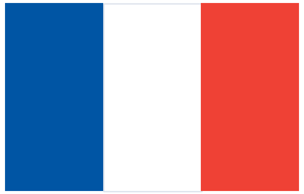
Brazil

The 6th French-Brazilian Symposium

I come from statistics and signal processing, and I moved into hearing research

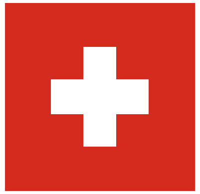
Hearing Loss: A Global Challenge

Worldwide, **over 1.5 billion people** live with hearing loss, and the burden rises steeply with age



16,2 million

France



1.9 million

Switzerland

48,9 million

Brazil

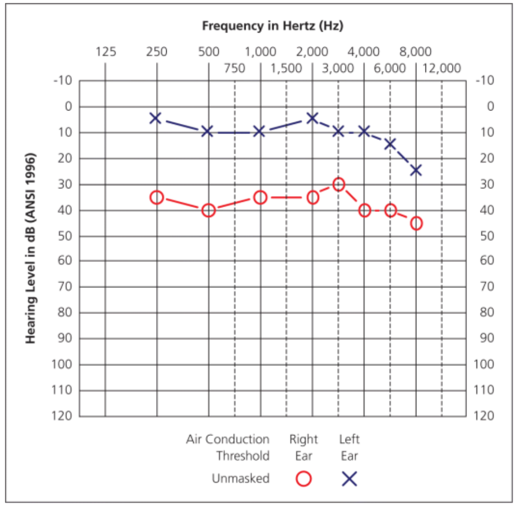
Prevalence of hearing loss, all ages, 2023

The same challenge in every country, even here in Brazil

Sources: Global Burden of Disease Study 2023 (IHME)

What the Clinic Measures

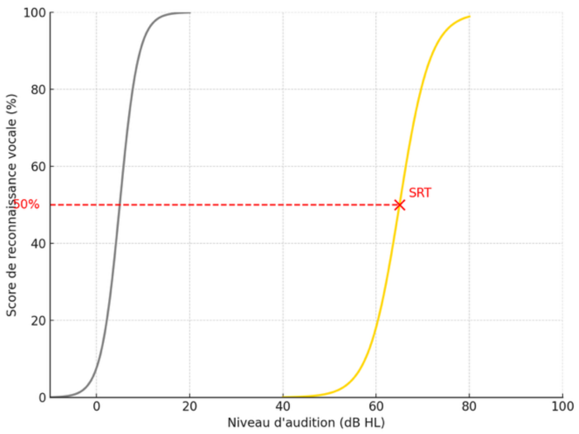
Pure-tone audiogram



The softest tones a patient can detect

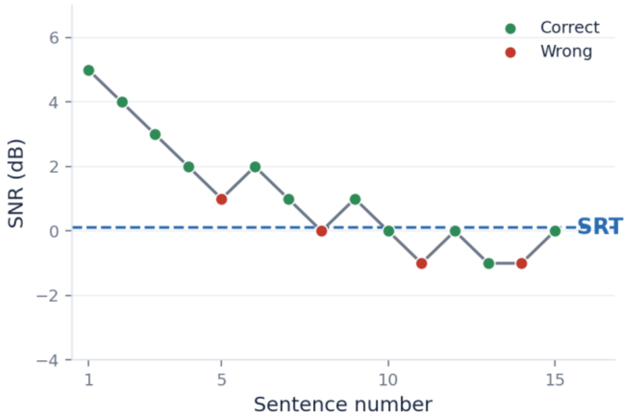
Pure-Tone Average (PTA) for degree of hearing loss

Speech in quiet



SRT: the level at which 50% of words are understood

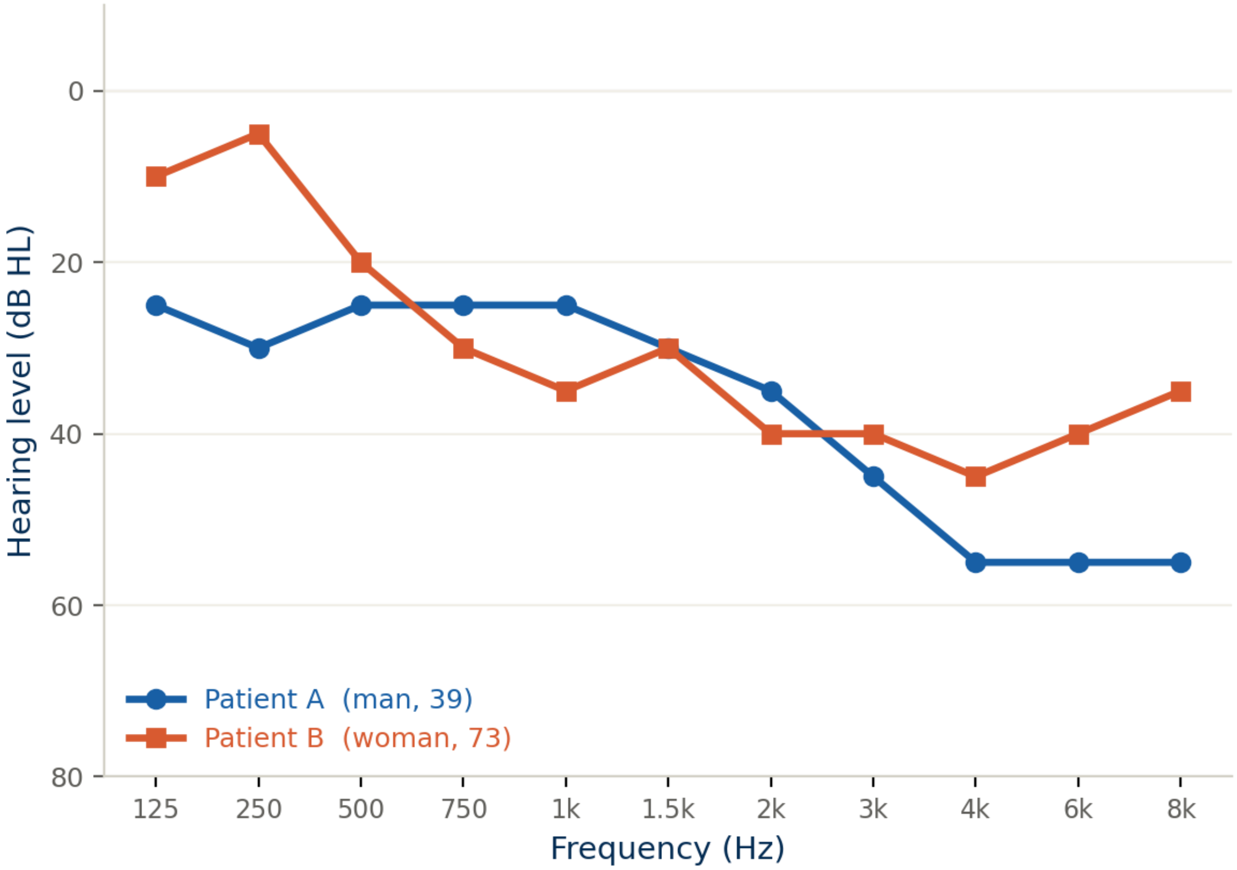
Speech in noise



SNR: the noise level at which 50% of speech is understood

The tests compresses a **performance into one number hiding information about the patient**

Same Number, Different Person



same PTA

35 dB

PATIENT A · man, 39 **−4 dB SNR**

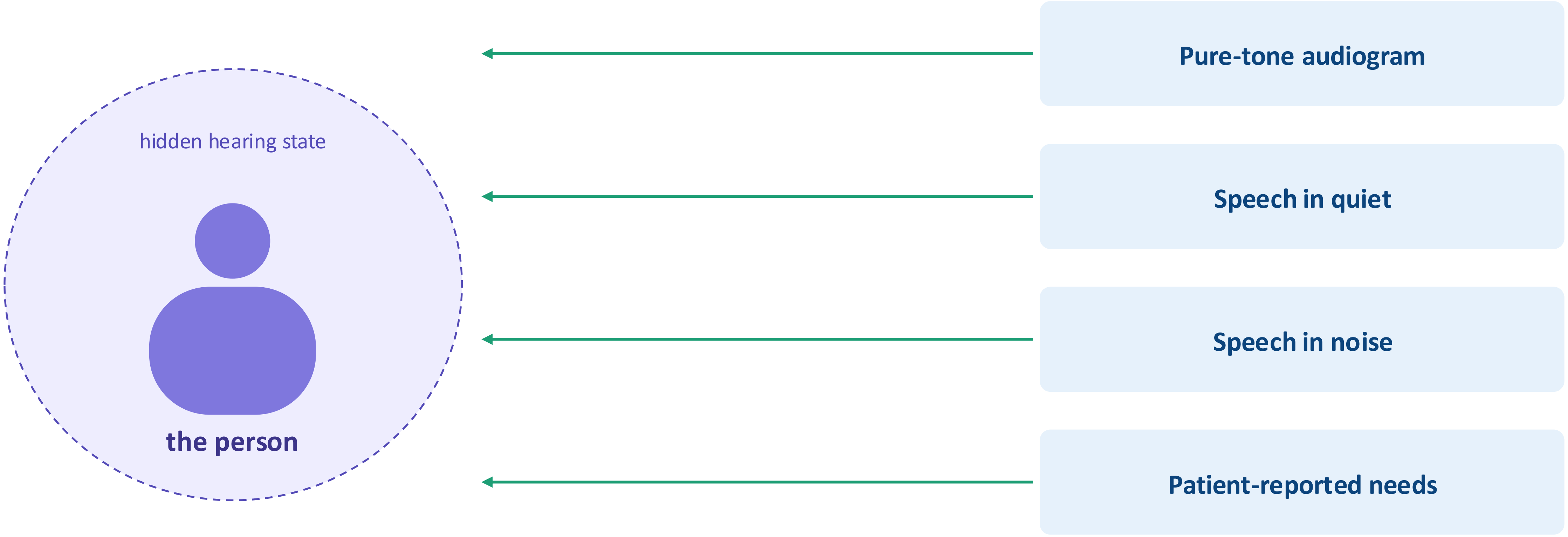
"I want to understand the TV better"

PATIENT B · woman, 73 **+19 dB SNR**

"I want to follow conversation, in quiet and in noise"

Same PTA. Different ear, different life. The average hides the patient

Behind the Numbers



Estimate the hidden state from the tests

Why Big Data (AI) Methods

- Behind the **same number** there are **different people**
- What separates tests performances **is not** the number, it is the **structure behind it**
- That structure is **hidden in combinations of variables**, not in any single measure
- To recover it we need **rich data**, and **models** that learn it and stay **interpretable**

Three Studies

With data at this scale, one question: **can we recover the patient the average hides?**

	SAMPLE		DATA SOURCE
1. Audiogram shape predicts who benefits from a hearing aid	**22k**		Amplifon, France
2. What matters to the patient, in their own words	**91k**		Amplifon, France
3. Who will struggle after a cochlear implant	**473**		CICH database

1. Audiogram Shape Predicts Who Benefits From a Hearing Aid

RESEARCH ARTICLES | MARCH 19 2026

From Hearing Patterns to Functional Outcomes: Quantifying Audiometric Profiles for Precision Hearing Care

Subject Area: Audiology and Speech , Further Areas , Neurology and Neuroscience

Perrine Morvan; Marta Campi; Guillaume Staerman; Pauline Roger; Fabien Auberger; Catherine Boiteux; Alexandre Osty; Paul Avan; Bernard Fraysse; Hung Thai-Van

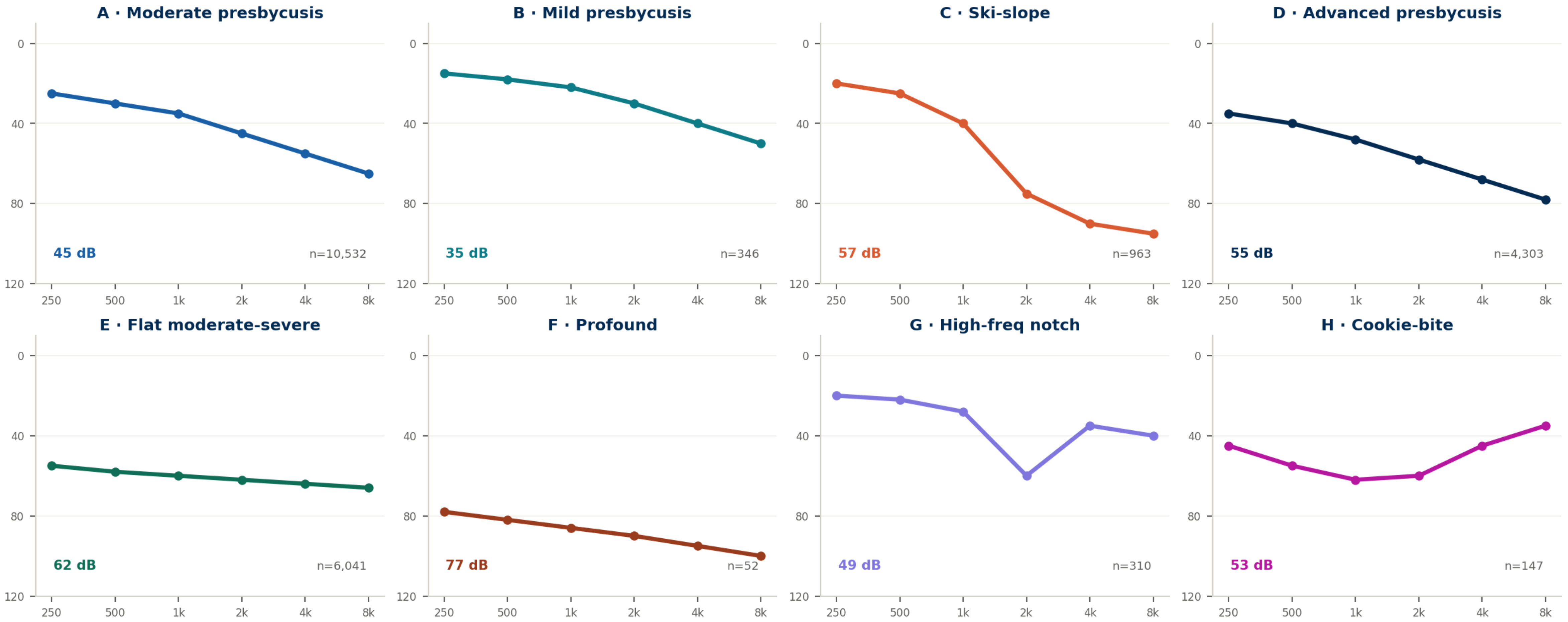
Audiol Neurotol (2026)

https://doi.org/10.1159/000551057 Article history

PubMed:41855361

Eight Audiometric Profiles

Defined from pathophysiology (Dubno, Pittman, Parthasarathy)



Data & Research Question

WHO

22,694

adults aged 50–90

780 audiology centres, France · 2018–2021

WHAT WE MEASURE

~19 variables

Tests: audiogram (PTA), speech in quiet (Lafon), speech in noise (HINT)

Patient: age, sex

Hearing aid & care: amplification, follow-up, daily use, technology

WHAT WE PREDICT

who improves

the gain after fitting

$\Delta SRT \geq 6$ dB (quiet) · $\Delta SNR \geq 2$ dB (noise)

Does the audiogram *shape* add information beyond the usual clinical variables?

Does the Audiogram Shape Add Anything?

THE AI MODEL:

We predict who improves after fitting (ΔSRT in quiet, ΔSNR in noise), and compare A model and B model:

A model (audiogram + patient + hearing-aid variables)

B model (same variables + **audiogram shape**)

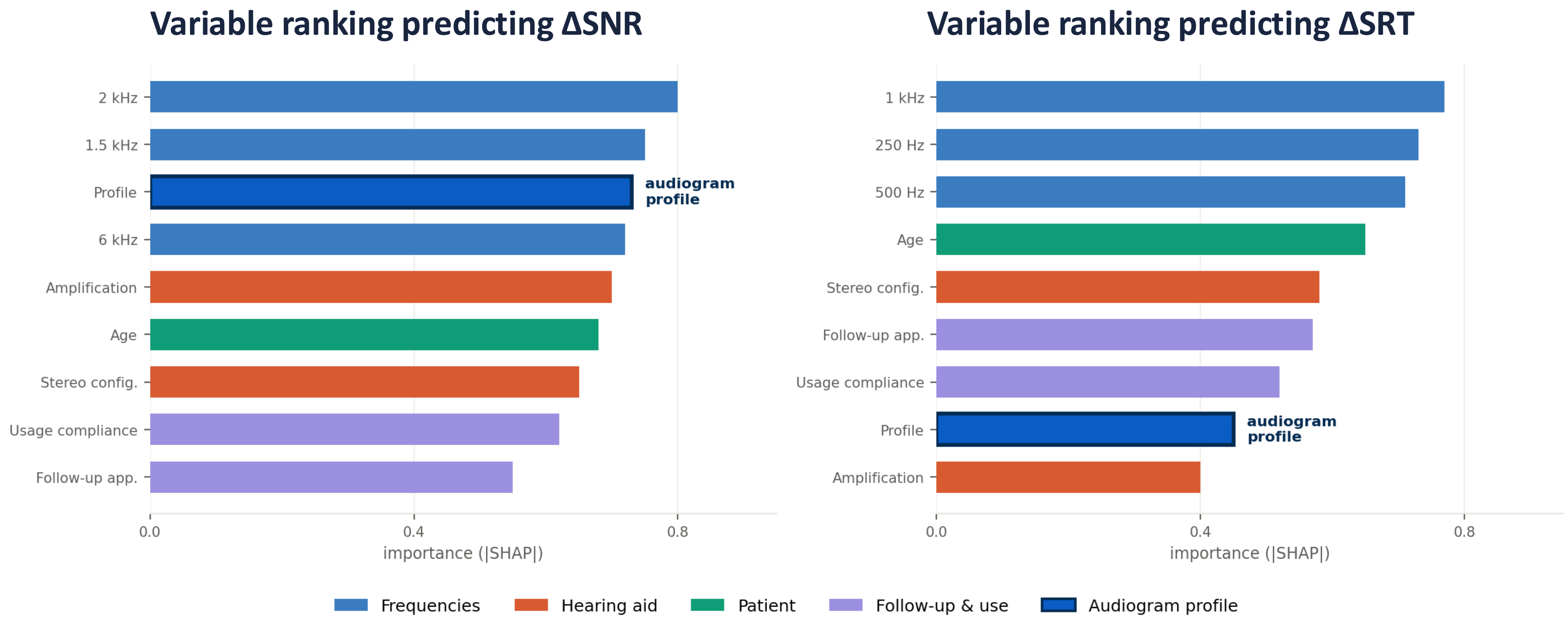
RESULTS:

Speech in quiet	**ΔSRT**	A **0.87**	→	B **0.92**	**+5%**	*accuracy*
Speech in noise	**ΔSNR**	A **0.76**	→	B **0.85**	**+9%**	*accuracy*

*Adding the shape improves prediction for both, **more in noise***

What Predicts Who Benefits

Different drivers in noise and in quiet, and the audiogram profile is among the strongest in noise

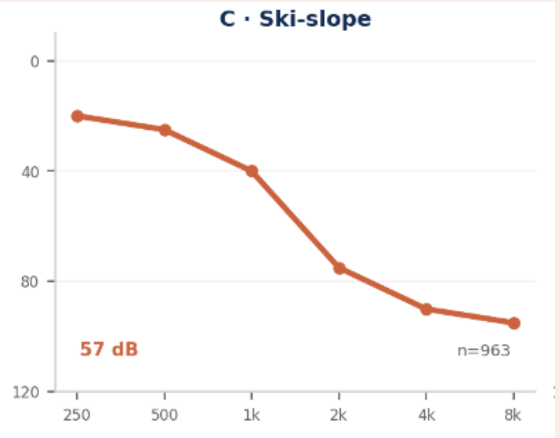


Same Severity, Opposite Levers

Two profiles with the same PTA respond to completely different interventions

Profile C · Ski-slope

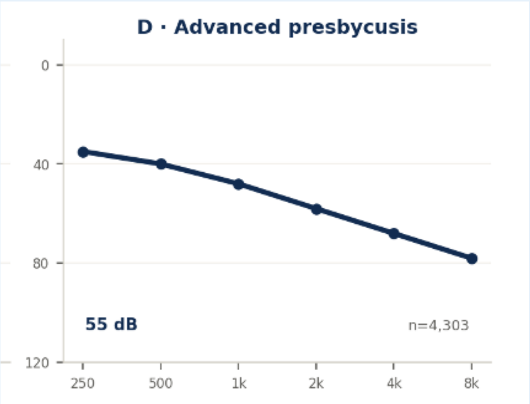
PTA 57 dB



needs **premium technology + high daily use**

Profile D · Advanced presbycusis

PTA 55 dB



needs **frequency-specific amplification only**

2. What Matters to the Patient, In Their Own Word

Beyond the Audiogram: Using PROMs and AI Modelling to Characterise Age- and Sex-Specific Hearing Loss Needs in a Nationwide Study

Perrine Morvan*
Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
pmorvan@pasteur.fr

Marta Campi*
Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
mcampi@pasteur.fr

Gareth W. Peters
Department of Statistics and Applied Probability
University of California Santa Barbara
Santa Barbara, CA 93106, United States
garethpeters@ucsb.edu

Hung Thai-Van
Institut Pasteur
Université Paris Cité
Inserm, Institut de l'Audition
IHU reConnect, F-75012, Paris, France
@pasteur.fr

But What Does the Patient Experience?

The same audiogram can mean very different difficulties in everyday life

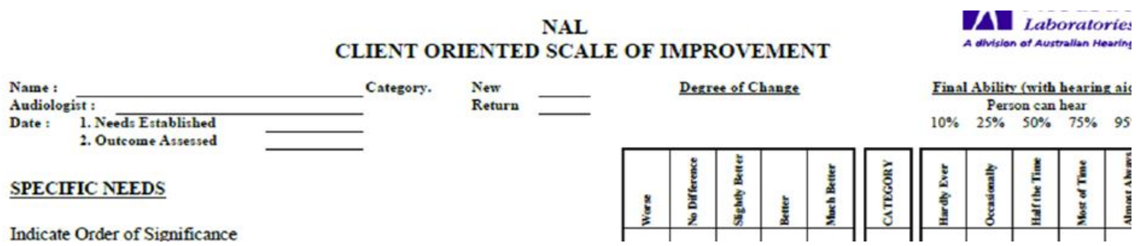
- **A conversation at home**
- **A noisy restaurant**
- **The telephone**
- **Following a group**
- **Meetings at work**
- **Television, music**
- **Grandchildren's voices**
- **Hearing in the street**

Standard tests do not capture this. **To see it, we have to ask the patient**

Patient Reported Outcome Measures (PROMs)

PROMs:

Questionnaires: the patient reports how they function in everyday life and what matters most to them


NAL

CLIENT ORIENTED SCALE OF IMPROVEMENT

Laboratories

A division of Australian Hearing

Name :

Audiologist :

Date : 1. Needs Established

2. Outcome Assessed

Category.

New

Return

Degree of Change

Final Ability (with hearing aid)

Person can hear

10% 25% 50% 75% 95

SPECIFIC NEEDS

Indicate Order of Significance

Worse	No Difference	Slightly Better	Better	Much Better	CATEGORY	Hardly Ever	Occasionally	Half the Time	Most of Time	Almost Always

the real COSI form (Dillon, NAL)

In Hearing Loss:

The Client Oriented Scale of Improvement (COSI):

The patient names the listening situations they want to improve

EXAMPLE GOALS, IN THE PATIENT'S WORDS

“Hear the TV at a normal volume”

“Understand my wife without asking to repeat”

“Follow a group conversation”

“My disabling tinnitus”

Dillon, James & Ginis (1997), J Am Acad Audiol · ***COSI*** | *Kelkar et al. (2016), J Am Coll Cardiol HF* · ***PROMs in medicine***

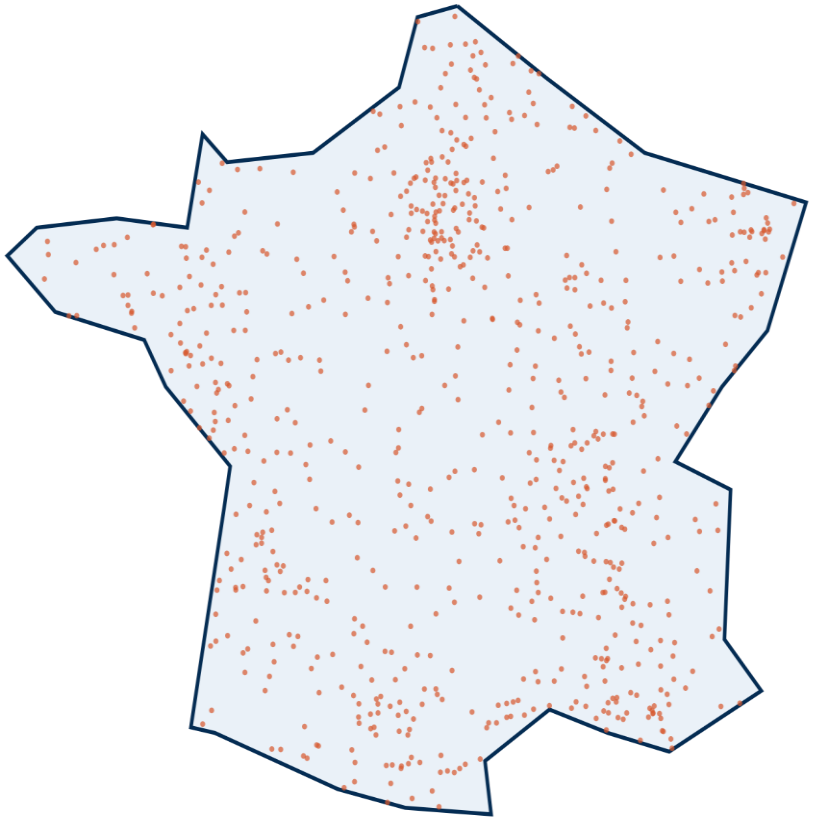
Data & Research Question

780

audiology centres in France

91,297

patients · COSI responses · Audiogram · speech in quiet (Lafon) · speech in noise (HINT)



The question: do the patient's own words add information beyond PTA and speech tests?

Reading the Patient's Words

An AI model maps each free-text answer to one of the 16 COSI need categories

THE PATIENT WROTE

"Baisser le son de la télé."

AI MODEL

CamemBERT

AI language model

THE 16 COSI CATEGORIES

1. Conversation, 1–2 in quiet
2. Conversation, 1–2 in noise
3. Conversation, group in quiet
4. Conversation, group in noise
5. **Television / radio, normal volume**
6. Familiar speaker on phone
7. Unfamiliar speaker on phone
8. Phone ring from another room
9. Front doorbell or knock
10. Hear traffic
11. Increased social contact
12. Feel embarrassed or stupid
13. Feel left out
14. Feel upset or angry
15. Church or meeting
16. Other

Hand-labelled on 1% (7,232 answers), then applied automatically to all 723,255, and verified

Do the Patient's words Add Anything?

THE MODEL:

We explain the degree of loss (PTA), and ask if the patient's words add to the speech tests:

A PTA ~ SNR + SRT

B PTA ~ SNR + SRT + PROMs

RESULTS:

ACROSS THE DEGREE OF LOSS

slight → moderate | mod.-severe → profound

PROMs add information | *speech tests already explain it*

AND WITHIN MILD LOSS, ACROSS AGE

younger | older

weaker contribution | *PROMs matter more with age*

3. Who Will Struggle After a Cochlear Implant

Prediction of Individual Speech Outcomes with Cochlear Implants: Limitations and Opportunities for Clinical Translation

A Preprint

Marta Campi[1,2] **Alexander Huber**[3] **Tobias Goehring**[1,3]

[1]Deep Hearing Lab, Department of Otorhinolaryngology,
University Hospital Zurich, University of Zurich, Switzerland

[2]Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS,
Fondation Pour l'Audition, Institut de l'Audition,
IHU reConnect, F-75012, Paris, France

[3]Department of Otorhinolaryngology,
University Hospital Zurich, University of Zurich, Switzerland

Outcome Prediction in Cochlear Implant (CI)

- ***"How well will I hear after the implant?"*** is a prediction question
- Needed for **counselling, surgery, monitoring**, and **rehabilitation**.
- Largest cohorts, every algorithm: error **17-25% points** [1,2]; variance explained **< 20%**[3]
- So **why** the ceiling and **what** can we predict?

[1] Shafieibavani et al. 2021, Trends in Hearing· [2] Demyanchuk et al. 2025, J Clin Med· [3] Anthonisen et al. 2026, Audiol Neurotol

Literature on Outcome Prediction in CI

All studies predict the same outcome: post-operative monosyllabic word score recognition, in quiet, scored as words or phonemes depending on the language

Study	N	Best result (validation)	Most important variables	Model
Kim 2018	120	**MAE 4.8→17.1** (LOO* → cross-site)	duration, age, HA use	random forest
Shafieibavani 2021	2,489	**MAE 18-22 pp** (cross-site)	demographics, audiometry, etiology	ensemble (XGB)
Demyanchuk 2025	2,571	**MAE 17 pp** (held-out + temporal)	age, duration, pre-op MS, PTA	decision tree
Stronks 2025	533	**R^2 0.10-0.19** (in-sample)	duration, age, aided speech	regression
Anthonisen 2026	2,251	**R^2 = 0.11** (held-out 10×)	age, duration, pre-op hearing	8 models (XGB)
Walia 2022	35	**R^2 = 0.59** (in-sample)	**post-op cochlear measure**	regression

MAE = mean absolute error (percentage points); *R^2* = variance explained (0 = no better than the group mean)

* LOO = Leave-One-Out validation

Two Research Questions

Prediction of speech outcomes

Regression: can we predict the score?

Dependent variable: Monosyllabic word score recognition (% words correct), continuous

Prediction of binary risk for poor outcomes

Classification: can we flag the risk of poor outcome before surgery?

Dependent variable: poor outcome, Monosyllabic word score recognition < 40% (yes / no)

Cohort: CICH Database, Zurich

- **FULL DATASET**

473 **adults, bilateral HL (single-sided deafness excluded)** median age 60 (18 to 92) · 53% female

- **PRE-OPERATIVE PREDICTORS** *79 routinely collected variables*

46 Acoustic measures	**33** Non-acoustic variables
audiometry · pre-operative speech tests	*demographic · developmental · clinical · socioeconomic*

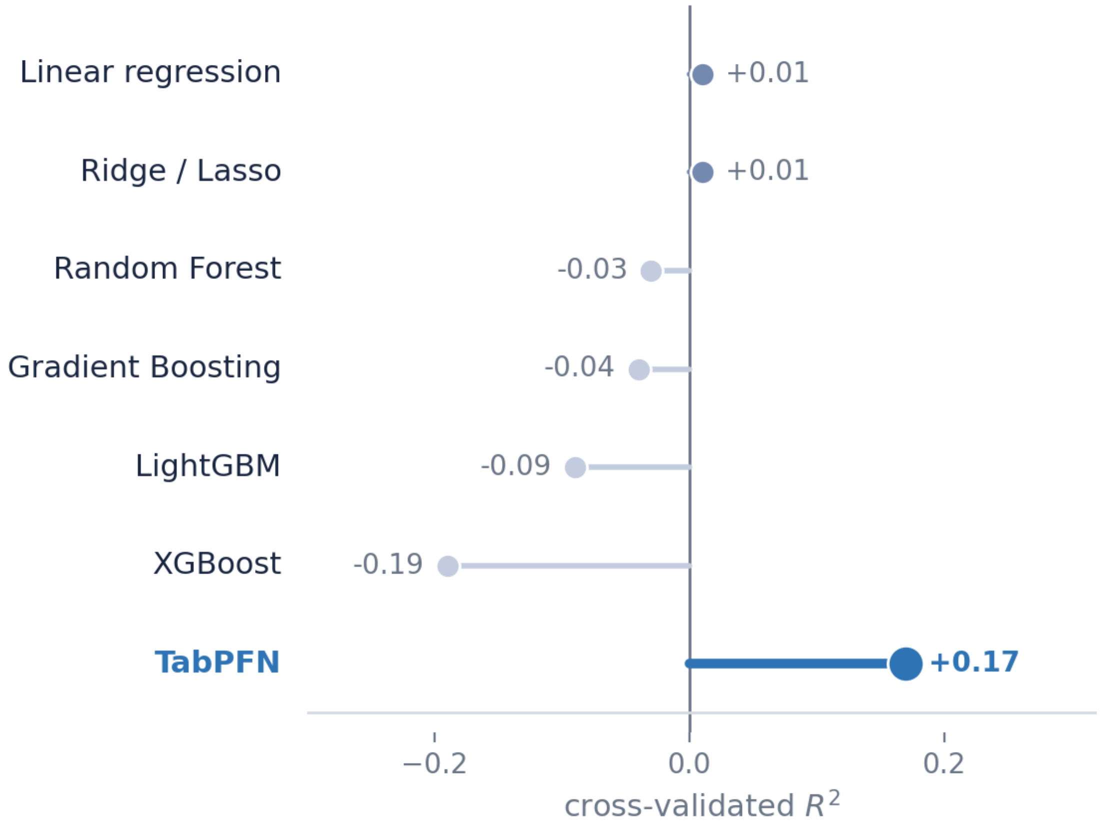
- **OUTCOME** *in quiet, 65 dB SPL, free field*

Mono1 **Freiburger monosyllables score measured 12 months after implantation**

Poor outcome: **Mono1 < 40%** (about 1 in 5) **measured 12 months after implantation**

Prediction of Speech Outcomes

EXPLAINED VARIANCE: *Best $R^2 = 0.17$*



±41 pp interval (60% means, from 19% to 100%)

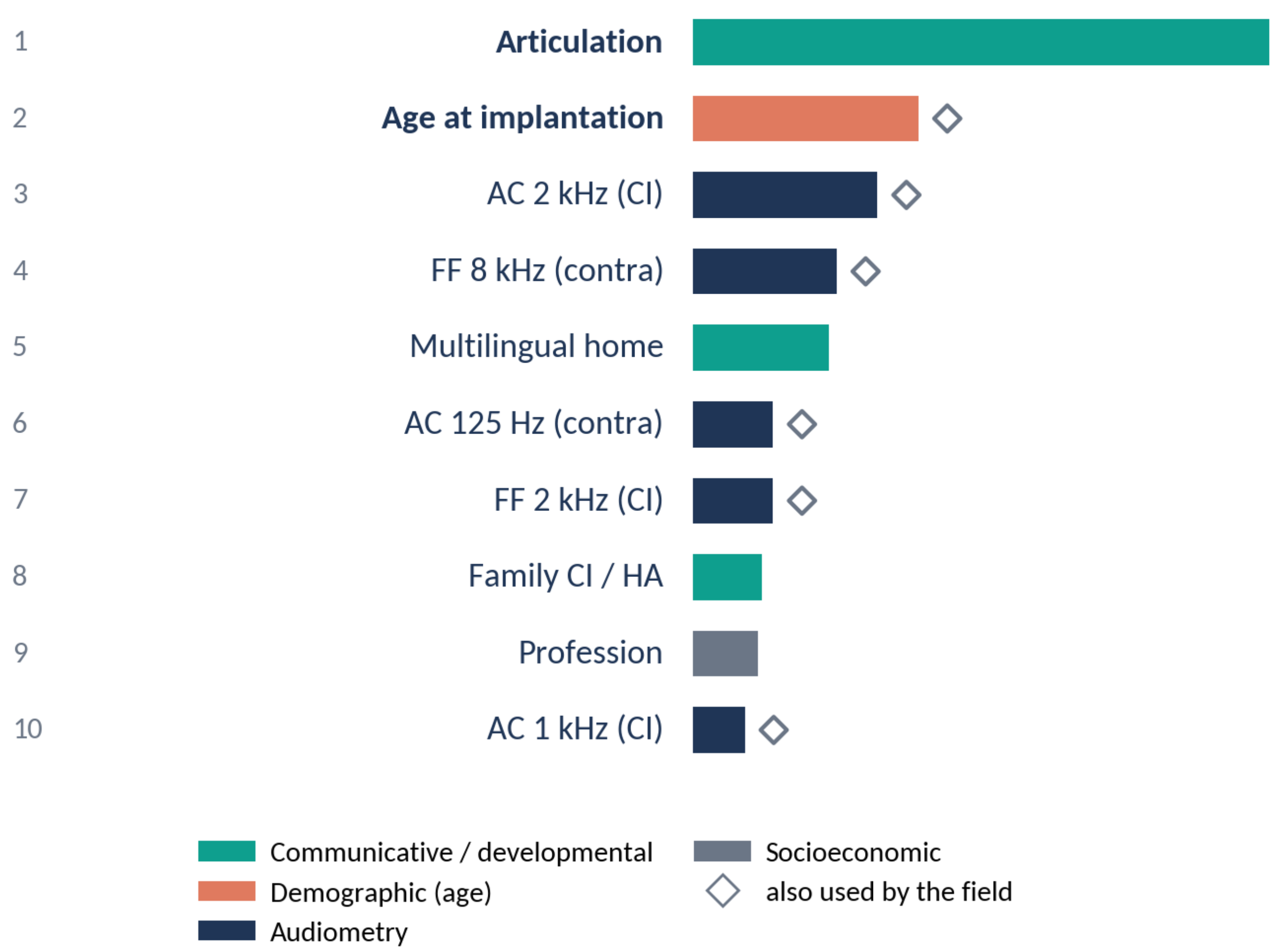
Our pre-op data predict post-op the monosyllabic score with a best $R^2 = 0.17$ with AI

Prediction of Binary Risk for Poor Outcomes

- Reframed as **binary risk of poor outcome** (Mono1 < 40%, about 1 in 5 patients)
- Achieved an Area Under the Curve **(AUC = 0.71)**

Its strongest driver is **articulation**

WHAT DRIVES IT *AI Model variable importance*



◇ = also a top predictor in prior work (Kim 2018, Demyanchuk 2025, Anthonisen 2026, Blamey 2013, Shew 2025). Bars normalized to the top feature; N = 378.

Summary & Conclusions

- The **audiogram shape predicts who benefits**; the **same PTA** can **need opposite levers**
- The **patient's own words add what the tests miss**, most in **mild loss and with age**
- In CI, **after** the implant **the score is hard to predict**, but we can still flag **who is at risk of a poor outcome**
- Behind the **same number**, **different people**; with **rich data & interpretable AI**, we **recover the individual the average hides**

INSTITUT PASTEUR

Institut de l'Audition

University of Zurich[UZH]

Deep Hearing Lab

FONDATION POUR L'AUDITION
À L'ÉCOUTE DE LA VIE

reConnect
Healing hearing and speech disorders

USZ Universitäts Spital Zürich

European Commission
Marie Skłodowska-Curie Actions

Prof Paul Avan,
Prof Hung Thai-Van,
Dr Mareike Buhl,
Dr Perrine Morvan,
Prof Gareth W. Peters

Prof Alexander Huber,
Prof Tobias Goehring

Thank you!