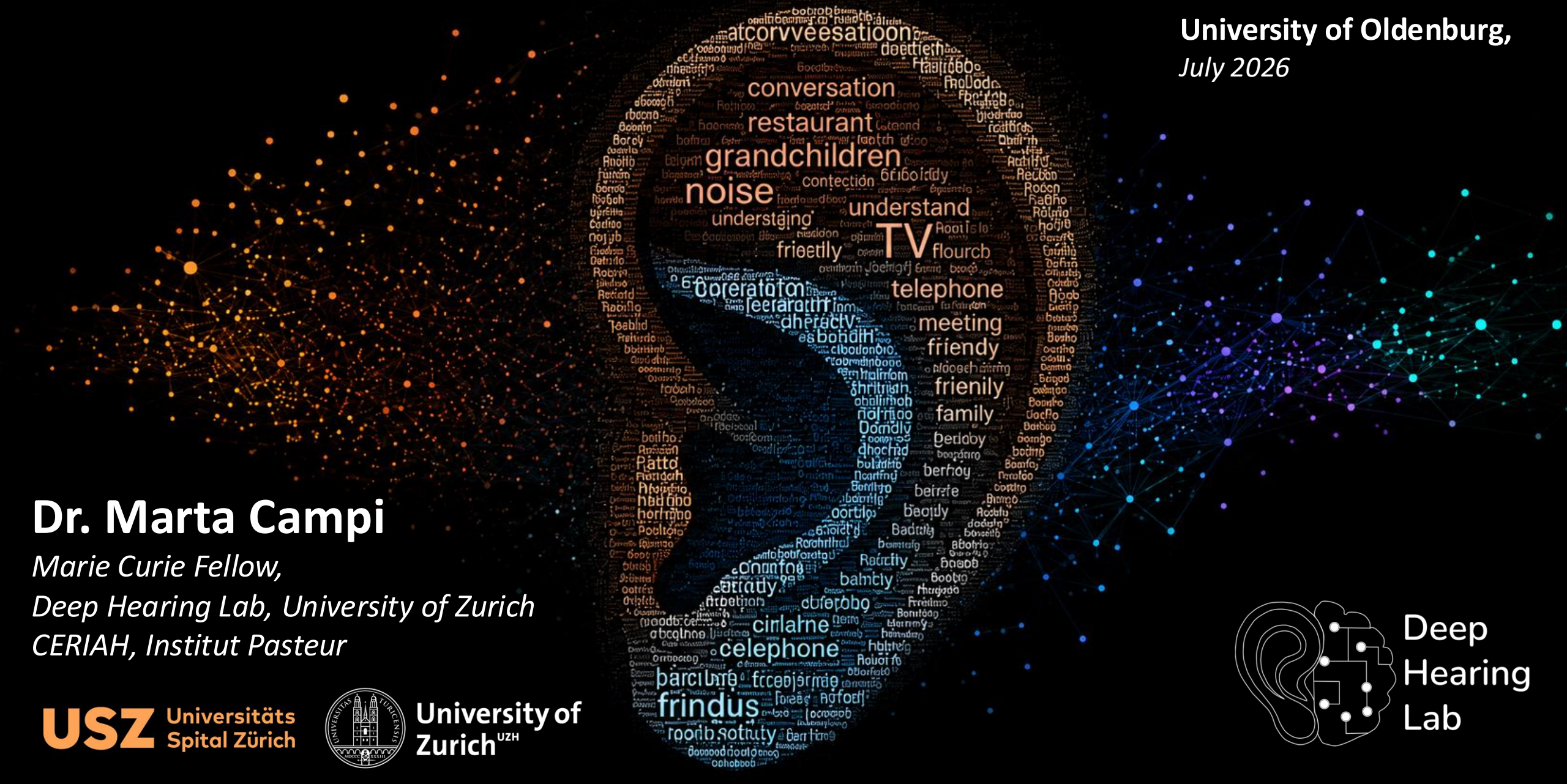
Big Data Approaches to Personalised Hearing Care

University of Oldenburg,
July 2026



Dr. Marta Campi

Marie Curie Fellow,
Deep Hearing Lab, University of Zurich
CERIAH, Institut Pasteur

USZ Universitäts Spital Zürich

University of Zurich[UZH]

Deep Hearing Lab

From Statistics to Hearing Care

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

GB

London

University College London

PhD

- Statistical Science & Signal Processing
- Speech Recognition for Parkinsons' Disease

FR

Paris

Institut Pasteur

Postdoc

- Big Data for audiological care
- Computational modelling for ANSD

CH

Zurich

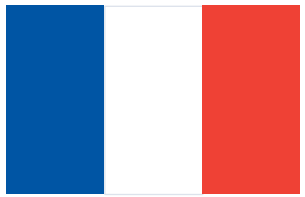
University Hospital Zurich

Marie Curie Fellow

- Outcome prediction in cochlear implants
- Computational modelling for CI

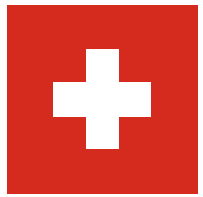
Hearing Loss: A Global Challenge

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

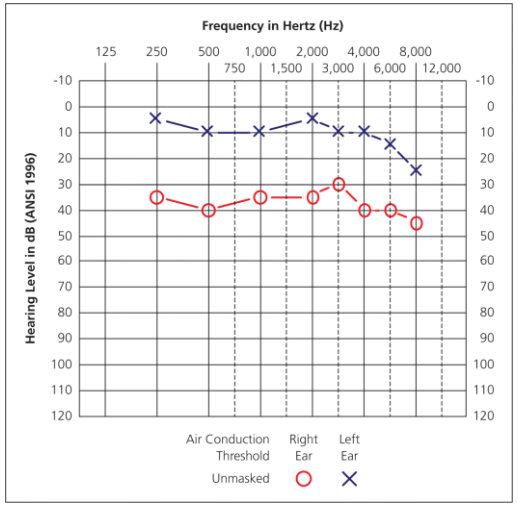
Prevalence of hearing loss, all ages, 2023

The same challenge in every country

Sources: Global Burden of Disease Study 2023 (IHME)

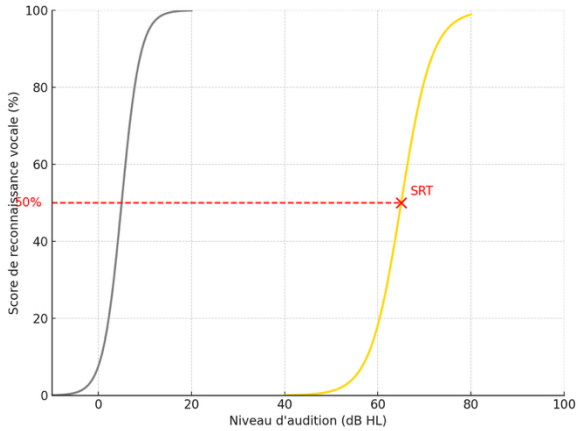
What the Clinic Measures

Pure-tone audiogram



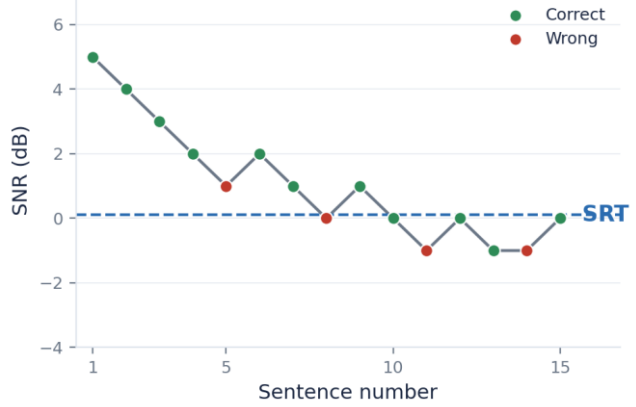
The softest tones a patient can detect

Speech in quiet



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

Speech in noise

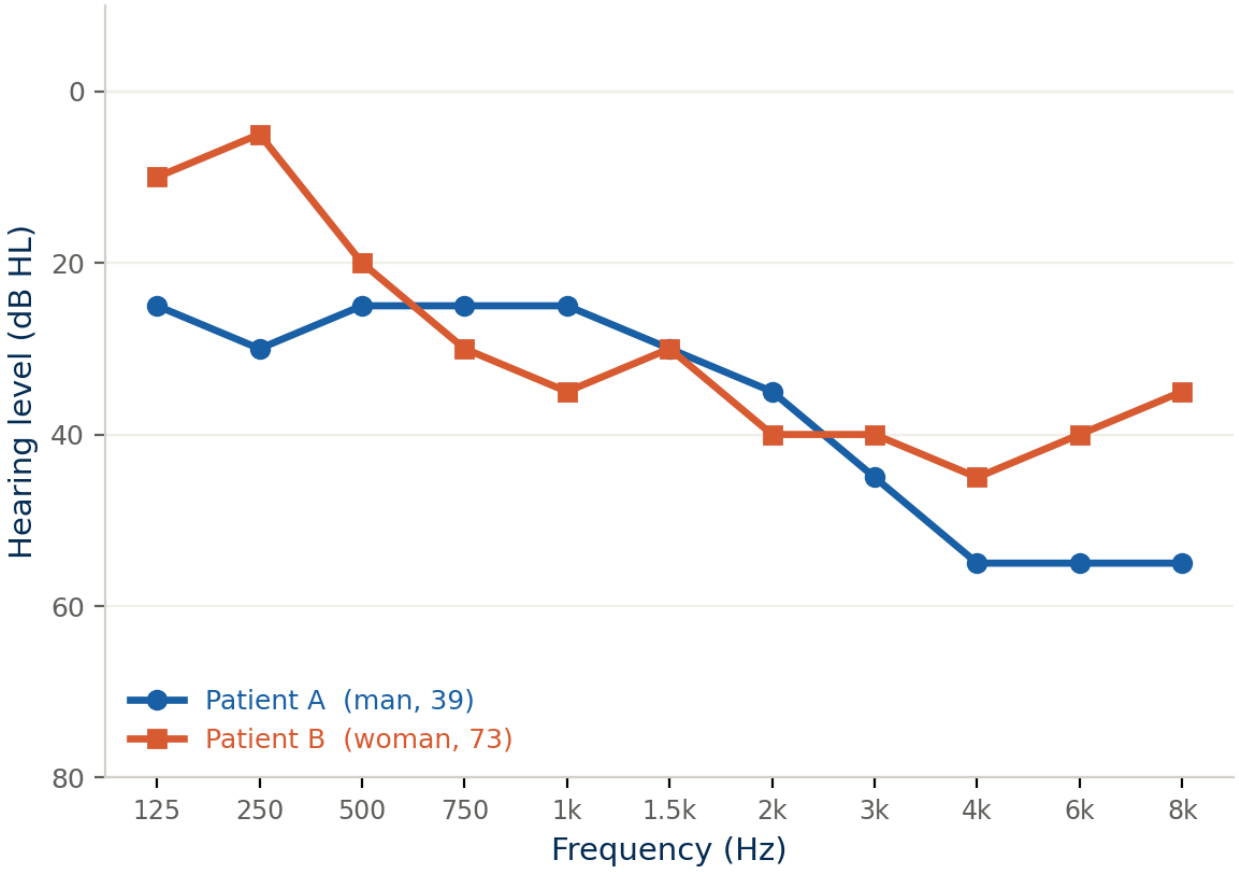


SNR: the noise level at which 50% of speech is understood

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

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**

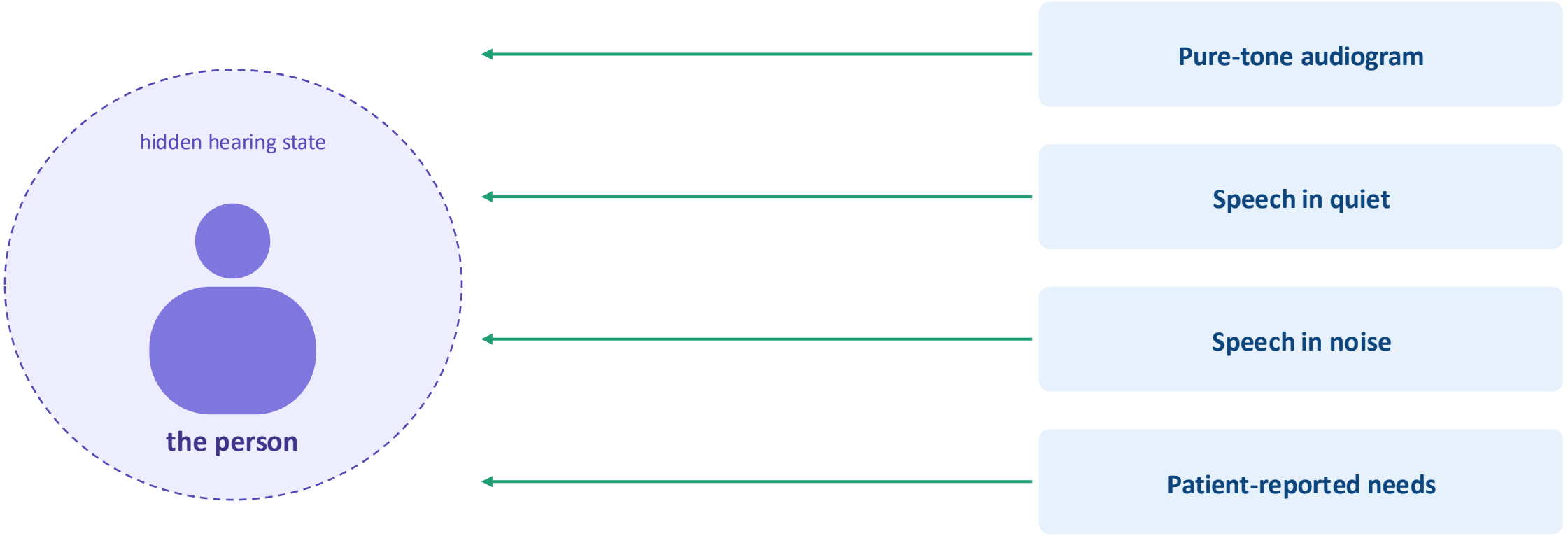
"I want to understand the TV better"

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 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 **the structure** and stay **interpretable**, from statistical models to machine learning

Three Studies

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

		SAMPLE		DATA SOURCE
1.	A reference model for hearing loss	48k		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. A Reference Model for Hearing Loss

MATHEMATICS IN MEDICAL AND LIFE SCIENCES
2025, VOL. 2, NO. 1, 2535979
https://doi.org/10.1080/29937574.2025.2535979

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Standardised hearing loss risk profiles with state-space models

M. Campi[a], G. W. Peters[b], Perrine Morvan[a], Mareike Buhl[a] and Hung Thai-Van[a]

[a]Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS, Fondation Pour l'Audition,Institut de l'Audition, HU reConnect, F-75012, Paris, France; [b]Department of Statistics & Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA

ABSTRACT

Hearing loss is a growing public health concern, affecting millions and contributing to impaired communication, social isolation, and reduced quality of life. As a hidden condition, hearing loss is typically diagnosed through behavioural tests like pure-tone audiometry, which often fail to capture the full extent of auditory deficits. Additional tests, such as speech-in-quiet and speech-in-noise, provide a more detailed understanding, yet they are underutilised due to equipment and time constraints. To explore test complementarity, we developed an advanced method for profiling hearing loss by integrating audiogram and speech test data using state-space models (SSMs). Our approach – the Campi-Peters-Morvan-Buhl-Thai-Van (CPMBT) model – accounts for unobservable aspects of hearing loss and infers common trends across population segments. We established a baseline model using only audiograms and an extended version incorporating speech tests to capture multiple dimensions of auditory function. Through rigorous statistical testing, we demonstrated that the extended model significantly outperforms the baseline model, particularly for individuals with slight to moderate hearing loss. Our analysis revealed distinctive frequency-specific patterns in how speech tests contribute to hearing assessment. The CPMBT model provides standardized risk profiles that quantify age-related and frequency-dependent patterns in hearing loss progression, establishing a national benchmark for clinical practice and public health strategies.

PLAIN LANGUAGE SUMMARY

Hearing loss is a major public health issue affecting millions of people worldwide, leading to difficulties in communication, social isolation, and reduced quality of life. Traditional hearing tests like pure-tone audiometry often miss important aspects of hearing problems, while speech tests that

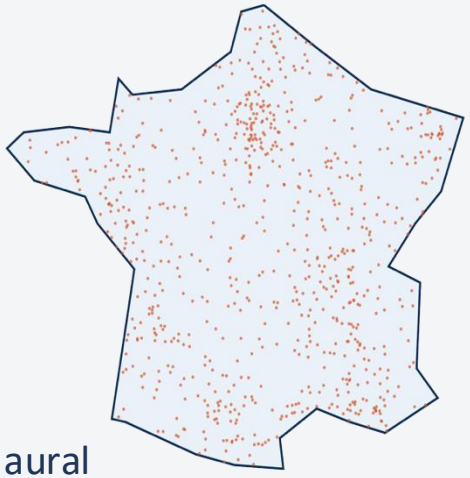
ARTICLE HISTORY
Received 11 March 2025
Revised 26 June 2025
Accepted 10 July 2025

KEYWORDS
State-space model; Kalman filter; likelihood ratio test; hearing loss

The Problem & the Data

Hearing loss is a hidden condition and, unlike cardiology (Framingham) or paediatrics (growth charts), it has **no standardised population benchmark**

THE DATASET



48,144 adults with **symmetric hearing loss** from **780** centers in France, **Amplifon** France

- **Audiogram (11 frequencies)** – headset, per year
- **Speech-in-quiet** - Lafon (French monosyllabic lists): level for 50% words correct, free field, binaural
- **Speech-in-noise** - HINT (European-French, adaptive): SNR for 50% correct (SRT50), fixed noise, free field, binaural

The Audiogram is Not Enough: Audibility vs Clarity

AUDIOGRAM (+ SPEECH in Quiet)

Audibility (attenuation)

"Can you detect the sound?"

The softest sound you need to detect (audiogram) or repeat in quiet

vs

SPEECH in Noise

Clarity (distortion)

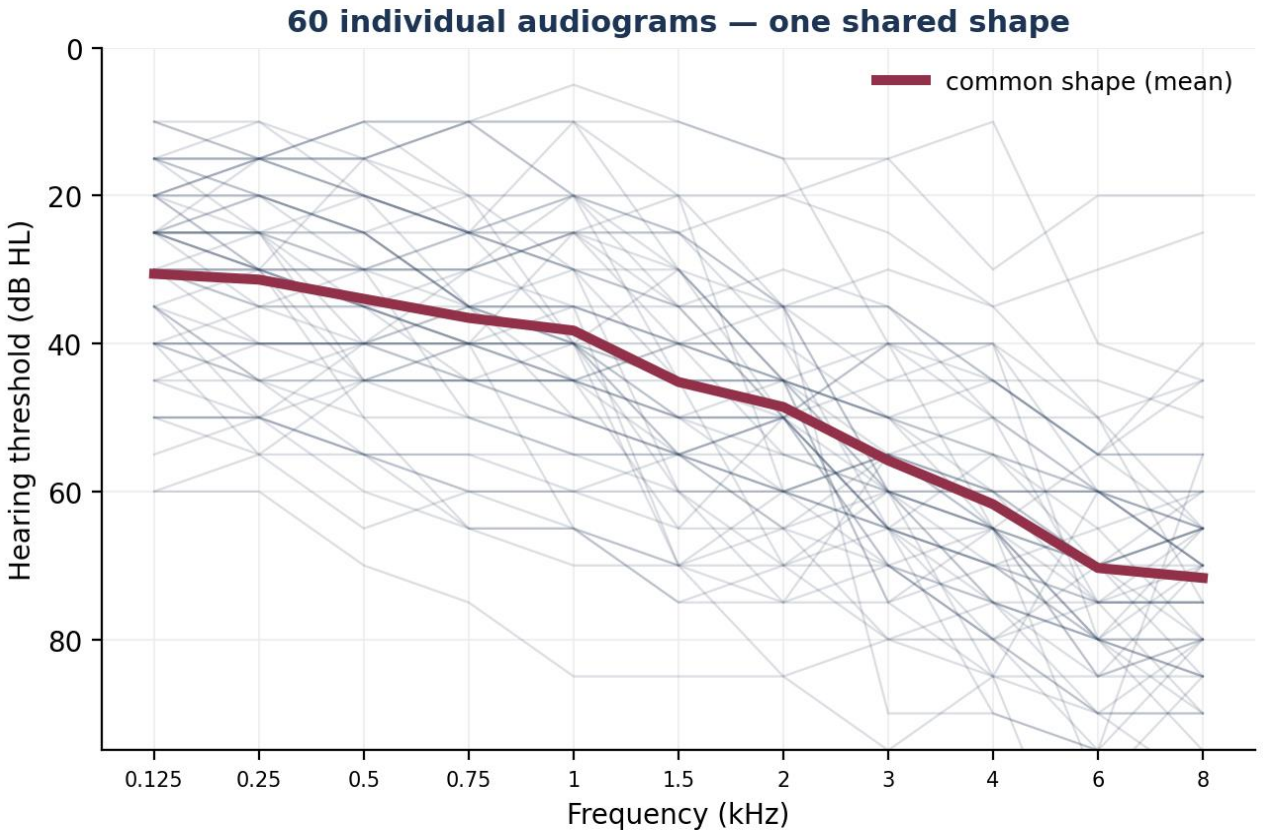
"Can you understand it, in noise?"

Recognising words when it is noisy, closer to daily listening demands

The question: do speech tests add what the audiogram misses, and can we turn it into a reference for hearing?

Attenuation vs distortion framework: Plomp (1978); the noise component as "SNR loss": Killion (2000).

One Trend Behind the Audiogram



The Hypothesis

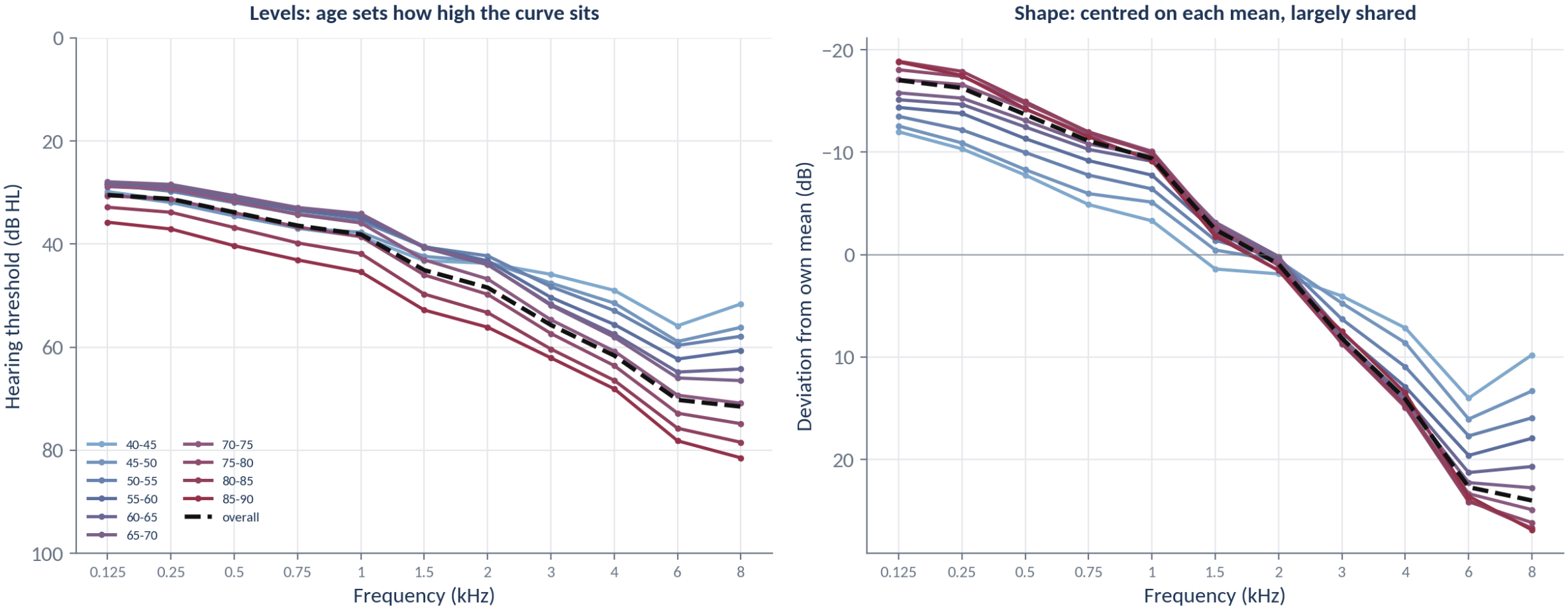
- A **common frequency pattern** underlies all audiograms
- One **latent trend (κf)** captures this pattern
- Age groups differ by **level (α)** and **trend sensitivity (β)**

α *the overall level of loss for that age group.*

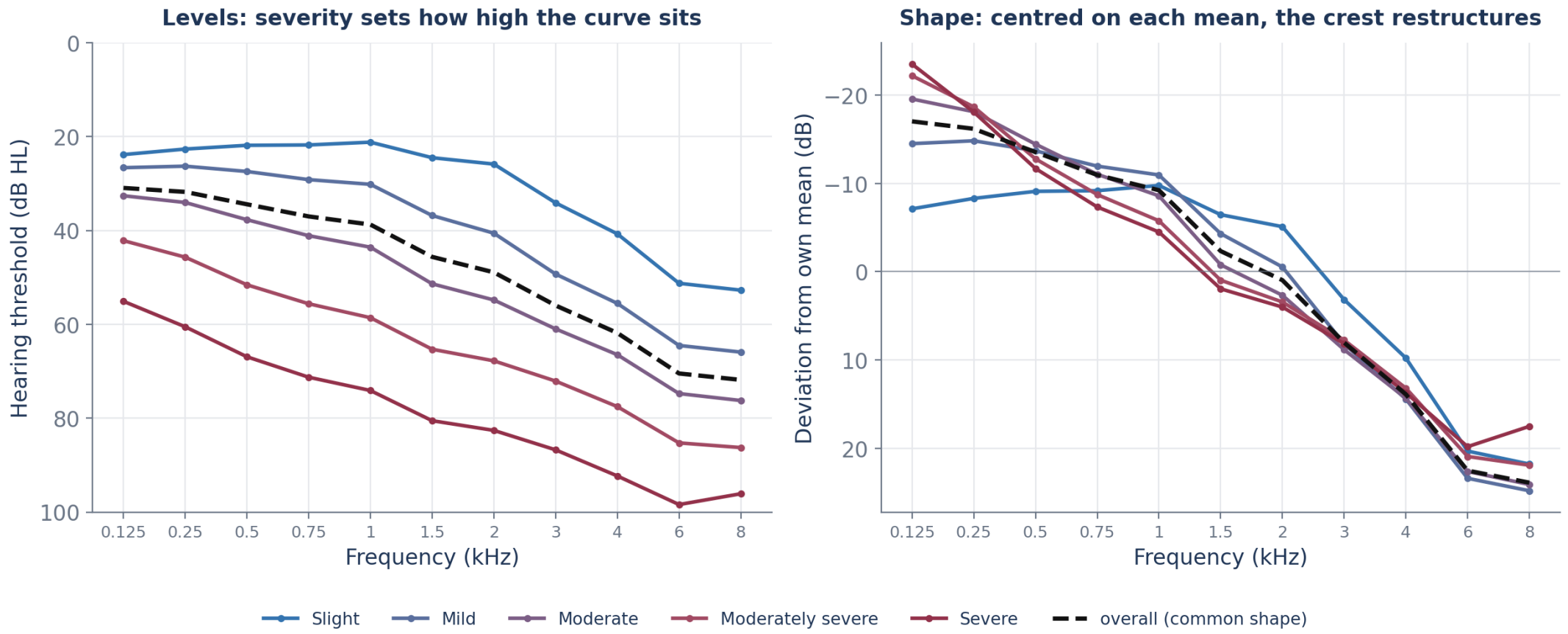
β *how strongly that age group follows the common trend.*

60 individual audiograms; bold line = population mean. Amplifon France, ages 40–90.

One Shape, the Level Shifts with Age



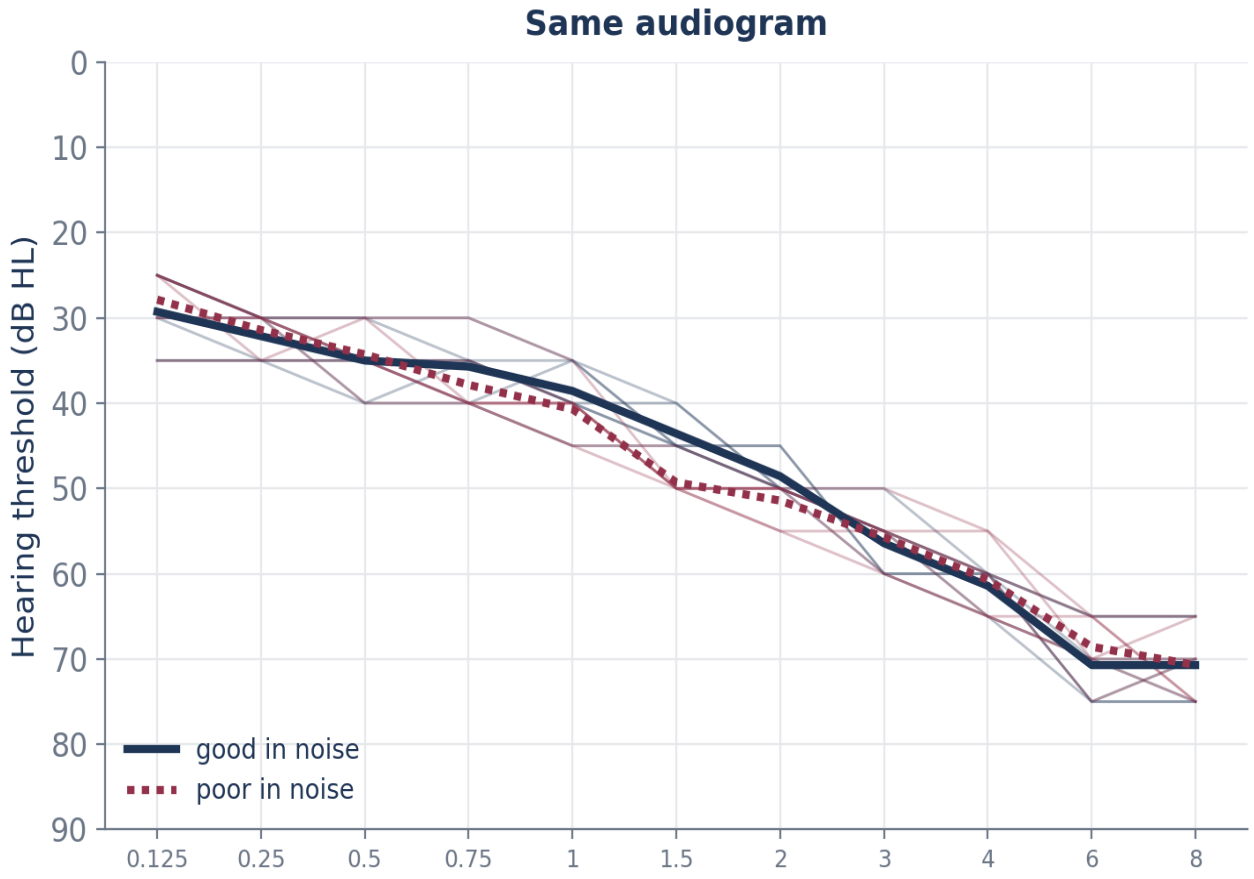
The Same Holds for Any Segment

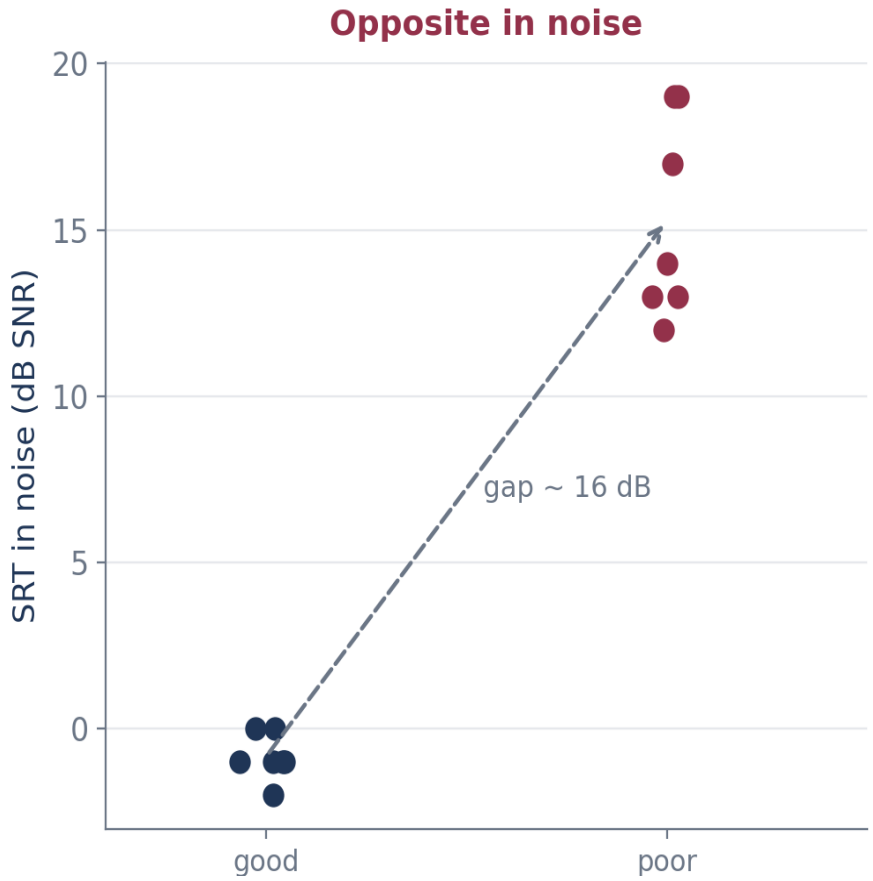


The same decomposition holds for any segment, here by severity: the level shifts (left), the shape stays largely shared (right)

Same Audiogram, Opposite in Noise

14 real patients, matched on audiogram: the case that motivates the model





The audiogram is silent about this. There is information beyond its shape; is it systematic? That question needs the model

Does Speech Add Beyond the Audiogram?

The same model, with and without the speech tests

Baseline
audiogram only

hearing loss across frequency = age level + latent trend

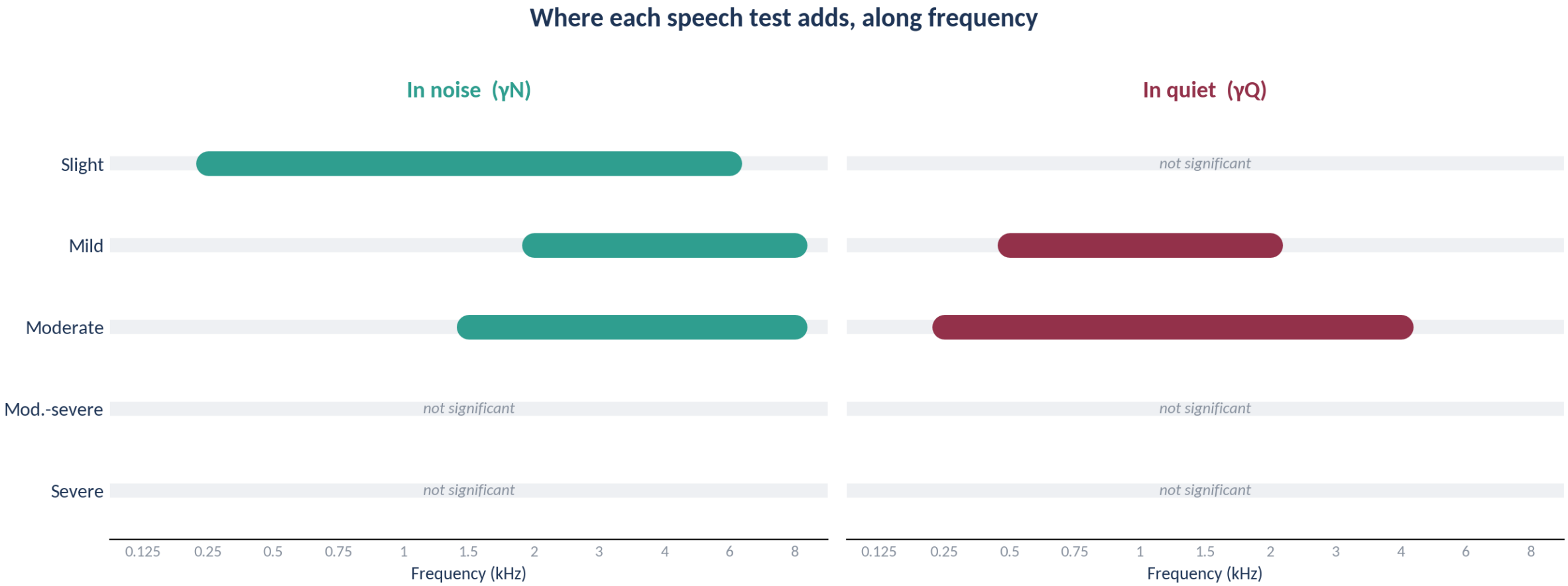
Extended
+ speech tests

hearing loss across frequency = age level + latent trend **+ speech-in-quiet + speech-in-noise**

(More formally: $\tilde{\pi}_{a,f} = \alpha_a + \beta_a \kappa_f + \gamma_Q \pi_Q + \gamma_N \pi_N + \varepsilon,$ *with* $\kappa f = \vartheta + \varphi_1 \kappa f_{-1} + \omega f$ *)*

- *formally, responses are proportions exceeding population quantiles, not raw dB;*
- *κf is a latent AR(1) trend over frequency, adapted from Lee-Carter mortality models (frequency plays the role of time), fit by a Kalman filter and tested by Log-likelihood ratio tests*

..Not the Same Test for Different Hearing Loss Degree



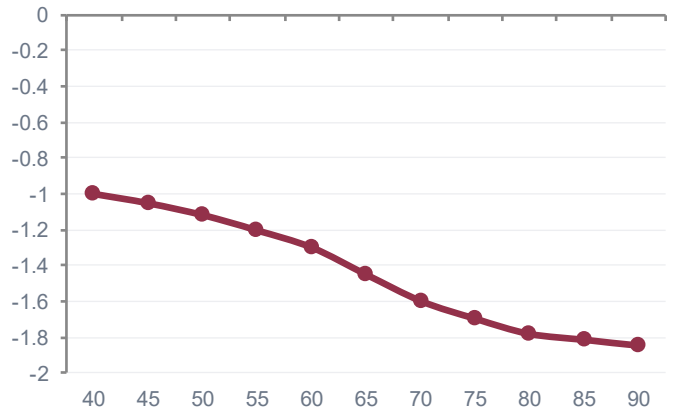
Schematic, read from the paper's text (20% level); exact frequency-by-frequency values are in the paper.

Noise adds early and moves to the highs; quiet fills the mid frequencies in moderate loss; neither adds in severe loss

A Reference Model for Hearing Loss

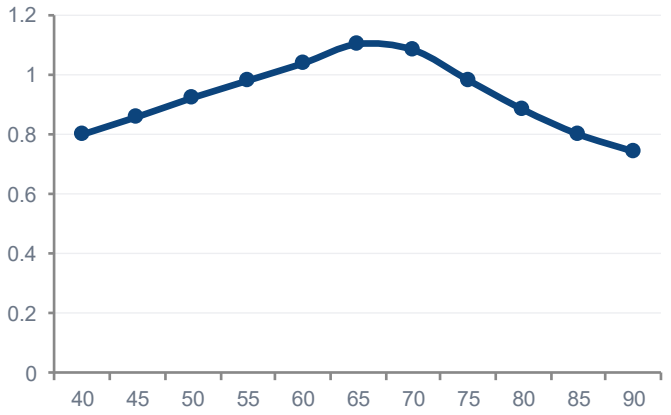
A growth chart for hearing loss: the three curves the model delivers, by age and frequency

The overall loss level at each age



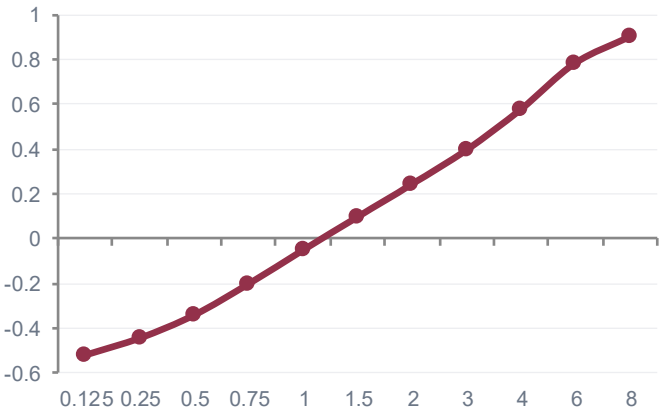
Loss worsens with age; more negative means poorer hearing

How strongly each age follows the shared pattern



Highest at 65 to 70: that age responds most to the trend, the critical window

The shared shape: how loss grows across frequency



Loss accelerates through the speech region, 1 to 3 kHz

Together, these are the reference curves for hearing loss: for screening and expected profiles by age, sex, and severity

Why a Latent Trend, Not Just the PTA?

The PTA collapses the audiogram into one number; κf keeps its shape

One number vs the whole shape

The PTA averages four frequencies into one value; two very different audiograms can share the same PTA. κf keeps the profileacross all eleven frequencies.

Data-driven, not a fixed convention

The PTA frequencies are a historical choice. κf is estimated from the data: the shape that explains how frequencies co-vary in this population.

Level and shape, not only a level

The PTA gives only severity, a level. The model separates level (α) from shape (κf) and sensitivity (β): κf contains the PTAand adds what it discards.

It builds a benchmark

A single number cannot place a patient on a reference. α, β and κf give population curves: the growth chart the PTA cannot build.

κf does not replace the PTA; it keeps the whole audiogram shape, which is what a reference needs.

Some Take Home from Study 1

1. **A common trend**

   11 audiometric frequencies collapse into one interpretable latent trend

2. **Speech adds where it's silent**

   Beyond the audiogram, speech-in-noise carries real information, mild-to-moderate loss

3. **A population benchmark for hearing**

   Standardised risk profiles: a growth chart for hearing

Method paper. To be applied to more complex scenarios

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

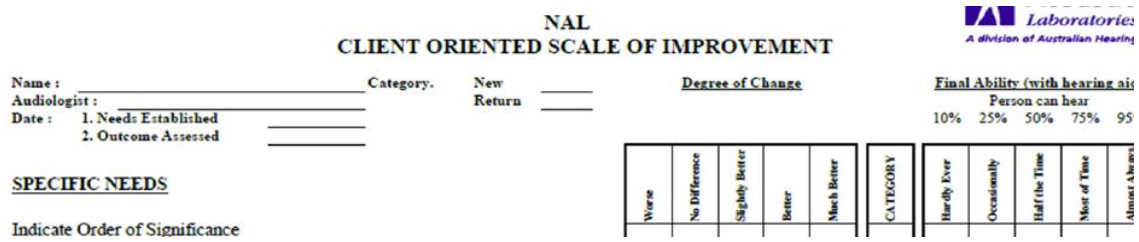
- A conversation at home
- A noisy restaurant
- The telephone
- Following a group
- Meetings at work
- Television, music
- Grandchildren's voices
- Hearing in the street

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 : _____ Category. New ____
Audiologist : _____ Return ____
Date : 1. Needs Established ____
2. Outcome Assessed ____

SPECIFIC NEEDS

Indicate Order of Significance

Degree of Change

Worse | No Difference | Slightly Better | Better | Much Better | CATEGORY

Final Ability (with hearing aid)
Person can hear
10% 25% 50% 75% 95%

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***

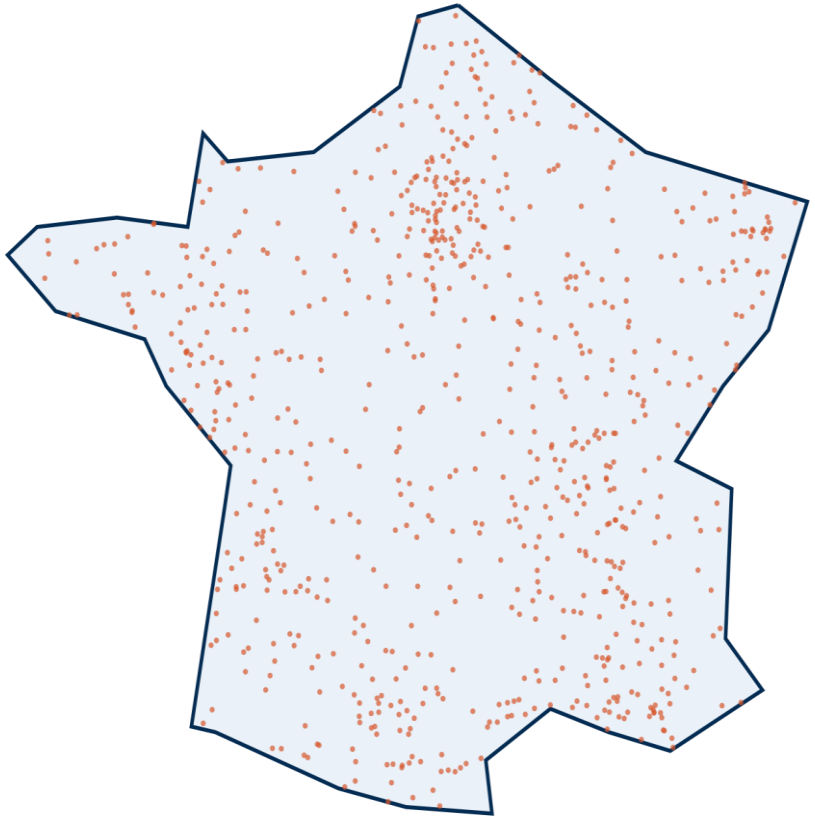
Data & Research Question

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?

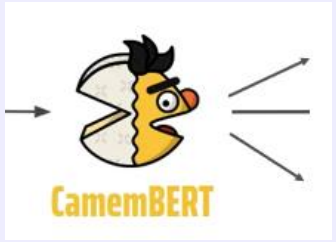
Reading the Patient's Words

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

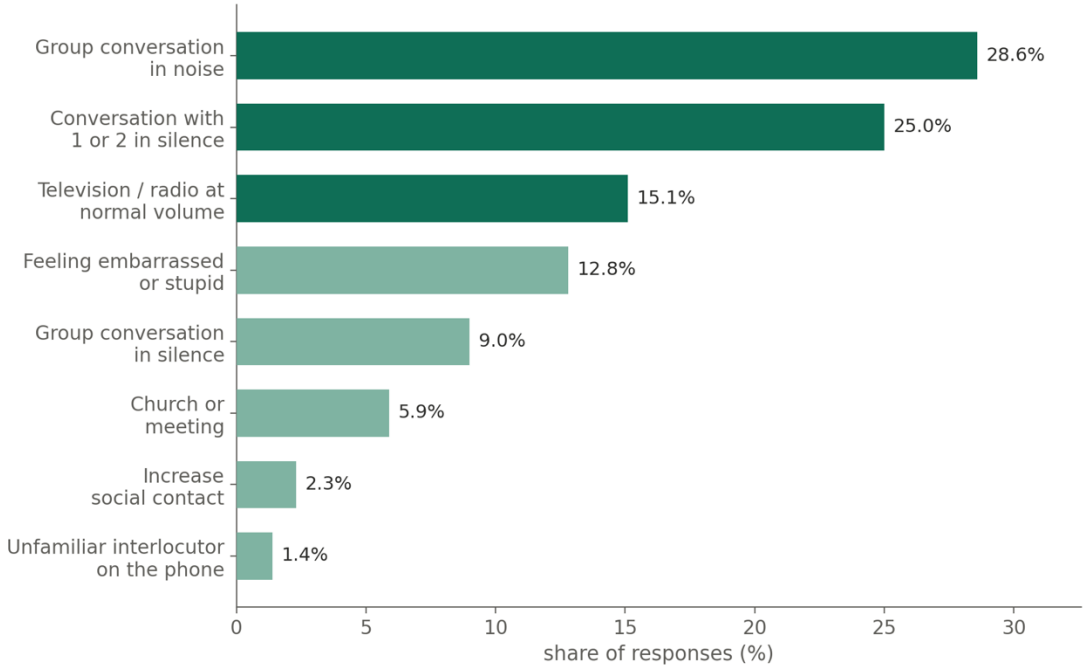
THE 16 COSI CATEGORIES

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

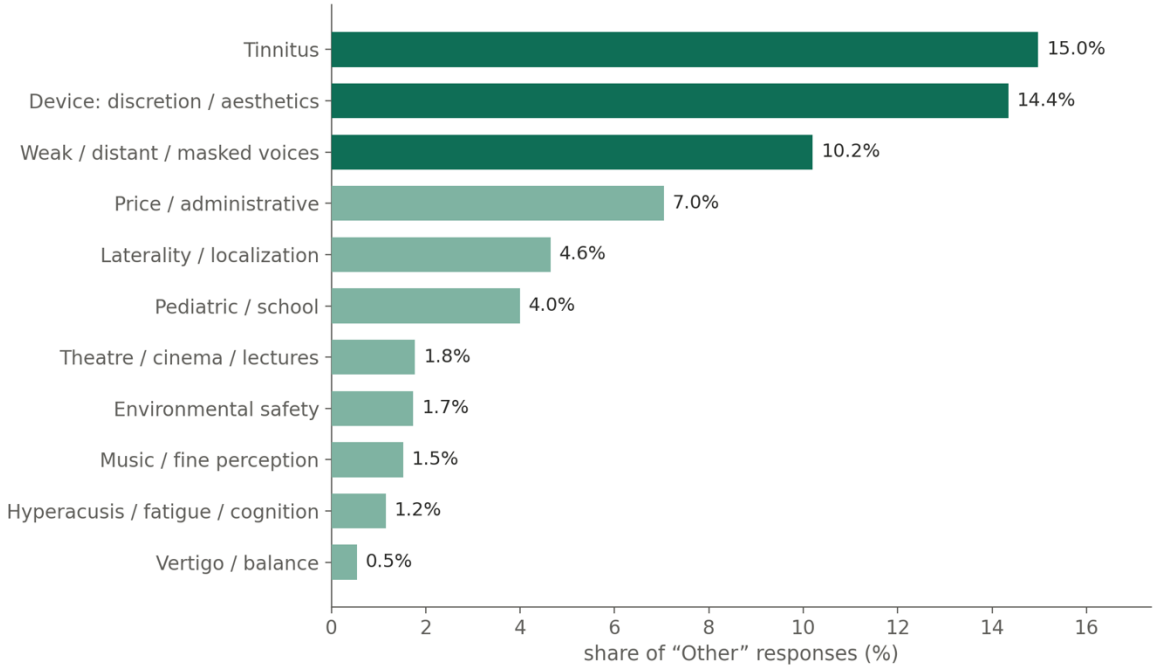
Hand-labelled on 1% (7,232 answers), then applied automatically to all 723,255, and verified

What patients ask for, and what the fixed form misses

THE 16 GIVEN CATEGORIES



INSIDE THE “OTHER” BUCKET



Nearly one in four COSI responses (23.8%, 71,961 of 302,204) falls into “Other”

Do the Patient's words Add Anything?

THE MODEL:

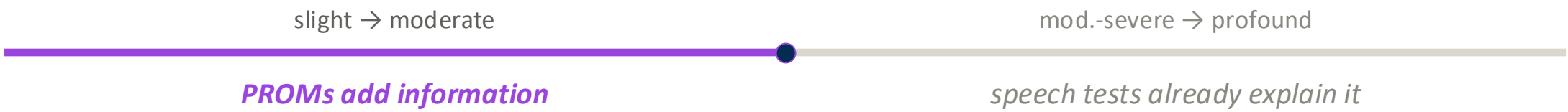
We explain the degree of loss (PTA), and ask if the patient's words add to the speech tests:

A **PTA ~ SNR + SRT**

B **PTA ~ SNR + SRT + PROMs**

RESULTS:

ACROSS THE DEGREE OF LOSS



AND WITHIN MILD LOSS, ACROSS AGE

Limitations and where this goes next

*Here the audiogram (PTA) is a severity **anchor**, not a prediction target; the conclusion does not depend on the direction. The stronger tests come next.*

01

Incremental validity toward a real outcome

Predict hearing-aid benefit (COSI improvement, final ability) from audiometry vs audiometry plus PROMs, out of sample. The gold standard for "words add beyond the tests."

02

Shared-information analysis

Canonical correlation and commonality between {PTA, SNR, SRT} and PROMs: shows non-redundancy with no dependent variable to argue about.

03

A latent two-factor model

Separate an audiometric dimension from a patient-experience dimension, correlated but distinct. The claim, stated directly: PROMs measure what audiometry cannot.

04

Need as the outcome

Model the stated need itself (ordinal or multinomial) from audiometry and demographics, in the natural causal direction.

The result stands either way: PROMs carry hearing-related information the tests miss, concentrated in mild loss

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 **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

Q1 Prediction of speech outcomes

Regression: can we predict the score?

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

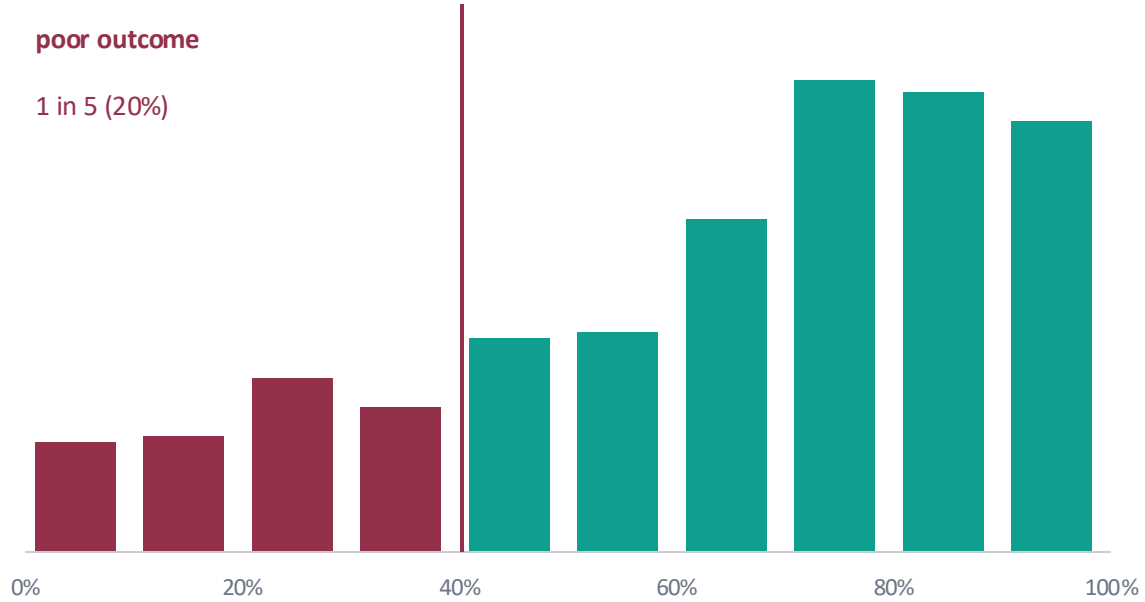
Q2 Prediction of binary risk for poor outcomes

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

Post-op speech outcome (Mono1)



Mono1 score (% words correct, in quiet, at 12 months)

464 adults

median age 57 (5–90) · bilateral · onset: 256 post-lingual, 55 pre-lingual

OUTCOME

Freiburg monosyllables (Mono1), % correct in quiet at 12 months post-implant

Poor outcome: Mono1 < 40%

About 1 in 5 — the tail we most want to predict before surgery

79 PRE-OPERATIVE PREDICTORS

46 acoustic

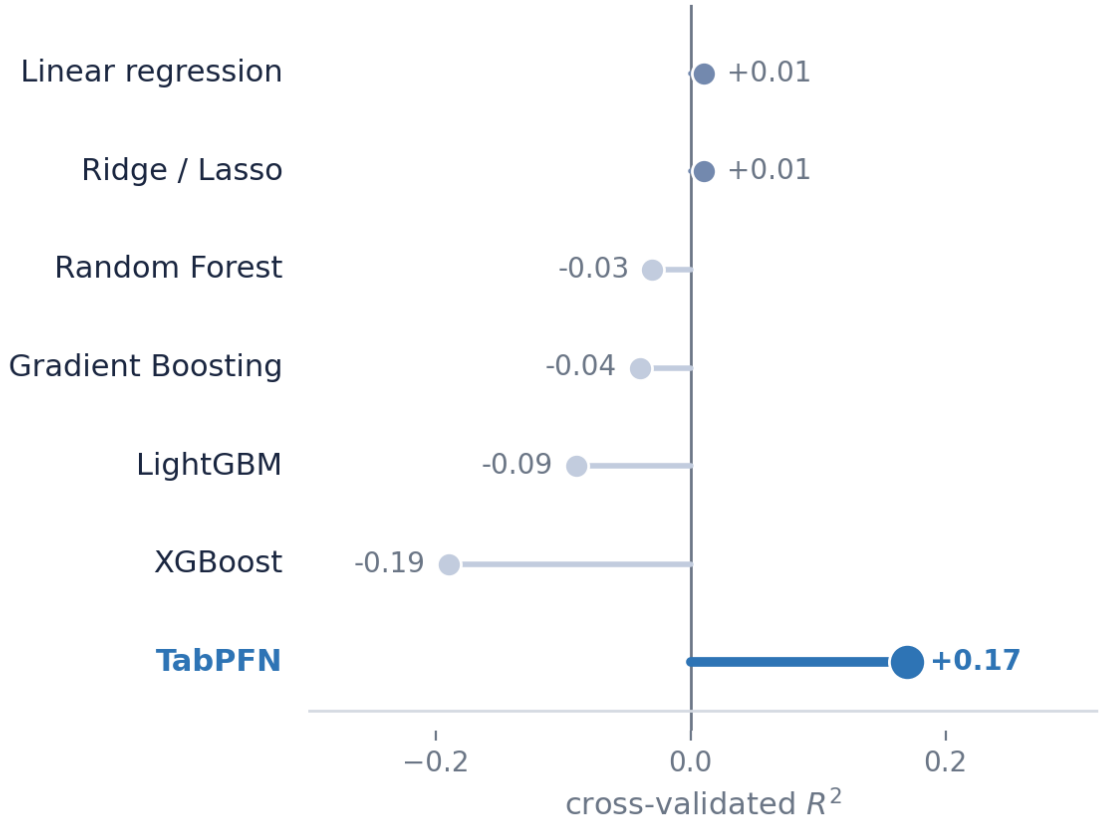
pure-tone audiometry (thresholds, PTA) · pre-op speech (Mono1 / Mono2) · residual hearing

33 non-acoustic

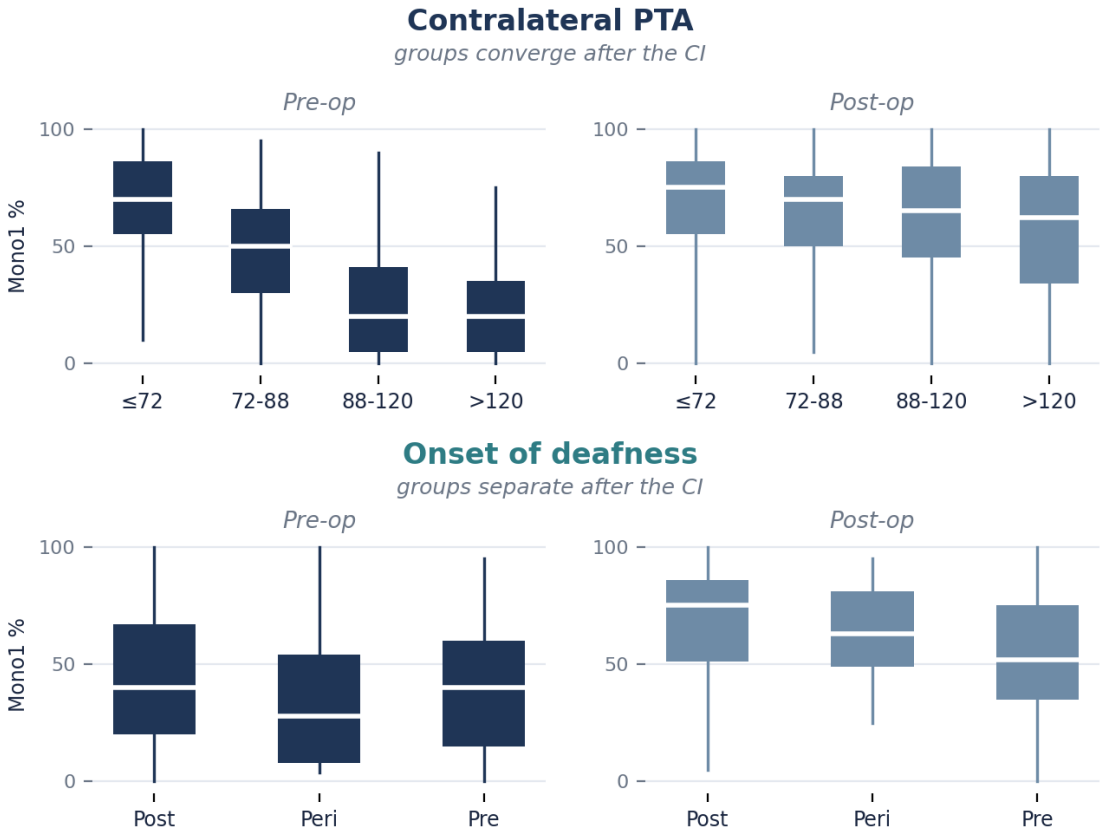
age at implant · onset (pre/post-lingual) · deafness duration · etiology · device & electrode · test language

Prediction of Speech Outcomes

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



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



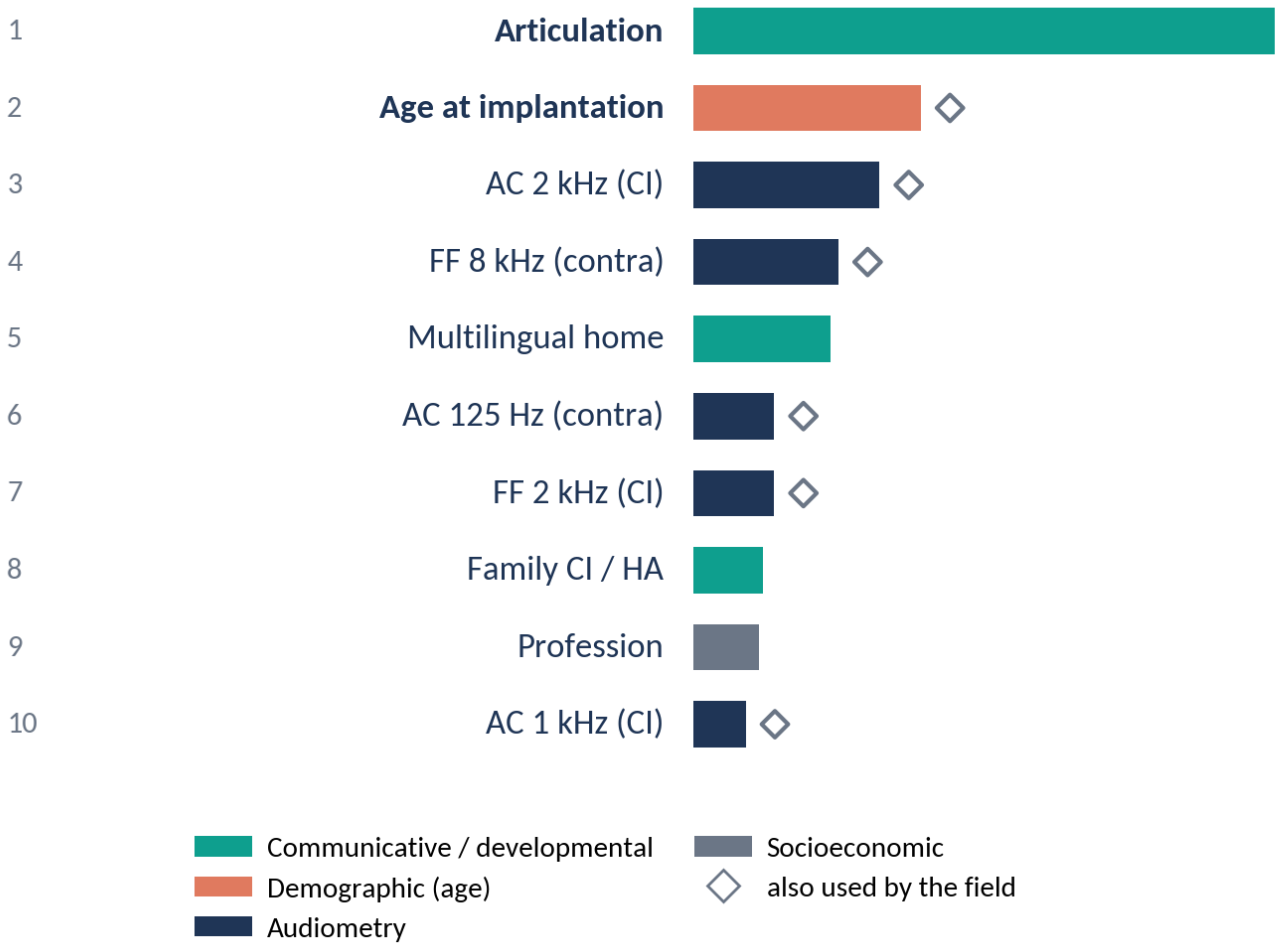
Our pre-op data predict post-op the monosyllabic score with a best $R^2 = 0.17$ with AI

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

- Behind the audiogram lies **one shared trend**, giving hearing a **standardized benchmark**, and **speech adds what it misses in early loss**
- 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**

University of Zurich[UZH]

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!