

HOW SUPERCOMPUTERS & HIGH-PERFORMANCE PROGRAMMING LANGUAGES ENABLE

Radical Medical Breakthroughs & Exponential Scientific Discoveries

Zak Allal, MD, MBA

Polymath

Polymath *def. A person whose knowledge spans many different fields.*

- Medical doctor
- Clinical researcher and neuroscientist
- Technology innovator and programmer
- Geopolitical and strategic risk expert
- Professional pianist and composer



THE CURRENT TREND

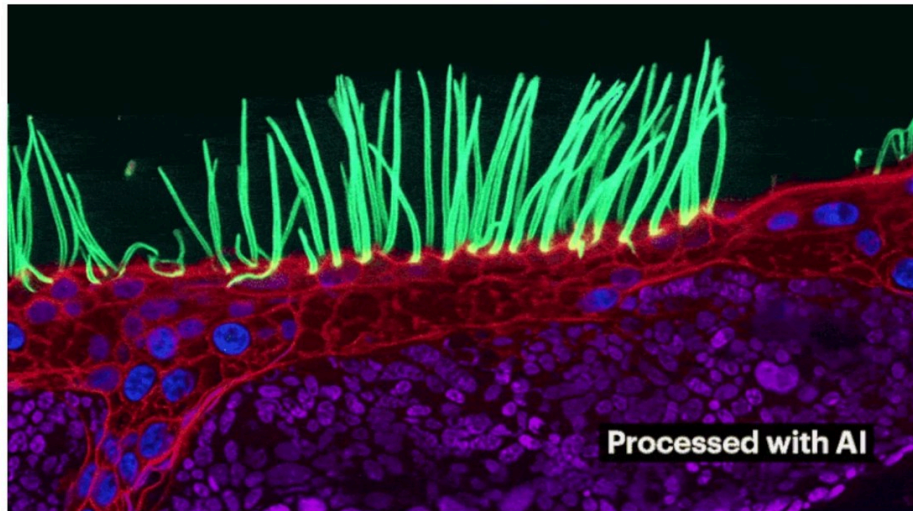
We talk about AI coding, vibe coding, and efficient software development, but rarely about AI and software engineering solving humanity's greatest challenges.

Award-winning microscopy image disqualified over AI use

The incident has prompted discussions about the use of artificial intelligence in scientific imagery.

By [Rachel Fieldhouse](#)

[Save article](#)



A winning video showing beating airway cilia has been disqualified from the Small World in Motion competition for failing to comply with competition rules. Credit: Ning Xu/Nikon Small World

AI

Palantir CTO says some AI researchers are taking up smoking because they expect AI to cure cancer

By [Thibault Spirlet](#) [+ Follow](#)



Palantir's CTO Shyam Sankar said taking up smoking has become a way for AI researchers to show how much they believe in AI. Tom Brenner/Bloomberg via Getty Images

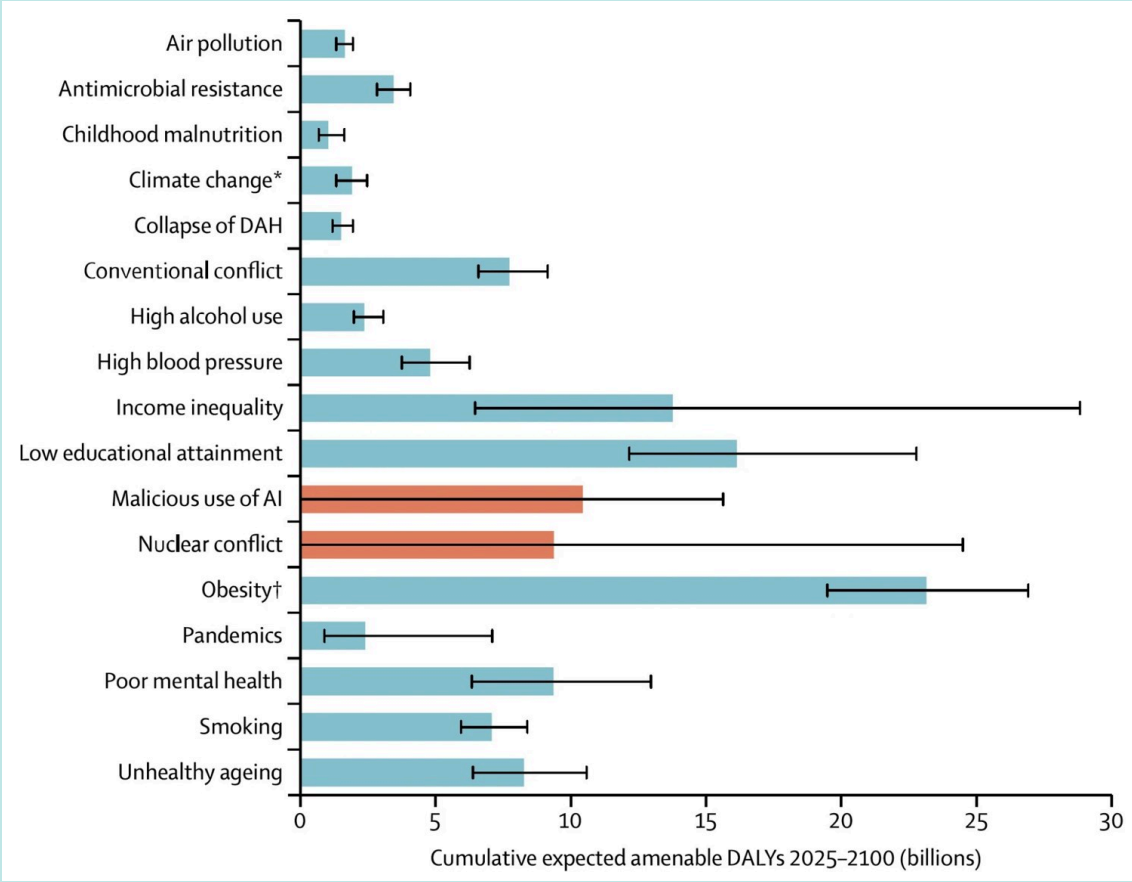
THE LANCET COMMISSIONS · Online first, October 10, 2026

The *Lancet* Commission on 21st century global threats to health

[Prof Christopher J L Murray, DPhil](#) ^{a,b,*} · [Natalia Kanem, MD](#) ^{c,*} · [Shameran Abed, Bar-at-Law](#) ^d · [Prof Ibrahim Abubakar, PhD](#) ^e · [Ifedayo M Adetifa, PhD](#) ^{f,g,h} · [Ximena P Aguilera, MD](#) ⁱ · et al. [Show more](#)

[Affiliations & Notes](#) ▾ [Article Info](#) ▾ [Linked Articles \(1\)](#) ▾

» This report from the *Lancet* Commission on 21st century global threats to health lays out a foresight-driven, health-focused approach to identifying and prioritising major threats to global health over the remainder of the century. We argue that recent global shocks (including COVID-19, Ebola Bundibugyo virus disease, major geopolitical conflict, accelerating climate impacts, rapid shifts in development assistance, and the emergence of powerful artificial intelligence [AI] systems) show how quickly assumptions about stability can be overturned—and why governments and multilateral institutions must counter short-termism and localism with systematic, repeatable risk assessment and preparedness planning.



WHAT IF ?

What if architecture and engineering
can lead to a major **medical
breakthrough** without blackbox AI ?

Blackbox AI *def.* An AI system that gives an answer without showing how it reached it, so the answer cannot be inspected, explained or audited.



THE **COST** OF REPETITION

Question:

WHAT WEATHER IS IT OUTSIDE ?

WHAT WEATHER IS IT OUTSIDE ?

STARTING POINT:

YOU ARE SITTING HERE INSIDE THE BUILDING

WHAT WEATHER IS IT OUTSIDE ?

STARTING POINT: YOU ARE SITTING HERE INSIDE THE BUILDING

METHOD & TOOLS PRIMARY RELIABLE METHOD: GO CHECK YOURSELF

WHAT WEATHER IS IT OUTSIDE ?

STARTING POINT:

YOU ARE SITTING HERE INSIDE THE BUILDING

METHOD & TOOLS

PRIMARY RELIABLE METHOD: GO CHECK YOURSELF

SECONDARY RELIABLE METHOD: ASK SOMEONE YOU KNOW

WHAT WEATHER IS IT OUTSIDE ?

STARTING POINT:

YOU ARE SITTING HERE INSIDE THE BUILDING

METHOD & TOOLS

PRIMARY RELIABLE METHOD: GO CHECK YOURSELF

SECONDARY RELIABLE METHOD: ASK SOMEONE YOU KNOW

THIRD LESS RELIABLE METHOD: ASK SOMEONE YOU DON'T KNOW

WHAT WEATHER IS IT OUTSIDE ?

HYPOTHESIS

The weather is rainy.

WHAT WEATHER IS IT OUTSIDE ?

HYPOTHESIS

The weather is rainy.

EXPERIMENT

Go and check in person.

WHAT WEATHER IS IT OUTSIDE ?

HYPOTHESIS

The weather is rainy.

EXPERIMENT

Go and check in person.

RESULTS

The objective physical observation.

WHAT WEATHER IS IT OUTSIDE ?

HYPOTHESIS

The weather is rainy.

EXPERIMENT

Go and check in person.

RESULTS

The objective physical observation.

COST

About 1 minute, walking, calories.

WHAT WEATHER IS IT OUTSIDE ?

COST

time, effort, energy.

| WHAT IF I ASKED YOU TO REPEAT IT ?

WHAT IF I ASKED YOU TO REPEAT IT ?

1 TIME

WHAT IF I ASKED YOU TO REPEAT IT ?

2 TIMES

WHAT IF I ASKED YOU TO REPEAT IT ?

3 TIMES...

WHAT IF I ASKED YOU TO REPEAT IT ?

4 TIMES...

WHAT IF I ASKED YOU TO REPEAT IT ?

| **COST & EFFORT BECOME A BARRIER**

WHAT IF I ASKED YOU TO REPEAT IT ?

**THE COST & EFFORT INFLUENCE
YOUR DECISION TO REPEAT IT**



**I INSIST THAT YOU REPEAT THE
EXPERIMENT**

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK A FRIEND
1 TIME

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK A FRIEND
2 TIMES

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK A FRIEND
3 TIMES...

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK A FRIEND
4 TIMES...

WHAT IF I ASKED YOU TO REPEAT IT ?

| **COST & EFFORT BECOME A BARRIER**

| WHAT IF I ASKED YOU TO REPEAT IT ?

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK SOMEONE YOU DON'T KNOW
1 TIME

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK SOMEONE ELSE YOU DON'T KNOW
2 TIMES

WHAT IF I ASKED YOU TO REPEAT IT ?

YOU ASK SOMEONE ELSE YOU DON'T KNOW
3 TIMES

WHAT IF I ASKED YOU TO REPEAT IT ?

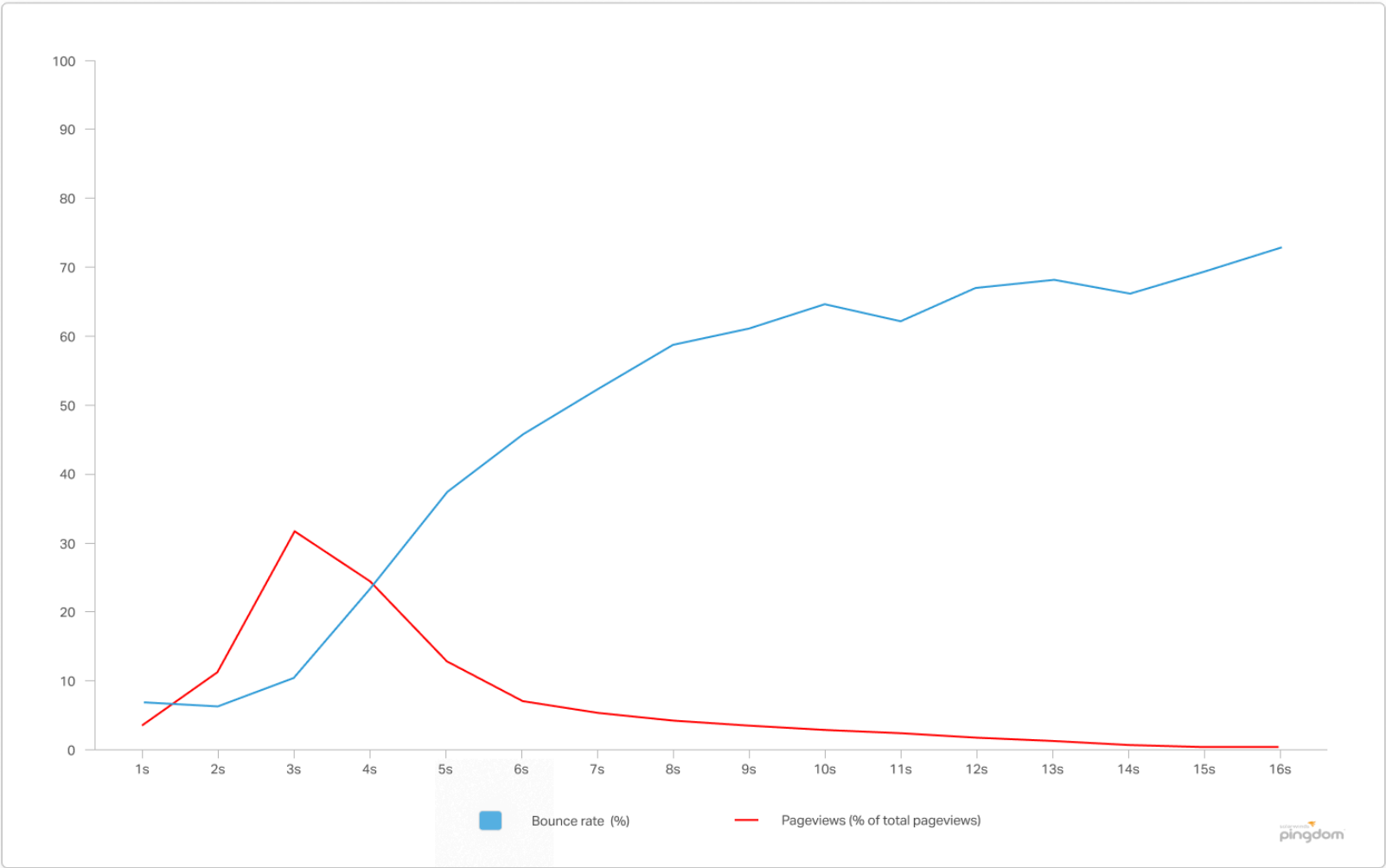
YOU ASK SOMEONE ELSE YOU DON'T KNOW
4 TIMES...

WHAT IF I ASKED YOU TO REPEAT IT ?

**COST & EFFORT BECOME A BARRIER
& ULTIMATELY IMPACT THE QUALITY**

COST & EFFORT BECOME A BARRIER

Web Page Load Time vs Bounce Rate



| WHAT IF WE USED A BETTER TOOL?

WHAT IF WE USED A BETTER TOOL?

CRITERIA 1

RELIABLE WEATHER CHECK.

CRITERIA 2

THE COST DOESN'T IMPACT REPETITION

WHAT IF WE USED A BETTER TOOL?

CRITERIA 1	RELIABLE WEATHER CHECK.
CRITERIA 2	THE COST DOESN'T IMPACT REPETITION
SOLUTION	AN API.

WHAT WEATHER IS IT OUTSIDE ?

HYPOTHESIS

The weather is rainy.

EXPERIMENT

USE AN API: PHONE APP, WEB...

RESULTS

A TRUSTED OBSERVATION YOU CAN BOTH REPEAT.

COST

ALMOST NONE: 5-10 SECONDS.

ALERT ME WHEN THE WEATHER CHANGES

COST

ALMOST NONE: PASSIVE NOTIFICATIONS.

WHAT IF I TOLD YOU THAT...



MEDICAL & CLINICAL RESEARCH IS STILL STUCK IN THE PAST

HOW ?

INADEQUATE TOOLS FOR THE CURRENT MEDICAL GRAND CHALLENGES

Clinical research is mostly developed in Python/Jupyter Notebook, Excel and GUI statistical tools on a personal or basic work computers.

- You formulate a hypothesis.
- You perform an experiment.
- You get the results.
- You publish them.
- One at a time...

HOW ?

INADEQUATE TOOLS FOR THE CURRENT MEDICAL GRAND CHALLENGES

Clinical research is mostly developed in Python/Jupyter Notebook, Excel and GUI statistical tools on a personal or basic work computers.

- The cost of one experiment is already too much.
- With pressure to design AHEAD of time.
- Just like walking outside to check the weather, you wouldn't DARE to REPEAT MORE EXPERIMENTS to afford getting a more granular, nuanced, refined image or answer MORE questions.

WHAT IF THERE WAS AN “API” ?

DECREASING THE COST OF DARING AND TRYING

- The cost of one experiment is already too much.
- The pressure to design AHEAD of time.
- Just like walking outside to check the weather, you wouldn't DARE to REPEAT MORE EXPERIMENTS to afford getting a more granular, nuanced, refined image or answer MORE questions.

WHAT IF THERE WAS AN “API” ?

DECREASING THE COST OF DARING AND TRYING

- Just like checking the weather, cost wouldn't impact REPETITION
- You would get notifications
- and can adjust on the go based on the emerging barriers and challenges

ENTER THE...



SUPERCOMPUTERS

Front page > News > Roihu supercomputer is officially inaugurated



News

28.05.2026

Roihu supercomputer is officially inaugurated

Roihu, Finland's newest supercomputer, was officially inaugurated on 22 May at the LUMI data center in Kajaani. The celebration literally lit up the occasion, as the data center's exterior glowed in Roihu's signature orange for the evening. A distinguished group of invited guests gathered to mark the milestone, with the Finnish Government represented by Minister of Science and Culture Mari Leena Talvitie, who brought official greetings to the ceremony.

Source: <https://csc.fi/en/news/roi-hu-supercomputer-is-officially-inaugurated/>

CSC Roihu: 384 CPU cores per node, up to 60 nodes at once, 784 GB of RAM per node and NVIDIA GH200 GPUs.



CSC office and data center in Renforsin Ranta Business Park, Kajaani.
Photo: Mikael Kanerva, © CSC.



Roihu supercomputer. Photo: Pekka Agarh, © CSC.



LUMI supercomputer, HPE Cray EX. Photo: Mikael Kanerva, © CSC.

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	LineShine - LingKun, LX2 304C 1.55GHz, LingQi, Kylin OS, Shenzhen Cloud Computing Center Co., Ltd. National Supercomputing Centre in Shenzhen (NSCS) China	13,789,440	2,198.40	2,735.82	42,220
2	El Capitan - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,340,000	1,809.00	2,821.10	29,685
3	Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
4	Aurora - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
5	JUPITER Booster - BullSequana XH3000, GH Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, RedHat Enterprise Linux, Bull EuroHPC/FZJ Germany	4,801,344	1,000.00	1,226.28	15,794
6	HPC7 - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, RHEL 9, HPE Eni S.p.A. Italy	3,461,472	571.50	861.13	8,735
7	Eagle - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure United States	2,073,600	561.20	846.84	
8	HPC6 - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A. Italy	3,143,520	477.90	606.97	8,461
9	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu Interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
10	Alps - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	574.84	7,124

Global top 10

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
11	LUMI - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107
91	Roihu-G - BullSequana XH3000, GH Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR, Redhat Enterprise Linux, Bull CSC (Center for Scientific Computing) Finland	107,712	23.50	27.80	381
193	Roihu-C - BullSequana XH3000, AMD EPYC 9965 192C 2.25GHz, Infiniband NDR200, Redhat Enterprise Linux, Bull CSC (Center for Scientific Computing) Finland	186,624	9.69	12.86	705
261	LUMI-C - HPE Cray EX, AMD EPYC 7763 64C 2.45GHz, Slingshot-10, HPE EuroHPC/CSC Finland	194,560	6.30	7.63	1,216
290	MAHTI - Bull Sequana XH2000 , AMD Rome 7H12 64C 2.6GHz, Mellanox HDR100, Bull CSC (Center for Scientific Computing) Finland	169,728	5.39	7.06	1,070

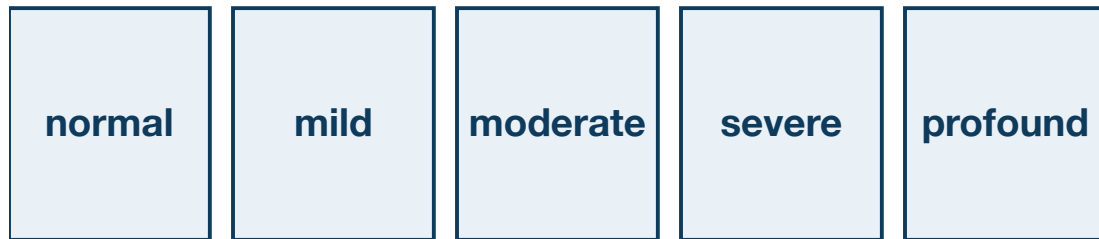
Finnish supercomputers, global ranking as of June 2026



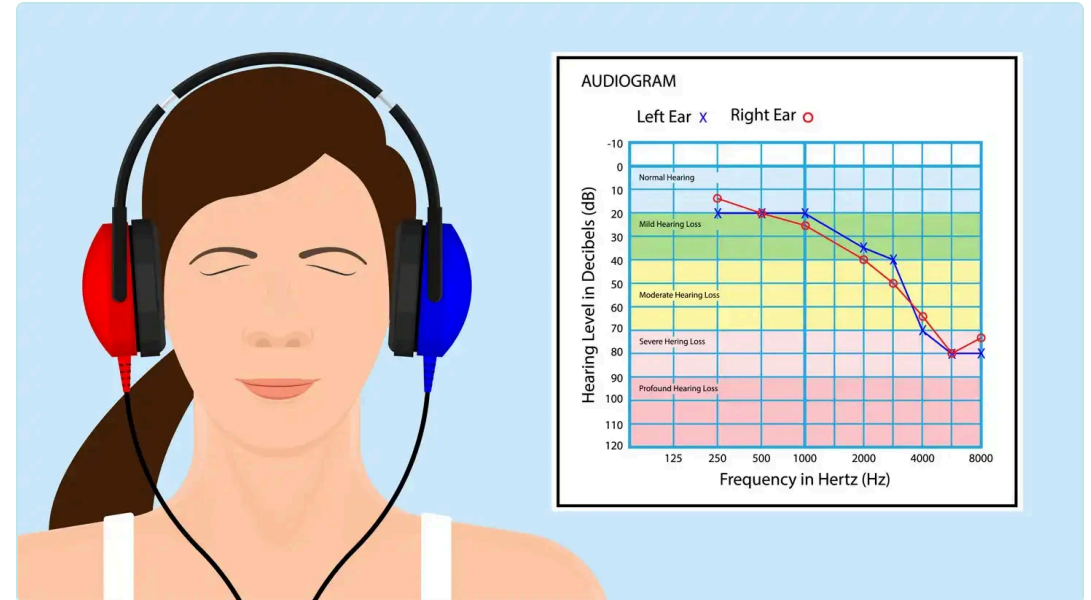
How it was used to make a major medical discovery

How hearing is assessed

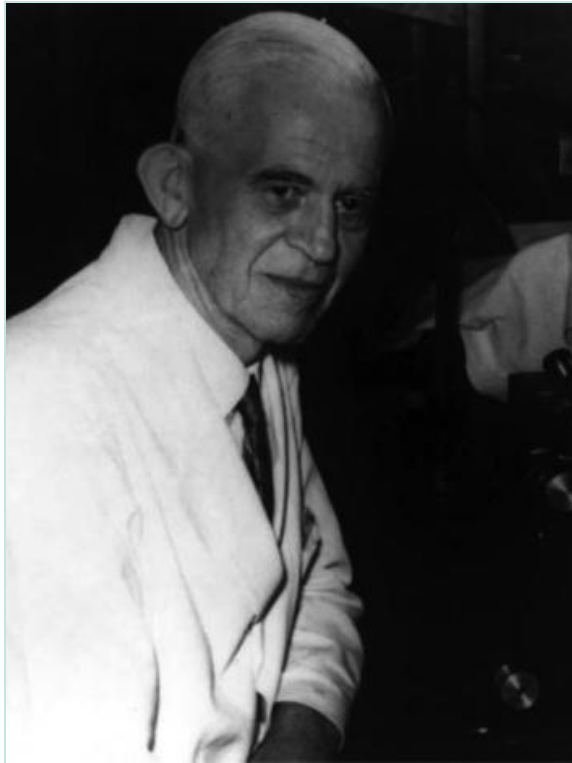
Described for about 140 years



Categories: cut-offs on one number, the pure-tone average.
Shapes: matched by eye to a few named curves.

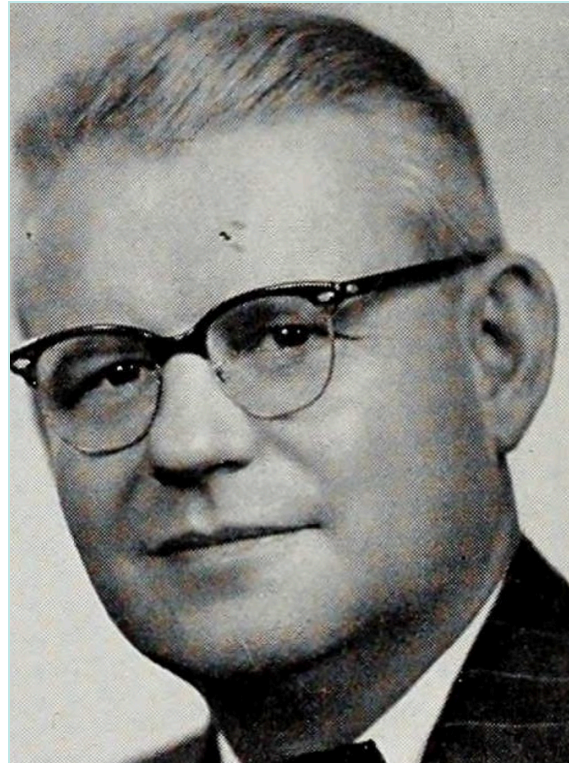


Since 1885 it is believed that human hearing comes in types: normal, slope, notch, flat... and it is ranked or classified as normal, mild, moderate, severe, profound...



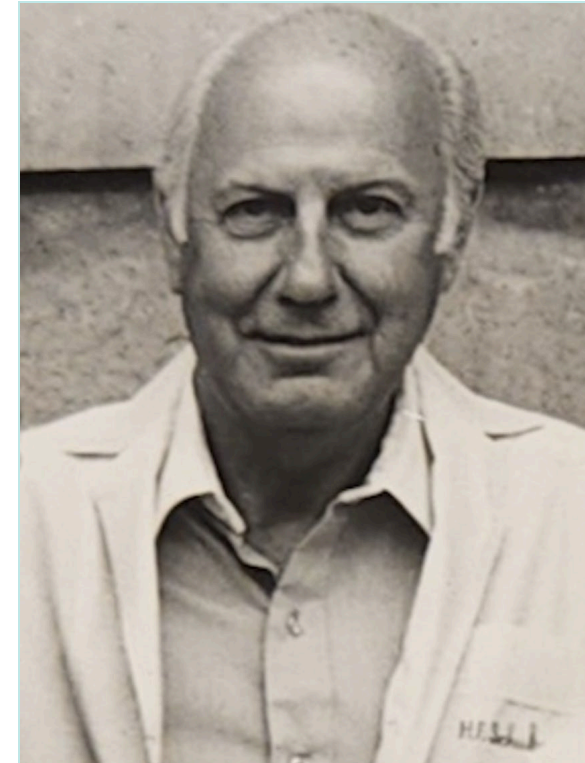
Guild

Stacy R. Guild, circa 1965. Photographer unknown. Courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; in Francis et al., *Laryngoscope* 2015.



Carhart

Raymond Carhart, yearbook photo, Northwestern University, 1958. Wikimedia Commons, CC BY-SA 4.0.



Schuknecht

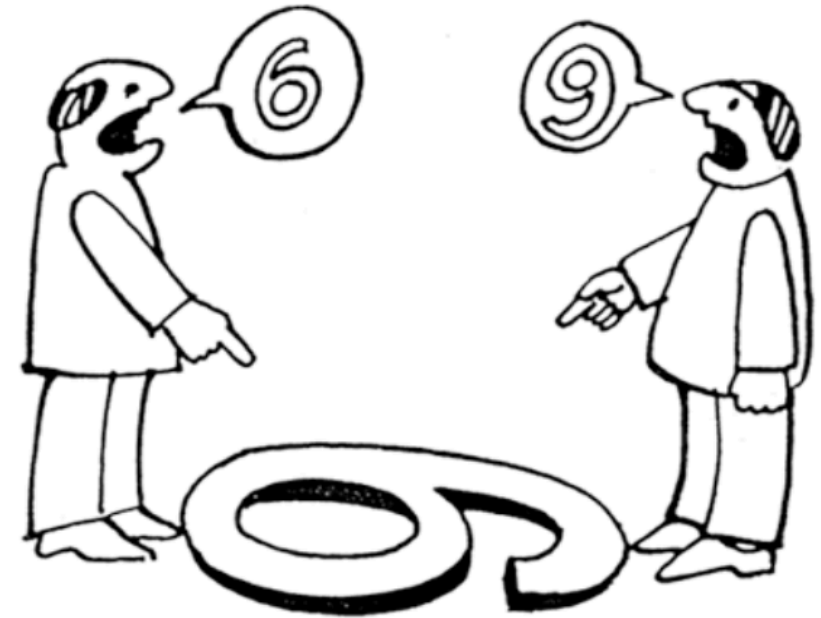
Harold Schuknecht at Massachusetts Eye and Ear Infirmary, undated. Detail by Miriam Redleaf; in Ark et al., *Ear, Nose & Throat Journal* 2025, CC BY-NC 4.0.

The AI era followed... and everyone disagreed: some found 9, some 4 and then 11, others found 6 and then 10, other 21...

| Each had their constraints.

- “All included studies evaluated a single algorithm, with no study comparing the clustering solutions to different algorithms.”
- “Two studies did not perform any validation of their clustering outcomes (Lee et al. 2010; Sanchez Lopez et al. 2018).”
- “However, only two studies attempted external validation (Parthasarathy et al. 2020; Wang et al. 2021).”
- “Notably, none of the included studies used a common dataset.”

**I ran an experiment
myself to figure out
how many**



FIRST TIME

14

SECOND TIME

21

ANOTHER TIME

6

ANOTHER TIME

29

Something was off... why is this **unstable**? Every time I slightly change the population, break it down, the count changes... it shouldn't be like that

I had access to supercomputer, and an architecture which allowed me to repeat and repeat and experiment.

Therefore I was able to ask more questions and investigate **WHY these results are changing**



**GREAT CHALLENGES NEED
GREATER POWER.**

I adopted a different approach, one I had already tried before

- In July 2013, with Chris Gibson, Ron Alfa, and a couple of other colleagues, we created a pharma start-up project called 4 Dimension Therapeutics or 4DT.
- The start-up project was about using robotics AI to perform synthetic computational exponential drug discovery.
- I left the project to go back to work and build the Organ Preservation Alliance, a more mature organ transplantation biotech startup which at the time was growing.
- 4 Dimension Therapeutics or 4DT became Recursion Pharmaceuticals, which owned the most powerful supercomputer of any pharmaceutical company, and is today traded on NASDAQ, and worth about \$2.15 billion.



From left to right: Keith Glover, Ron Alfa, Ming Cheah, Chris Gibson, Zak Allal, Sung-Yon Kim, July 2013.

Recursion Pharmaceuticals was doing industrial scale computational drug and molecules discovery for rare diseases

- To run massive foundation models for drug discovery like Phenom-1, Recursion built the BioHive-2 supercomputer in May 2024.
- It is an NVIDIA DGX SuperPOD powered by 504 H100 GPUs, making it the fastest supercomputer wholly owned by any pharmaceutical company worldwide,
- and ranked #35 in the TOP500 list of the most powerful supercomputers in the world across all industries as of May 2024, and 76th as of November 2025.
- This obviously takes millions of dollars in investment, engineering, and tooling, as well as massive electrical bills.



Recursion has completed its BioHive-2 supercomputer, showcased by several of the company's staffers and executives. [L.-R.] Paige Despain, DevOps high-performance compute (HPC) engineer; John Durkin, Joshua Fryer, senior HPC engineer; Jesse Dean, principal, real estate and workplace experience partner; Ganesh Jagannathan, chief information security officer and VP of IT; co-founder and CEO Chris Gibson, PhD; Lindsay Ellinger, VP of procurement; chief financial officer Michael Secora, PhD; Alex Timofeyev, associate director, HPC engineering; and Ben Mabey, chief technology officer. [Recursion] Source: genengnews

Recursion Pharmaceuticals was doing industrial scale computational drug and molecules discovery for rare diseases

- Just like a high-frequency trading firm, instead of trading stock, they perform high frequency biological experiments.
- They treat drug discovery as a massive, brute-force computational search problem.
- Their core philosophy is that human scientists are too slow to guess the right molecule. Instead, they built an Operating System for Biology to test millions of hypotheses simultaneously.

How they spawn experiments

1

Intent

A scientist writes a config (YAML/Python): knock out Gene X, test 100,000 compounds against it.

2

Fan-out

The orchestrator builds a DAG and spawns 100,000 parallel jobs, one container per trial arm.

3

Prune

AI models score every arm. Dead branches are dropped, the top 100 move on.

4

Wet lab

The top 100 are compiled into robot instructions. Robots mix, microscopes take millions of pictures.

5

Feedback

The pictures go back into the model. The ledger is updated. The model gets smarter.

The stack: Kubernetes and Argo Workflows for orchestration, Kafka for streaming, BigQuery for the ledger, BioHive-2 for compute.

List	Rank	System	Vendor	Total Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
06/2026	93	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400	Nvidia	40,320	23.32	34.22	
11/2025	76	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400	Nvidia	40,320	23.32	34.22	
06/2025	69	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400	Nvidia	40,320	23.32	34.22	
11/2024	54	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400	Nvidia	40,320	23.32	34.22	
06/2024	35	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400	Nvidia	40,320	23.32	34.22	

BIOHIVE-2 - NVIDIA DGX H100, XEON PLATINUM 8480C 56C 2GHZ, NVIDIA H100, INFINIBAND NDR400

Site:	Recursion
System URL:	https://www.recursion.com/
Manufacturer:	Nvidia
Cores:	40,320
Processor:	Xeon Platinum 8480C 56C 2GHz
Interconnect:	Infiniband NDR400
Installation Year:	2024
Performance	
Linpack Performance (Rmax)	23.32 PFlop/s
Theoretical Peak (Rpeak)	34.22 PFlop/s
Nmax	2,150,400
Software	
Operating System:	Ubuntu 22.04
Compiler:	NVIDIA NVCC 12.4
Math Library:	NVIDIA cuBLAS 12.4.2.65
MPI:	OpenMPI 4.1.4+

What Recursion Pharmaceuticals was leading, was also followed by the sovereign/offline/off-the-grid workstations

- In August 2025, Marco Mascorro developed a16z Personal AI Workstation. He packed four NVIDIA RTX 6000 Pro Blackwell GPUs with 384 gigabytes of VRAM into a wheeled, \$39,500 under-desk powerhouse that ran on a standard wall outlet.
- Just months later, the industry followed his lead. In late 2025 and early 2026, we saw the \$3,999 NVIDIA DGX Spark, the Dell GB10, the MSI Edge Expert, and the Asus Essent GX10 miniaturize this concept into desktop devices.
- I used my own workstation, but at some point, needed more power.



RunPod isn't that powerful, and Azure is expensive.

Compute

GPU

CPU

3 GHz

5 GHz

General Purpose

Network volume

2 vCPU

8 GB RAM

High

\$0.08/hr

4 vCPU

16 GB RAM

High

\$0.16/hr

8 vCPU

32 GB RAM

High

\$0.32/hr

16 vCPU

64 GB RAM

High

\$0.64/hr

32 vCPU

128 GB RAM

High

\$1.28/hr

Showing 57 of 951 VM sizes | Subscription: Azure subscription 1 | Region: Switzerland North | Current size: Standard_M832ds_12_v3 | Image: Ubuntu Server 24.04 LTS | [Learn more about VM sizes](#) | Group by series

VM Size ↑↓	Type ↑↓	vCPUs ↑↓	RAM (GiB) ↑↓	Data disks ↑↓	Max IOPS ↑↓	Local storage (GiB) ↑↓	Premium disk ↑↓	Cost/hour ↑↓
M-Series v3								
The 3rd generation M series sizes for extremely large databases or applications								
M832ds_12_v3	Memory optimized	832 (Customize)	11400	64	130000	400 (SCSI)	Supported	\$154.60
M832s_12_v3	Memory optimized	832	11400	64	130000	N/A	Supported	\$147.24
M832ds_16_v3	Memory optimized	832	15200	64	130000	400 (SCSI)	Supported	\$228.31
M832s_16_v3	Memory optimized	832	15200	64	130000	N/A	Supported	\$217.44
M624ds_12_v3	Memory optimized	624	11400	64	130000	400 (SCSI)	Supported	\$154.11
M624s_12_v3	Memory optimized	624	11400	64	130000	N/A	Supported	\$146.77
M416ds_6_v3	Memory optimized	416	5696	64	130000	400 (SCSI)	Supported	\$77.05
M416s_6_v3	Memory optimized	416	5696	64	130000	N/A	Supported	\$73.38
M416ds_8_v3	Memory optimized	416	7600	64	130000	400 (SCSI)	Supported	\$102.74
M416s_8_v3	Memory optimized	416	7600	64	130000	N/A	Supported	\$97.85
M416ds_9_v3	Memory optimized	416	8568	64	130000	400 (SCSI)	Supported	\$116.09
M416s_9_v3	Memory optimized	416	8568	64	130000	N/A	Supported	\$110.56
M416ds_10_v3	Memory optimized	416	9496	64	130000	400 (SCSI)	Supported	\$128.66

What Recursion did on an industrial scale in 2015 and on... is today possible with good engineering AND accessible supercomputers such as Roihu

- Just like we have an MVP and Symfony/Laravel revolutionized the web, I create a structure which decreased cost of repetition
- to the point that I would spawn a container which has the scientific kernel and orchestrator
- tune my environment and parameters
- run it
- forget it
- get notified when results come back
- spawn spawn spawn spam spam spam
- and by exhaustion I would get the nuance.

**I LEVERAGED ROIHU WITH 60 NODES,
47TB RAM + RUST AND JULIA AND C++**

Clinical Sense OS

- The Clinical Sense OS. A clinical research operating system for clinical translational research.
- Rust kernel as a daemon, orchestrator and DAG.
- Julia for the science.
- C++ for any powerhorse needed.
- Spawn using Apptainer.
- DuckDB registry.
- Clinical Experiments as Git: We formulate a hypothesis: spawn, run and do.

```
CSC - Tieteen tietotekniikan keskus - IT Center for Science

ROIHU

Roihu.csc.fi - Bull BullSequana XH3000
486 AMD Turin CPU nodes - 132 Nvidia GH200 GPU nodes

Links
Documentation:      https://docs.csc.fi
Account management: https://my.csc.fi
Getting started guide: https://docs.csc.fi/support/tutorials/roihu
Web interface:      https://roihu.csc.fi

Info
All GPU jobs should be launched from roihu-gpu login node, and
all CPU jobs should be launched from roihu-cpu login node

Available modules can be listed with module avail and module spider

News
19.09.2026 There will be a maintenance break on Roihu
6.10.2026 08:00 - 7.10.2026 22:00 EEST. During the break it will not be
possible to access login nodes, the web interface, or the storage system.

For more information see
https://research.csc.fi/service-break/roihu-maintenance-2026-10-06/

Last login: Mon Sep 28 18:49:51 2026 from 2001:999:3002:9aff:549f:23b2:f9b7:6df0
```

What Clinical Sense OS + super-performant programming languages + supercomputer allowed?

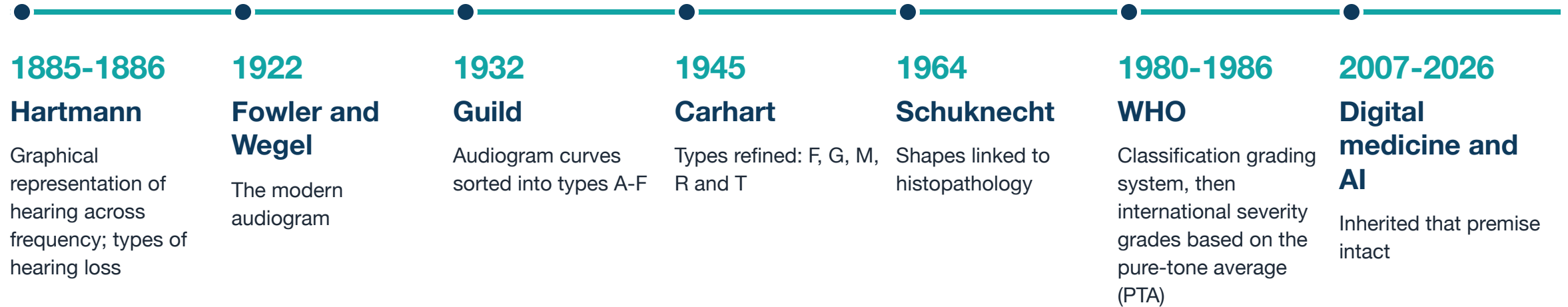
- **Four million** hearing tests, **a million and a half** people, **sixty-two years**. That is about **a quarter of Finland**.
- It was **8 countries** with different ages and ethnic backgrounds, and generations.
- Reading them one per second would take **fifty days**.
- The computation would have kept a laptop busy for **thirteen years**. We did it in **two months**.
- On the busiest day, we had the power of **seven thousand laptops** running at once.
- About **850,000 core-hours** on Roihu, plus about **50,000** on Mahti. Together that is about **0.9 million core-hours**, or **102 core-years**.
- The peak was **58,288 cores** working at the same instant in July.

**THE supercomputer ENGINEERING for
ultraperformant experimentation lead to medical
discoveries**

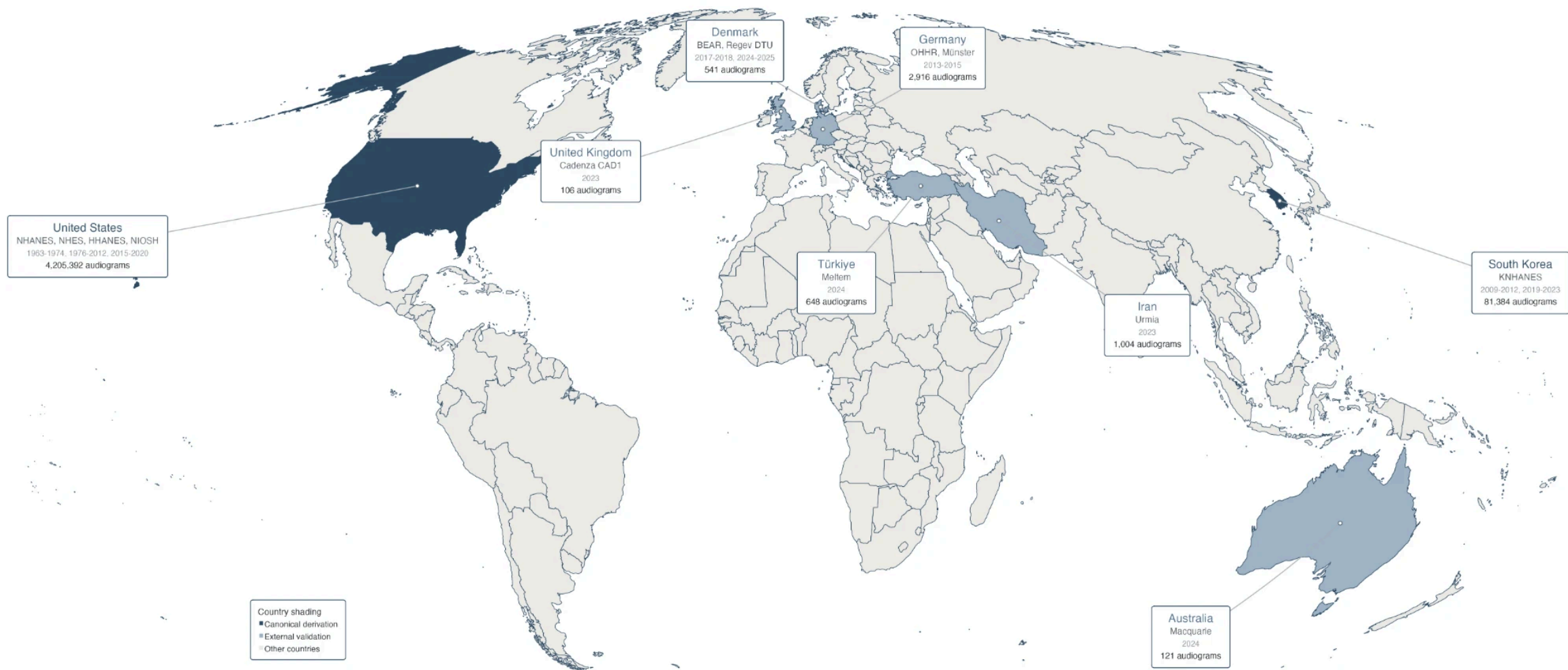
Question: what if... what if... the data is not categorical? It's not flat, normal, ... it's continuous.

I found out... from 1885 to 2026, it was carried from the founding father of hearing science and otology, to WHO to AI... without... checking the premise.

No one checked the premise.



We collected and dug out, including reconstructing old data files from the 60s, to assemble the largest heterogeneous human audiogram dataset ever published, made of 4.29 million audiograms, 8 countries, 62 years, and different measurement standards





Ran the experiment: an out-of-sample competition

- A discrete model, which assigned audiograms to mutually exclusive groups, against a continuous model, which allowed variation without forced assignment.
- Both were fitted on development data and compared by prediction in held-out participants.
- 195,262 ears from 97,631 people, across 25 survey waves.
- 100 participant-grouped outer validation folds.
- A cohort was considered discrete only when the subtype model consistently outperformed the continuous model out of sample.
- Independent derivation: 1,365,737 US workers, decomposed without access to the Atlas coordinates, rank or labels, regenerated the same coordinates.
- Transfer without refitting: applied unchanged to people who contributed nothing to the derivation, five coordinates reconstructed their audiograms to 1.35 dB (22,438 ears), and to 1.018 dB across 4,092,824 occupational ear-examinations.

Results: on every large adult population, in every country and every era, the conclusion was: continuous.

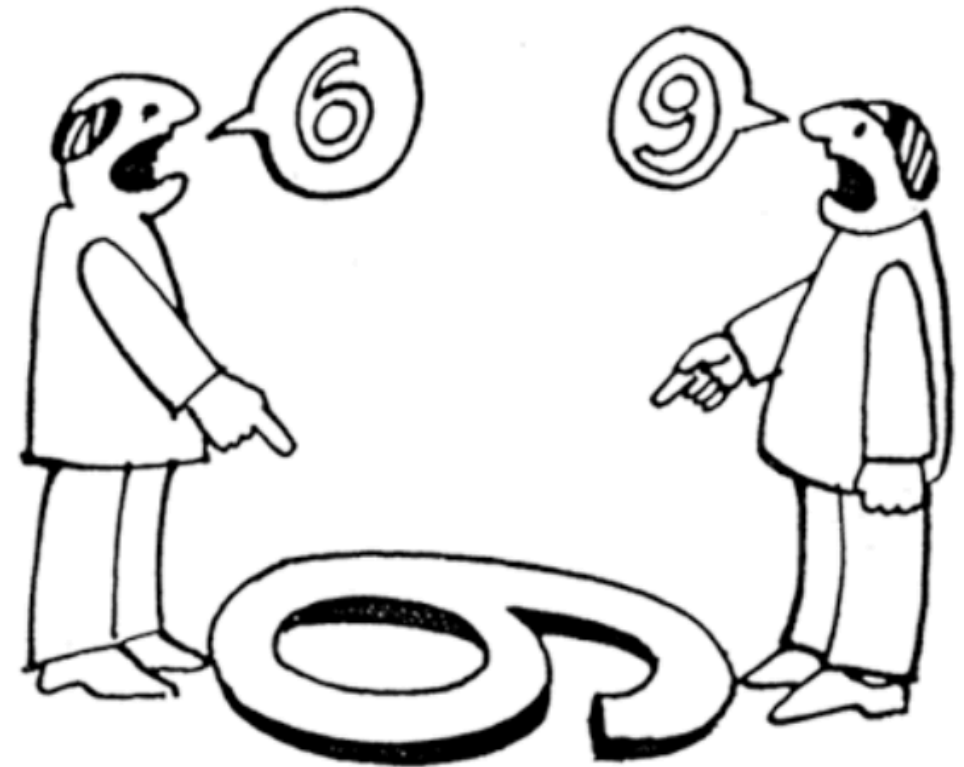
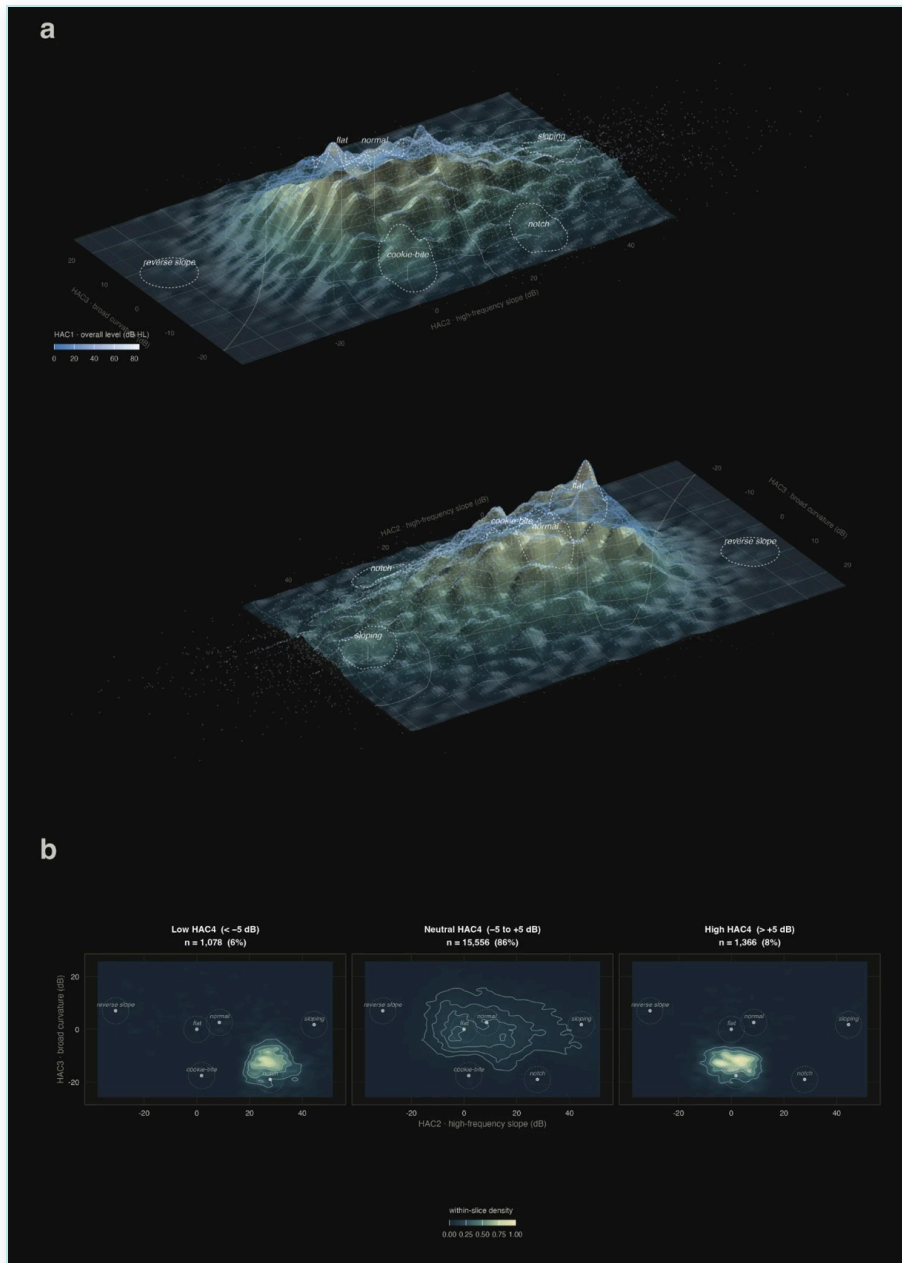
In traditional academia: this would stop there. This is itself a finding. But since I have the framework, I continue to dig more.

| Why were others wrong, and for so long?

Why were others wrong, and for so long?

- **A statistics error described in 1898.** Hearing tests are recorded in 5 dB steps, so different ears land on identical tracings and look like groups.
- **Leaking ears.** The left and right ears of the same person were counted as independent observations.
- **Pooling.** Groups that appear in one cohort or generation dissolve when populations are pooled.
- **Unstable counts.** In all 29 analyses that favoured types, the number of types changed when the analysis was repeated.

All of it was already reported in the literature. They just didn't connect the dots.



Deep learning and AI, misunderstood, with academic pressure and conformity or orthodoxy, led to 15 years of disagreement and debates, where in fact the answer was without AI, and not even part of the debate over the number of categories.

If it's continuous, how ?

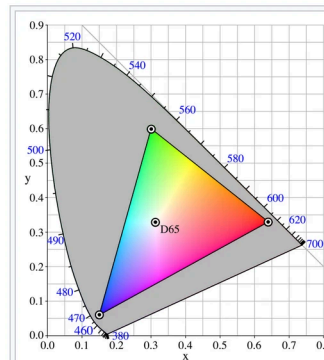
Physical principles for the choice of red, green, and blue [\[edit \]](#)

The choice of primary colors is related to the physiology of the [human eye](#); good primaries are stimuli that maximize the difference between the responses of the [cone cells](#) of the human retina to light of different [wavelengths](#), and that thereby make a large [color triangle](#).^[7]

The normal three kinds of light-sensitive [photoreceptor cells](#) in the human eye (cone cells) respond most to yellow (long wavelength or L), green (medium or M), and violet (short or S) light (peak wavelengths near 570 nm, 540 nm and 440 nm, respectively^[7]). The difference in the signals received from the three kinds allows the brain to differentiate a wide [gamut](#) of different colors, while being most sensitive (overall) to yellowish-green light and to differences between hues in the green-to-orange region.

As an example, suppose that light in the [orange](#) range of wavelengths (approximately 577 nm to 597 nm) enters the eye and strikes the retina. Light of these wavelengths would activate both the medium and long wavelength cones of the retina, but not equally—the long-wavelength cells will respond more. The difference in the response can be detected by the brain, and this difference is the basis of our perception of orange. Thus, the orange appearance of an object results from light from the object entering our eye and stimulating the different cones simultaneously but to different degrees.

Use of the three primary colors is not sufficient to reproduce *all* colors; only colors within the color triangle defined by the [chromaticities](#) of the primaries can be reproduced by additive mixing of non-negative amounts of those colors of light.^[7][\[page needed\]](#)



A set of primary colors, such as the [sRGB](#) primaries, define a [color triangle](#); only colors within this triangle can be reproduced by mixing the primary colors. Colors outside the color triangle are therefore shown here as gray. The primaries and the [D65 white point](#) of sRGB are shown. The background figure is the [CIE xy chromaticity diagram](#).

Colour vision: three primaries (RGB color model)

Einthoven's triangle

[Article](#) [Talk](#)

[5 languages](#) ⌵

[Read](#) [Edit](#) [View history](#) ⋮

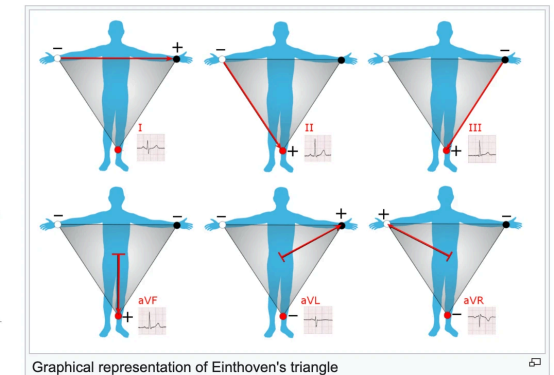
From Wikipedia, the free encyclopedia

Einthoven's triangle is an imaginary formation of three limb leads in a triangle used in the [electrocardiography](#), formed by the two shoulders and the pubis.^[1] The shape forms an inverted equilateral triangle with the heart at the center. It is named after [Willem Einthoven](#), who theorized its existence.^[2]

Einthoven used these measuring points, by immersing the hands and feet in pails of salt water, as the contacts for his [string galvanometer](#), the first practical ECG machine.^[3]

Lead placements [\[edit \]](#)

- Lead I** — This axis goes from shoulder to shoulder, with the negative electrode placed on the right shoulder and the positive electrode placed on the left shoulder. This results in a 0-degree angle of orientation.^[4]
 $I = LA - RA$
- Lead II** — This axis goes from the right arm to the left leg, with the negative electrode on the shoulder and the positive one on the leg. This results in a +60 degree angle of orientation.^[4]
 $II = LL - RA$
- Lead III** — This axis goes from the left shoulder (negative electrode) to the right or left leg (positive electrode). This results in a +120 degree angle of orientation.^[4]
 $III = LL - LA$



Cardiac electrical axis: Einthoven's triangle

Result: Human Hearing is 5-Dimensional. Human hearing has 5 dimensions.

- Just like colors are RGB, hearing has HAC1-HAC5.
- We literally falsified and refuted a 140-year medical dogma. And defined the law of human hearing states: Human's hearing has states. NOT types. It is a 5 coordinate system in a map. And we defined the human hearing atlas.

Natural law of human hearing

Human hearing occupies one continuous five-dimensional coordinate space. We call the empirical regularities the Law of Human Hearing States.

HAC1	HAC2	HAC3	HAC4	HAC5
Overall level	Slope	Curvature	2 kHz contrast	3 kHz contrast

Colour vision

Three RGB coordinates describe any perceived colour.

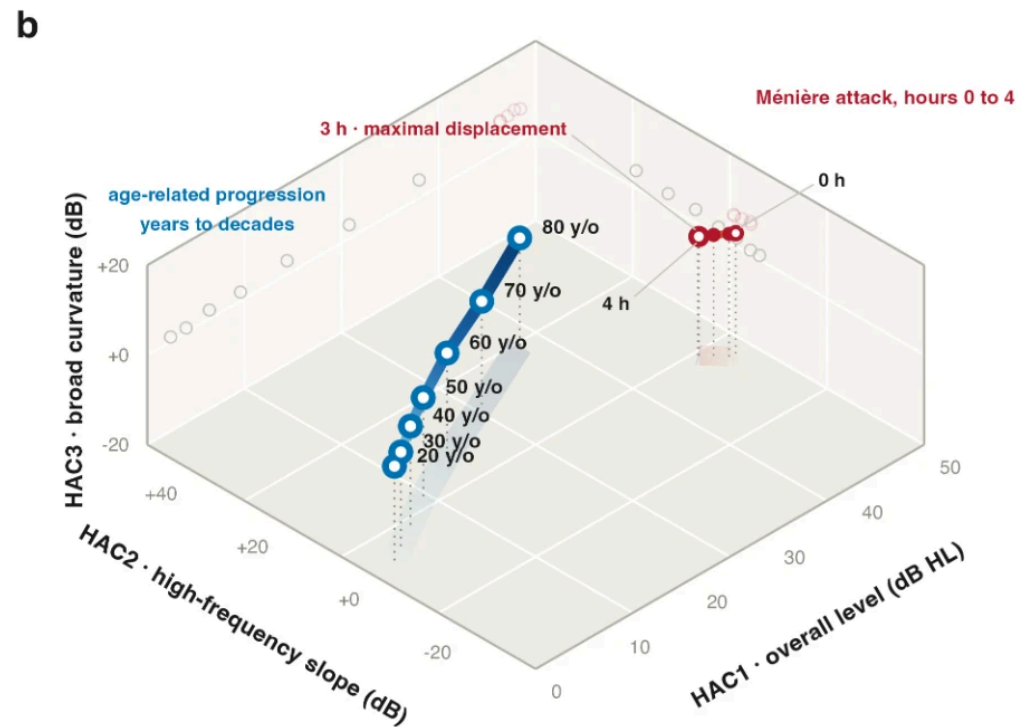
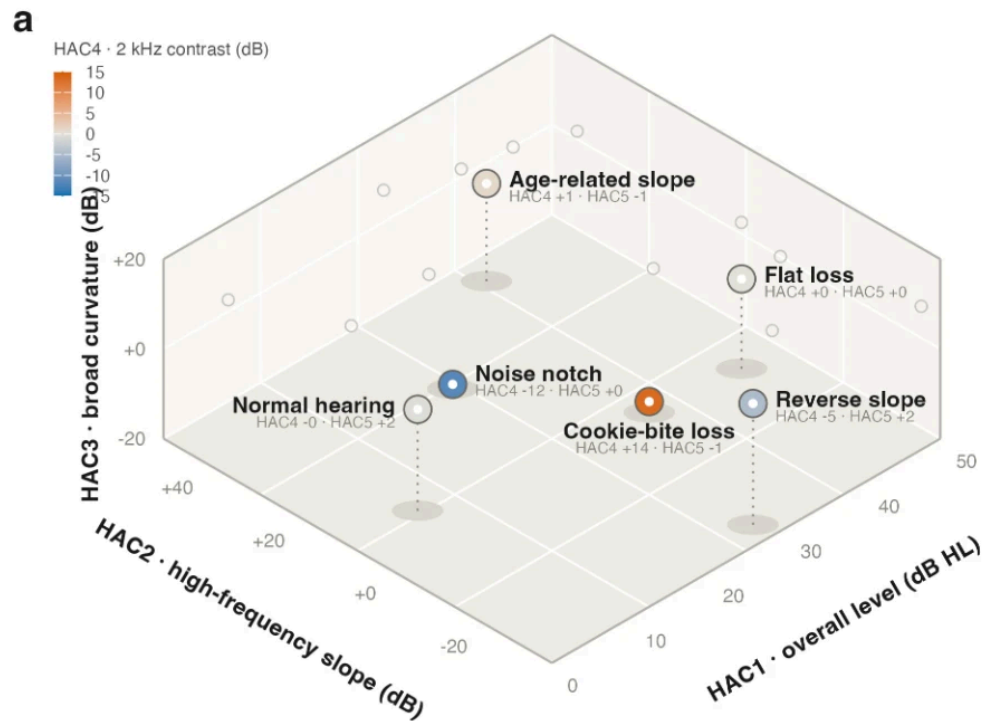
Cardiac electrical axis

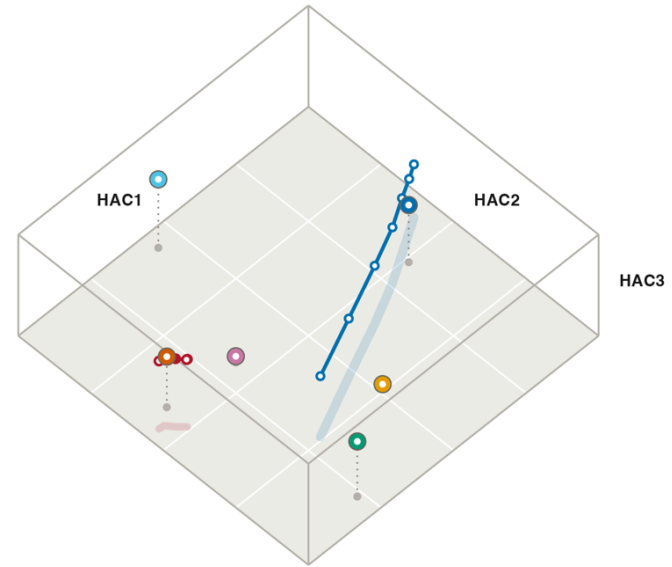
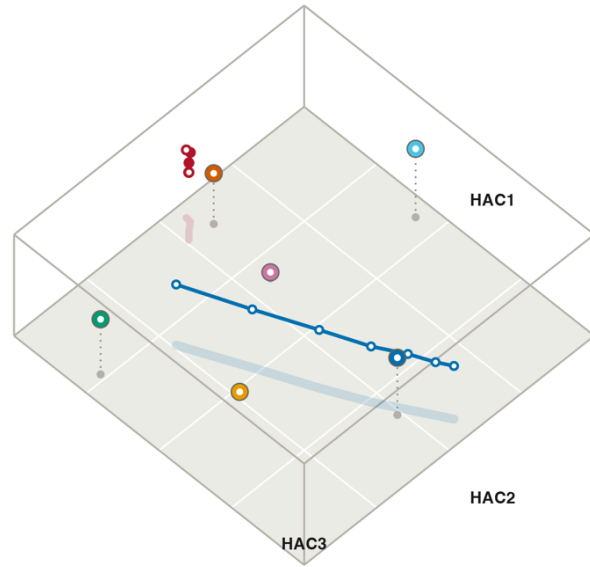
Einthoven's triangle describes the heart's electrical vector from the limb leads.

Human hearing

Five Hearing Atlas Coordinates describe any audiogram, in every population studied.

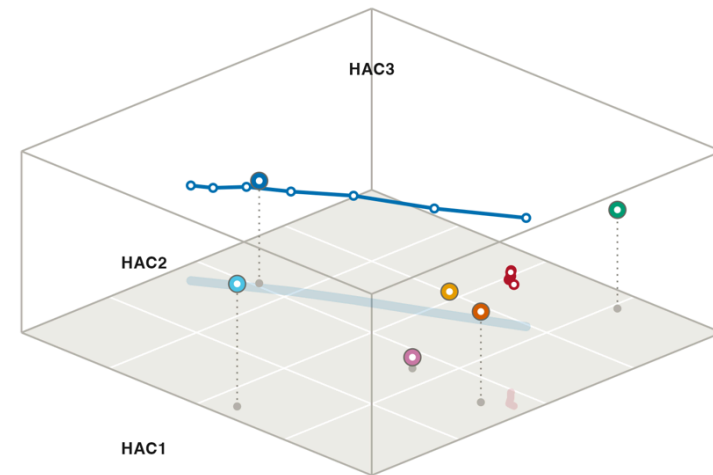
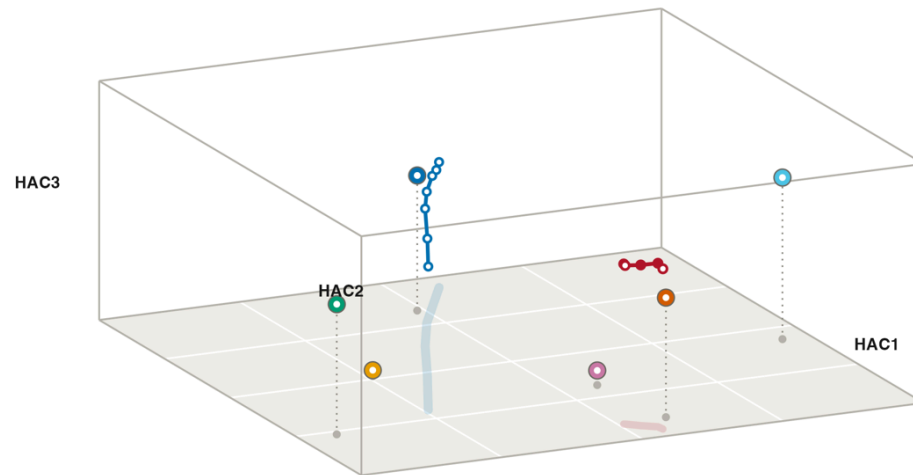
The same five coordinates recur across countries, decades, calibration standards, and air and bone conduction. To confirm this is an invariant mathematical property and not an artefact of one algorithm, the theorem system was formalised in Lean 4: 607 theorems, each machine-checked.





View from below the HAC1-HAC2 floor

Low oblique view from the far corner



PRE-PRINT AVAILABLE. CURRENTLY IN REVIEW IN A MAJOR SCIENTIFIC JOURNAL.

Co-authors and team: Pr. Antti Aarnisalo, and Pr. Antti Mäkitie, Helsinki University Hospital.



THE PREPRINT SERVER FOR HEALTH SCIENCES

 [Follow this preprint](#)

The Human Hearing Atlas: a canonical map of human hearing

 Zakarya Hadj-Allal,  Antti Aarnisalo,  Antti A. Mäkitie

doi: <https://doi.org/10.64898/2026.09.14.26363004>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract

Full Text

Info/History

Metrics

 Preview PDF

Abstract

For 140 years, hearing science and otology have rested on a premise treated as a principle of human physiology: that the sense of hearing, in health and disease, comes in distinct kinds. Its standard measure, the audiogram, was collapsed into a severity average and a subjective shape label. The reduction carried the premise from otology's founding fathers into the digital and artificial intelligence era. We tested this premise and falsified it. Human hearing does not resolve into a reproducible catalogue of types; it occupies a continuous, five-dimensional coordinate system. We call it the Human Hearing Atlas. Established across more than 4.29 million audiograms spanning 33 datasets, eight countries, 62 years and two calibration standards, the Atlas reconstructs unseen audiograms to near-decibel accuracy and preserves its clinical details. It maps within-ear hearing change over hours or decades, half of which is

Companion mathematical manuscript also in review

The first formal mathematical axiomatization (Lean 4) of a human sensory system and the third of a human physiological system, following cardiac electrical conduction (2011/2012) and endocrine glucose homeostasis (2014)

Formal Mathematical Axiomatization and Verification of Human Hearing

Zakarya Hadj-Allal¹, Antti A. Mäkitie^{1*}, Antti Aarnisalo¹

¹ Department of Otorhinolaryngology-Head and Neck Surgery, University of Helsinki and HUS Helsinki University Hospital, Helsinki, Finland

* Corresponding author: Antti A. Mäkitie, antti.makitie@helsinki.fi

Keywords: human hearing; sensory system; physiological law; axiomatization; formal verification; Human Hearing Atlas

a			
Invariant	Physiological principle	Experiment	Results
I. Continuity	Human hearing does not resolve into fixed types; familiar flat, sloping, rising and notched audiograms occupy neighbouring regions of one continuous space.	Discrete and continuous models were required to predict audiograms in people neither had seen. Classical configuration boundaries were then tested during same-ear movement.	Every large pooled adult population favoured the continuum. Across 1,361,350 same-ear transitions, 27.5% crossed a clinical configuration boundary. Among transitions with an Atlas displacement of 5 dB or less, 3.68% changed label. Ears that changed label had a median Atlas displacement of 19.4 dB, compared with 13.2 dB for those that did not.
II. Low dimensionality	Most audiogram variation can be represented at clinical resolution by a few independent dimensions of hearing level and shape.	Intrinsic dimension was estimated directly from raw thresholds across audiometric frequency grids. The resulting coordinates were then required to reconstruct audiograms from people excluded from derivation.	Intrinsic dimension was 3.39, 3.95, 4.39 and 4.46 on grids of four, six, seven and eight frequencies. Five Atlas coordinates reconstructed unseen six-frequency audiograms to 1.35 dB, with 95.7% of thresholds within 5 dB and 99.93% of audiogram pairs differing by more than 5 dB at any frequency still distinguishable.
III. Universality	The same hearing coordinates recur when population, generation and audiometric standard change.	A frozen Atlas was transferred without refitting across independent countries, eras, clinical settings and calibration standards, and compared with coordinate systems fitted locally to each cohort. Independent cohorts were then decomposed without access to Atlas coordinates, rank or labels, and a coordinate system built from NIOSH alone was applied unchanged to further populations.	The coordinates transferred across eight countries, 62 years and two calibration standards. Across 2,731,474 occupational worker-ear series, rebuilding the coordinate system locally improved reconstruction by only 0.190 dB (95% CI 0.189 to 0.191), 3.8% of one 5 dB recording step. Blind independent regeneration recovered HAC1 in all six testable cohorts (0.8958 to 0.9916), and 1,365,737 NIOSH workers regenerated HAC2 to HAC6 completely (0.8455 to 0.9801). The NIOSH-built system then transferred zero-shot to NHES II, NHES III and Basel on all six axes ($r = 0.6636$ to 0.9787).
IV. State-motion correspondence	The dimensions along which one patient's hearing differs from another's are the dimensions along which a single ear changes over time.	Between-ear audiogram-shape structure and within-ear longitudinal change were estimated independently after overall hearing level was removed.	Across 1,294,932 within-ear changes in 421,868 workers, the two shape spaces overlapped by 0.920 (95% CI 0.918 to 0.922), exceeding the 0.663 95th percentile of randomly oriented spaces. Acute Ménière and chronic hearing-change directions also overlapped by 0.804. Applied unchanged to 1,422 acute Ménière transitions, the chronic geometry captured 55.0% of acute shape-change variance, 94.5% of the variance captured by a geometry fitted directly to the acute data.



**TODAY WE SUBMITTED OUR INTENTION TO GET
REGULATORY BIOMARKER QUALIFICATION, IT
TAKES MANY MONTHS**

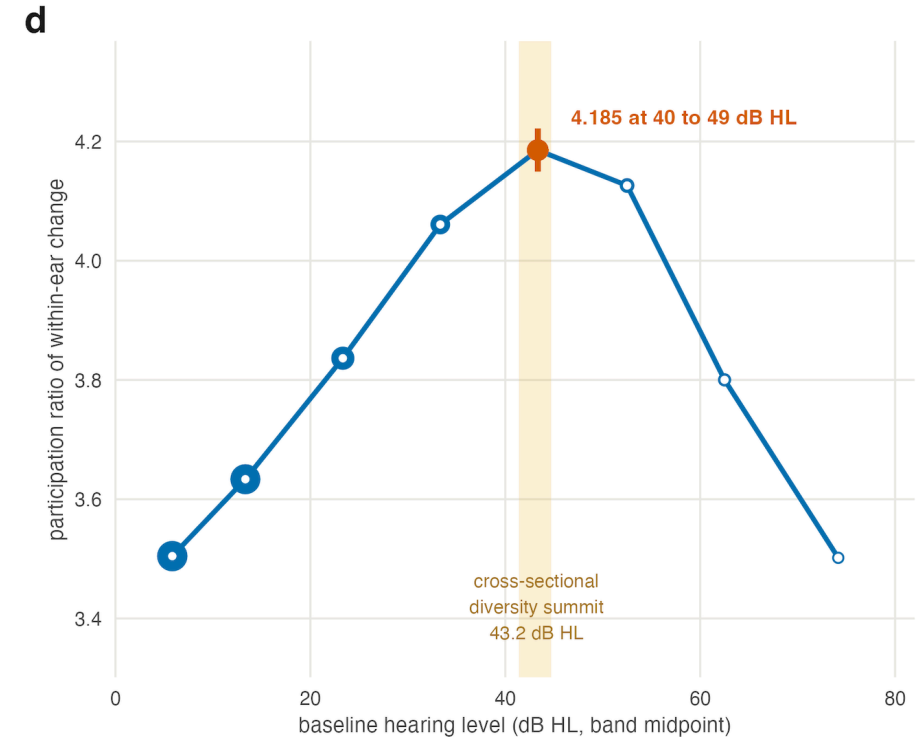
**THIS DISCOVERY HAS POTENTIAL TO
Changing the Lives of 1.5 Billion People.**

HOW IT IMPACTS 1.5 BILLION PEOPLE

- 1 Prospective Therapeutic Trials**
- 2 Retrospective Studies**
 - Retrospective Interventional Clinical Trials
 - Retrospective Observational, Surgical, and Real-World Evidence (RWE) Studies
 - Re-Commissioning Multi-Decade Longitudinal & Historical Cohorts
- 3 Otological and Combination Interventions**
- 4 Trial Enrichment, Natural History Registries, and Rare Disease Programs**
- 5 Ototoxicity Monitoring, Prevention, and Hearing Conservation**

How ?

- WHO projects that by 2050 nearly 2.5 billion people will have some degree of hearing loss, and more than 700 million will require hearing rehabilitation.
- Unaddressed hearing loss costs almost US\$ 1 trillion every year.
- About 75% of people with hearing loss are counted in a single band, mild: 1.17 of 1.57 billion in 2019.
- We found that this is where hearing becomes most chaotic and complex: its diversity climbs through the mild band and peaks at 43.2 dB, just above it.
- Yet WHO treats it as one dimension: a single average.



THE RENAISSANCE POLYMATHY

Piloted by Curiosity

- We are living in a renaissance age, and in a renaissance polymaths thrive.
- This medical breakthrough was not because of autonomous AI, it was agentic enabled using Clinical Sense OS with containers using Rust/Julia/C++.
- AI was piloted using clinical/medical knowledge, musical knowledge, coding engineering knowledge, and curiosity (the "but why"/"why not").
- Intentional curiosity: wanting to understand why something is the way it is, or why it is broken, and then throwing exhaustive brute force at it.

THE RENAISSANCE POLYMATHY

The Key Message

- We are living in the most exciting time of humanity.
- Sovereignty and the dangers of AI are stealing the attention economy.
- BUT AI is nothing without good human intuition, curiosity, and intentional directionality and piloting.
- Mixing this with good practice of architecture and engineering, fantastic programming languages (Rust/Julia/C++/C), supercomputers, and cross-disciplinary knowledge leads to revolutionary findings that will change the lives of billions of people.

Thank you.

me@zakallal.com

X: @ZakAllaIMD

Collaboration, science, tech, business, or chat... don't hesitate to reach out!



www.zakallal.com