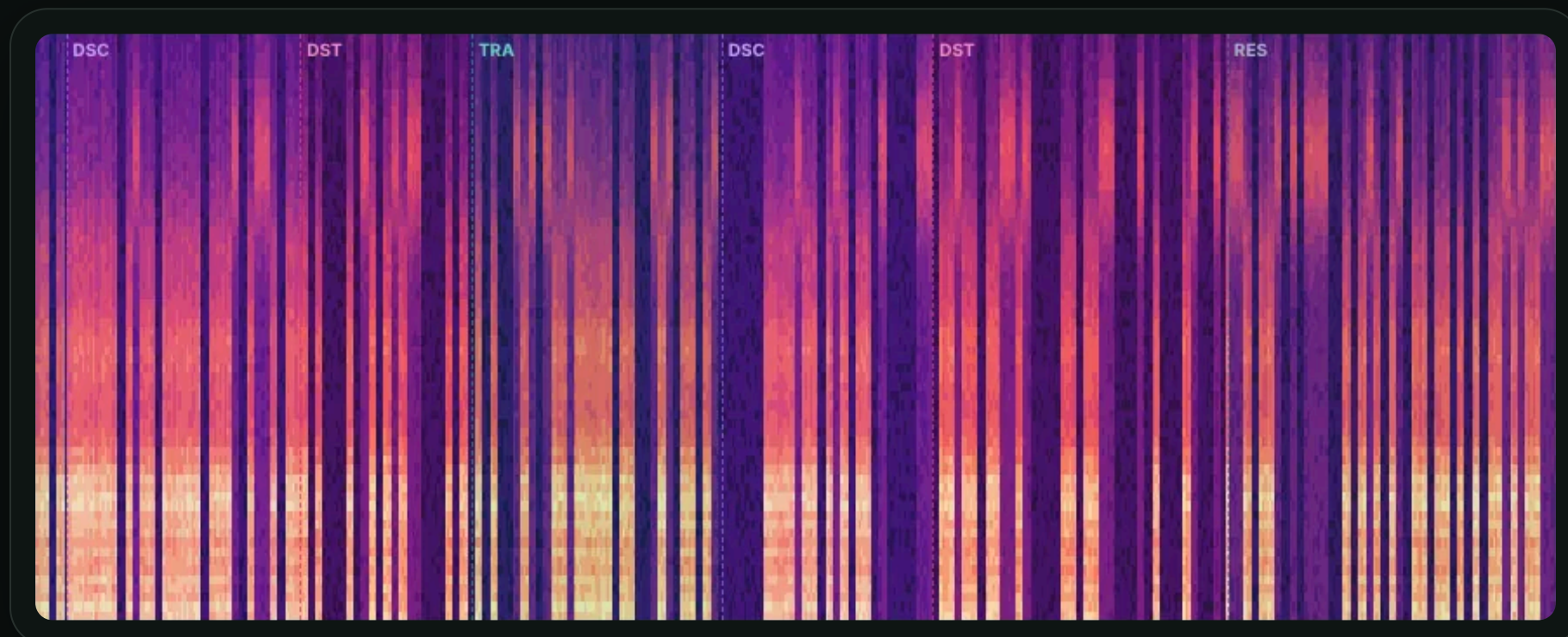


HAIVA • VOICE

Voice, measured carefully.

What haiva measures in the speech of a therapy session, what it will never claim, and the five gates in between.



14 minutes of patient speech • sample record

haiva

WHY VOICE

Clinicians have always listened to how people speak

The mental-state examination

records speech: its rate, volume, rhythm, and how long a reply takes to start.

Depression

Slowed speech is part of psychomotor slowing, one of the criteria for a depressive episode in DSM-5 and ICD-11.

Research

Across studies, more severe depression goes with slower speech, longer and more variable pauses, and flatter, quieter voice.

The idea

Measure what a clinician already hears, the same way every session, and keep it next to the transcript.

Cummins et al., Speech Communication, 2015 (review) · Mundt et al., Biological Psychiatry, 2012 · Cummins et al., Journal of Affective Disorders, 2023 (585 participants, three languages)

WHY CAREFULLY

Voice is easy to over-claim

$r \approx 0.25$

Pause time and depression severity in a well-run study: a real link, but a modest one.

Groups,
not people

The same study could not classify individual patients from their voice.

Voice also changes with

tiredness

a cold

medication

age

language

microphone

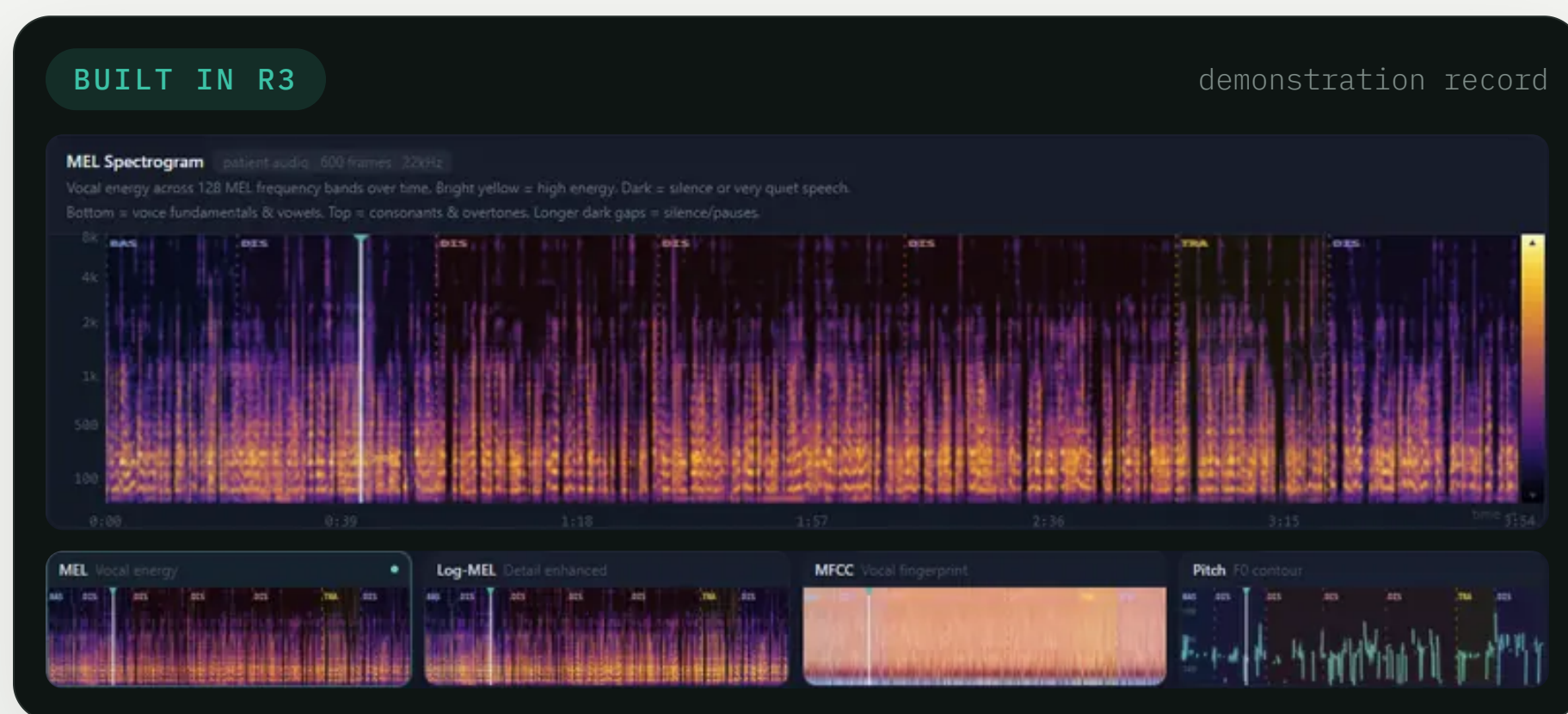
room

A vocal measure is not a validated vocal biomarker.

Mundt et al., Biological Psychiatry, 2012 · VOCAL Initiative expert consensus, 2025 (preprint)

WHERE WE STARTED

haiva R3 pictured the voice of every session

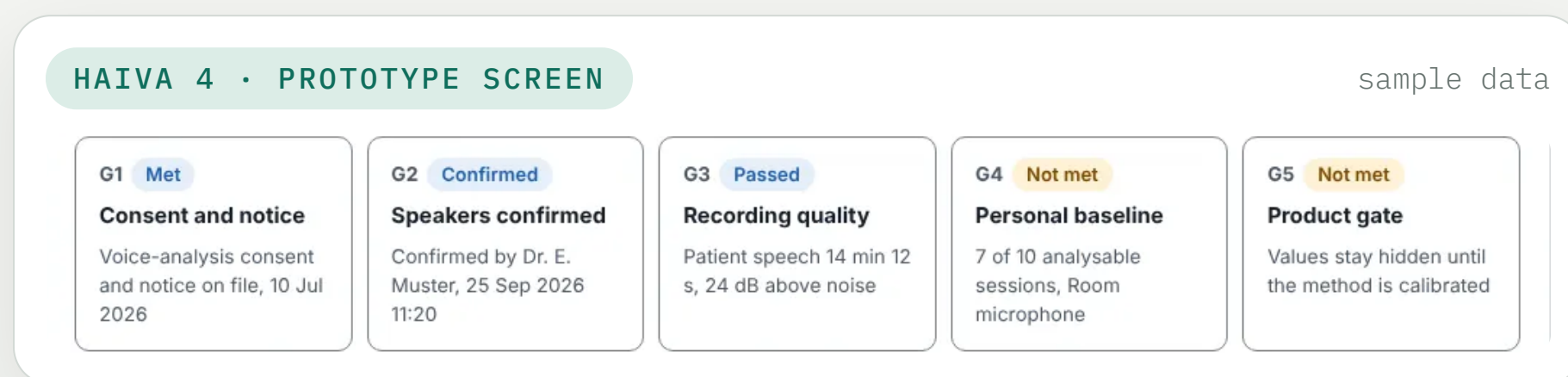


Recording, speaker separation, four views of the same voice and thousands of acoustic features per part of the session. It proved the whole chain works.

It also taught us that a number on a screen is read as a **finding**, validated or not. haiva 4 starts from that lesson.

IN PROGRESS IN HAIVA 4

Five gates before any voice value



G1 **Consent and notice**, in the patient's language, on file before anything runs.

G2 **Speakers confirmed** by the practitioner, never assumed by the machine.

G3 **Recording quality**: enough clear speech. A failed check is a gap, not a result.

G4 **Personal baseline**: sessions in the same setup, so each value meets the patient's own range.

G5 **Validation**: until then, values stay hidden, or labelled exploratory in the research edition.

GATES 1 AND 2

Consent first, then: who is speaking?

G1

Consent

read at 09:44

On file

Audio recording of sessions

Given 3 Jul 2026, version 1

On file

Transcription of recordings

Given 3 Jul 2026, version 1

On file

AI-assisted drafting of reports

Given 3 Jul 2026, version 1

Not on file

Analysis of the patient's voice

Voice analysis will not run for this session.

If the patient withdraws recording consent during the session, for example in Sona, recording stops and you are told here.

No voice-analysis consent on file: **voice analysis does not run** for this session.

G2

sample data

Who is speaking?

Waiting for your confirmation

Roles are never assigned automatically. Listen to a short clip of each voice, then say which one is the patient. Nothing downstream uses the transcript until you confirm.

① The voices overlap at 04:12 and 09:37. Check those turns; you can swap the speaker of a single turn.

Speaker A

Not assigned

6 turns · 49 percent of the words in this sample · first heard at 00:00:04

Guten Morgen, Herr Beispiel. Die Aufnahme läuft, wie wir es zu Beginn vereinbart haben. Wie war Ihre Woche?

▶ Play a 5-second clip

Next: clip 1 of 3, from 00:00:04

This voice is

The patient

Me

Someone else

Speaker B

Not assigned

6 turns · 51 percent of the words in this sample · first heard at 00:00:15

Gemischt. Bei der Arbeit war wieder viel los, und ich habe schlecht geschlafen, meistens vielleicht vier oder fünf Stunden.

▶ Play a 5-second clip

Next: clip 1 of 3, from 00:00:15

This voice is

The patient

Me

Someone else

↔ Swap A and B

Choose a role for both voices.

✓ Confirm roles

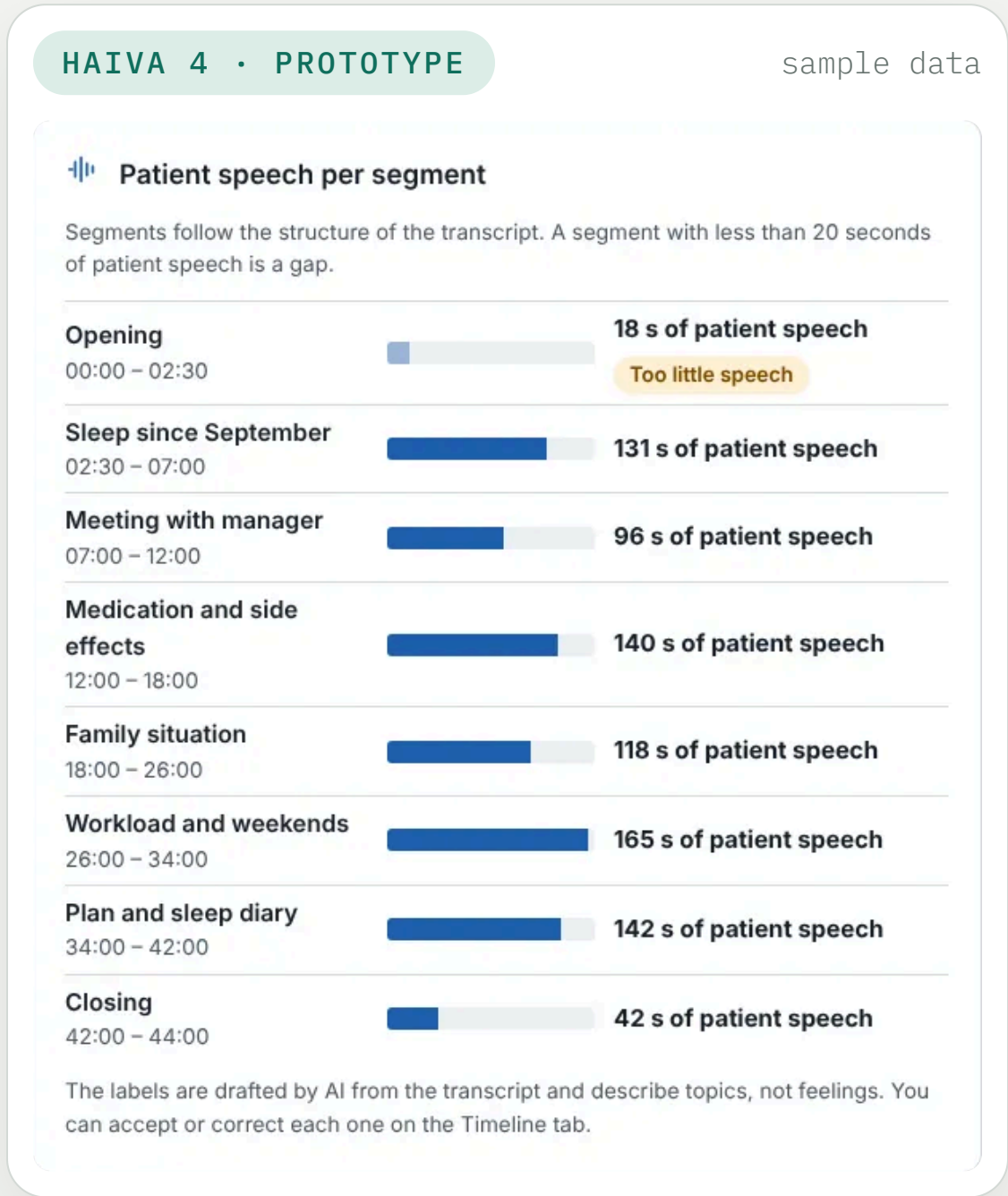
Roles are never assigned automatically. The practitioner plays a short clip of each voice, then confirms.

InnerEcho Initiative · Geneva

Swipe →

GATE 3 • RECORDING QUALITY

A failed check is a gap, not a result



Each part of the session needs at least **20 seconds** of patient speech.

The opening had 18: it is shown as a gap, and nothing is computed from it.

Noise, clipping and dropouts are checked the same way.

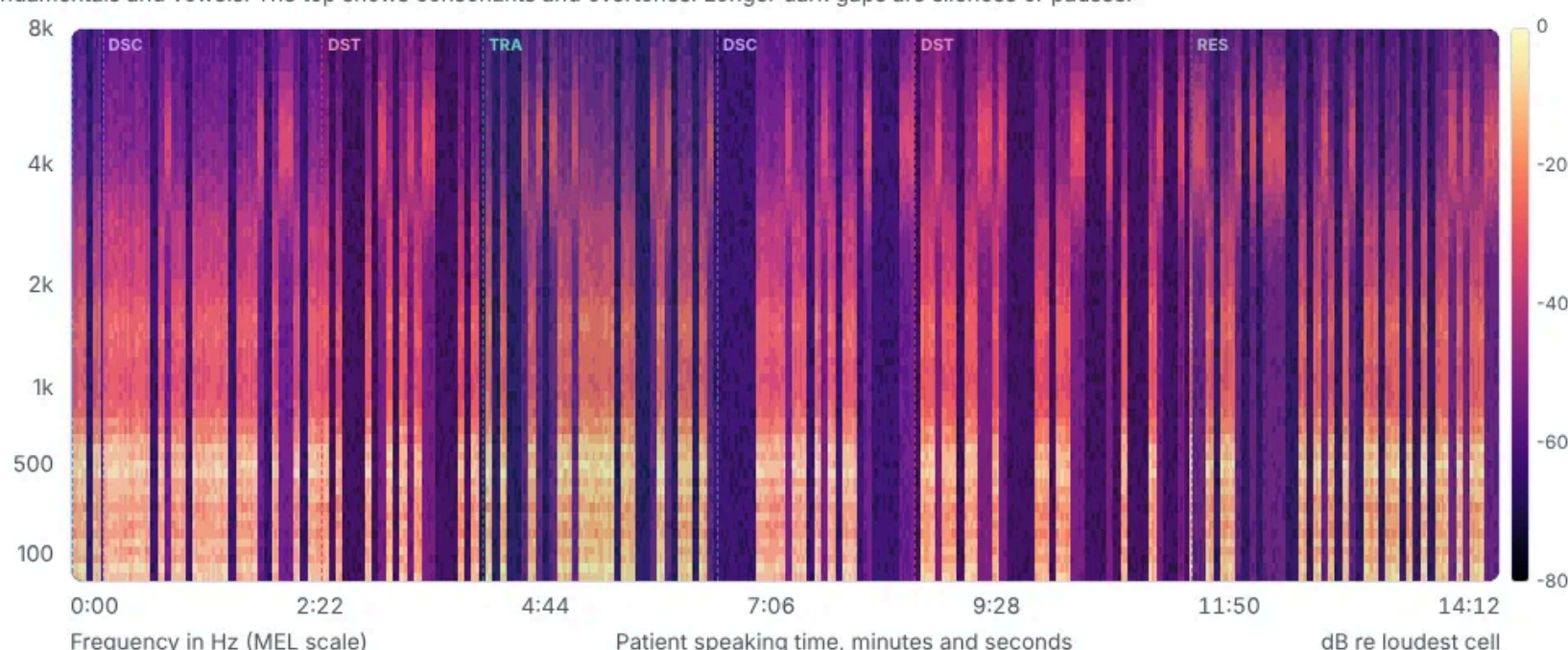
WHAT THE PRACTITIONER SEES FIRST

A picture of the sound, not a score

HAIVA 4 • PROTOTYPE SCREEN

sample data

Vocal energy across 128 MEL frequency bands over time. Bright means high energy. Dark means silence or very quiet speech. The bottom shows voice fundamentals and vowels. The top shows consonants and overtones. Longer dark gaps are silences or pauses.



A spectrogram, with loudness, pitch and pauses on the same time axis, next to the transcript. Where the patient spoke, paused, rose or fell. It does not say how they feel.

GATES 4 AND 5

Only against the patient's own history

HAIVA 4 · PROTOTYPE SCREEN

sample data

🔒

Voice trends

Locked

7

of 10 calibration sessions

Values are hidden from you until calibrated.

What unlocks this view

This patient's personal baseline: 10 analysable sessions in the same setup

Counts only recordings in the same setup that passed every quality check, with speakers confirmed by you. Room microphone and video call are counted separately.

7 of 10

Model validation against PHQ-9, GAD-7 and a structured interview

Pre-registered go or no-go threshold, checked by the scientific team.

0 of 50 validated sessions

Regulatory clearance and a clinical investigation (pilot 2)

Requires the regulatory specialist and an approved investigation plan.

Not started

All three are needed. No values are computed for display before that, so there is nothing hidden behind this tile.

What it will show

Speech rate, pause length and pitch range, each against this patient's own earlier sessions. Never emotions, moods, disorders, risk or a combined score.

- Compared with the patient's own earlier sessions, never with population norms.
- Never shown to the patient. Never a diagnosis. Never written into notes, reports, billing or the safety path.

InnerEcho Initiative · Geneva

Swipe →

OPEN, AND LOOKING FOR PARTNERS

The measuring code **will**
be open.

captavoce[™]

the recording: microphone, phone or
call, with consent

tramavoce[™]

the transcript, with who said what

meravoce[™]

the measures: pauses, timing, pitch,
loudness, each with unit, definition and
citation

Under MPL-2.0. Researchers, universities, speech
scientists and clinicians: **we are looking for validation
partners.**

jb.decorzent@innerecho.ch

innerecho.ch