

OPEN VOICE MODULES · MPL-2.0

Three modules for listening carefully.

How a therapy session's sound becomes a recording, a transcript and measures that can be checked, by anyone, line by line.

RECORD

captavoce™

the recording, kept
 safe and then erased
 with proof

TRANSCRIBE

tramavoce™

who said what, and
 how sure the machine
 was

MEASURE

meravoce™

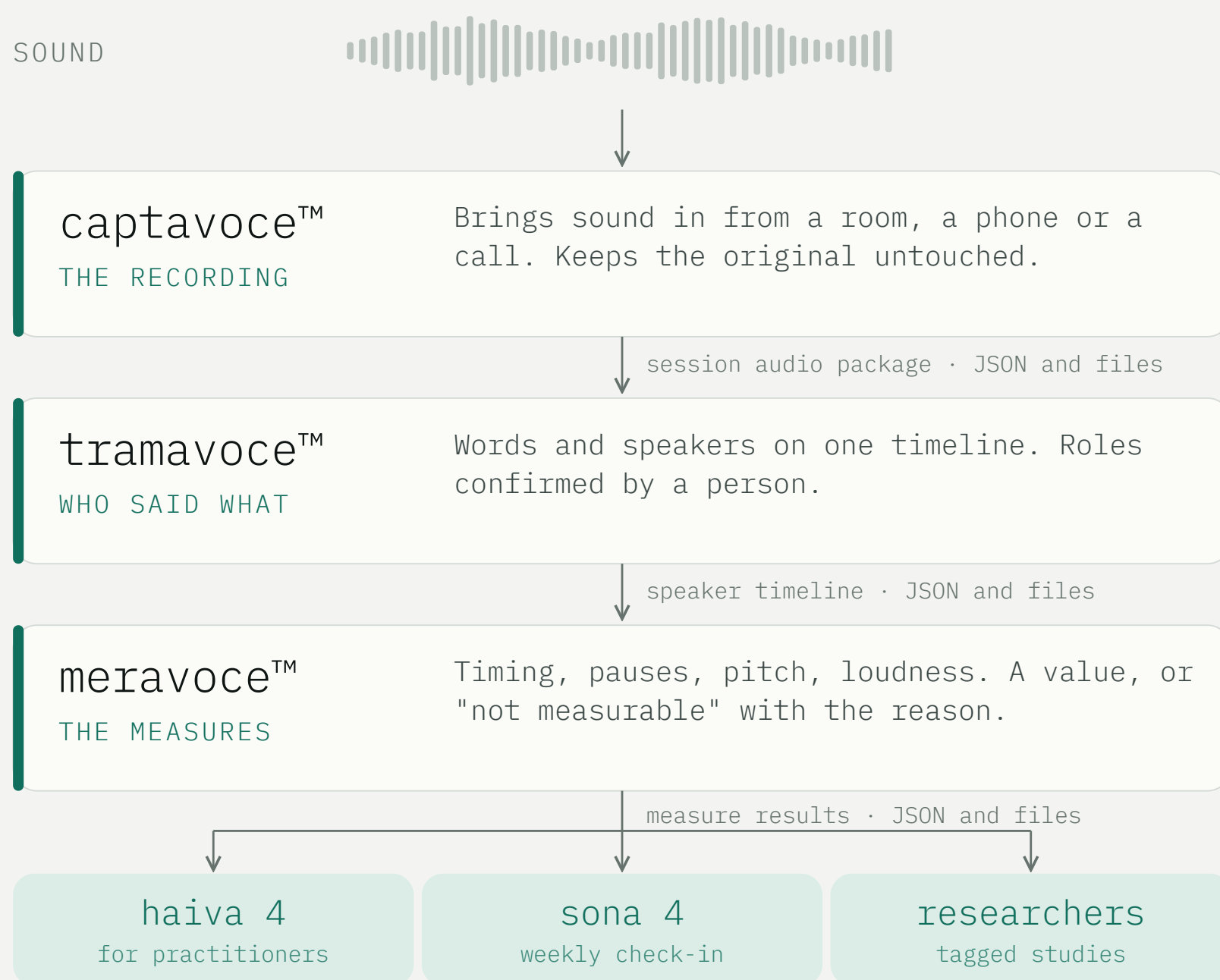
timing, pauses,
 pitch, loudness,
 each with its source



Four views of one voice, from haiva R3 · demonstration record

ONE CHAIN, THREE CONTRACTS

Each module does one job,
and hands on files



No code dependency between the modules, one clock for the whole session, and roles never guessed. Each can be tested, replaced or reused on its own.

CAPTAVOCE™ • THE RECORDING

Many ways in, one honest package out



A recording left on a device is a hazard: an unsafe home, a shared phone, a checkpoint. So captavoce removes it, and proves it did.

CAPTAVOICE™ • WHAT IT HANDS ON

Nothing hidden in the package

SESSION PACKAGE

Original

untouched, with its fingerprint

Working copy

declared recipe, own fingerprint

One session clock

offset, drift and gaps per track

Indicators

level, clipping, noise, dropouts

Verdict per use

yes · limited · no, with the reason

History

every processing step recorded

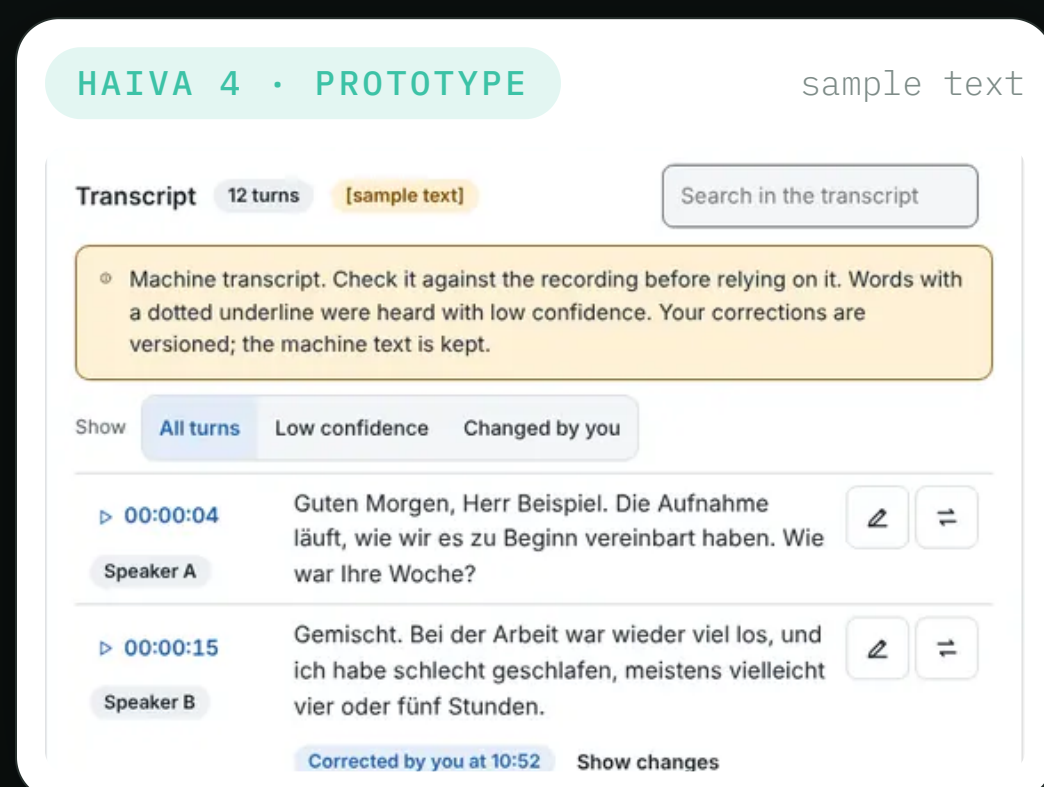
A use not checked is "limited", never "yes". Erasing a package writes an erasure report: what was removed, what could not be reached, and no names or paths.

IN PROGRESS • THE FURTHEST ALONG

Built so far, private: intake into a sealed package, checked erasure, quality indicators, reports in six languages, more than 200 automated tests.

TRAMAVOCE™ · WHO SAID WHAT

A transcript that says how sure it is



Status per layer

If the words fail, the speaker timeline survives.

Roles by a person

The machine separates voices; the practitioner says who is who.

Quiet speech kept

Whispered or tearful words are held back for review, never deleted.

Corrections versioned

The machine text is always kept.

Swiss German is written as a clearly labelled Standard German rendering, never passed off as a word-for-word record.

TRAMAVOCE™ • MODELS

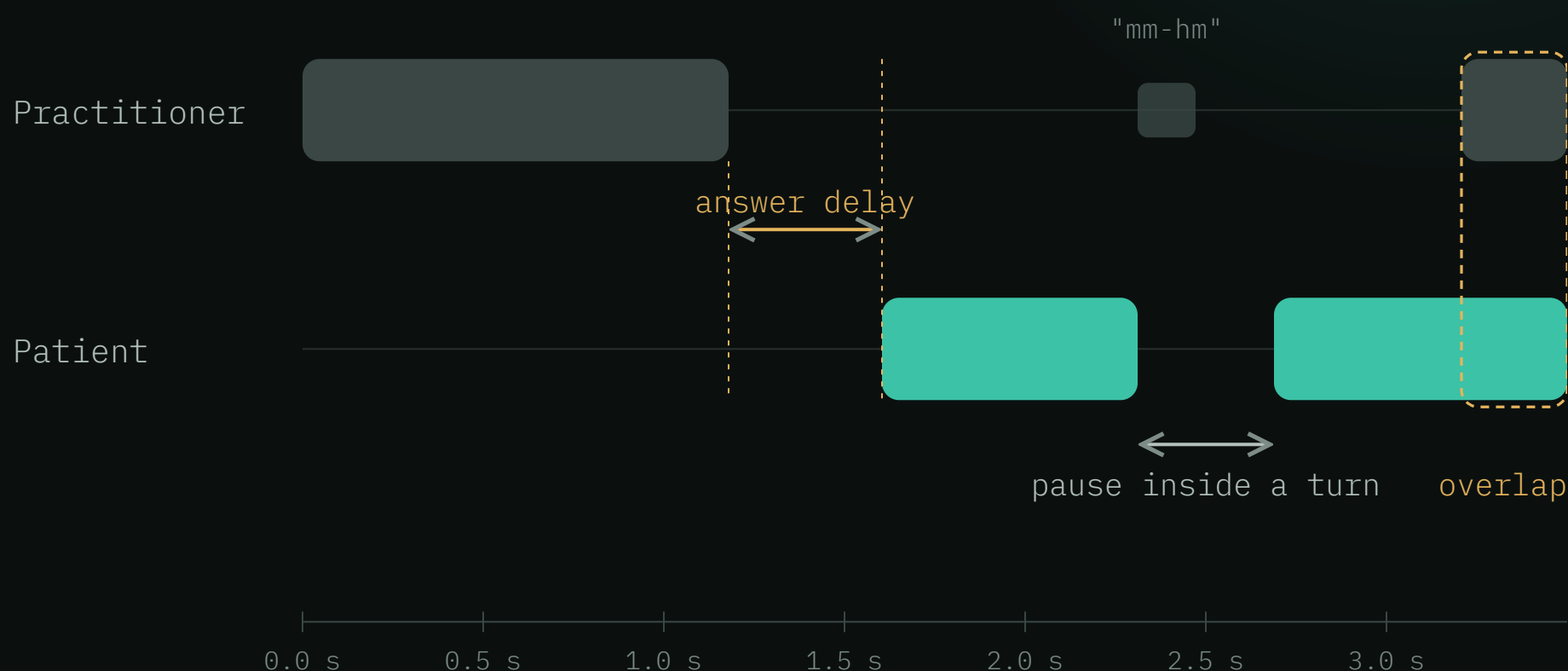
Ships no models. Lists every one.

A licence gate	Every model is named in a lock file with its licence and the status of its training data. Non-commercial, research-only or unlisted: refused.
A bench with known answers	Test conversations built from public speech corpora, with truth files. Each engine is scored; the report scores a perfect timeline at zero errors.
A yardstick, not a wrapper	Voxtral, a strong hosted service, is the bar to reach. Open engines compete on the bench; the winner is chosen by its results.
Third opinion only	Recognisers known to drop fillers and add words during silence are never primary.

On words added during silence: Koenecke et al., ACM FAccT, 2024.

MERA VOCE™ • TIMING

A conversation has a rhythm



Answer delay is always reported with its direction, and a listener's "mm-hm" never starts one. The usual gap between turns is about 200 ms, the same size as the error in deciding where a turn ends. So the threshold and the setting travel with every value, and video-call delays are never compared with room sessions.

Levinson & Torreira, Frontiers in Psychology, 2015

MERAVOCE™ • THE CONTRACT

Every number carries its own explanation

ONE MEASURE RESULT • SCHEMA

measure	pause time inside a turn (example)
unit	seconds
definition	silence $\geq$ threshold inside one speaker's turn
threshold	the silence threshold, stated in every result
citation	published source for the definition
setting	room • video call • phone call • alone
value	– or "not measurable" + reason code
gate	verdict for this measure, this recording
evidence grade	E0 to E5, changed only after review
version	library, settings fingerprint, input fingerprint

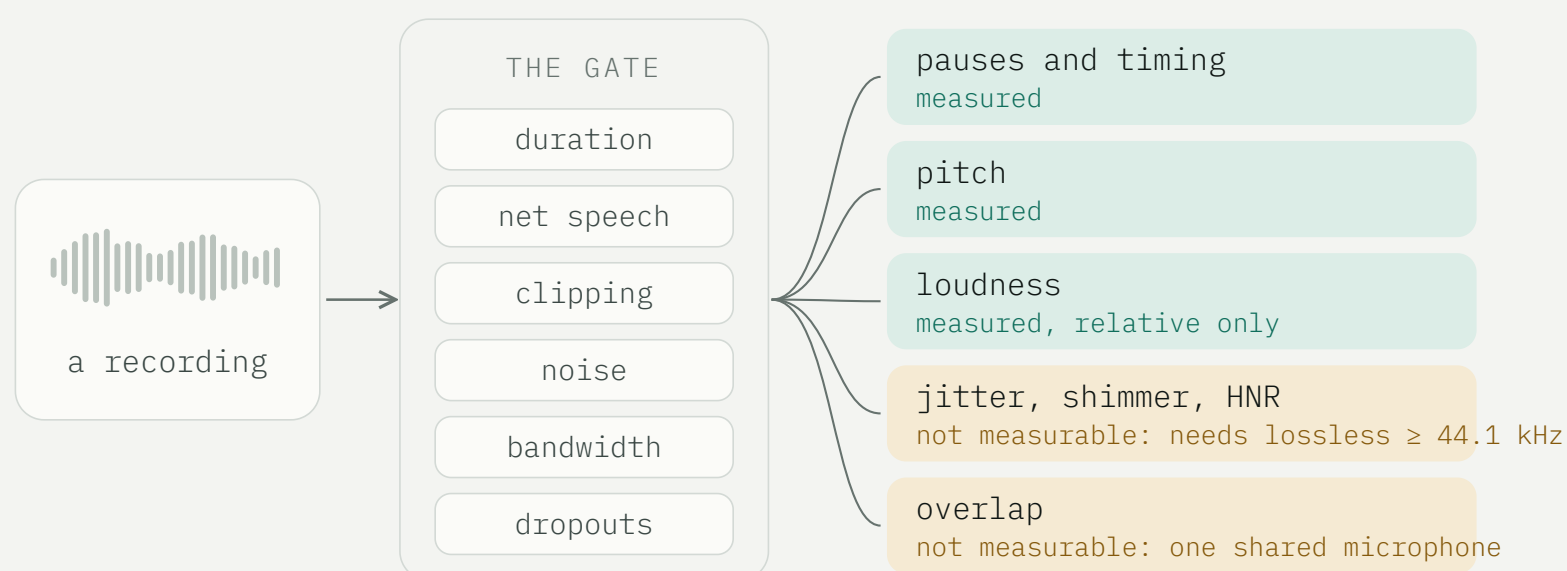
an acoustic measure
meravoce outputs this

a listener's rating
never mixed in

what the person feels
never inferred

MERAVOCE™ • THE GATE

Each measure earns its place



HAIVA 4 • PROTOTYPE

sample data

Companion tracks

Same time axis as the picture above.

Energy contour (RMS) How loud the patient channel is. Peaks are speech, valleys are pauses.



Pitch contour (F0) Estimated fundamental frequency in voiced stretches, 60 to 400 Hz. Gaps are unvoiced or silent.



Voiced and paused Blue is speech or breath activity. Grey marks pauses of 2.5 seconds or more.



Written from published papers only, no borrowed code. 88 standard parameters (Eyben et al., 2016). Validation in stages: verification, analytical, clinical (Goldsack et al., 2020; Robin et al., 2020).

WHERE THEY ARE, AND WHO WE ARE LOOKING FOR

Open when they are ready to be checked

captavoce™	In progress: the first build tasks are done and tested, private for now.
-------------------	--

tramavoce™	In progress: design adopted, first build tasks under way.
-------------------	---

meravoce™	In progress: design adopted, first build task prepared.
------------------	---

Each will be published under MPL-2.0 with its validation and repeatability reports. **Speech scientists, universities, contributors, validation partners:** we would like to hear from you.

jb.decorzent@innerecho.ch

innerecho.ch