



Toward Shareable Clinical Speech: A Fully Local De-Identification Pathway for Clinical Speech Recordings

Peter S. Pressman, MD

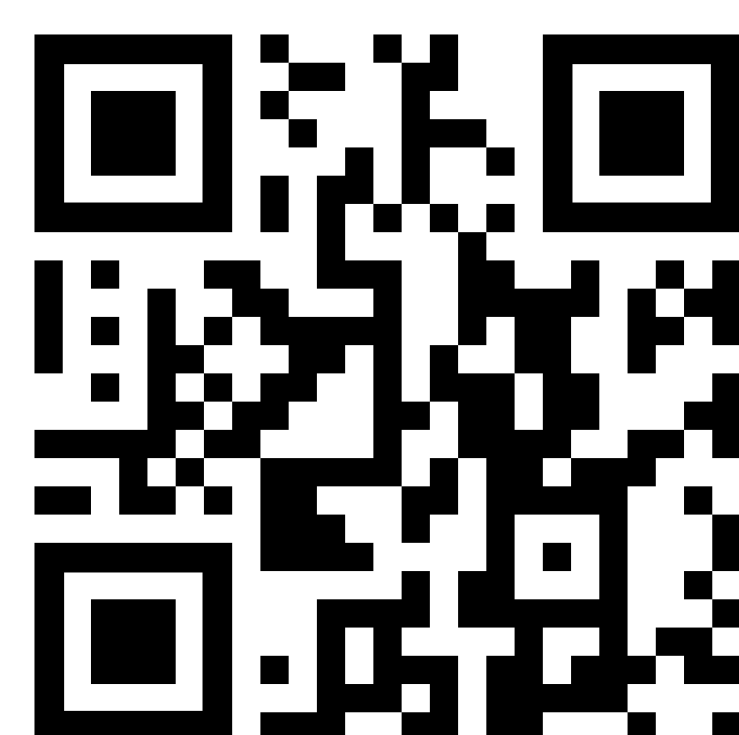
Oregon Health & Science University, Layton Aging and Alzheimer's Disease Research Center

Introduction

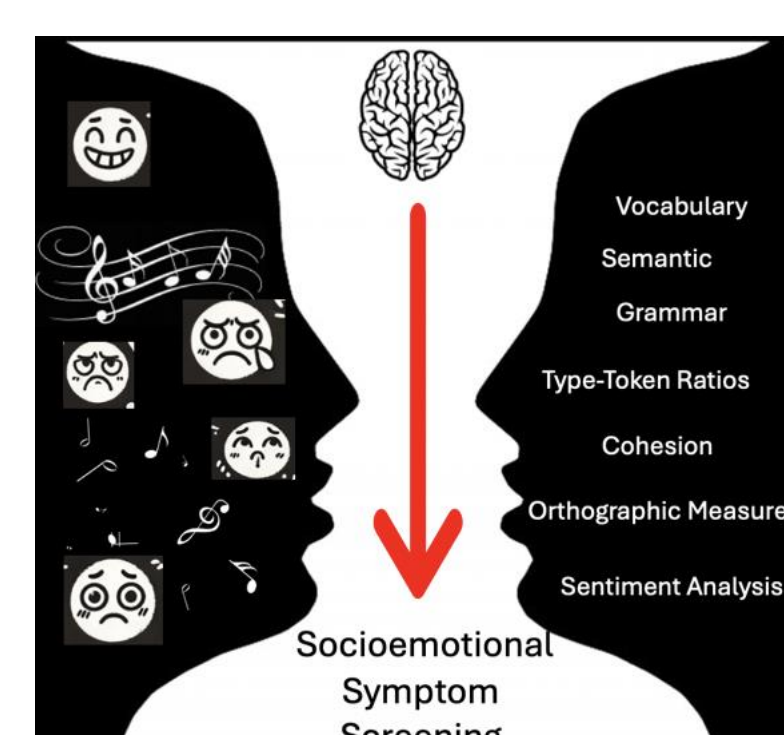
- Across the AD research center network, sites now record patient speech during cognitive testing, a low-cost early signal of decline. However, almost no recordings can now be easily shared, as raw speech is inherently identifiable information. A validated de-identification standard is needed. Meanwhile, centers store recordings and wait.
- The potential benefit.** Under widely reproducible and replicable methods, speech features reach AUC 0.84 for Alzheimer's, on par with testing and structural MRI.¹
- Three problems.** 1) Removing spoken PHI and identifiable acoustic properties. 2) Retaining research-relevant acoustics and linguistic properties. 3) Must be trustworthy and be able to run locally to avoid third-party sharing of identifiable information.
- We built a working, fully local de-identification pipeline, and release it openly.

Methods

- All processing is local. No cloud APIs.
- Transcription.** WhisperX yields word-level timestamps that anchor every downstream redaction.²
- Two-layer voice.** VoicePrivacy B2 at a fixed 0.8, then a randomized pass, so the effective coefficient is [0.75–0.90] drawn per recording.³ A fixed value would let one breach expose the corpus; per-recording randomization defeats cross-session linkage. Speaker re-ID error rises from 4.5% to 76–86% with this step alone.
- Text de-ID, two layers.** CliniDeID removes rule-based PHI; timestamp-driven redaction scrubs the matching audio, replacing with speech-like noise to preserve timing. A local Large Language Model (LLM) then narrows residual identifiability, estimating the population a biographical detail implies and flagging combinations below a re-identification threshold.⁴
- "Anchor" mitigation.** Diarization and role identification find examiner turns; each is replaced with LibriSpeech filler, 251 speakers, run through the identical transform, so filler is acoustically indistinguishable from de-identified speech.

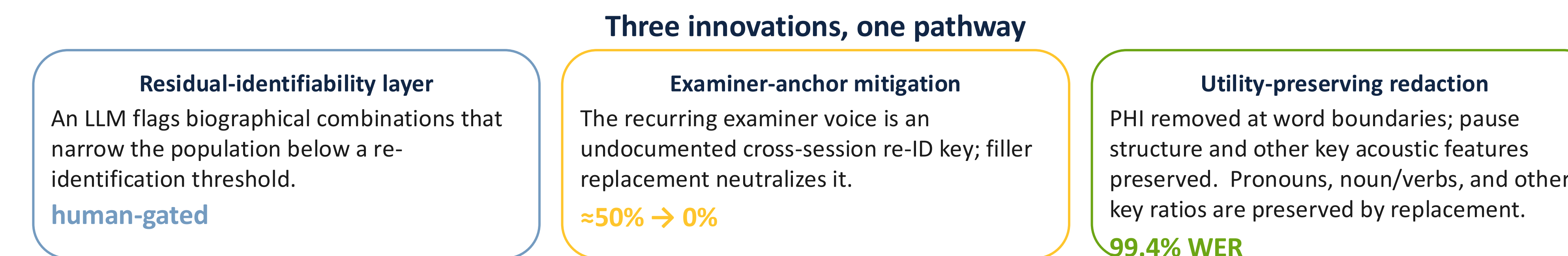
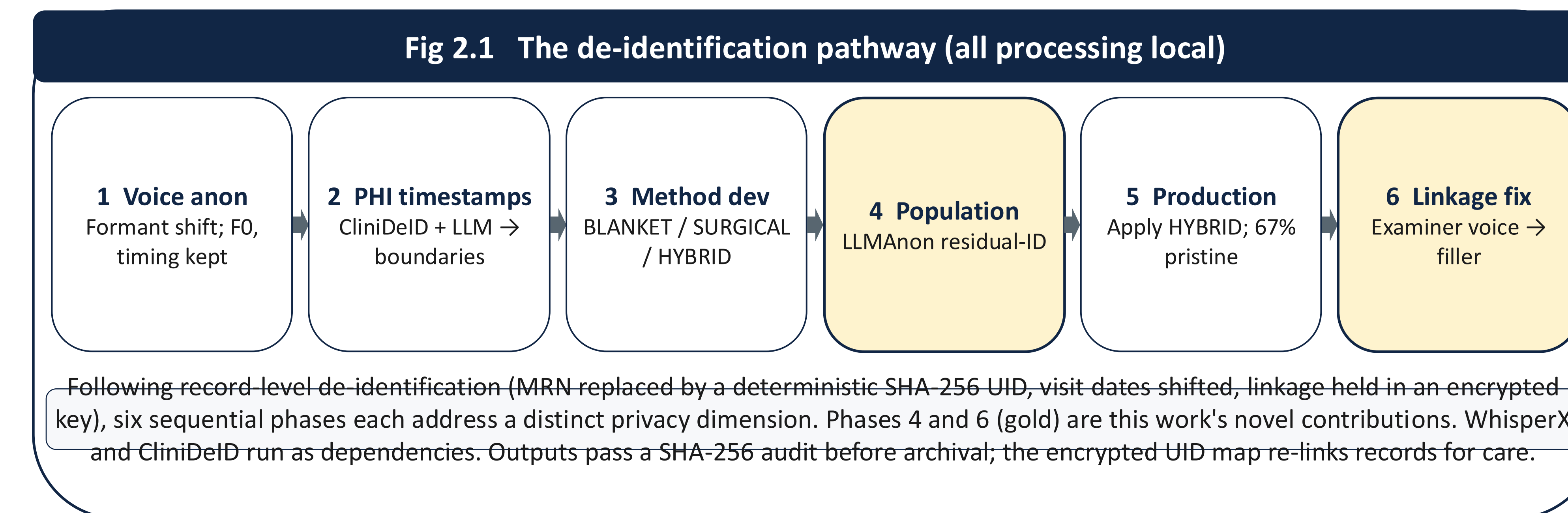


www.csandlab.org



Results

A working, fully local de-identification pathway, validated on clinical-speech recordings and tasks of the kind the National Alzheimer's Coordinating Centers are now collecting.

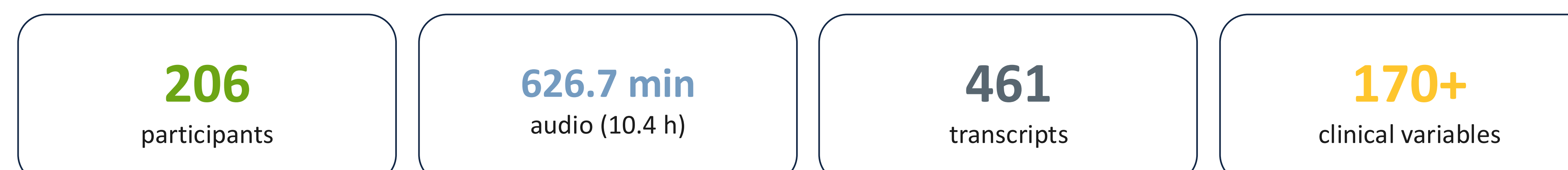


Linguistic redaction bake-off (ASR word error rate; higher = more secure)

Method	Mean WER	Median	Verdict
Original (ref)	5.2%	4.8%	baseline
BLANKET	100%	100%	secure; destroys pause structure
SURGICAL	60.9%	59.6%	FAILED (177 files < 90%)
HYBRID (shipped)	99.4%	100%	secure; pauses kept; 0 files < 90%

SURGICAL's failure is why HYBRID ships. ASR confidence 0.221 vs 0.923; 33.6% of audio modified, 66.4% left pristine.

Corpus (validated this session)



Future Directions

F0 correlation > 0.95 in unmodified regions • 66.4% pristine • only 2.1% redacted at PHI tokens.

Full validation grades every acoustic feature by paired agreement, the original recording against its de-identified twin. The paired jitter, shimmer, and HNR figure is forthcoming.

Grade	Rule	Use on de-identified data
Robust	ICC ≥ 0.90	use freely
Rank-preserved	Spearman ≥ 0.90	re-derive the clinical cutoff
Degraded	0.50 ≤ ICC < 0.90	use with caution
Broken	ICC < 0.50	not a de-identified biomarker

ICC(2,1) absolute agreement, Lin's concordance, and Spearman rank, FDR-corrected. Absolute agreement fails a fixed shift; Spearman still passes it, so a biased-but-ordered feature stays usable. The biomarker panel is the intersection of preservation grade and recording-condition robustness.

Open, citable pipeline code: github.com/pspressman/ASCEND_CODE_REPOSITORY (MIT licensed)

Discussion

- LLMAnon fills a named gap.** Wiest flagged residual identifiability, that removing common identifiers can still leave enough to re-identify, and deferred it. We built that population-narrowing layer for conversational clinical speech. It over-flags on purpose, about 97% false positives, so a human reviewer never misses. Human-gated by design.
- The examiner anchor is a new attack surface.** A single clinician's voiceprint survives participant anonymization and re-links sessions, allowing realignment of original acoustics. Mitigation drops attack success from about 50% to near 0%.
- Filler, not silence.** Silence merges the pauses around a removed turn, erasing turn-taking and latency that carry diagnostic signal, and a gap itself advertises what was cut. Filler preserves timing and hides the seam.
- Preservation, kept honest.** Redacting only PHI tokens, followed by replacement considering key ratios, leaves the word-property substrate, shown robust for AD speech, almost fully intact.¹ Timing preservation rests on engineering logic.⁵
- A standard, not a product.** The pathway runs on consumer hardware, is NIH-funded and openly released, and offers validated de-identification for sharing recordings between centers.

Examiner replacement: design space

Option	Result
HYBRID on examiner	Anchor survives — rejected
Silence	Timing lost — rejected
Flat noise (-35 dB)	Works, but an obvious marker
LibriSpeech filler	Shipped: indistinguishable, timing kept

Compared on privacy vs timing. One measured endpoint (~50%→0%), not a per-option rate.

References

- Caro I, ..., Garcia AM. Benchmarking speech biomarkers of Alzheimer's against cognitive and neural measures. *Alzheimers Dement*. 2026;22(4):e71365. (industry-affiliated; open access)
- Bain M, Huh J, Han T, Zisserman A. WhisperX: time-accurate speech transcription of long-form audio. *Proc INTERSPEECH*. 2023.
- Tayebi Arasteh S, et al. Addressing challenges in speaker anonymization to maintain utility while ensuring privacy of pathological speech. *Commun Med*. 2024;4:182.
- Wiest IC, et al. LLM-Anonymizer: local privacy-preserving de-identification of clinical text. *NEJM AI*. 2025. doi:10.1056/AIdbp2400537.
- RedactOR: transcript-timestamp audio de-identification. *arXiv:2505.18380*. 2025.

Acknowledgements

- Work was funded via NIH K23 AG 063900.
- Thanks to staff and faculty of OHSU, CU, and CSU.