

²Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH. Switzerland.

Data Pipeline Direction

Continuous Side

Data Ingestion Layer

- Microphone
 - ppFunction
 - ...
- Camera
 - ppFunction
 - ...
- ...
 - ppFunction
 - ...

Bridge Layer

MQTT Broker

- "mic"
- "cam"
- "..."

Processing Layer

Metrics Computation

- "..." Service
 - Extractor
 - Extractor
- Speech Service
 - Extractor
 - Extractor

User Profiling

Voice Profiling

Triggered Side

Temporal Context Modeling Layer

- Aggregate Strategies
 - Daily Avg.
 - ...
- Contextualize Strategies
 - EMA
 - HMM
 - ...

Analysis Layer

- Analyse Strategies
 - z-scores
 - Baseline adjustment
 - User Baseline
- Compute Indicators
 - Indicator Computation Service
 - Pyramid Specification

Dashboard Layer

- Visualize Metrics
 - Raw
 - Aggregates
 - Contextual
 - Analyzed
- Visualize Indicators
- Visualize User Baseline
- PHQ-9 Submission

Persistence Layer

- User Embeddings
- Raw Metrics
- Aggr. Metrics
- Context. Metrics
- Analyz. Metrics
- Indicator Scores
- Baseline

[illegible]

The system robustly identifies and ranks acoustic variations. Key features like fundamental frequency (F0) and pause metrics show significant deviations, providing the raw, objective data needed for the linkage model.

Feature	Robustness Score (approx.)
f0_avg	-3.8
glottal_pulse_rate	-3.2
f0_range	-2.8
f0_std	-2.5
spectral_modulation	-2.2
formant_f1_frequencies_mean	-2.0
jitter	-1.8
rate_of_speech	-1.5
f2_transition_speed	-0.5
articulation_rate	-0.5
shimmer	-0.5
psd-4	-0.5
voice_onset_time	-0.5
pause_duration	-0.5
psd-5	-0.5
psd-7	-0.5
pause_count	-0.5
snr	-0.5
mis_energy_range	-0.5
i13	-0.5
voiced16_20	-0.5
mis_energy_std	-0.5
spectral_flatness	-0.5
temporal_modulation	-0.5
hnr_mean	-0.5

- Statistical testing shows the framework's core translation layer.
- Specific acoustic changes predictably map to recognized behavioral states, successfully bridging the gap between sound and insight.

- **Bridging the explainability gap:** our Linkage Framework successfully translates raw, low-level acoustic features into transparent DSM-5 behavioral indicators
 - Transparent acoustic analysis and linking
- **Emphasis on a decision support system (not diagnosis):** passive, "in-the-wild" sensing can yield statistically robust data, providing medical professionals with a reliable and continuous supplementary data stream for patient evaluation.
- **Multi-modal expansion:** extend the framework beyond acoustic data (network traffic, presence sensors)