Universität St.Gallen

Ambient Intelligence: Are We There Yet?

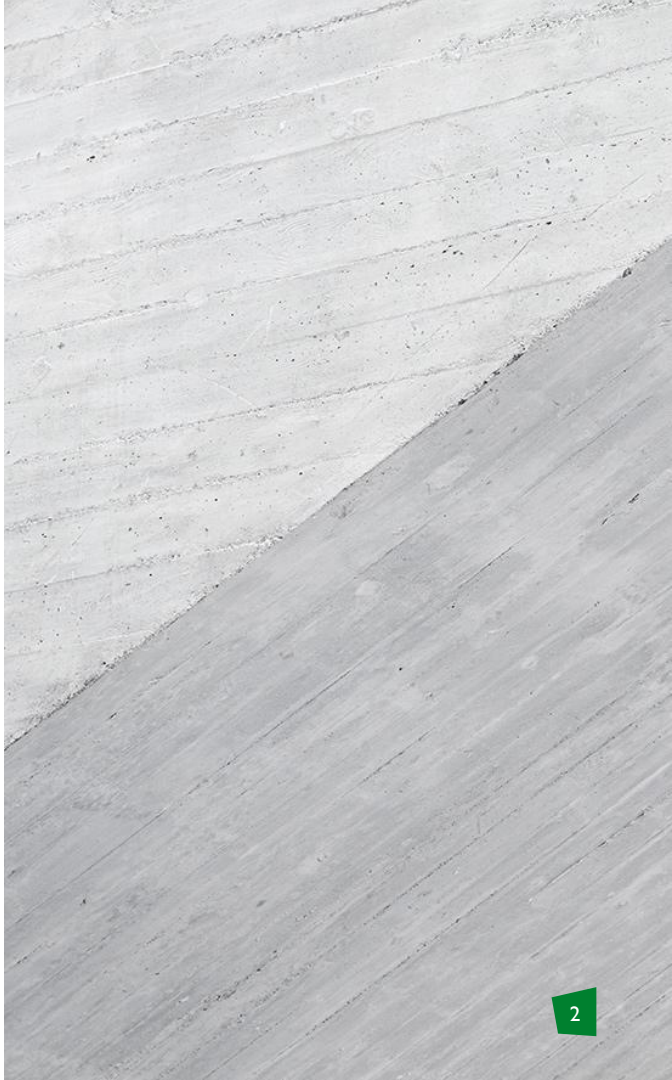
Prof. Dr. Bruno Rodrigues

Institut für Computer Science in Vorarlberg
(ICV-HSG)

From insight to impact

Agenda

1. Ubiquitous Computing and Ambient Intelligence
2. Building Automation vs Intelligence
3. Core Challenges
4. Arbitration of Conflicts in Shared Spaces
5. Case Study: Home as a Diagnostic Tool
6. Summary and Outlook



Computers and Computing are Everywhere

In 1991, Mark Weiser (chief scientist at Xerox PARC) coined the term "**Ubiquitous computing**" or "**Pervasive Computing**"

Computation will be integrated in our daily lives in a way that they will disappear

- Pagers, personal digital assistants, CD players
- First commercial mobile phones
- Mostly associated to mobile, wearable devices


The Computer for the 21st Century

"The most profound technologies are those that disappear" Mark Weiser, 1991.

Ubiquitous or Pervasive Smart Devices

You probably walked by one of this today

Presence sensors

You probably have seen by, have seen or own one of these

IP Cameras

You probably have seen or own one of these

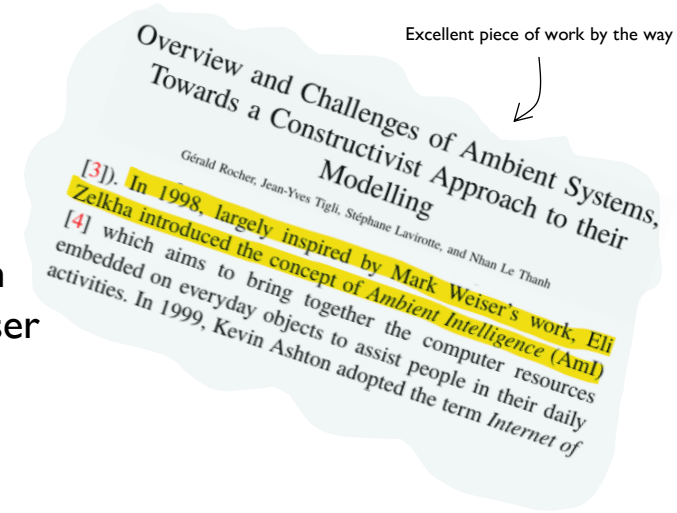
Vacuum cleaners

Smart lock

Smart thermostats

Ambient-intelligence (AmI)

Excellent piece of work by the way



Overview and Challenges of Ambient Systems, Towards a Constructivist Approach to their Modelling

Gérald Rocher, Jean-Yves Tigli, Stéphane Lavirotte, and Nhan Le Thanh

[3]). In 1998, largely inspired by Mark Weiser's work, Eli Zelkha introduced the concept of *Ambient Intelligence* (AmI) [4] which aims to bring together the computer resources embedded on everyday objects to assist people in their daily activities. In 1999, Kevin Ashton adopted the term *Internet of*

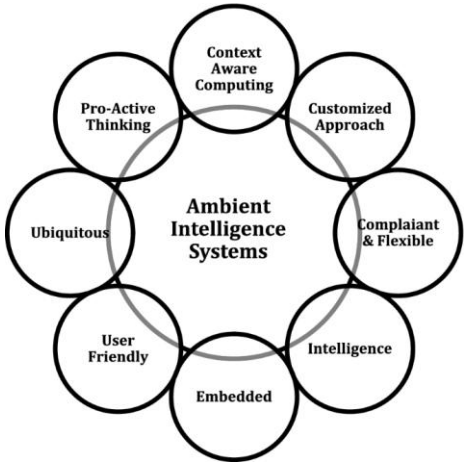
In 1998, Eli Zelkha (and his team) coined the term "**Ambient Intelligence**," inspired by Mark Weiser

CONCEPT

Refers to electronic environments that are sensitive and responsive to the presence of people.

CHARACTERISTICS

- **Ubiquitous**: it is integrated into the environment
- **Context-aware**: able to categorize actions within an environment
- **Personalized**: adapt the environment to personal needs (individual or group)
- **Adaptive**: respond to environmental changes to achieve a goal (typically increasing well-being)



The “Smart Conference Room” Example

HVAC may operate in a fixed time-schedule or by activated by a remote. No awareness on real occupancy.

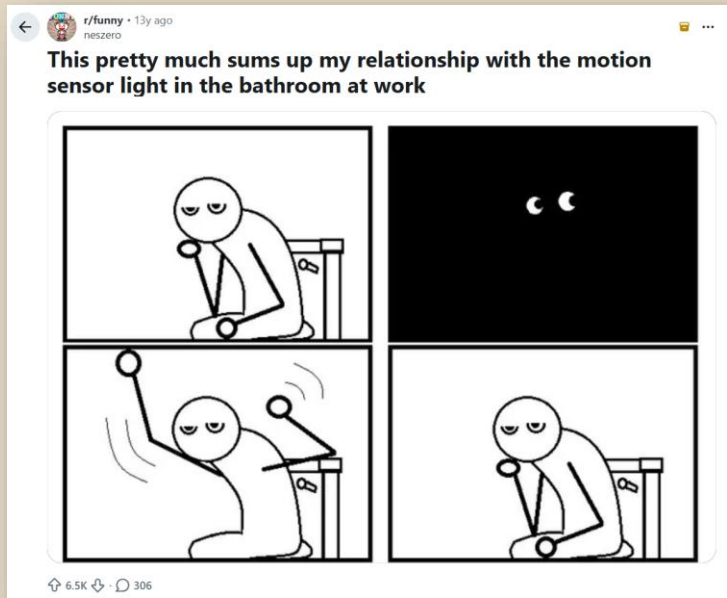
Booking system synchronized with a calendar but is not integrated with other room systems.

Lighting system may use local motion sensors. Knows whether the room is empty, but does not share any data.

Conference system that has camera/microphones but is not able to integrate with other room systems.

HVAC: Heating, ventilation, and air conditioning

Are ubiquitous devices surrounding us closer to **ambient intelligence** or to just **automation**?



Less context may be better in this specific case

The Data Paradox

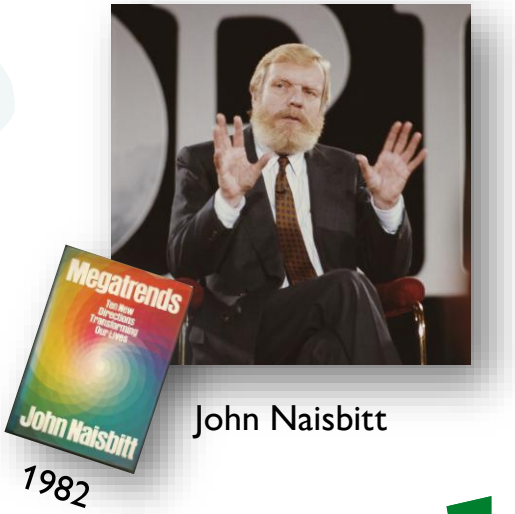
THE DATA FLOOD

More sensors than ever but isolated systems creating massive amounts of data that largely remain siloed (and underutilized)

THE PERCEPTION GAP

Connected systems does not lead to perception.
A security system might know one entered a building, but the HVAC waits for local motion to be triggered.

"We are drowning in information but starving for knowledge"



John Naisbitt

And in our context: "We are drowning in data but starving for context"

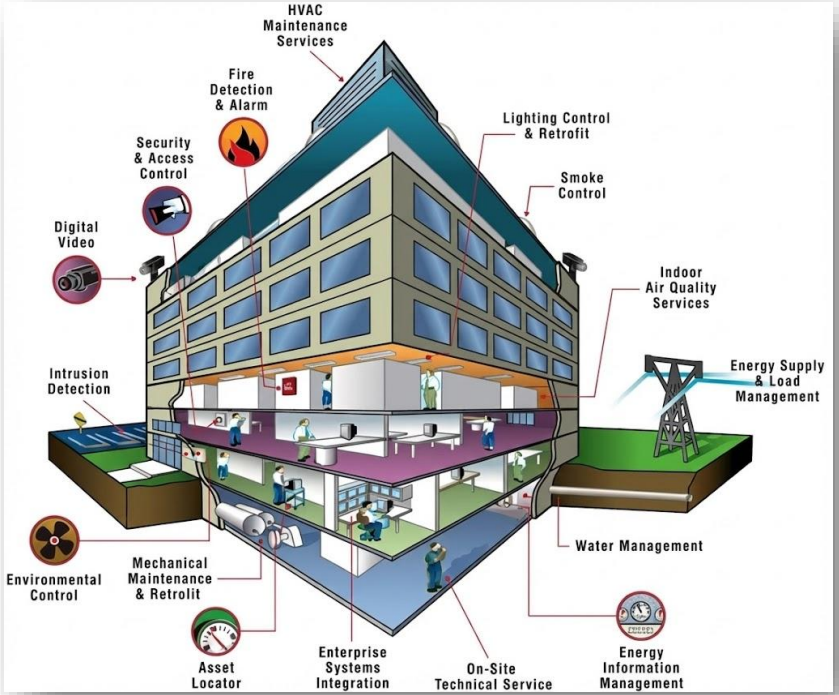
Smart Buildings - Where We Stand

ISOLATED AUTOMATION (PAST)

- Initial era of "smart buildings" with systems operating in isolation
- Examples: Doors opening via motion sensors, HVAC operating based on time schedules
- Functional but "dumb" → no communication

CONNECTED AUTOMATION (PRESENT)

- Systems connect to a central point at the edge or cloud
- Automation routines, remote control via Web or Mobile application
- Still…connected systems operating in their **own data silos**



Source: https://www.csemag.com/integration-building-automation-and-fire-alarms/

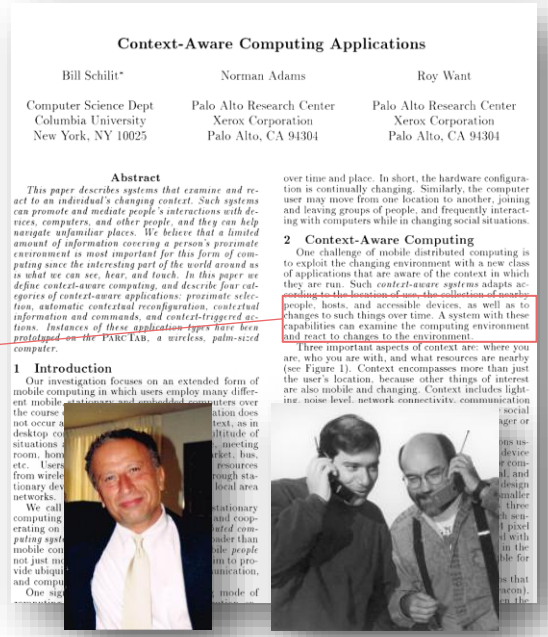
Smart Buildings - Where We Want to Go

AMBIENT INTELLIGENCE (THE FUTURE)

> *"…A system with these capabilities can examine the computing environment and react to changes to the environment..."*

- Requires perception, contextualization, and continuous adaption of the environment
- Shift from scenario-based templates to goal-driven environments

RULE-BASED (past)	SCENARIO-BASED (present)	GOAL-BASED (future)
"If This, Then That." Rigid, brittle logic. It fails when nuances arise that weren't hard-coded by the programmer.	**"Set Scene: Presentation."** A pre-set static combination of settings. Better, but still requires manual activation or rigid triggers.	**"Goal: Maximize Focus."** The AI acts as an agent, continuously adjusting diverse systems to maintain the ideal state for the user's task.



Context-Aware Computing Applications

Bill Schilit* Norman Adams Roy Want

Computer Science Dept Columbia University New York, NY 10025

Palo Alto Research Center Xerox Corporation Palo Alto, CA 94304

Palo Alto Research Center Xerox Corporation Palo Alto, CA 94304

Abstract

This paper describes systems that examine and react to an individual's changing context. Such systems can promote and mediate people's interactions with devices, computers, and other people, and they can help navigate unfamiliar places. We believe that a limited amount of information covering a person's proximate environment is most important for this form of computing since the interesting part of the world around us is what we can see, hear, and touch. In this paper we define context-aware computing, and describe four categories of context-aware applications: proximate selection, automatic contextual reconfiguration, contextual information and commands, and context-triggered actions. Instances of these application types have been prototyped on the PARCTAB, *a wireless, palm-sized computer.*

1 Introduction

Our investigation focuses on an extended form of mobile computing in which users employ many different mobile, stationary, and embedded computers over the course of

2 Context-Aware Computing

One challenge of mobile distributed computing is to exploit the changing environment with a new class of applications that are aware of the context in which they are run. Such *context-aware systems* adapts according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time. A system with these capabilities can examine the computing environment and react to changes to the environment.

Three important aspects of context are: where you are, who you are with, and what resources are nearby (see Figure 1). Context encompasses more than just the user's location, because other things of interest are also mobile and changing. Context includes lighting, noise level, network connectivity, communication

Eli Zelkha

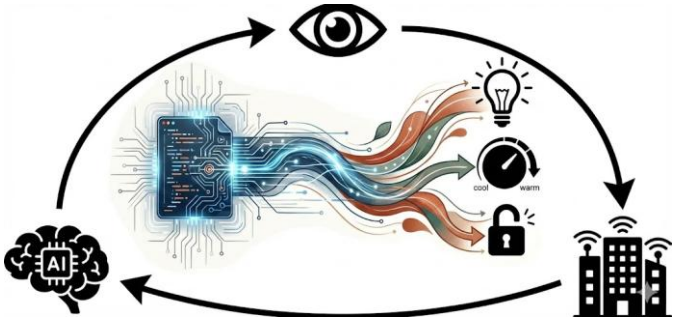
Bill Schilit and Mark Weiser

What these gentlemen have in common?

Ambient Intelligence Requires Mastering **Three Core Challenges**

PERCEPTION

How do we sense the environment? Involves acquiring raw data from a multitude of sensors to understand *what* and *where*



REACTION

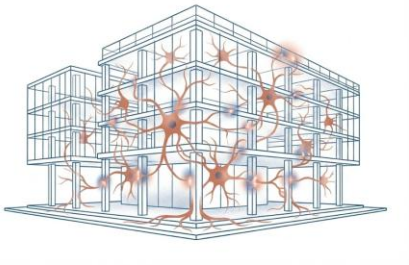
How do we act intelligently on our understanding? Involves synthesizing context to achieve **goals**, arbitrate conflicts, and adapt the environment

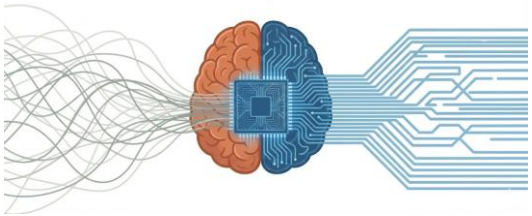
CONTEXTUALIZATION

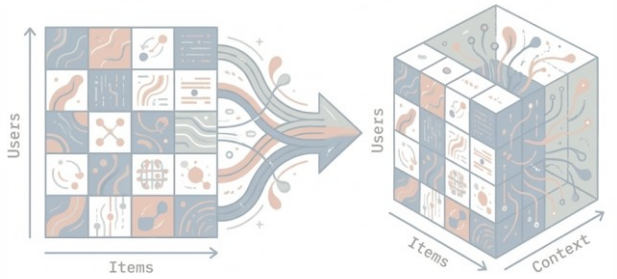
How do we digitally model the environment?

Involves modeling the environment as a virtual space that can be updated in real-time

Perception and Contextualization – A Generic Pipeline







Stage 1: Data Acquisition

- Raw sensor readings collected in segments via a sliding window

Stage 2: Preprocessing

- Data is filtered and fused to correct "gaps" and noise

Stage 3: Feature Extraction

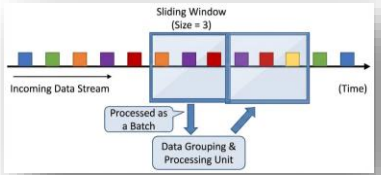
- Statistical characteristics are extracted to create a "feature vector"

Stage 4: Classification

- An in-house LLM maps the feature vector to a known context class

Stage 5: Postprocessing

- Smoothing technique to eliminate inconsistencies

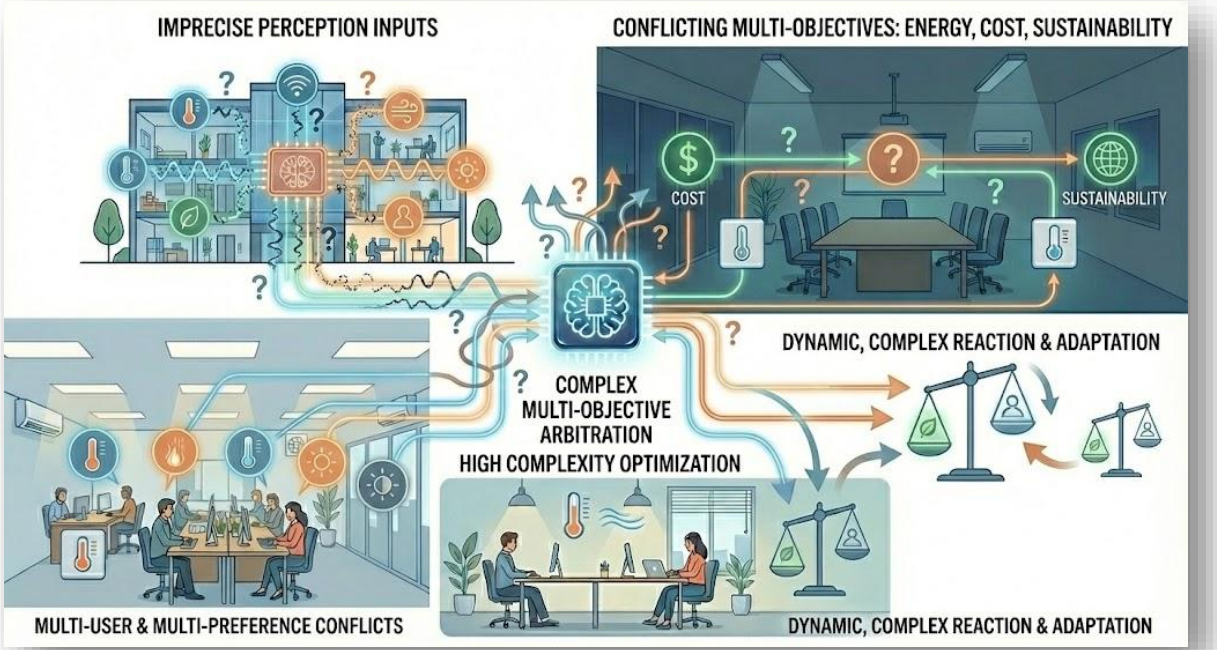


We figured out a context,
which is only half the battle.
What comes next?

Contextualization and Reaction: Not Straightforward

CHALLENGES

- Perception may not be precise
- Opposing goals: multi-user, multi-objective
- Requires solving multi-objective optimization in real-time



A truly smart system must arbitrate these conflicts to find the optimal human-centric trade-off.

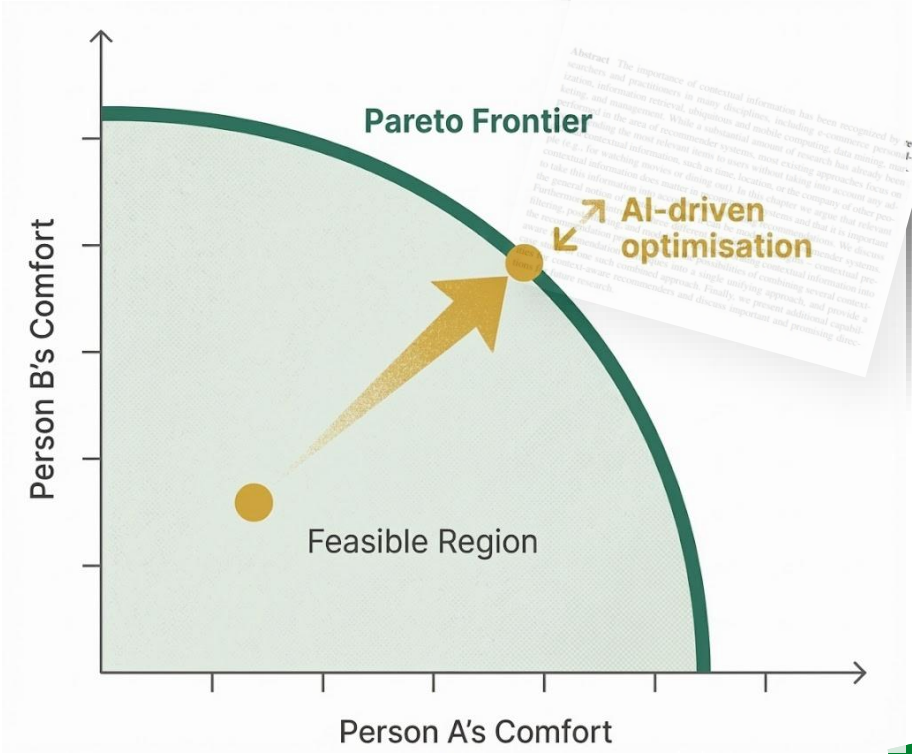
Context-aware Recommender System

THE PROBLEM IS KNOWN

- Shared spaces are often zero-sum games: Person A wants 20°C; Person B wants 24°C.
- Maximizing one person's comfort directly negatively impacts the other.
- Pareto-optimal solution is key

NOVELTY: EDGE COMPUTING AND LLMs

- LLMs map qualitative human feedback into quantitative utility functions
  - Drawing area under "**feasible region**"
- Classic ML deployed at the edge perform a constrained optimization to find the **pareto frontier**



Current State-of-Research: The House as a Perception Layer

PASSIVE MONITORING

- No need for wearables, charging, or manual mood journaling
- Reduced “anxiety” once people are already in their living environment

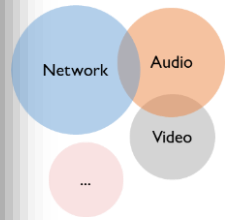
LONGITUDINAL BASELINE

- Possible to analyze behavioral baseline over months of data
- Compare and detect subtle behavioral drifts
- Multi-modality

PRIVACY BY DESIGN

- No data is up-streamed to a cloud server or leave the premises

  (well, at least in the system proposed here)



IHearYou: Linking Acoustic Features to DSM-5 Depressive Behavior Indicators

Jonas Länzlinger[1], Katharina O.E. Müller[2], Bruno Rodrigues[1]
[1]Embedded Sensing Group ESG, School of Computer Science SCS, University of St. Gallen HSG, Switzerland
[2]Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH, Switzerland
E-mail:jonas.laenzlinger@student.unisg.ch|mueller@ifi.uzh.ch|bruno.rodrigues@unisg.ch

CareNet: Linking Home-router Network Traffic to DSM-5 Depressive Behavior Indicators

Stephan Nef, Bruno Rodrigues
Embedded Sensing Group ESG, School of Computer Science SCS, University of St. Gallen HSG, Switzerland
E-mail: stephan.nef@student.unisg.ch|bruno.rodrigues@unisg.ch

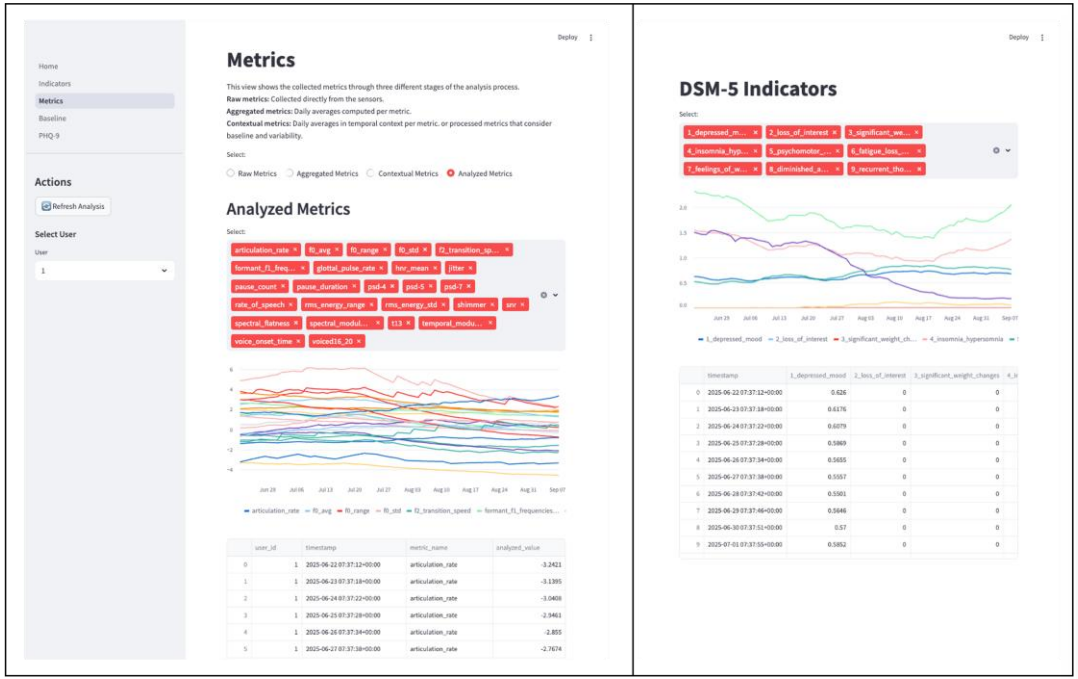
IHearYou – Linking Speech Features to Depression Symptoms

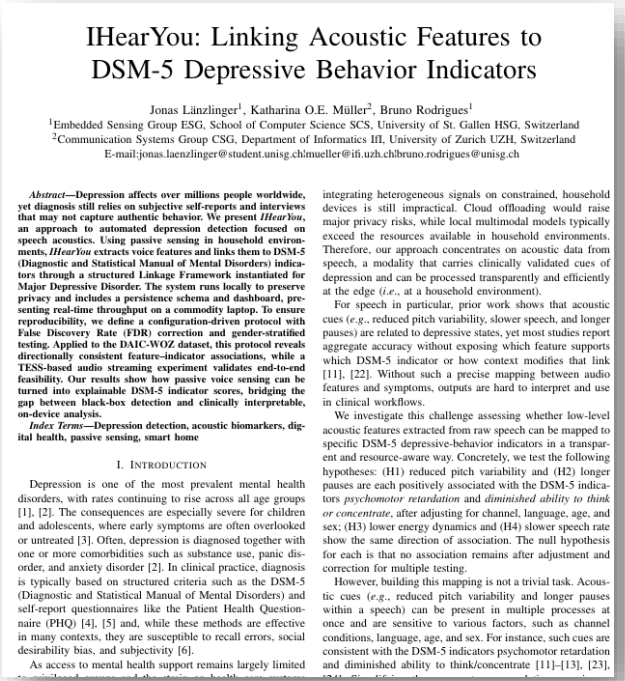
APPROACH

- Explainable translation rules between **voice** features and depression symptoms
- Important to measure against own personal baseline (longitudinal)
- All data is computed at the household

Video trimmed ~ 2:30 mins

Demonstration





IHearYou: Linking Acoustic Features to DSM-5 Depressive Behavior Indicators

Jonas Länzlinger[1], Katharina O.E. Müller[2], Bruno Rodrigues[1]
[1]Embedded Sensing Group ESG, School of Computer Science SCS, University of St. Gallen HSG, Switzerland
[2]Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH, Switzerland
E-mail:jonas.laenzlinger@student.unisg.ch|mueller@ifi.uzh.ch|bruno.rodrigues@unisg.ch

***Abstract*—Depression affects over millions people worldwide, yet diagnosis still relies on subjective self-reports and interviews that may not capture authentic behavior. We present *IHearYou*, an approach to automated depression detection focused on speech acoustics. Using passive sensing in household environments, *IHearYou* extracts voice features and links them to DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) indicators through a structured Linkage Framework instantiated for Major Depressive Disorder. The system runs locally to preserve privacy and includes a persistence schema and dashboard, presenting real-time throughput on a commodity laptop. To ensure reproducibility, we define a configuration-driven protocol with False Discovery Rate (FDR) correction and gender-stratified testing. Applied to the DAIC-WOZ dataset, this protocol reveals directionally consistent feature–indicator associations, while a TESS-based audio streaming experiment validates end-to-end feasibility. Our results show how passive voice sensing can be turned into explainable DSM-5 indicator scores, bridging the gap between black-box detection and clinically interpretable, on-device analysis.**

***Index Terms*—Depression detection, acoustic biomarkers, digital health, passive sensing, smart home**

I. Introduction

Depression is one of the most prevalent mental health disorders, with rates continuing to rise across all age groups [1], [2]. The consequences are especially severe for children and adolescents, where early symptoms are often overlooked or untreated [3]. Often, depression is diagnosed together with one or more comorbidities such as substance use, panic disorder, and anxiety disorder [2]. In clinical practice, diagnosis is typically based on structured criteria such as the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) and self-report questionnaires like the Patient Health Questionnaire (PHQ) [4], [5] and, while these methods are effective in many contexts, they are susceptible to recall errors, social desirability bias, and subjectivity [6].

As access to mental health support remains largely limited

integrating heterogeneous signals on constrained, household devices is still impractical. Cloud offloading would raise major privacy risks, while local multimodal models typically exceed the resources available in household environments. Therefore, our approach concentrates on acoustic data from speech, a modality that carries clinically validated cues of depression and can be processed transparently and efficiently at the edge (*i.e.*, at a household environment).

For speech in particular, prior work shows that acoustic cues (*e.g.*, reduced pitch variability, slower speech, and longer pauses) are related to depressive states, yet most studies report aggregate accuracy without exposing which feature supports which DSM-5 indicator or how context modifies that link [11], [22]. Without such a precise mapping between audio features and symptoms, outputs are hard to interpret and use in clinical workflows.

We investigate this challenge assessing whether low-level acoustic features extracted from raw speech can be mapped to specific DSM-5 depressive-behavior indicators in a transparent and resource-aware way. Concretely, we test the following hypotheses: (H1) reduced pitch variability and (H2) longer pauses are each positively associated with the DSM-5 indicators *psychomotor retardation* and *diminished ability to think or concentrate*, after adjusting for channel, language, age, and sex; (H3) lower energy dynamics and (H4) slower speech rate show the same direction of association. The null hypothesis for each is that no association remains after adjustment and correction for multiple testing.

However, building this mapping is not a trivial task. Acoustic cues (*e.g.*, reduced pitch variability and longer pauses within a speech) can be present in multiple processes at once and are sensitive to various factors, such as channel conditions, language, age, and sex. For instance, such cues are consistent with the DSM-5 indicators psychomotor retardation and diminished ability to think/concentrate [11]–[13], [23],

Major Takeaways

"The most profound technologies are those that disappear"

Mark Weiser, 1991

- **Context-perception is not far from reality**, but not possible without real exchange of data between systems of multiple vendors
- **Context-reaction is still far from reality** as requires precise perception and continuous goal adaptation

AI allows the hardware to finally recede into the background, making the building a silent, supportive partner in our daily lives.

Questions?

Universität St.Gallen

Prof. Dr. Bruno Rodrigues
Assistenzprofessor

bruno.rodrigues@unisg.ch

EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

"From insight to impact"