Towards Accessible Interactive Visual Elements on the Web: An AI-Based Descriptive and Auditory Input/Output Approach for Screen Reader Users

Toni Barth — Leonie Weinert — Prof. Dr. Frank Heckel
Anhalt University of Applied Sciences
toni.barth@hs-anhalt.de

03/09/2026 — WEB 2026

Speaker

Toni Barth

PhD Student, game development & accessibility

Research focus

- Web accessibility
- AI-supported interfaces

Goal

- Reducing dependency of blind users in visual workflows

Introduction

- The modern web is increasingly visual and interactive
- Digital whiteboards, dashboards, collaborative canvases
- Screen readers are linear — visual systems are spatial & parallel

⟶ Interactive visual content remains largely inaccessible

Example: Digital Whiteboard

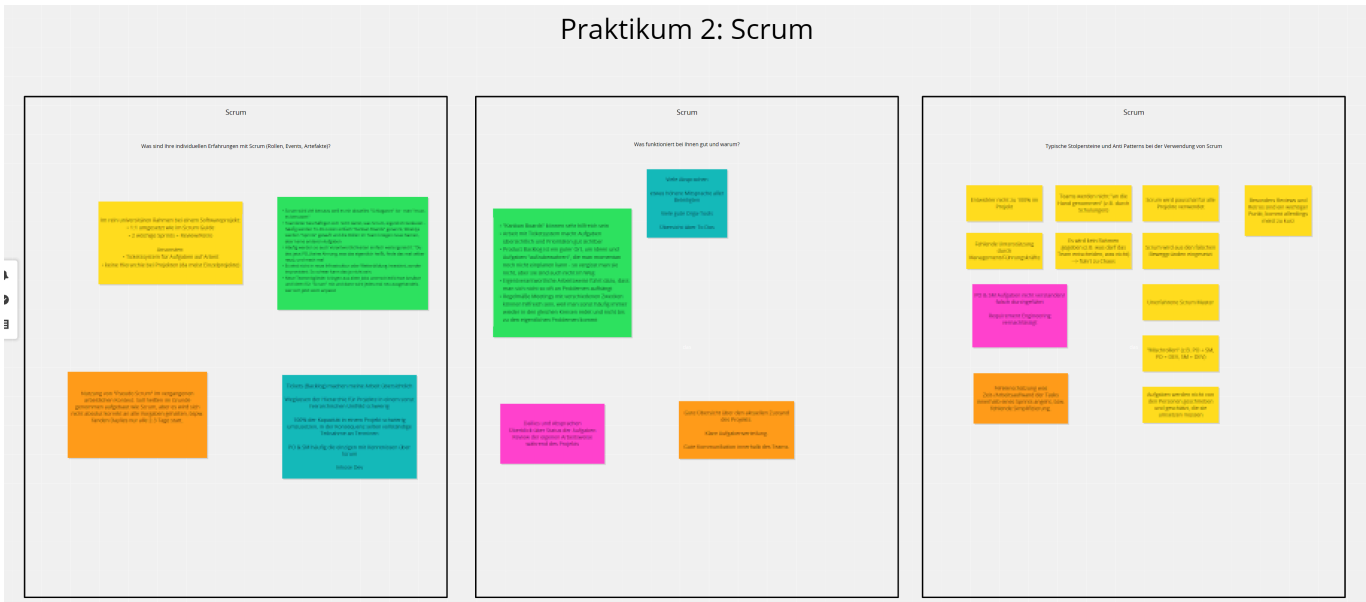

Figure 1: Digital whiteboard used in a workshop where students wrote down their thoughts on Scrum.

Example: Digital Whiteboard

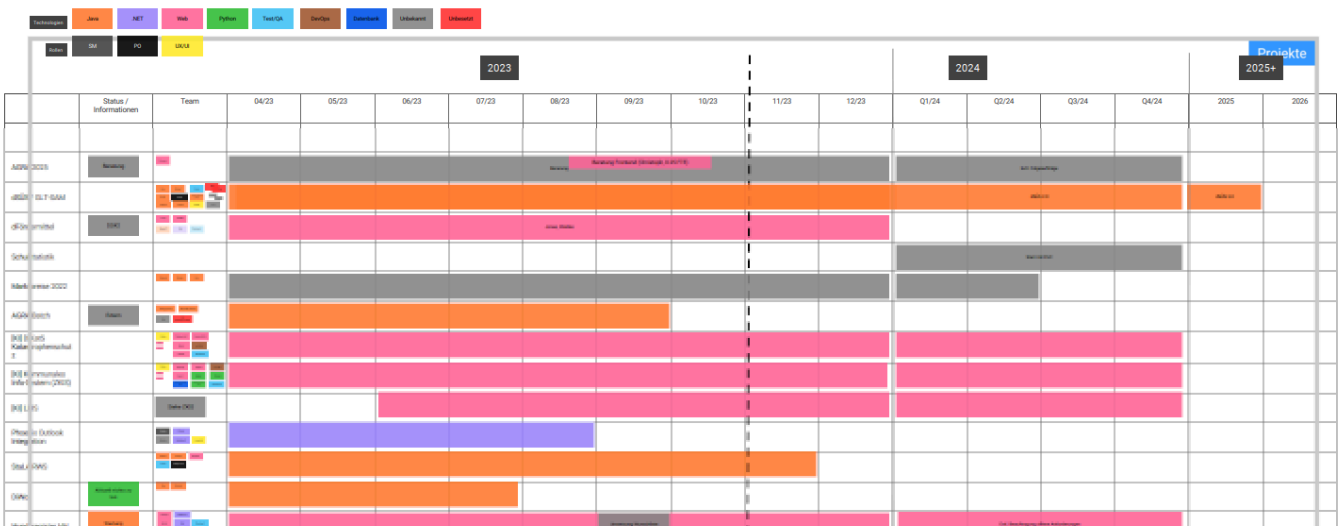

Figure 2: Digital whiteboard used for multi project planning in a company with multiple teams and tech stacks.

Use Case: Collaborative Digital Whiteboard

Scenario:

- Team brainstorming session
- Visual participants draw & move objects
- Blind user receives spatial audio updates
- User contributes via voice or text commands

Use Case: Collaborative Digital Whiteboard

Beneficiaries:

- Screen reader users
- Any person who cannot look at a screen continuously
- Developers of collaboration tools
- Organizations aiming at inclusive workflows

Why Is This Relevant?

Real Problems:

- Dependency on intermediaries
- Reduced participation in workshops
- Limited access to modern collaboration platforms

Broader Impact:

- Equal contribution in education
- Inclusion in professional environments
- Potential new accessibility standards

Core Idea

How it works conceptually:

- AI detects semantically meaningful visual events
- Sound transports spatial information in parallel to speech
- Natural language enables non-visual manipulation

What makes it better?

- Not just description — but interaction
- Filters noise ⟶ reduces cognitive load
- Supports independent collaboration

Core Idea

One sentence:

An AI-based bidirectional system that translates visual interaction into spatialized audio and natural-language interaction — in real time.

State of the Art

- ARIA & WCAG ⟶ focus on static / moderate dynamic content
- Screen readers ⟶ sequential speech output
- Sonification research ⟶ spatial audio reduces cognitive load
- AI-based description ⟶ mostly offline, not real-time

Gap:

No integrated system for real-time, interactive visual collaboration.

Problem 1: Communicating Visual Changes

Challenges:

- Rapid parallel updates
- Spatial information
- Identifying relevant vs. transient events
- Avoiding cognitive overload

Key question:

How to summarize dynamic visual changes meaningfully in real time?

Problem 2: Enabling Interaction

Challenges:

- No spatial mouse-based interaction
- Keyboard insufficient for complex spatial tasks
- Requires machine-consistent mental model

Requirement:

Non-visual specification of objects, properties, and location.

Proposed System Architecture

HOCHSCHULE ANHALT University of Applied Sciences

Three Components:

1. AI-driven visual change analysis
2. Spatialized auditory representation
3. AI-based natural-language manipulation

$\longrightarrow$ Closed perception-action loop

Component 1: AI Change Detection

- Monitor visual workspace continuously
- Detect low-level events (add, move, modify)
- Aggregate into semantic interaction events
- Filter by contextual relevance

Output:

What changed?
Where?
Who initiated it?

Component 2: Auditory Representation

- Spatial audio mapping (left/right, distance)
- Sound textures represent object types/actions
- Parallel perception beyond linear speech
- Focus-based announcements to reduce overload

Optional: Detailed speech on demand via selection

Component 3: Natural Language Input

Examples:
“Draw a blue rectangle in the top right corner.”
“Connect node A to node B.”

AI extracts:

Intent, Object type, Attributes, Spatial reference

⟶ Translates into graphical operations

Open Questions & Risks

- Real-time performance constraints
- Efficient change detection intervals
- Local vs. cloud AI processing
- Privacy & data ownership

Open Questions & Risks

Evaluation:

- Iterative prototyping
- User studies with screen reader users
- Task performance & workload metrics