Reviewer 1 Decision: Weak Accept

Evaluation Criteria:

Relevance & Originality:	(7.0/10)
Technical Quality & Rigor:	(6.5/10)
Clarity & Organization:	(6.5/10)
Literature Review & References:	(8.0/10)
Impact & Practicality:	(7.0/10)
Reviewer's Familiarity:	(6.5/10)

Comments to Author:

The paper presents a cool idea: a smart hat that helps visually impaired users "hear" their surroundings using depth cameras and gestures. The concept works, and early tests look promising, but it needs some areas to be improved -

1. All 20 participants were sighted individuals simulating visual impairment, which limits generalizability.
2. No confidence intervals, variances, or significance testing.
3. No latency measurements, depth estimation accuracy metrics, or failure-case study.
4. Even though filtering the information helps reduce overload, Task 6 shows that users still get tired after about 9.5 minutes, meaning that listening to audio instructions alone is still mentally tiring.
5. System power consumption is not reported - Wearable device usability depends heavily on battery life, thermal performance, and power draw.

Reviewer 2 Decision: Strong Accept

Evaluation Criteria:

Relevance & Originality:	(8.5/10)
Technical Quality & Rigor:	(7.5/10)
Clarity & Organization:	(8.0/10)
Literature Review & References:	(8.0/10)
Impact & Practicality:	(8.5/10)
Reviewer's Familiarity:	(9.0/10)

Comments to Author:

The work demonstrates strong practical relevance, particularly in low- to middle-income contexts. The prototype shows promising task success rates (75–95%) and directly aligns with SDG 10 (Reduced Inequalities). The modular architecture and extensibility (IoT, NodeJS, API integration) increase real-world applicability.

The system design is technically coherent, combining stereo vision (OpenCV), IMU-based gesture control (MPU6050), MediaPipe hand recognition, and audio synthesis. Mathematical formulations for stereo calibration, gesture detection, and audio generation are included and generally sound.
However, rigor could be strengthened by:
- Including statistical significance analysis (like for instance in the confidence intervals or hypothesis testing)
- Justifying sample size (n = 20) more explicitly
- Clarifying baseline comparisons against existing SSDs beyond descriptive discussion
Consider too some minor foxes:
- Standardize figure captions (“Fig.” vs “Figure”)
-Check equation alignment and punctuation for IEEE compliance

Reviewer 3 Decision: Neutral

Evaluation Criteria:

Relevance & Originality:	(7.5/10)
Technical Quality & Rigor:	(7.0/10)
Clarity & Organization:	(7.0/10)
Literature Review & References:	(7.0/10)
Impact & Practicality:	(6.5/10)
Reviewer's Familiarity:	(8.0/10)

Comments to Author:

This paper presents a wearable AIoT-based navigation aid in the form of a head-mounted “World Navigation Hat” designed to support visually impaired users through controllable sensory substitution. The work is well motivated and clearly addresses a critical limitation of existing sensory substitution devices, namely sensory overload and fatigue caused by continuous and unfiltered feedback.

A key strength of the paper is its human-centered design approach, inspired by selective

attention in human perception. The integration of stereo vision, IMU-based head gestures, MediaPipe hand gesture recognition, and modular IoT connectivity is thoughtfully executed. The system architecture and processing pipeline are clearly described, and the proposed mode-based interaction (quiet, general outline, and advanced controls) is intuitive and practically relevant.

The experimental evaluation with twenty participants demonstrates promising results, with task success rates ranging from 75% to 95% across eight navigation and perception tasks. The inclusion of prolonged-use testing and subjective fatigue assessment is particularly valuable and often missing in related work. Results suggest that controllable substitution improves usability and integration compared to continuous-output approaches.

However, the study has some limitations. The evaluation relies on sighted participants simulating visual impairment, which limits direct applicability to real-world visually impaired users. Additionally, while improvements over continuous-output methods are reported, a clearer baseline comparison or quantitative ablation study would strengthen the technical contribution. Further clarification of the audio encoding algorithm and its perceptual implications would also benefit readers.

Overall, the paper offers a meaningful applied contribution with strong social relevance. With expanded validation involving visually impaired participants and deeper comparative analysis, the work could be further strengthened.