

# A TAXONOMY FOR D-SAV360, A DATASET OF GAZE SCANPATHS ON 360° AMBISONIC VIDEOS

Edurne Bernal-Berdun, Maria Plaza, Daniel Martin, Sandra Malpica, Pedro J. Perez, Belen Masia, Ana Serrano, and Diego Gutierrez

Universidad de Zaragoza, I3A



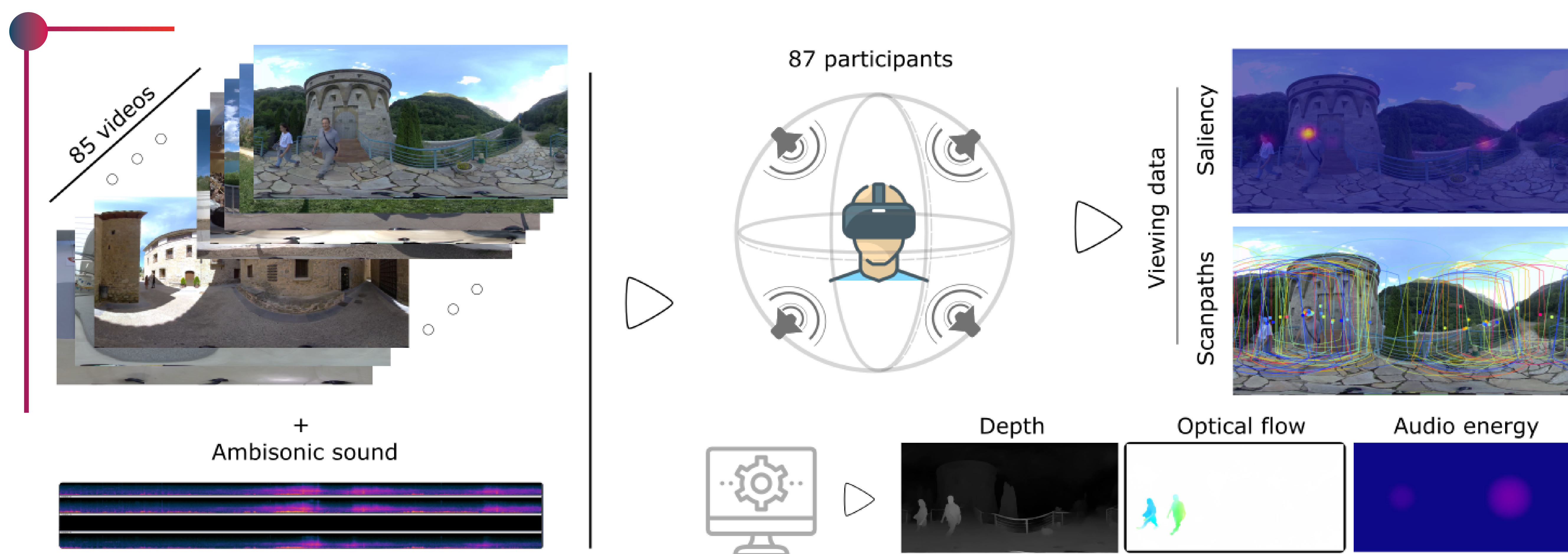
## ABSTRACT

**Understanding** how humans **visually behave** within virtual reality environments is essential to fully leverage its potential in creating engaging experiences. Therefore, it is necessary a **broad dataset** containing a wide range of scenarios, along with corresponding gaze data collected from a diverse and extensive group of observers.

Addressing this, the dataset **D-SAV360** (1) comprises the most extensive collection to date that captures **gaze data in 360° videos with directional audio**. Its breadth of configurations and content variety is summarized within our proposed taxonomy.

## RELATED WORK

The growing interest in 360° videos has generated the creation of multiple datasets in recent years. Despite their usefulness, some datasets, like the one collected by Morgado et al. (2), lacked **variety** due to batch downloading from YouTube. Others were found to be incomplete, not including crucial head and gaze data or directional audio, which has been shown to affect participants' visual behavior and is fundamental for immersion and understanding users' perception. Furthermore, some datasets were limited to a small sample of participants and videos.



## D-SAV360

### Video collection:

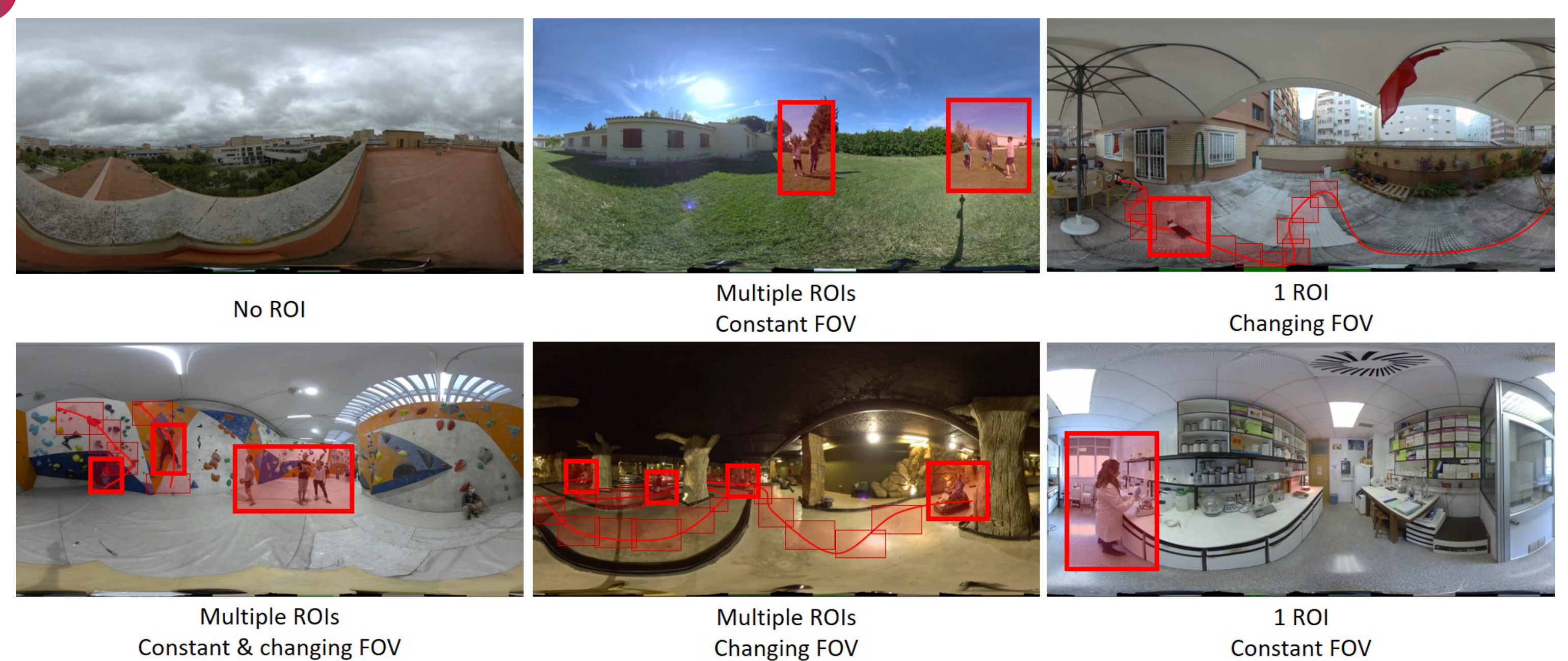
- 50 stereoscopic and monoscopic videos recorded using a Kandao Obsidian S, equipped with six fish-eye lenses.
- 35 monoscopic videos curated from Morgado (2) carefully selected
- Varied content, offering a rich diversity of visual, auditory, and audiovisual stimuli, which span various configurations along several fields of view.

### Gaze data collection:

To collect eye-tracking data, for each of the 85 videos, the SRanipal Unity SDK 4, developed for the Tobii eye-tracker integrated into the HTC Vive Pro Eye, was used.

The **collected tracking data includes:**

- 4,609 head and gaze trajectories
- Head position and orientation
- Pupil diameter
- Eye openness
- Eye gaze vector
- Gazed image coordinates
- Audio energy maps, depth, and optical flow estimations



Our taxonomy encompasses the dimensions of Content, Acoustic, and Spatial Regions of Interest (ROIs).

## CONCLUSIONS

Besides providing a rich source of information for future research, an analysis of the collected scanpaths reveals that various factors such as viewing mode, content type, and gender significantly impact eye movement statistics.

Our dataset jointly with the proposed taxonomy may be used as a benchmark and valuable resource for training novel audiovisual attention prediction models, as well as for studying and modeling human visual behavior in immersive environments.

## Acknowledgments:

This work has received funding from Grant PID2022-141539NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU; by the European Union's Horizon 2020 research and innovation program through the PRIME project (MSCA-ITN, grant agreement No. 956585); and by the Government of Aragon's Departamento de Ciencia, Universidad y Sociedad del Conocimiento through the Reference Research Group "Graphics and Imaging Lab (ref T34 23R). Additionally, Sandra Malpica, Daniel Martin, and Edurne Bernal-Berdun were supported by a Gobierno de Aragon predoctoral grant (2018-2022, 2020-2024, and 2021-2025, respectively).

## References:

1. Bernal-Berdun E, Martin D, Malpica S, Perez PJ, Gutierrez D, Masia B, et al. D-sav360: A dataset of gaze scanpaths on 360° ambisonic videos. TVCG (2023) 1-11.
2. Morgado P, Li Y, Vasconcelos N. Learning representations from audio-visual spatial alignment. Advances in Neural Information Processing Systems (2020).



Scan me

