

Jacob Chalk

Number: +44 (0) 7899346230 | Email: jacob.chalk@bristol.ac.uk | LinkedIn: <https://www.linkedin.com/in/jacob-chalk/> | Website: <https://jacobchalk.github.io/>

Research Focus

My research focus is on audio-visual video understanding, particularly in egocentric videos. Here, my aim is to leverage both audio and video modalities to their maximum potential for solving challenging problems such as action recognition and detection. More recently, I have also begun exploring video understanding in 3D.

Education

PhD in Computer Vision

University of Bristol

September 2021 – Current

Bristol, UK

- Thesis Title: *Audio-Visual Video Understanding*
- Anticipated Graduation Date: September 2025 (09/25)

MEng in Computer Science

University of Bristol

September 2017 – September 2021

Bristol, UK

- Degree Classification: *First Class Honours*
- Dissertation Title: *Video GANS for Human-Object Interactions (83%)*

Teaching

Teaching Assistant

University of Bristol

September 2019 – Current

Bristol, UK

Software Product Engineering (Y2), Games Project (Y3), Computer Graphics (Y3) Image Processing and Computer Vision (Y3), Applied Deep Learning (Y4)

Publications

- Plizzari, C., Goel, S., Perrett, T., **Chalk, J.**, Kanazawa, A. and Damen, D., 2025. *Spatial Cognition from Egocentric Video: Out of Sight, Not Out of Mind*. *International Conference on 3D Vision (3DV)* [\[Project Webpage\]](#) | [\[Paper\]](#) | [\[Code\]](#) (Pending Release)
- **Chalk, J.***, Huh, J.*, Kazakos, E., Zisserman, A. and Damen, D., 2024. *TIM: A Time Interval Machine for Audio-Visual Action Recognition*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18153-18163)* [\[Project Webpage\]](#) | [\[Paper\]](#) | [\[Code\]](#)
- Huh, J.*, **Chalk, J.***, Kazakos, E., Damen, D. and Zisserman, A., 2023, June. *Epic-sounds: A large-scale dataset of actions that sound*. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5)*. IEEE. [\[Project Webpage\]](#) | [\[Paper\]](#) | [\[Code\]](#)

*: Equal Contribution

Datasets

EPIC-Sounds

Audio-Centric Annotated Dataset

Released January 2023

Paper Published in ICASSP 2023

EPIC-Sounds is an audio-only dataset, gathered across 100 hours of untrimmed audios from the videos of EPIC-KITCHENS. This dataset was motivated by the observation that audio-visual networks tend to use both temporally and semantically identical annotations for both modalities, yet audio and video can significantly differ even when describing the same action.

Roles:

- Post-processed the labelled segments from annotators, this included:
 - Manually correcting typos in the free-form text descriptions.
 - Removing erroneous/redundant annotations.
 - Categorising the free-form descriptions into the 44 classes in the dataset.
 - Decomposing challenging segments containing multiple overlapping sounds into distinct annotations.

- Training binary classifiers on the head classes of the dataset to clean the training set.
- Trained and ran quantitative analysis on the baseline models.
- Visualisation of the class distribution for the dataset.
- Ran analyses on the gathered data, such as the duration distribution of all classes and interplay between the visual and audio labels.
- Set and currently manage two challenges on the dataset: Audio-Based Interaction Recognition and Audio-Based Interaction Detection

Honours and Awards

EgoVis 2022/23 Distinguished Paper Awards

June 2024

First Joint Egocentric Vision Workshop

IEEE/CVF Conference (CVPR)

Awarded to our 2023 ICASSP paper: "EPIC-Sounds: A large-scale dataset of actions that sound."

EPIC-KITCHENS Challenges Winner

June 2024

First Joint Egocentric Vision Workshop

IEEE/CVF Conference (CVPR)

Audio-Based Interaction Recognition (2nd), Audio-Based Interaction Detection (2nd), Action Detection (3rd)

Top 5 Third Year MEng Computer Science/Computer Science with Maths Student

October 2020

Netcraft

University of Bristol

Research Activities

Presentations

- EPIC-KITCHENS Challenges. Presented at the First Joint Egocentric Vision Workshop (CVPR 2024).
- Oral Presentation for EPIC-Sounds. Presented at ICASSP 2023.

Reviewing

- Reviewer for ECCV 2024 and TPAMI (2024)

Technical Knowledge and Programming

Programming Languages

- Python (Strongest)
- C++ (Intermediate)
- JavaScript (Intermediate)

Frameworks

- PyTorch (Strongest)
- TensorFlow (Intermediate)

Technical Fields

- Machine Learning, particularly Deep Learning
- Computer Vision
- Data Science