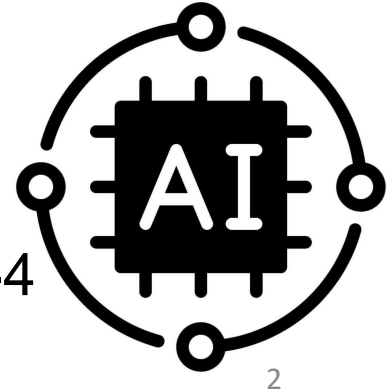
MANCHESTER 1824
The University of Manchester

# Eventsets are all you need

**Luca Peres**
luca.peres-2@manchester.ac.uk

# Rationale

- Amount of compute has been growing exponentially since 2012
- Traditional models and algorithms require more resources
- GPT-3 training is an example
  - Thousands of petaflops
  - 35 tonnes of $CO_2$
  - Cost $12M
  - 50x energy and 100x compute increase for GPT-4

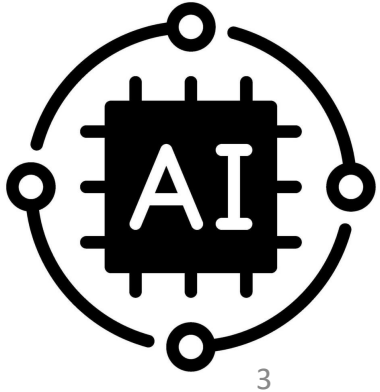





https://openai.com/blog/ai-and-compute/

# Rationale

- von Neumann bottleneck limits performance
- Neuromorphic can be the solution
  - Mimick the brain
  - High connectivity
  - Sparsity
  - Low energy
- Aim to produce targeted training sources
  - Still rely on traditional datasets



MANCHESTER 1824
The University of Manchester

# Events

- Asynchronous occurences
- Signal something has happened/changed
- Resemble action potentials
- Well suited for Neuromorphic hardware

MANCHESTER 1824
The University of Manchester

# Data vs Events

- Regular sampling + quantization of analog variables
- Dense
- Fine-grained temporal correlations are lost
- Higher abstraction from the physical phenomenon
- Large in size

- Threshold crossing in the analog variable
- Sparse
- Fine-grained temporal correlations are preserved
- Closer to hardware and the physical phenomenon
- Space-efficient

MANCHESTER 1824
The University of Manchester

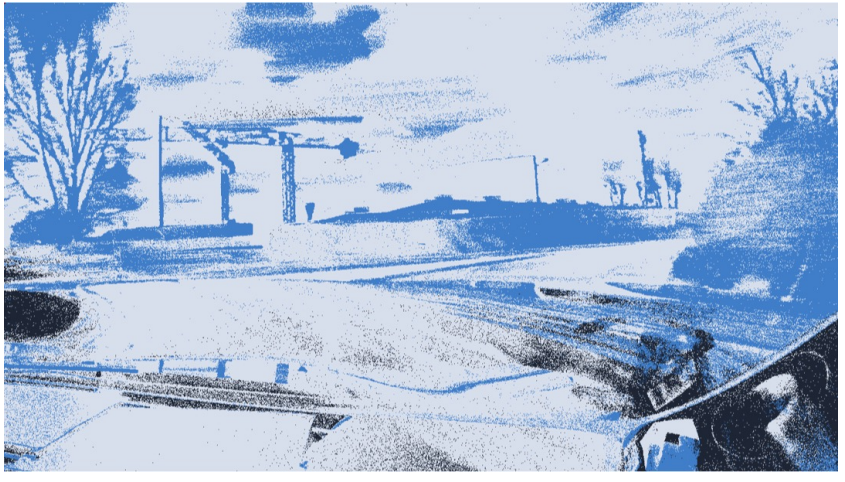
# Eventsets

- Collections of events
  - ID
  - polarity
  - timestamp
- Allow controlled level-based oversampling of the environment for better sensing
- Digital but not numeric

MANCHESTER 1824
The University of Manchester

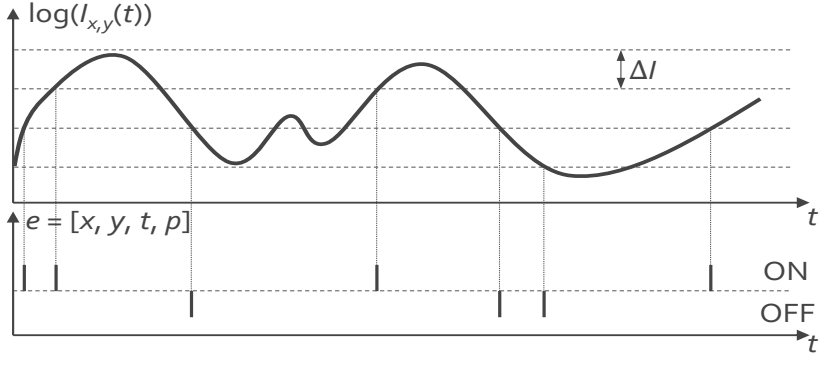
# Event Cameras

- Vision sensors inspired by the retina – event-based, sparse, low-latency sensing
- Detect changes of individual pixel intensity rather than sampling full frames
- High dynamic range (> 120 dB)
- High temporal resolution ($\approx 1\ \mu s$)
- Low power consumption (mW range)



# Event Cameras

- Visual Events
  - ID (x,y)
  - polarity
  - timestamp
- Events generated upon interval crossing



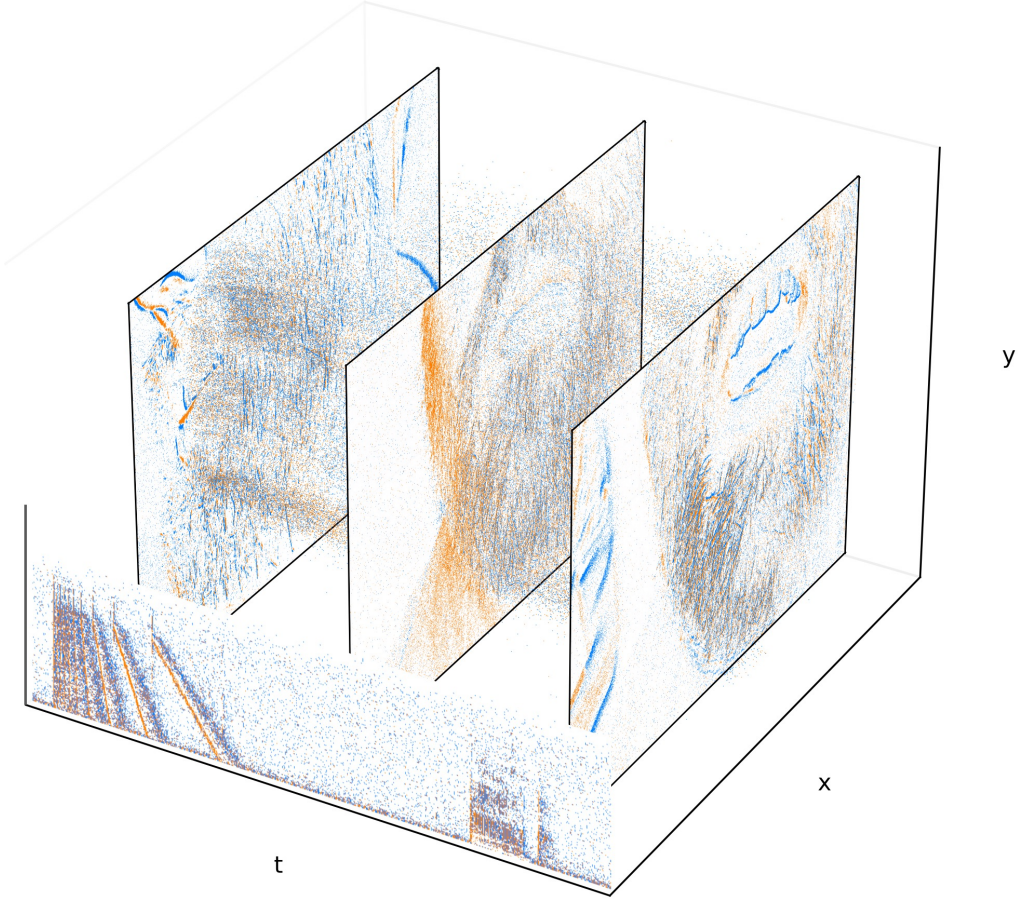
MANCHESTER 1824
The University of Manchester

# Audio Event Generation

- Audio events
  - frequency channel ID
  - polarity
  - timestamp
- Detect changes in the intensity of spectrogram
- Analogous to event cameras
  - 1D event camera viewing the output of an STFT changing over time

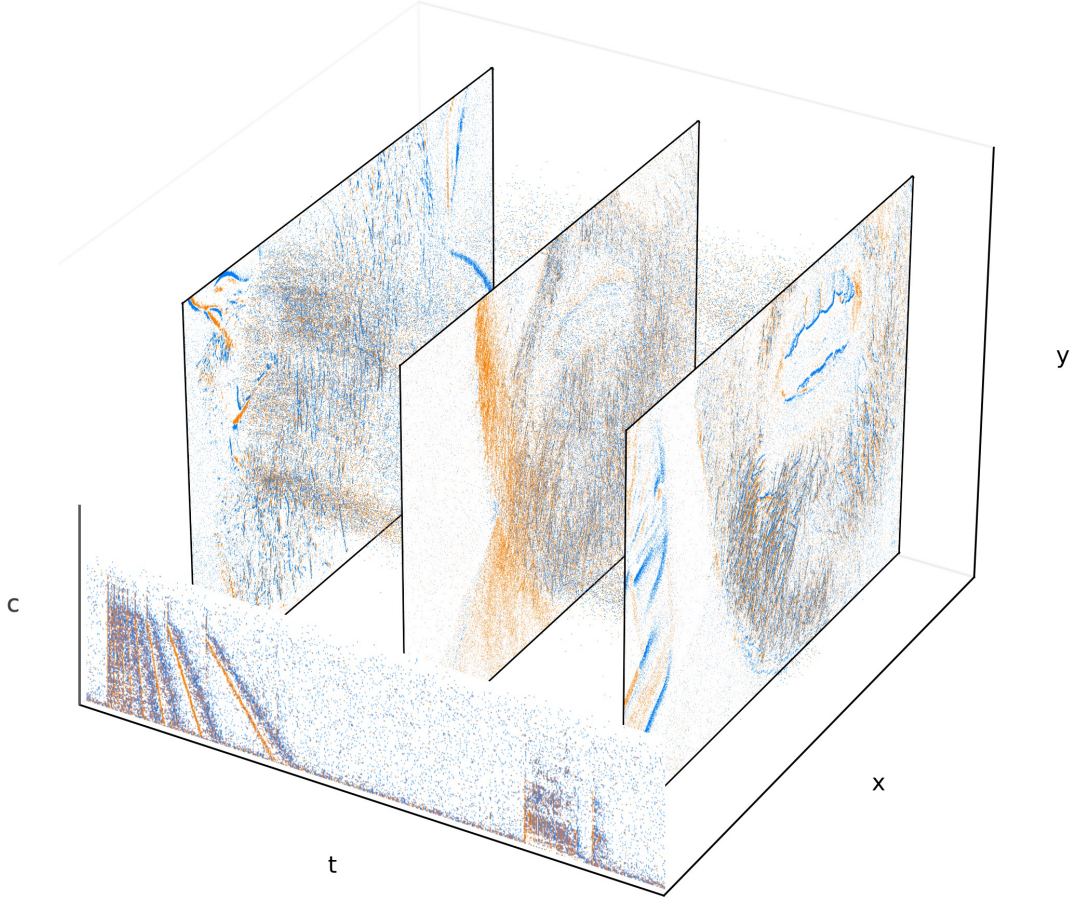
# Neuromorphic Eventset

- 150 subjects
  - Balanced gender split
  - 10 different nationalities
  - Released in phases
- MOCHA-TIMIT dataset
  - 460 sentences
  - Phonetically balanced
- Approximately 1.5 hours per subject



# Neuromorphic Eventset

- Recorded in a recording studio
  - Sentences recorded independently
- Only lip movements are recorded
  - Comply with ethics regulations



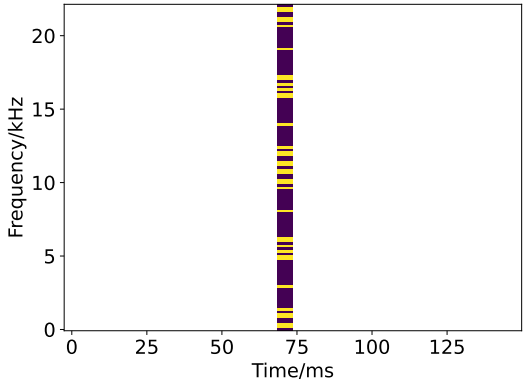
MANCHESTER 1824
The University of Manchester

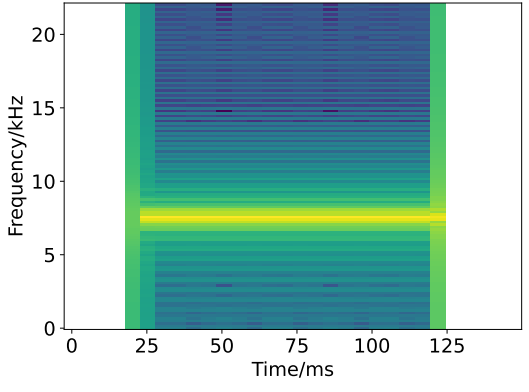
# Recording Protocol

- High resolution Prophesee event camera
  - 1280x720
- Stereo microphone
- Modalities synchronized to a common stimulus
  - Chirp sound
  - Scalable to multiple sensors
- Software synchronization allows for more scalability

# Chirp Comparison

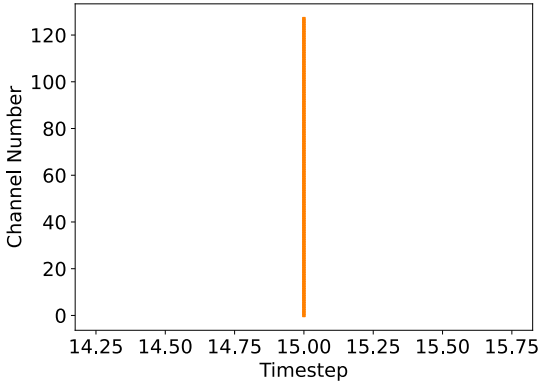
- Clearly identifiable sounds
  - Click sound
  - 7.5 kHz sine wave

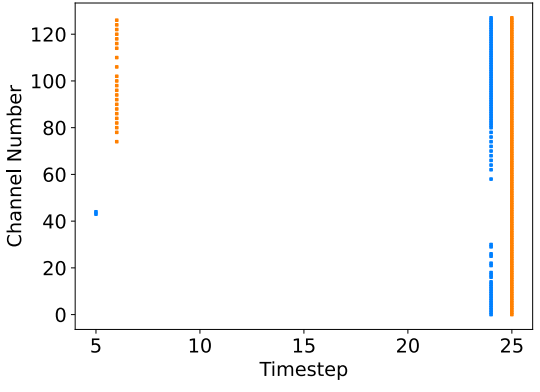


# Chirp Comparison

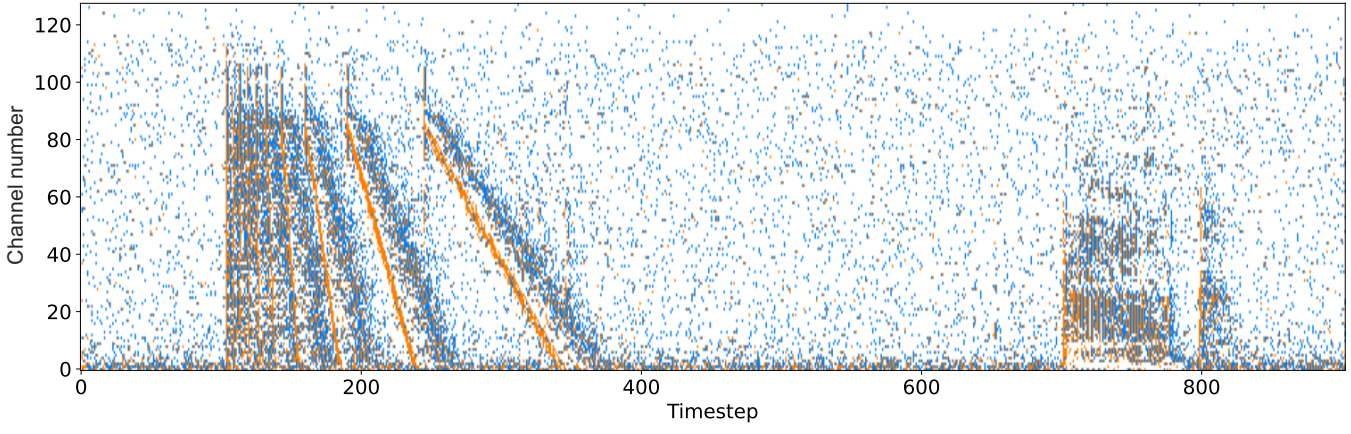
- Clearly identifiable sounds
  - Click sound
  - 7.5 kHz sine wave
- Low event count
- Not distinctive



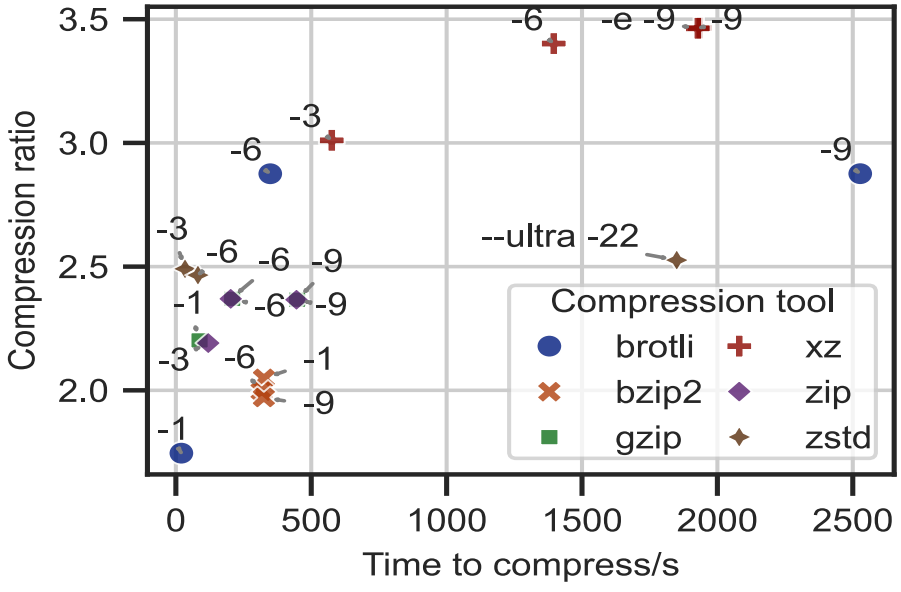
# Chirp Comparison

- Multichirp
  - Varying frequency
  - Little self-similarity
  - Good for moving sources



# Event Compression

- Large-scale eventsets are information-dense
  - 2.2 TB Uncompressed
  - 23 hours recording
- Events carry fine-grained temporal information
- Hard to make these accessible



MANCHESTER 1824
The University of Manchester

# Summary

- Framework for a phonetically rich multimodal eventset
- Various domains can benefit
  - Cross-modal compensation
  - Noisy applications
- Step towards reducing data bottleneck