Utilising Large Quantities of Found Audio Data

Per Fallgren

perfall@kth.se

Overall aim

Produce methods able to browse, and possibly annotate, large quantities of unexplored audio data of varied quality.

Found data

- Found data is data that was not recorded with the purpose of being used in research.
- National archives contain unmanageable quantities of audio, often millions of hours of unexplored data.
- Manual exploration and annotation of data is too laborious - adequate automatic methods do not exist.

Motivation

- High value of cultural worth in archive data.
- Found data can be more valuable compared to data collected in controlled settings as they provide for a higher ecological validity.
- Large sets of data without labels is of limited use.
- Our method may facilitate the work of researchers in other fields, which falls in line with Clarin's goals.

Proposed method

Human-in-the-loop framework based upon an approach we have named Temporally Disassembled Audio (TDA).

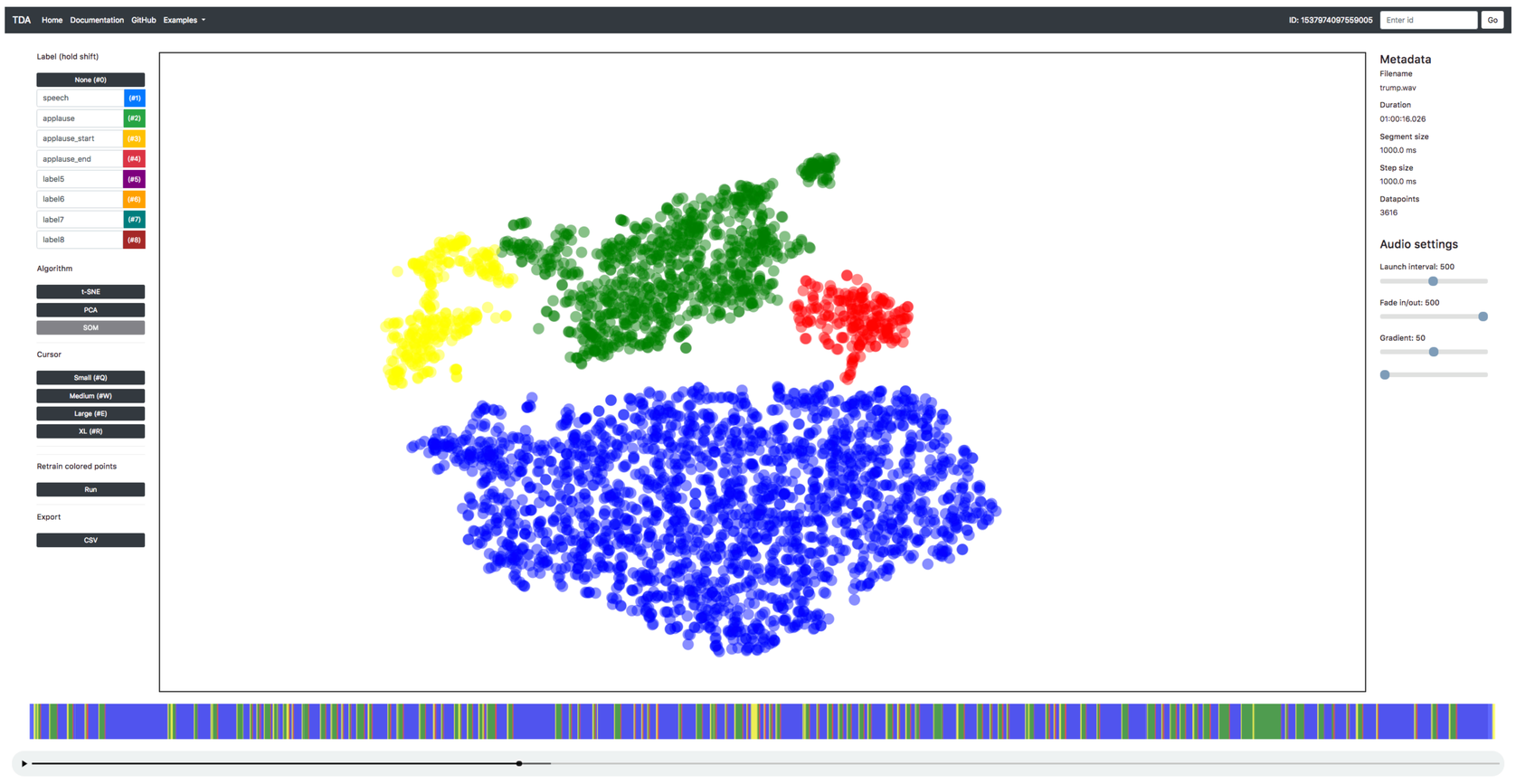
Results

The approach has a broad application area and has already demonstrated usefulness for many different types of audio, including long-term and short-term sound events.

- Male and female speech.
- Applause and cheering.
- Noisy and silent segments.
- Singing and instruments.
- Vowels, fricatives and discourse markers.

Future work

Investigate feature extraction and dimensionality reduction methods, and conduct further experiments with the TDA framework.



TDA interface

Process of TDA

1. Segment audio in chunks ranging from 25 to 1000 milliseconds.
2. Extract a feature vector for each chunk, e.g. MFCC or spectrogram.
3. Project vectors onto 2D space (t-SNE, PCA, SOM).
4. Let user listen, explore and annotate the processed data in the TDA framework.
5. Revert to original audio timeline with the new information.

Acknowledgements
The project is funded in full by the Riksbankens Jubileumsfond funded project TillTal (SAF16-0917: 1). Its results will be made more widely accessible through the national infrastructure Nationella språkbanken and Swe-Clarin (Swedish research Council 2017-00626).

KTH ROYAL INSTITUTE OF TECHNOLOGY

Per Fallgren
PhD Student
KTH, Royal Institute of Technology
Department of Speech, Music and Hearing

Lindstedtsvägen 24
114 28 Stockholm
SWEDEN

Tel: +46709140807
Email: perfall@kth.se