ACOUSTIC SCENE CLASSIFICATION USING SPATIAL FEATURES

Marc Ciufo Green and Damian Murphy

Audio Lab

Department of Electronic Engineering

University of York

OUTLINE

- 4th-order Ambisonic acoustic scene recordings
- Spectral and spatial analysis
- Machine listening system
- Spatial features outperform spectral

[1]

PREVIOUS WORK

Soundman OKM II Klassik Electret
http://www.soundman.de/en

- Use of mono/stereo recordings
- Easier to record (maybe)
- Inheritance of work from other fields
- Applications in smart devices / robotics
- Limited to spectral / binaural spatial features
- Can work well (83.3 % accuracy in latest DCASE)

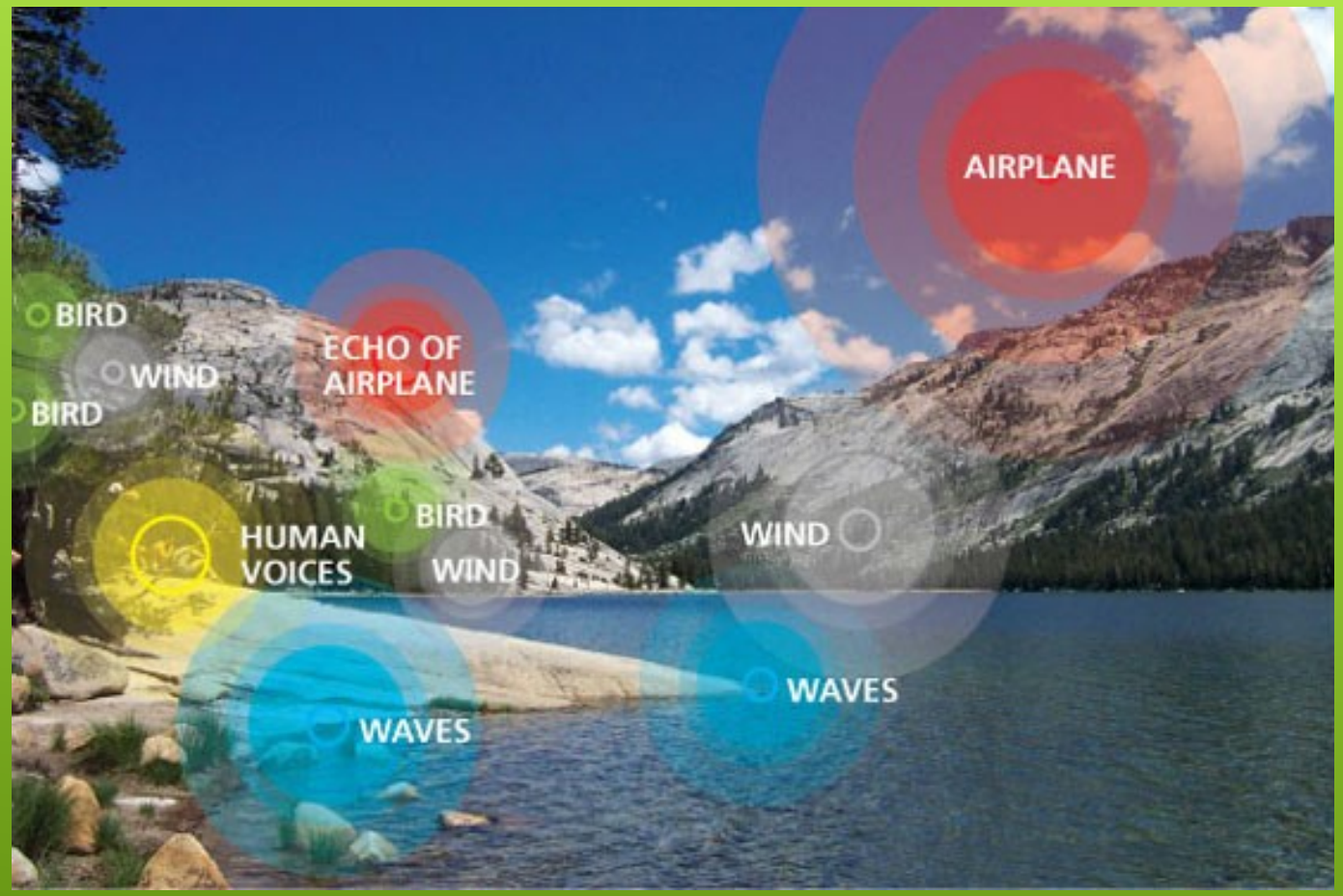
Aircraft takeoff: 110 dB(A)

http://www.flydayton.com/

Rock concert: 110 dB(A)

http://ajr.org/2014/11/13/music-critics-role-changing/

ALTERNATIVE APPLICATION – ENVIRONMENTAL SOUND

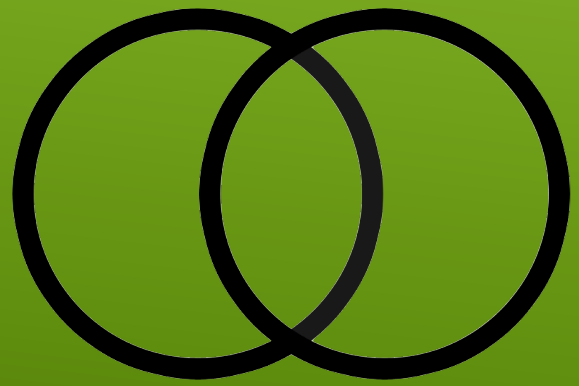


https://www.nps.gov/yose/learn/nature/soundscape.htm

- Goal: event detection
- Scene classification as first step

- Lots of examples of each scene needed
- TUT database – stereo
- DEMAND database – not enough examples
- **New database** required

DATABASE



https://en.wikipedia.org/w/index.php?curid=41429748

[4, 6]

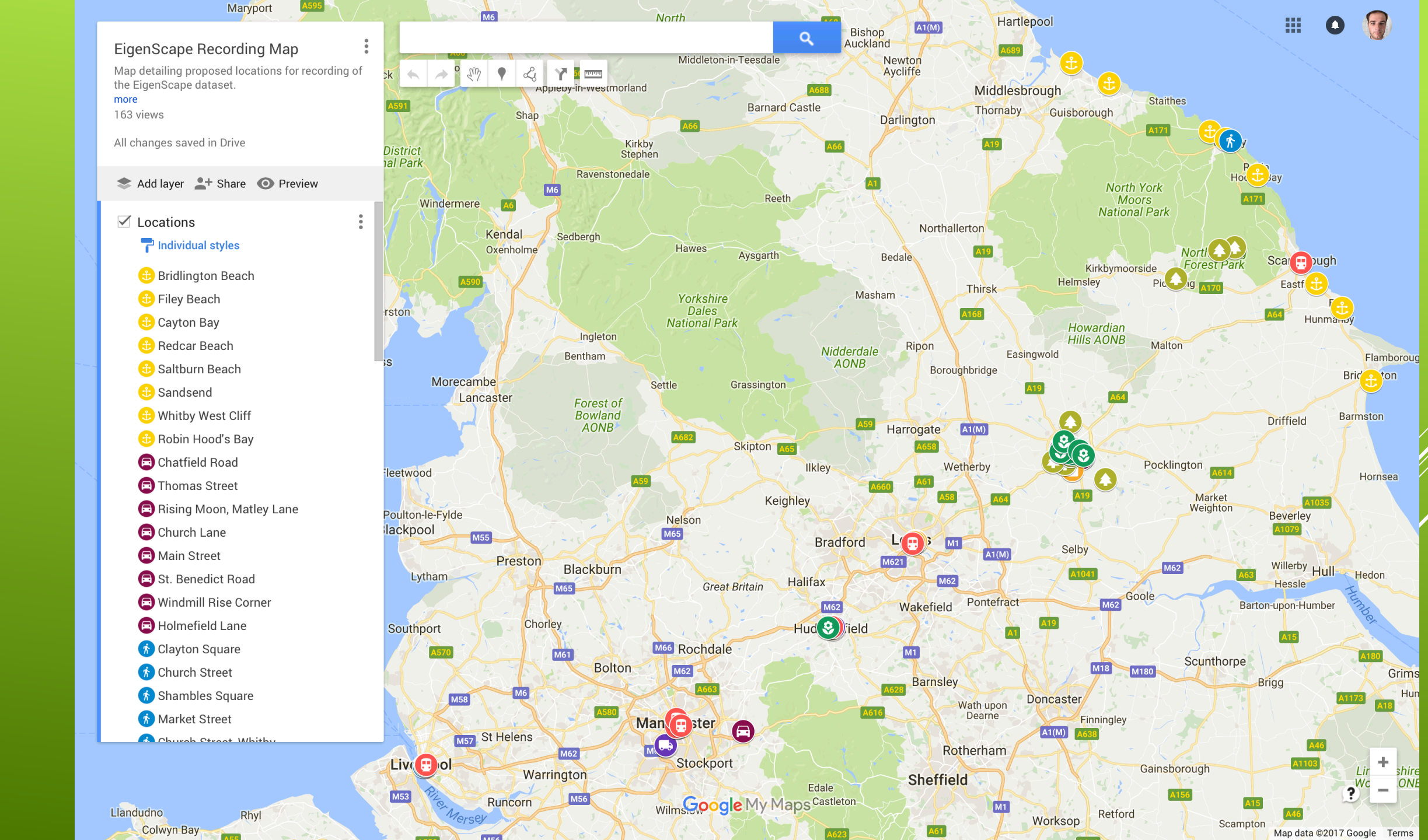
EIGENSCAPE DATABASE

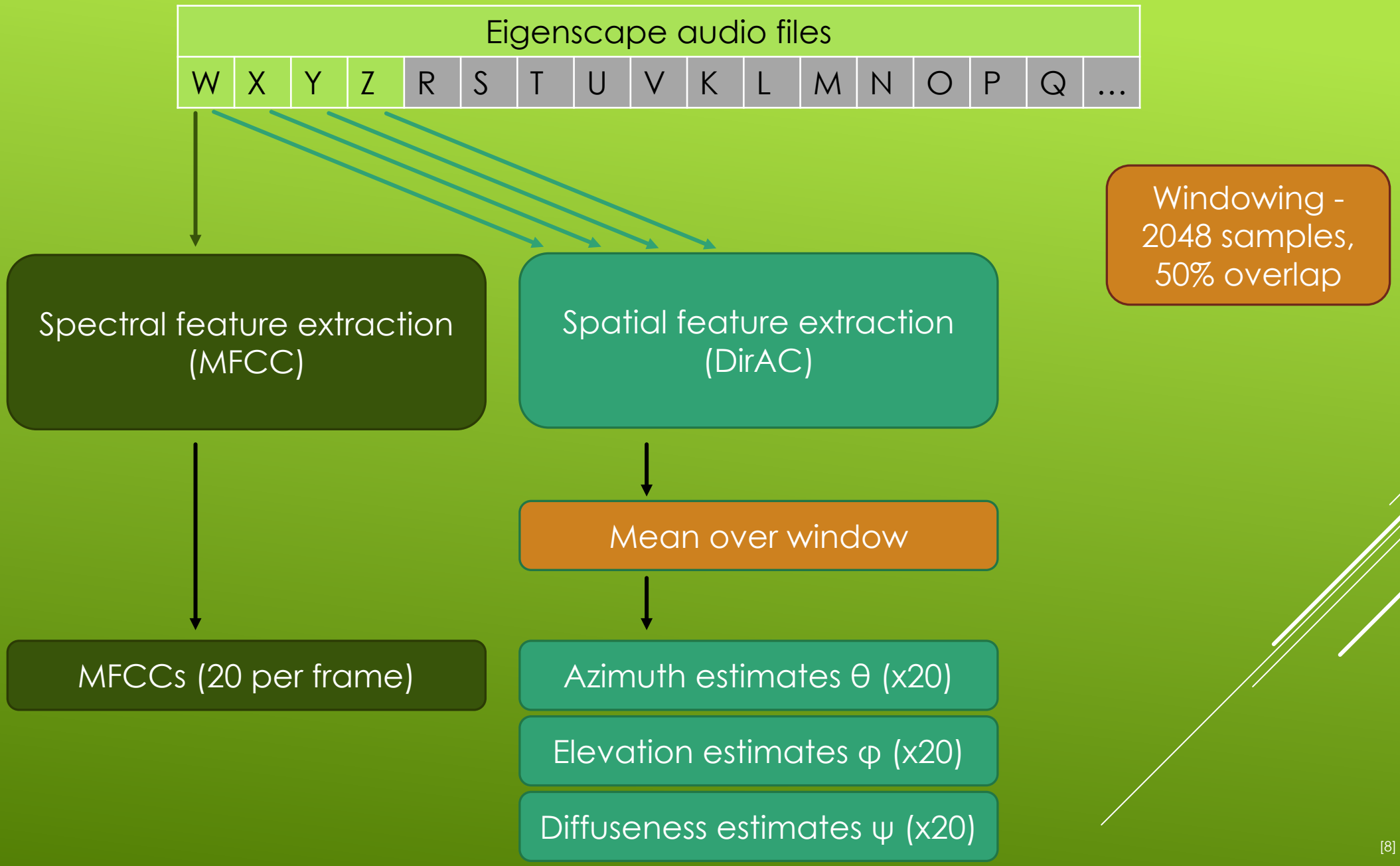
- mh Acoustics Eigenmike
- 4th-order Ambisonics – high spatial resolution
- 8 examples of 8 different scene classes
- 10 minutes per clip
- 24-bit / 48 kHz
- Available now!

DOI 10.5281/zenodo.1012809

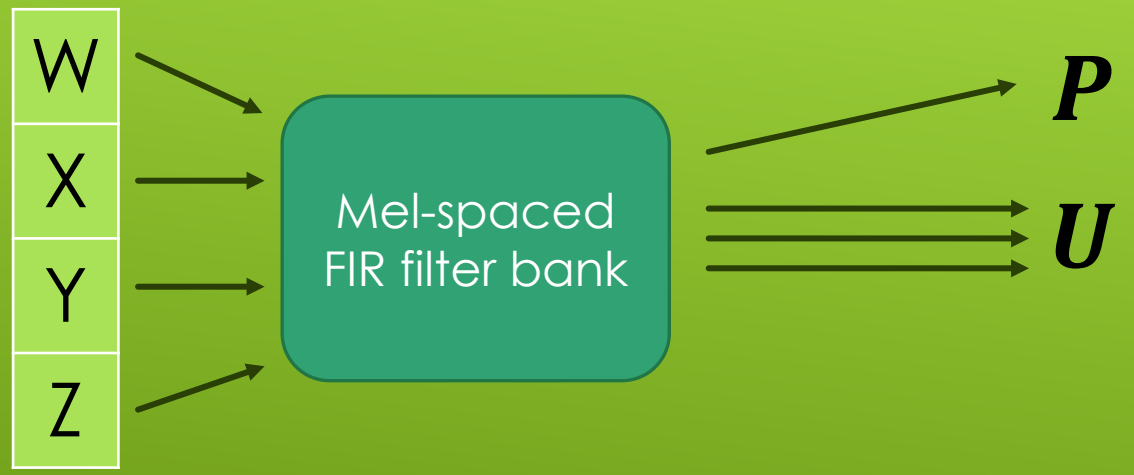
Beach
Busy Street
Park
Pedestrian Zone
Quiet Street
Shopping Centre
Train Station
Woodland


Inverness
Aberdeen
SCOTLAND
Dundee
Edinburgh
Glasgow
United Kingdom
NORTHERN IRELAND
Isle of Man
Dublin
Galway
Ireland
Limerick
Cork
ENGLAND
Cambridge
WALES
Oxford
Bristol
London
Cardiff
Brighton
Southampton
Plymouth
English Channel


EigenScape Recording Map
Map detailing proposed locations for recording of the EigenScape dataset.
more
163 views
All changes saved in Drive
Add layer
Share
Preview
Locations
Individual styles
Bridlington Beach
Filey Beach
Cayton Bay
Redcar Beach
Saltburn Beach
Sandsend
Whitby West Cliff
Robin Hood's Bay
Chatfield Road
Thomas Street
Rising Moon, Matley Lane
Church Lane
Main Street
St. Benedict Road
Windmill Rise Corner
Holmefield Lane
Clayton Square
Church Street
Shambles Square
Market Street
Google My Maps
Map data ©2017 Google
Terms


Eigenscape audio files
W
X
Y
Z
R
S
T
U
V
K
L
M
N
O
P
Q
...
Spectral feature extraction (MFCC)
Spatial feature extraction (DirAC)
Windowing - 2048 samples, 50% overlap
Mean over window
MFCCs (20 per frame)
Azimuth estimates θ (x20)
Elevation estimates φ (x20)
Diffuseness estimates ψ (x20)

DIRECTIONAL AUDIO CODING (DIRAC)



$$\boldsymbol{D} = -\boldsymbol{P}\boldsymbol{U}$$

Azimuth

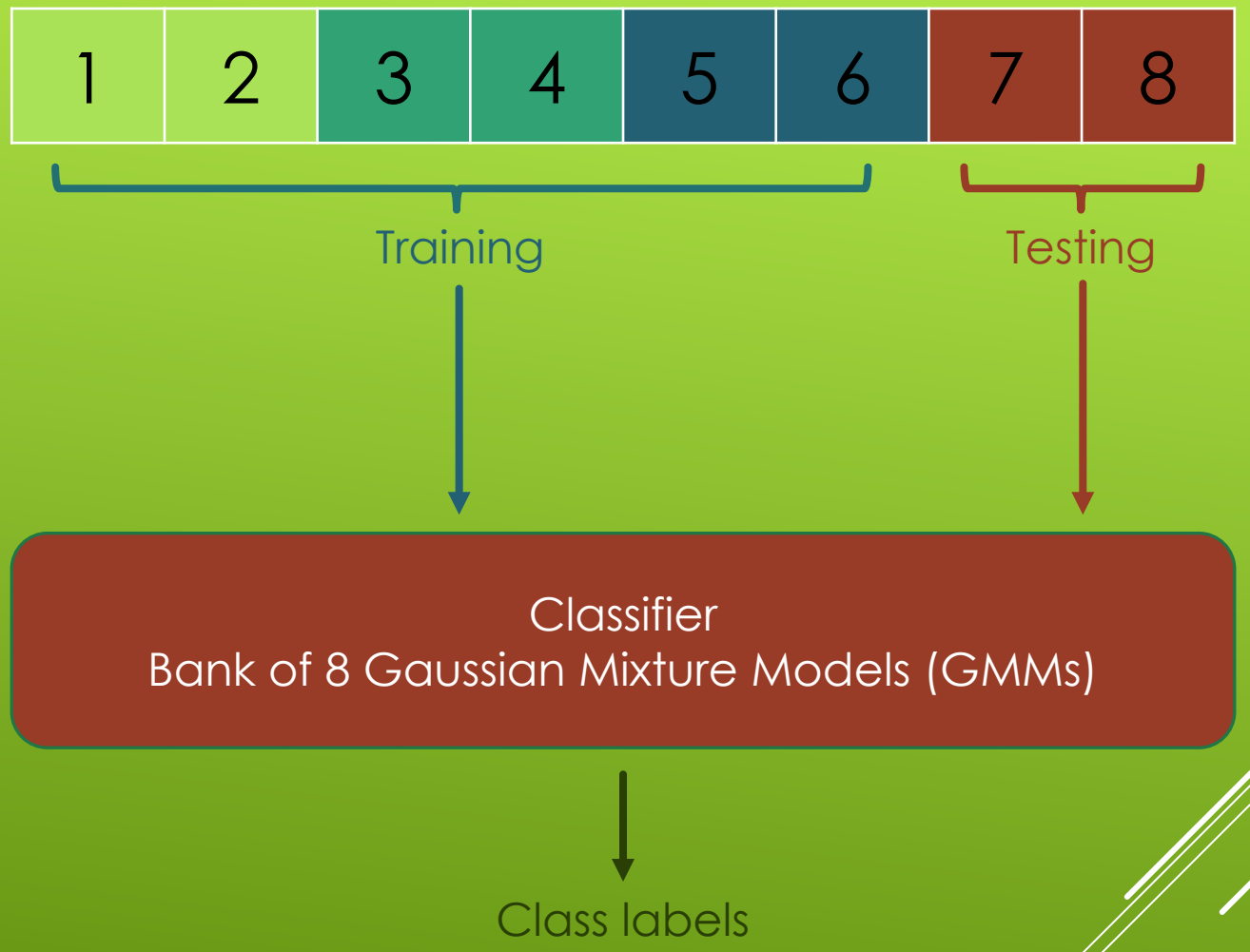
Elevation

$$\boldsymbol{\varphi} = 1 - \frac{\|-\boldsymbol{D}\|}{\boldsymbol{c}\{\boldsymbol{E}\}}$$

$$\boldsymbol{E} = \frac{1}{2} p_0 \left(\frac{\boldsymbol{P}^2}{Z_0^2} + \|\boldsymbol{U}\|^2 \right)$$

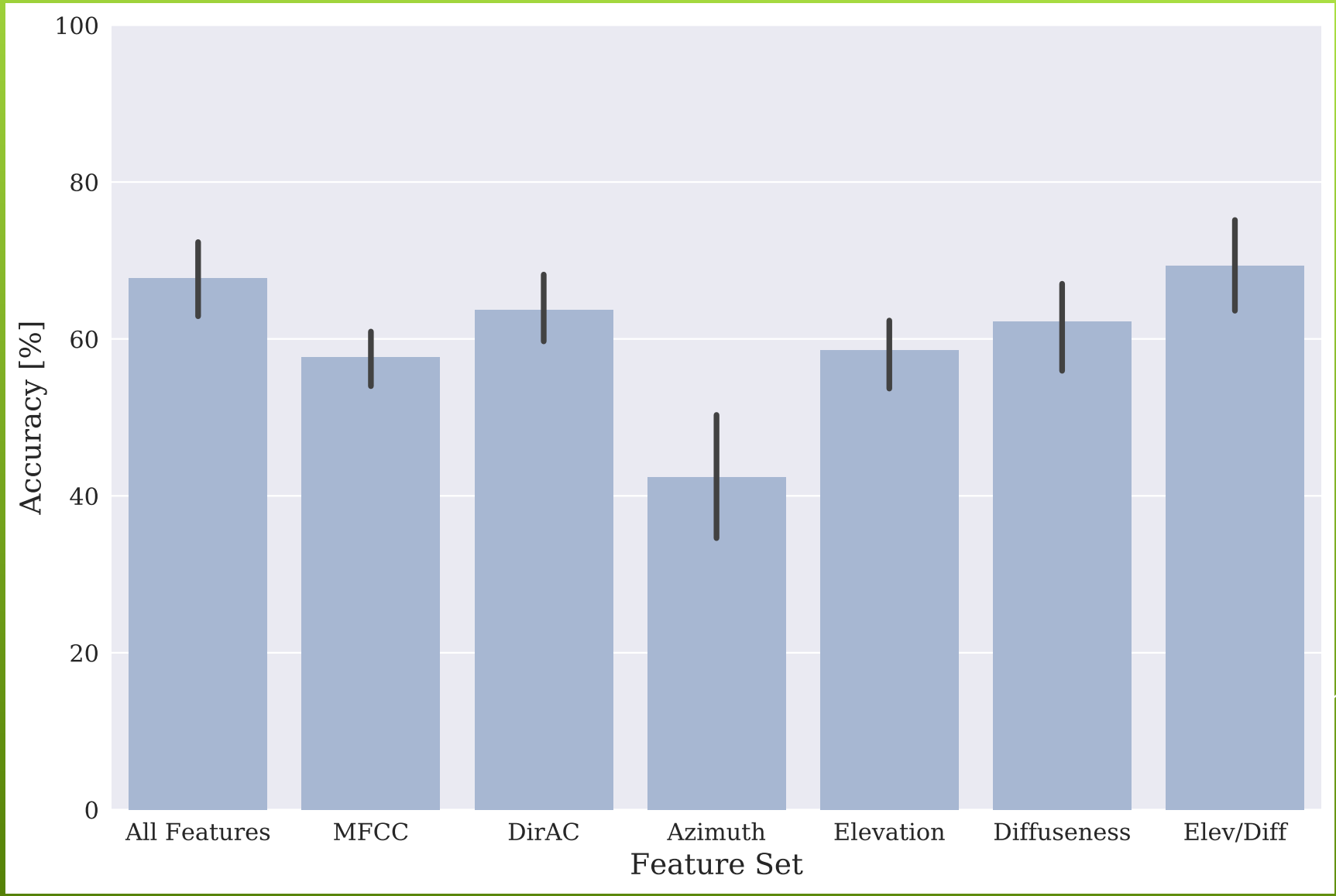
Diffuseness

4-fold cross-validation e.g.

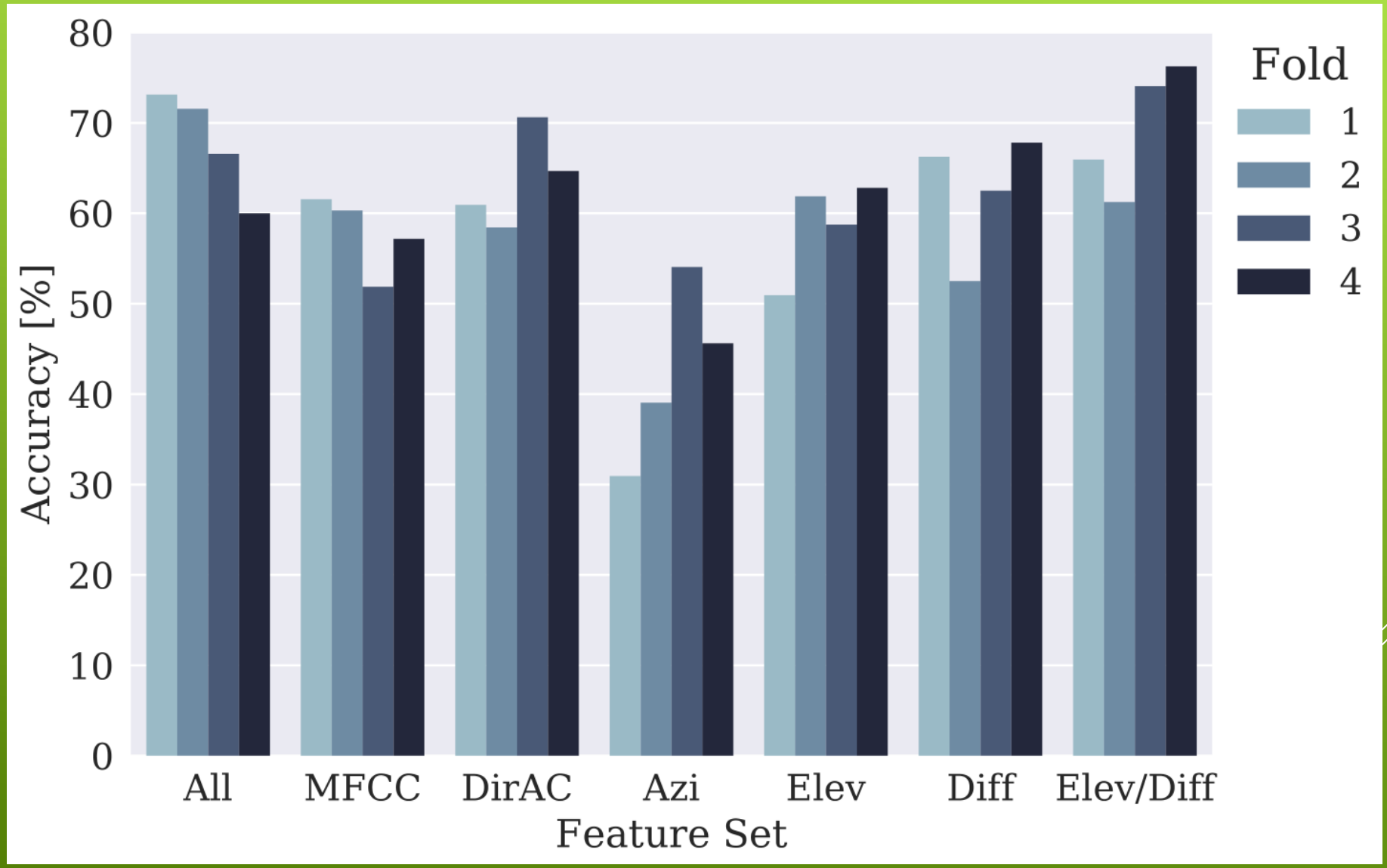


- Probabilities summed across 30 seconds of segments
- Highest probability returned determines label selected
- Classifiers trained and tested across all four folds with results aggregated

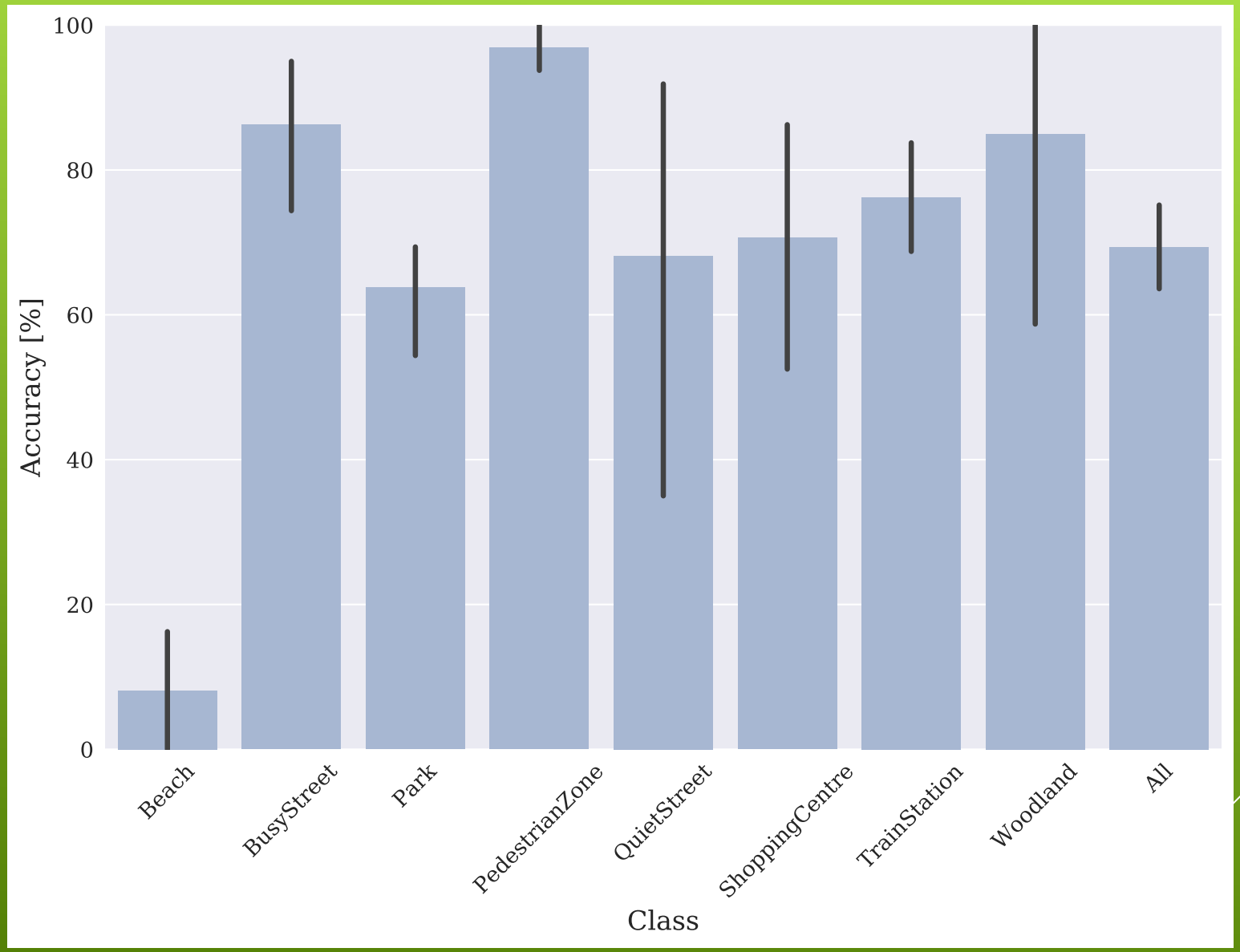
Mean classification accuracies using MFCC and DirAC features



Classification accuracies using MFCC and DirAC features

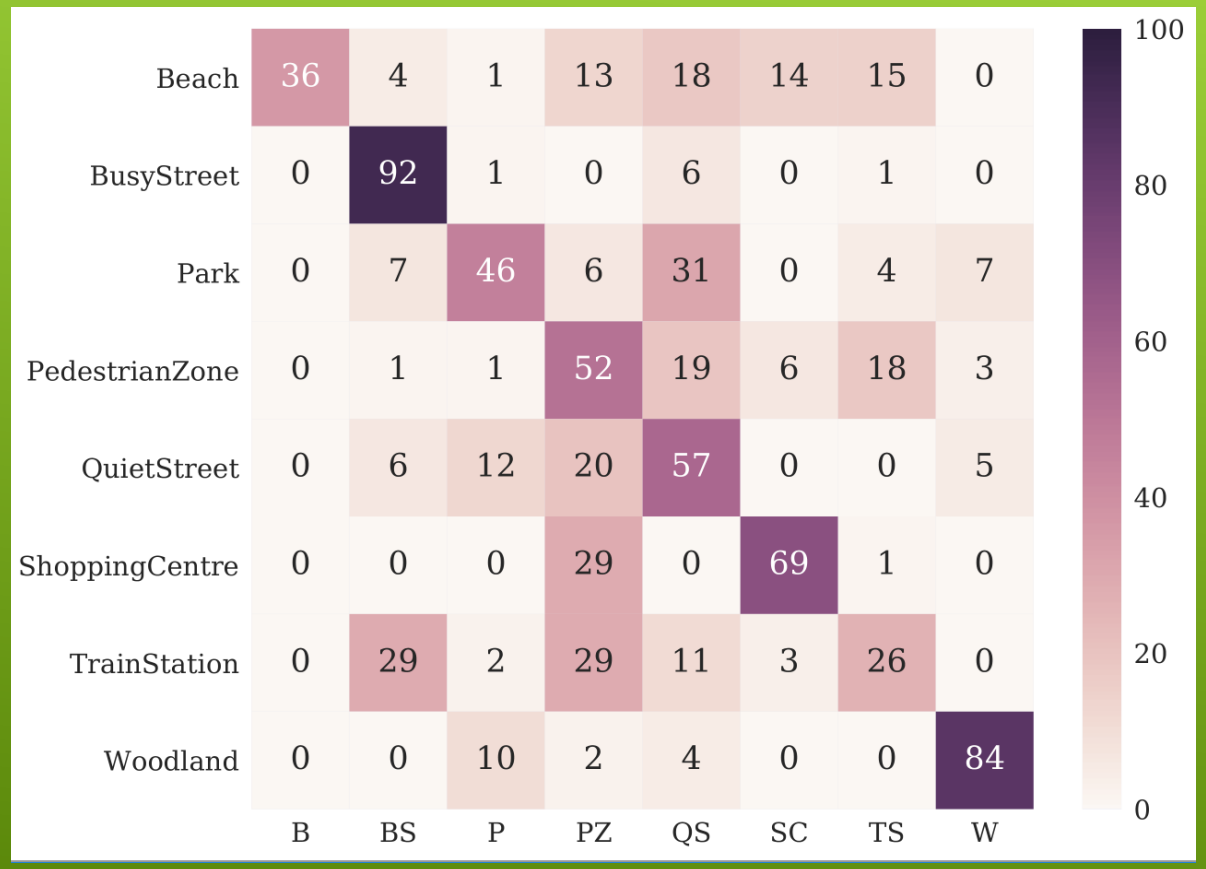


Per-class accuracies using Elevation/Diffuseness features

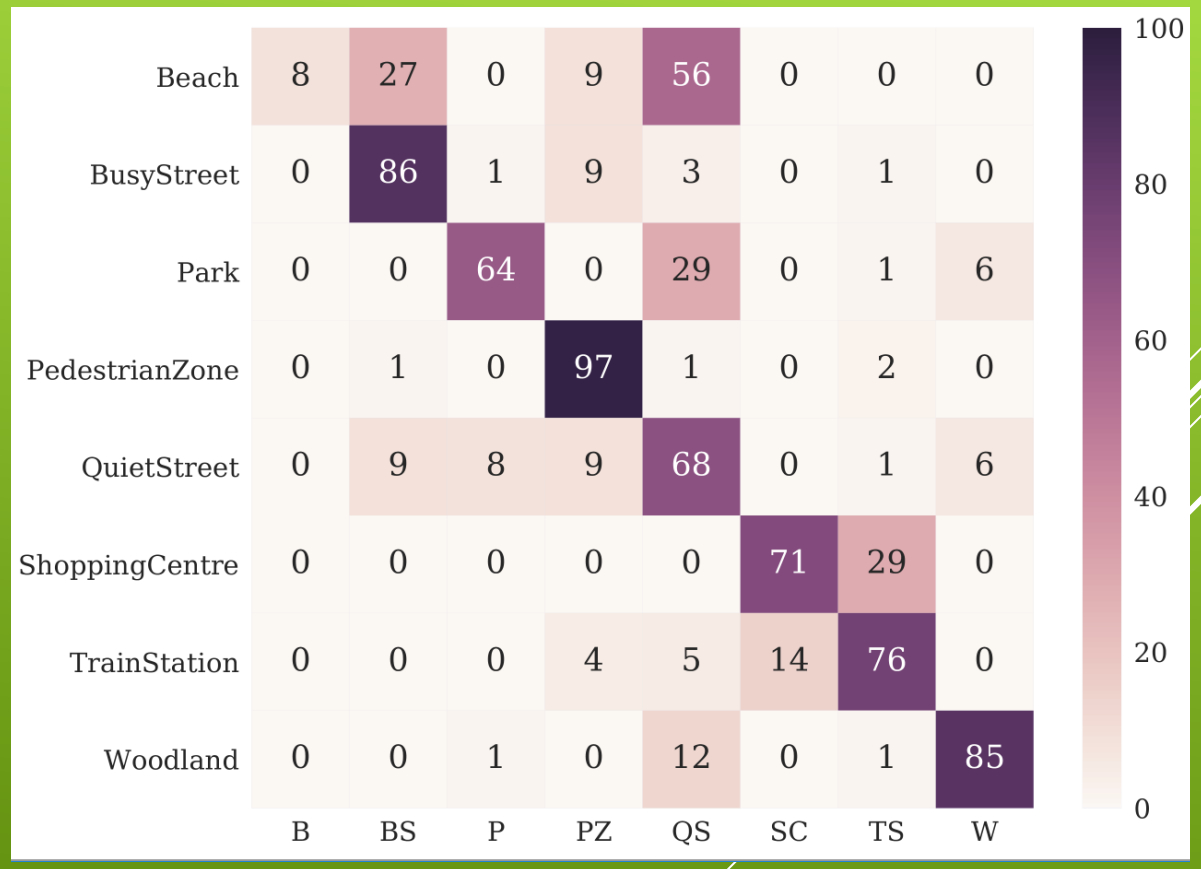


Confusion Matrices

MFCC features



Elevation/Diffuseness features



CONCLUSIONS

- Accurate classification with spatial features
- E/D features outperform MFCC
- Important initial result – spatial features valuable
- *Spectral* similarity and *spatial* similarity not the same
- Good, not perfect accuracy validates EigenScape

- **Next steps:**
  - Use of higher-order channels
  - Event detection

REFERENCES

[1] mh Acoustics, *em32 Eigenmike microphone array release notes*, mh acoustics, 25 Summit Ave, Summit, NJ 07901, April 2013. [Online]. Available: https://mhacoustics.com/sites/default/files/EigenmikeReleaseNotesV15.pdf

[2] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1733–1746, October 2015.

[3] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, "Acoustic scene classification: Classifying environments from the sounds they produce," *IEEE Signal Processing Magazine*, May 2015.

[4] J.-J. Aucouturier, B. Defreville, and F. Pachet, "The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music," *Journal of the Acoustical Society of America*, vol. 122, no. 2, 2007.

[5] M. Lagrange and G. Lafay, "The bag-of-frames approach: A not so sufficient model for urban soundscapes," *Journal of the Acoustical Society of America*, vol. 128, no. 5, November 2015

[6] A. Mesaros, T. Heittola, and T. Virtanen, "Tut database for acoustic scene classification and sound event detection," in *24th European Signal Processing Conference (EUSIPCO)*, August 2016.

[7] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, June 2007.

[8] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proc. of the 14th Python in Science Conference (SciPy 2015)*, 2015.

[9] A. L. Brown, "Soundscapes and environmental noise management," *Noise Control Engineering Journal*, vol. 58, no. 5, pp. 493 – 500, 2010.

[10] Marc Ciufo Green, & Damian Murphy. (2017). EigenScape [Data set]. Zenodo. http://doi.org/10.5281/zenodo.1012809

LINKS

- EigenScape Database: http://doi.org/10.5281/zenodo.1012809
- Code: https://github.com/marc1701/EigenScape
- Map: http://bit.ly/EigenSMap