Oregon State University

Identifying Sentiment from Crowd Audio

Patrick Donnelly & Alex Prestwich

soundbendor lab

College of Engineering | Computer Science

Motivating Use Case

Oregon State University

Imagine:

- multi-model classroom response system
- estimates student sentiment or *class temperature*

Reaction

- ☐ emoji recognition

Text

- ☐ NLP sentiment analysis

Movement

- ☐ gesture recognition[1-2]

Audio

- ?? *this work explores this task*

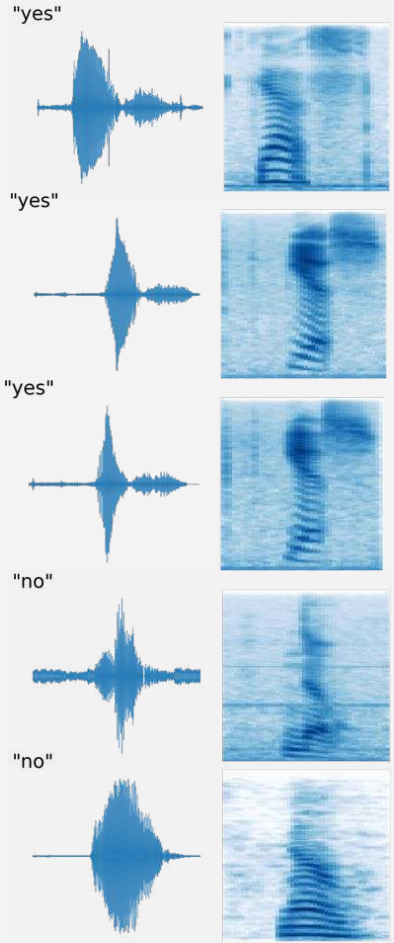
Related Work

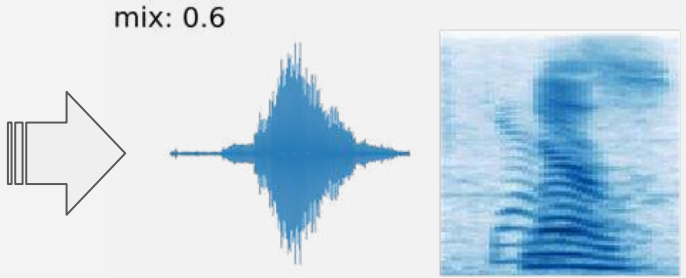
Oregon State University

- Overlapping speech
  - diarization is hard[3]
  - counting speakers[4]
- Emotion from audio
  - mix CNN and RNN[5-6]
  - multi-party calls[7]
- Crowd Sentiment
  - Superbowl winners[8]
  - stock market[9]
  - video game reviews[10]
  - crisis event detection[11-12]

Dataset

- TF Speech Recognition
  - “yes” / “no”
  - n = 4753
- Random mixtures
  - v = 1 to 10 voices
  - 10k per mixture





Experimental Design

Oregon State University

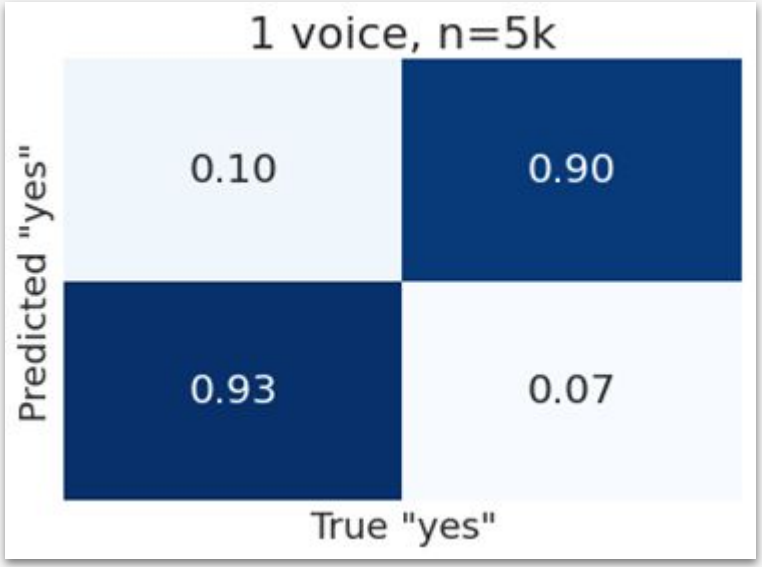
- Bidirectional LSTM
  - input layer: **256**
  - 3 × hidden layer: **128**
  - output layer: **1**

Hidden	MSE	Pearson's
2	.09	.62
3	.08	**.65**
4	.09	.62

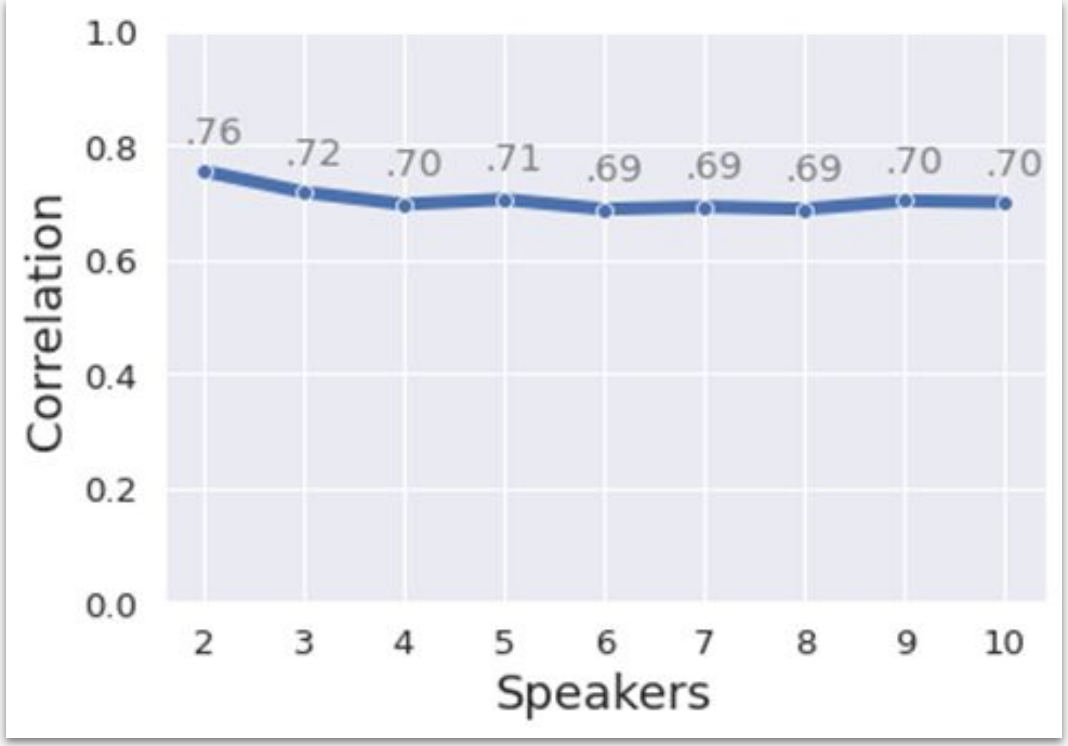
	MSE	Pearson's r
Sequential	.15	.33
Conv1D	.11	.38
RNN	.08	.51
GRU	.10	.54
LSTM	.10	.48
BLSTM	.08	**.70**
CNN+LSTM	.08	.54

Single Voice Baseline

- Experiment
  - “*yes*” or “*no*”
  - train on 10k
  - test on 5k
- Result
  - $r = 0.84$
  - accuracy = 0.91

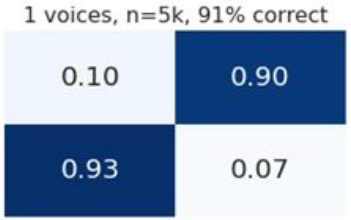


Number of Voices

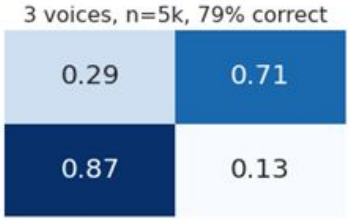


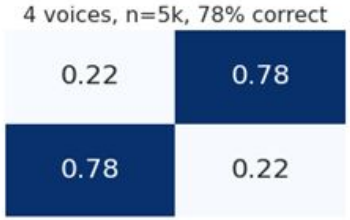
Number of Voices

- Varying from 1 to 10

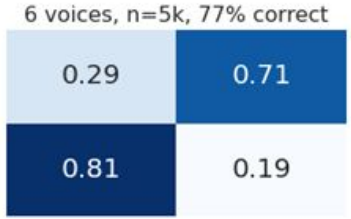


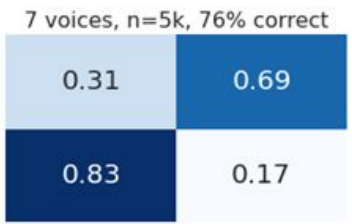


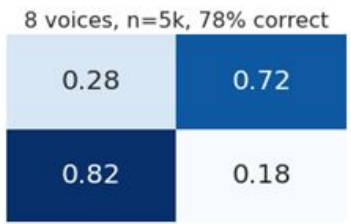




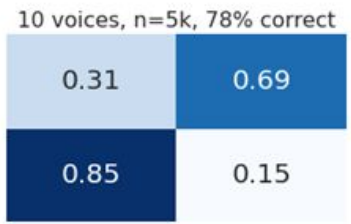




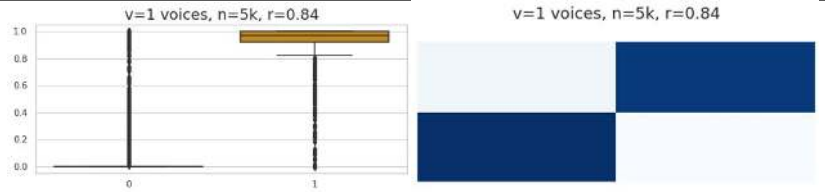




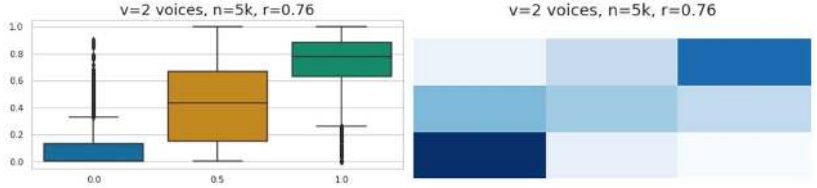




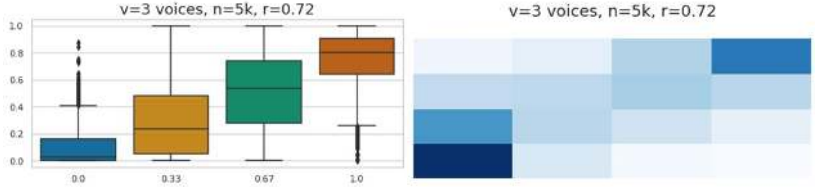
1



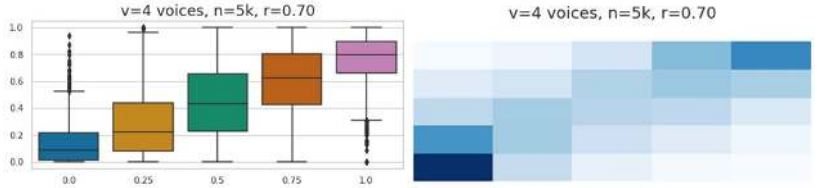
2



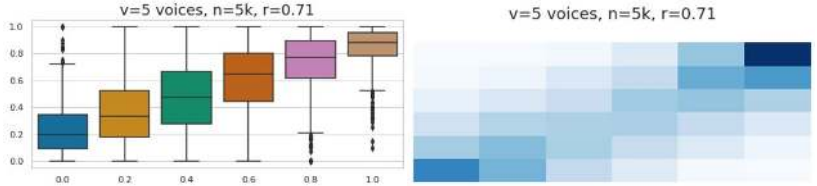
3



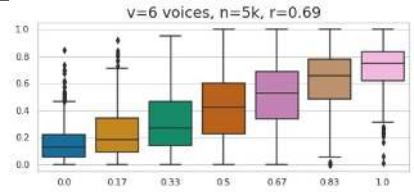
4

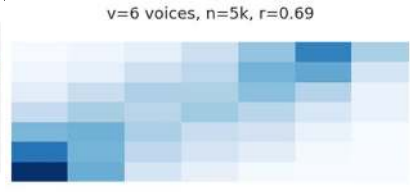


5

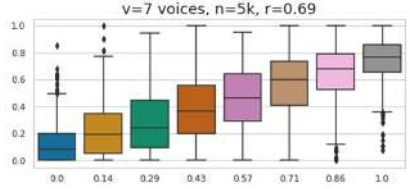


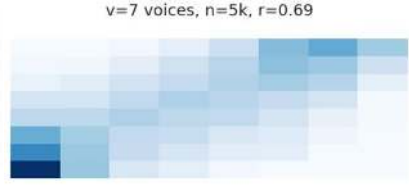
6



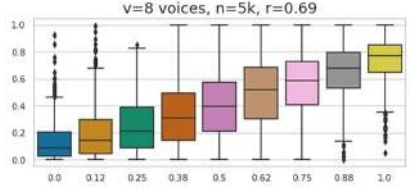


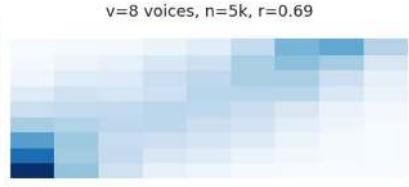
7



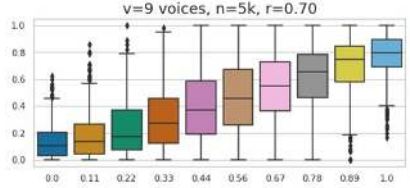


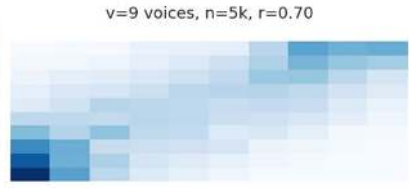
8



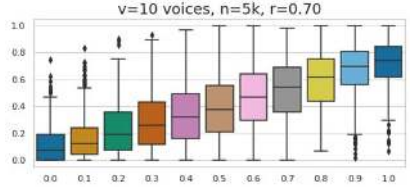


9



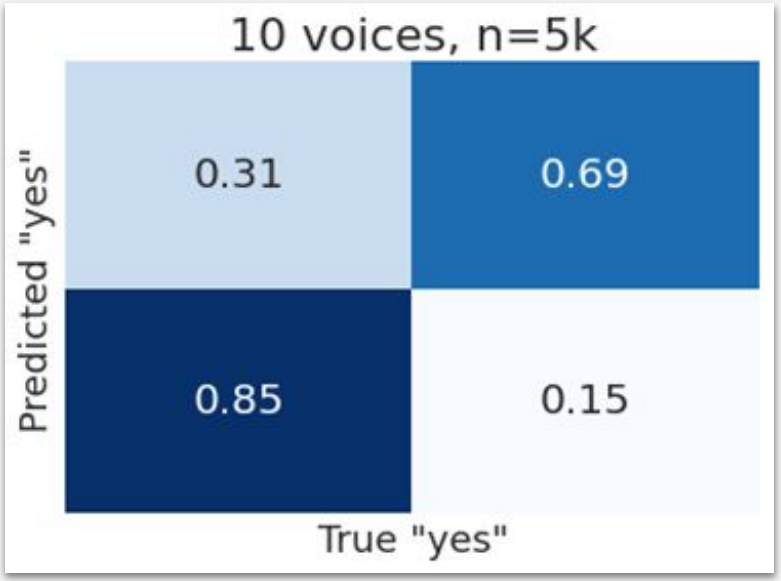


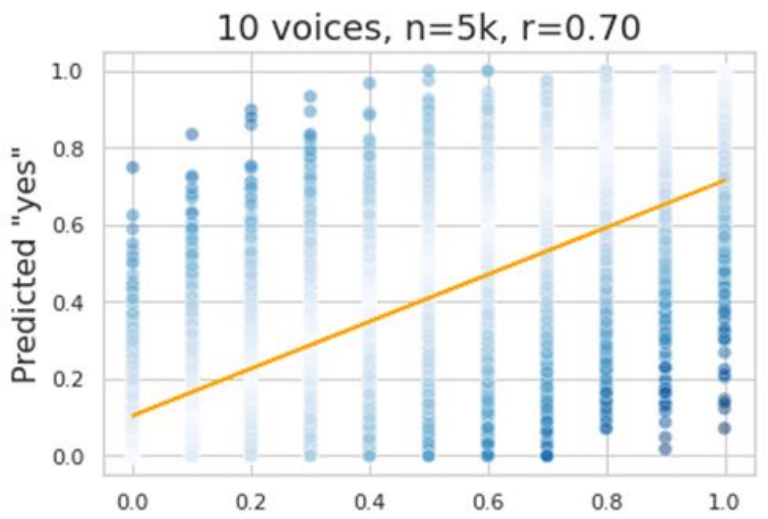
10

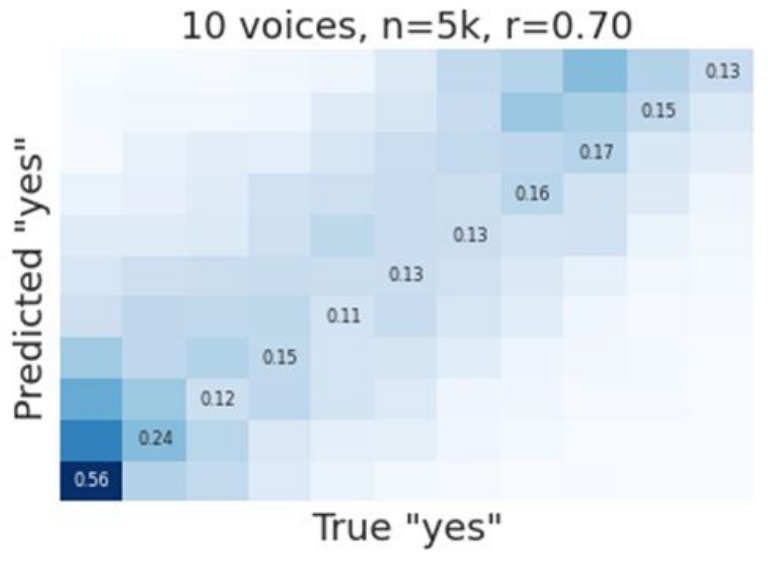




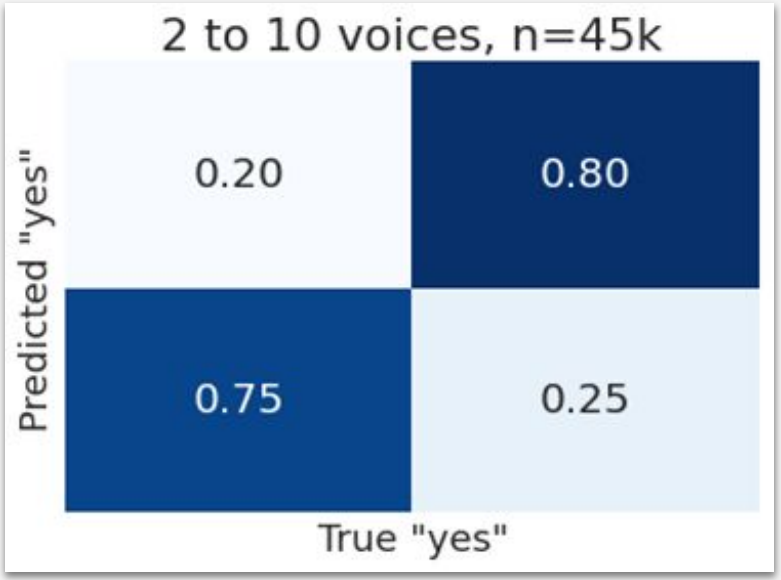
10 Simultaneous Voices

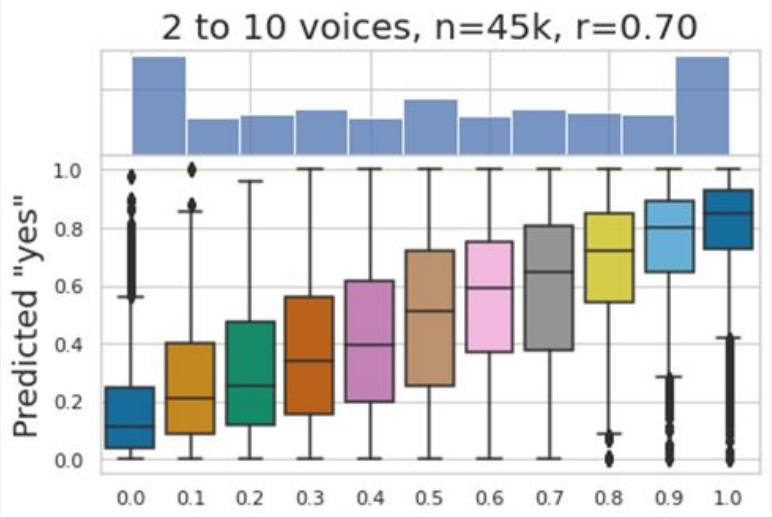


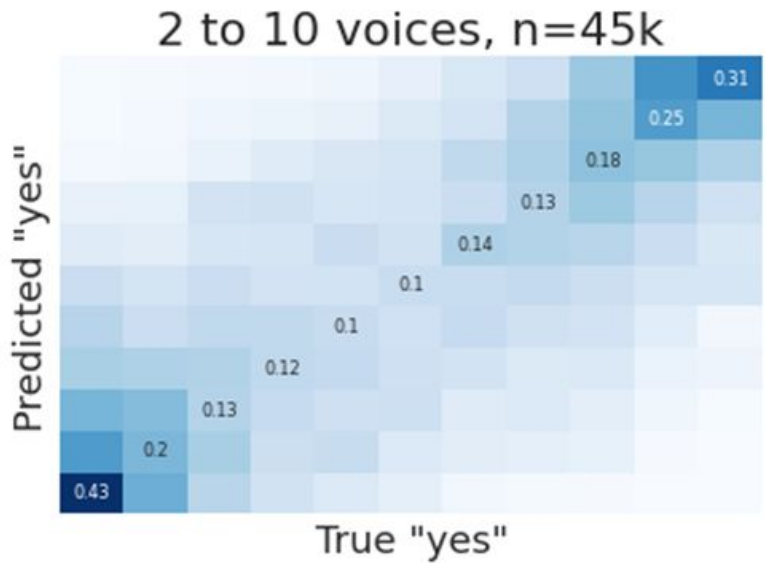




All Mixtures: 2 to 10







Summary

Oregon State University

Contributions

- novel problem
- able to predict majority
  - but limited: yes or no
- raw audio (no preprocessing)
  - potential for real-time

Future Work

- test different words
- add distractors, noise
- collect real-world data
- examine outliers

References

[1] Zhou, H., Jiang, F. and Shen, R. 2018. Who are raising their hands? Hand-raiser seeking based on object detection and pose estimation. Asian Conference on Machine Learning (2018), 470–485.

[2] Jiang, B., Xu, W., Guo, C., Liu, W. and Cheng, W. 2019. A classroom concentration model based on computer vision. Proceedings of the ACM Turing Celebration Conference-China (2019), 1–6.

[3] Park, T.J., Kanda, N., Dimitriadis, D., Han, K.J., Watanabe, S. and Narayanan, S. 2022. A review of speaker diarization: Recent advances with deep learning. Computer Speech & Language. 72, (2022), 101317.

[4] Andrei, V., Cucu, H. and Burileanu, C. 2019. Overlapped Speech Detection and Competing Speaker Counting—Humans Versus Deep Learning. IEEE Journal of Selected Topics in Signal Processing. 13, 4 (2019), 850–862.

[5] He, N. and Ferguson, S. 2020. Multi-view neural networks for raw audio-based music emotion recognition. 2020 IEEE International Symposium on Multimedia (ISM) (2020), 168–172.

[6] Orjesek, R., Jarina, R., Chmulik, M. and Kuba, M. 2019. DNN based music emotion recognition from raw audio signal. 2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA) (2019), 1–4.

[7] Jia, Y. and SungChu, S. 2020. A deep learning system for sentiment analysis of service calls. arXiv preprint arXiv:2004.10320. (2020).

[8] Iyer, R.R., Zheng, R., Li, Y. and Sycara, K. 2019. Event outcome prediction using sentiment analysis and crowd wisdom in microblog feeds. arXiv preprint arXiv:1912.05066. (2019).

[9] Wu, X., Ye, Q., Hong, H. and Li, Y. 2020. Stock selection model based on machine learning with wisdom of experts and crowds. IEEE Intelligent Systems. 35, 2 (2020), 54–64.

[10] Jeffrey, R., Bian, P., Ji, F. and Sweetser, P. 2020. The Wisdom of the Gaming Crowd. Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (2020), 272–276.

[11] Kruspe, A., Kersten, J. and Klan, F. 2021. Detection of informative tweets in crisis events. Natural Hazards and Earth System Sciences (NHESS). (2021).

[12] Hassan, S.Z., Ahmad, K., Al-Fuqaha, A. and Conci, N. 2019. Sentiment analysis from images of natural disasters. International Conference on Image Analysis and Processing (2019), 104–113.