

- with
- tasks. making calls.



## DETECTION OF CALLS FROM SMART SPEAKERS Vinay Maddali, David Looney, Kailash Patil Pindrop, Atlanta, GA, USA Pindrop EMEA, London, UK

$$\frac{T}{\sum_{t=1}^{T} \frac{1}{S(k,t)}}$$

Cell phone usage mode	User count
Normal(to-ear)	70
Speakerphone	55
Earphones	6
Bluetooth Headset	7
Smart speaker distance	User count
<10 cm	42
10 cm-100 cm	77
100 are 200 are	31
100 cm-200 cm	

# Performance was assessed using five-fold cross validation with 80-20 train-test split with 10s samples with no overlap in speakers. Compared with baseline of Call provenance features [3] and MFCC features. Best performance from various classifiers was used. EER: Proposed: 12.6% Call provenance: 17.6% MFCC: 23.8% provenance and 47% over MFCC features. 15 and spectral characteristics.

14th Python in Science Conference, 2015. communications security, 2010, pp. 109–120.



GHmini AEcho

- Performance for cell is best in normal mode as this has least reverberation and noise.
- Performance for smart speakers is best when used from far away, >100 cm.
- Worst performance in all cases was 21%.

Features are not overly biased on mode-of-usage for either devices.

### Conclusions

- Approach to differentiate smart speaker calls from cell phone calls. Proposed audio feature set to detect differences in reverberation, noise
- A dataset was collected through crowdsourcing with participants using both devices in different modes.
- Proposed system differs from previous works as it detects smart speaker calls and uses a dataset that closely resembles real-world call scenarios. Outperforms baseline systems.

## References

- [1] Patrick A. Naylor and Nikolay D. Gaubitch. *Speech Dereverberation*. Springer, 2010. [2] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. *librosa: Audio and music signal analysis in Python*. Proceedings of the
- [3] Vijay A. Balasubramaniyan, Aamir Poonawalla, Mustaque Ahamad, Michael T. Hunter, and Patrick Traynor. *PindrOp: Using single-ended audio features to determine call provenance*. CCS 10 Proceedings of the 17th ACM conference on Computer and