Deep Learning Features for Robust Detection of Acoustic Events in Sleep-Disordered Breathing

Hector E. Romero, Ning Ma, Guy J. Brown, Amy V. Beeston and Madina Hasan
Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK
{heromeroramirez1, n.ma, g.j.brown, a.beeston, m.hasan}@sheffield.ac.uk

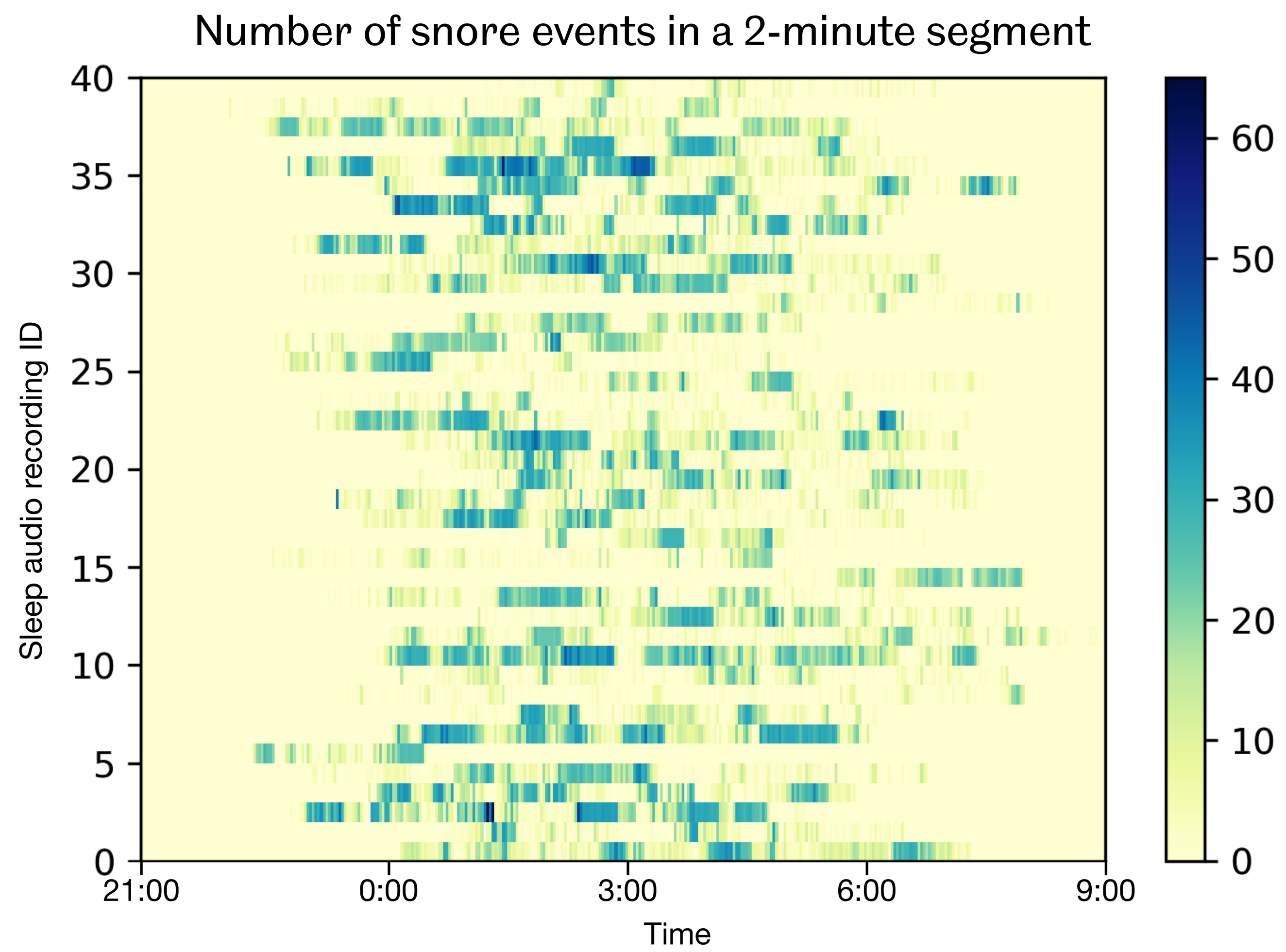
The University Of Sheffield.

1. Introduction

- **Sleep-disordered breathing (SDB)** is caused by the partial or complete collapse of the upper airway during sleep.
- The most prevalent forms of SDB are **snoring**, and obstructive sleep apnoea (OSA).
- The gold standard for diagnosing SDB is the **polysomnography (PSG) test**.
- PSG involves sleeping for a complete night in a laboratory while physiological parameters are measured via wired attachments to the body.
- PSG is expensive, time consuming, and uncomfortable for the patient.
- Alternatives to the diagnosis of SDB have been explored including at-home PSG, and smartphone-based solutions using acoustic analysis.

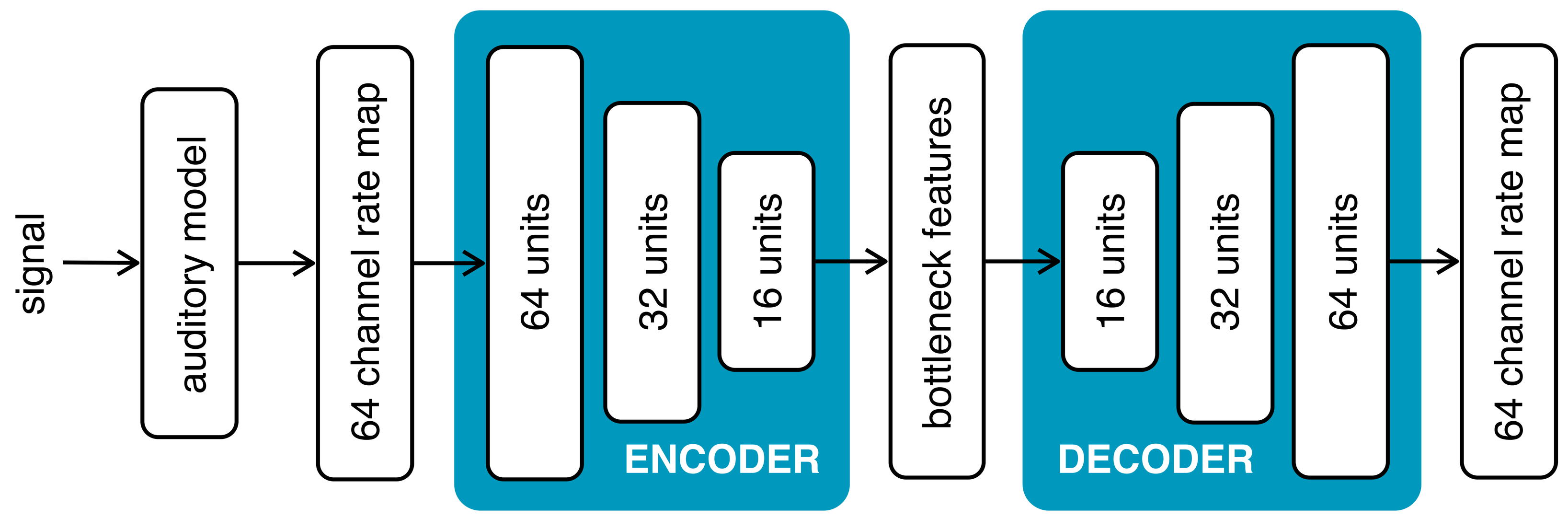
2. Sleep Breathing Sound Corpus

- Acoustic analysis of SDB is a data-scarce field.
- We created a corpus consisting of **6 hours of manually annotated sleep audio recordings** from 6 male participants.
- The recordings were made with a smartphone in the home.
- The annotation scheme considered “snore”, “breath”, “noisy in-breath”, “wheezing”, and “other”.

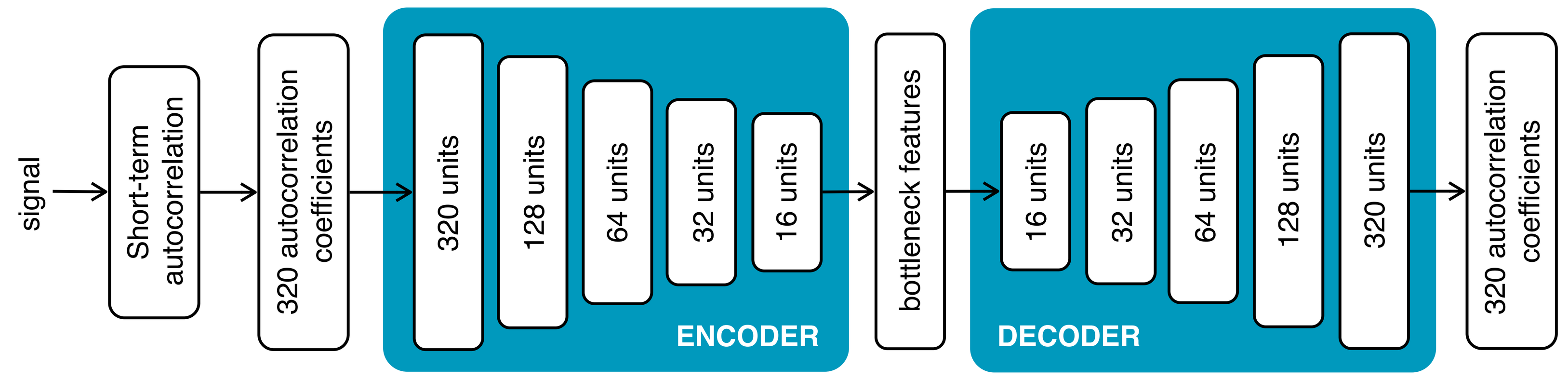


3. System Description

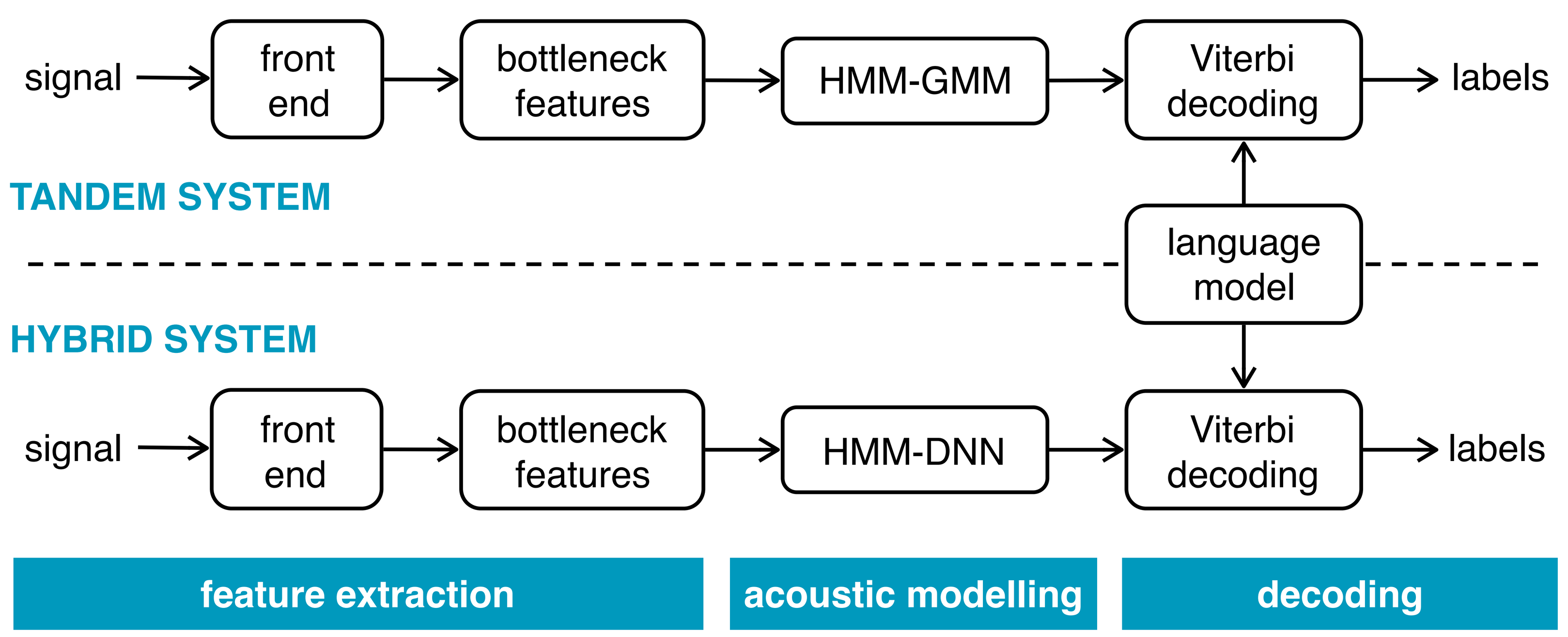
- We leveraged large amount of unlabelled data using unsupervised learning.
- A bigram **language model (LM)** was applied during the decoding process to exploit the breathing patterns.
- **Bottleneck features** from auditory nerve firing **rate maps (RM)**:



- **Snoring is a pitched acoustic event** produced by the vibration of structures like the soft palate, epiglottis, and pharyngeal walls.
- **Bottleneck features** from **short-term autocorrelation function (ACF)** encode pitch information:



- Given the limited amount of training data, 2 snore detection architectures were investigated.
- **Tandem** and **hybrid snore detection systems**:



4. Evaluation

- The systems are **‘snorer-independent’**.
- At event level, the **snore event error rate** was calculated, similar to the word error rate commonly used in ASR.
- At frame level, the **snore F-measure** was computed to evaluate the segmentation quality.

5. Results

- Snore event error rate:

	Tandem		Hybrid	
	No LM	LM	No LM	LM
MFCCs	19.94%	17.59%	9.40%	9.52%
RM bottleneck	12.00%	12.13%	13.40%	13.40%
ACF bottleneck	15.24%	14.48%	14.92%	14.92%
RM and ACF bottleneck	10.86%	**8.89%**	10.22%	9.90%

- Frame-based snore F-measure:

	Tandem		Hybrid	
	No LM	LM	No LM	LM
MFCCs	90.78%	91.67%	93.60%	93.45%
RM bottleneck	**95.29%**	95.23%	90.74%	90.74%
ACF bottleneck	88.34%	88.47%	86.96%	86.96%
RM and ACF bottleneck	94.43%	94.36%	94.73%	94.75%

6. Conclusions

- Robust snore detection in a home environment, from recordings made using a smartphone, is a challenging task.
- **The best performance was obtained using bottleneck features that encode both spectral shape and pitch information.**
- **The LM enforces realistic snore event durations.**
- In the future we will focus on building systems to detect other forms of SDB, such as OSA.

H. E. Romero was funded by a PhD studentship from Passion for Life Healthcare (PFLH) and the Department of Computer Science, University of Sheffield. A. V. Beeston was funded by KTP award 9905 with PFLH.