

---

# LIPS: Let's think of a better title...

---

**John C. Merfeld**  
Department of Computer Science  
Boston University  
Boston, MA 02215  
jcmurf@bu.edu

**Allison Mann**  
Department of Computer Science  
Boston University  
Boston, MA 02215  
ainezm@bu.edu

## Abstract

The abstract paragraph should be indented 1/2 inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 Introduction

## 2 Related work

## 3 Method

### 3.1 Data

The Oxford-BBC Lip Reading Sentences 2 (LRS2) Dataset was selected for training. Details of its creation can be found in [1]. A brief summary of the data will follow.

The atomic units of raw data are video files, each around ten seconds long. The video clips come from BBC news and talk shows, but each frame's viewing window is shrunk to track the face of an individual speaker. Accompanying each video is a metadata text file; this file contains an ordered list of words spoken in the clip along with their start and end time, down to the hundredth of a second.

We did not use the entirety of LRS2; we limited our study to the "pre-train" segment because the "main" segment's text files do not include word timestamps. Still, the model's potential training examples encompassed over 2 million utterances of 41,427 unique words

### 3.2 Feature Processing

### 3.3 Model

## 4 Experimental Results

## 5 Conclusion

### Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

## References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] T. Afouras, J. S. Chung, A. Senior, O. Vinyals, A. Zisserman Deep Audio-Visual Speech Recognition  
arXiv:1809.02108