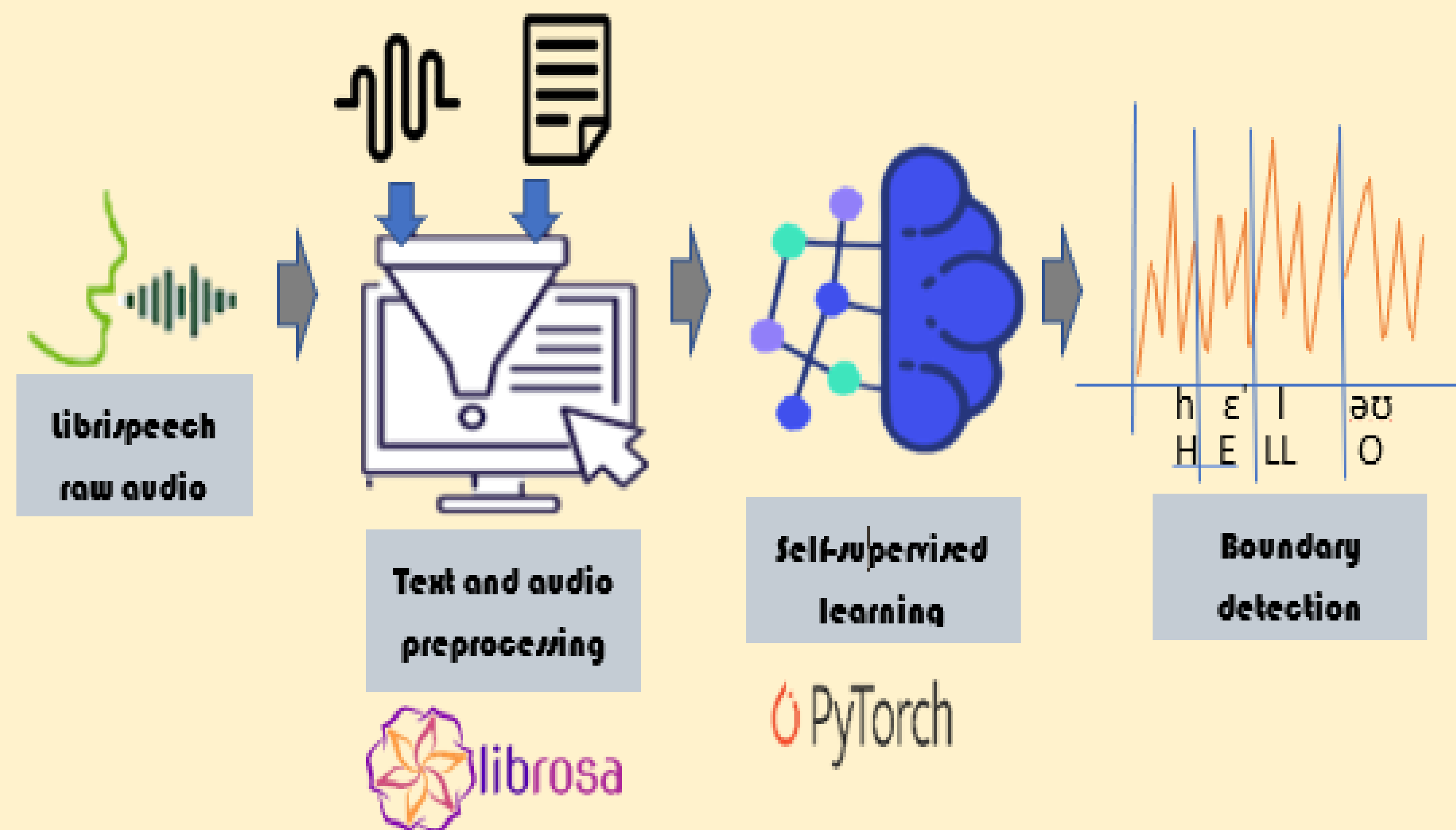


Self-Supervised Learning for Automatic Speech Recognition



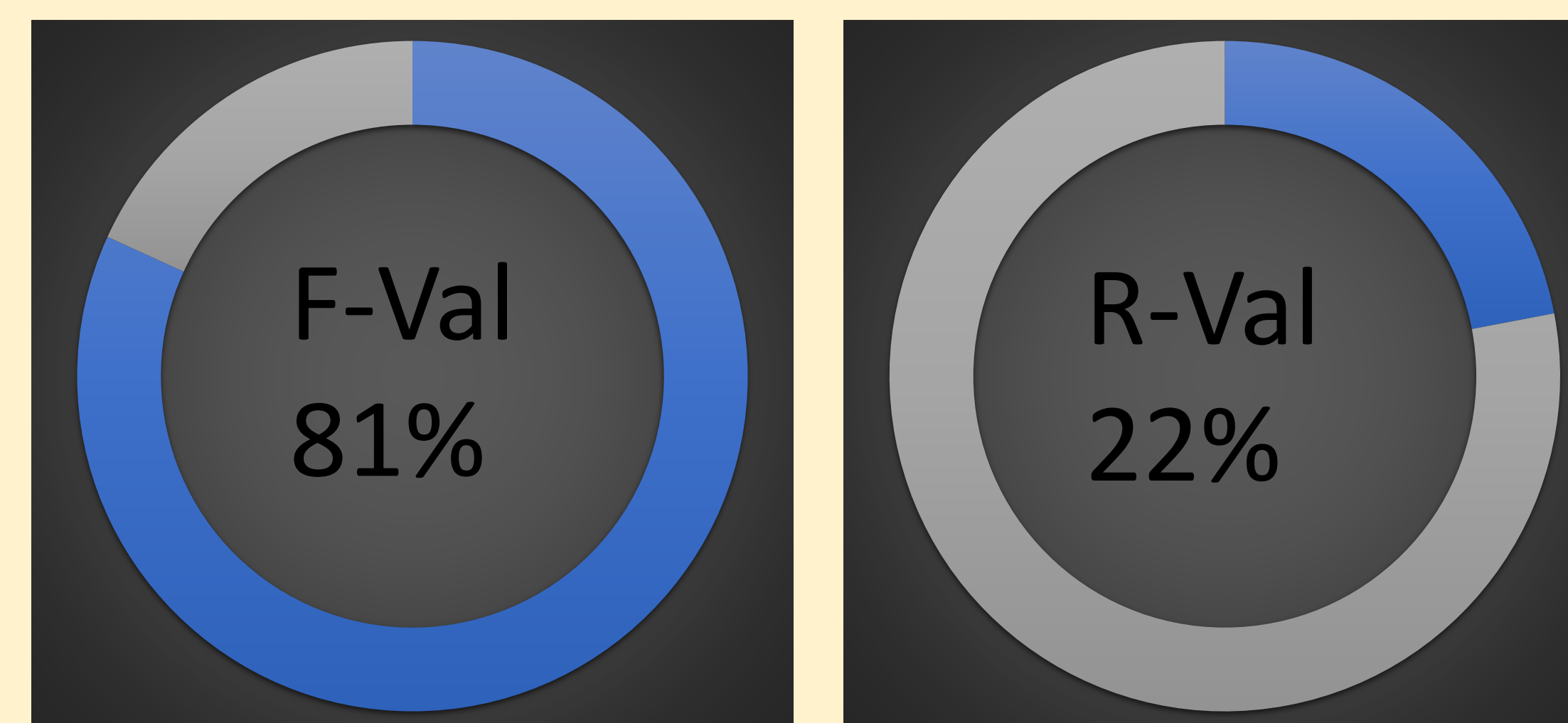
Discussion

Our efforts are geared toward the use of LSTM to perform a next frame prediction and peak tracking over the error curve, because of the positive correlation between those brut peaks and the presence of the potential boundary. Our efforts were guided according to the hypothesis mentioned in the works cited in the original document of blind phoneme segmentation

Methods and Materials

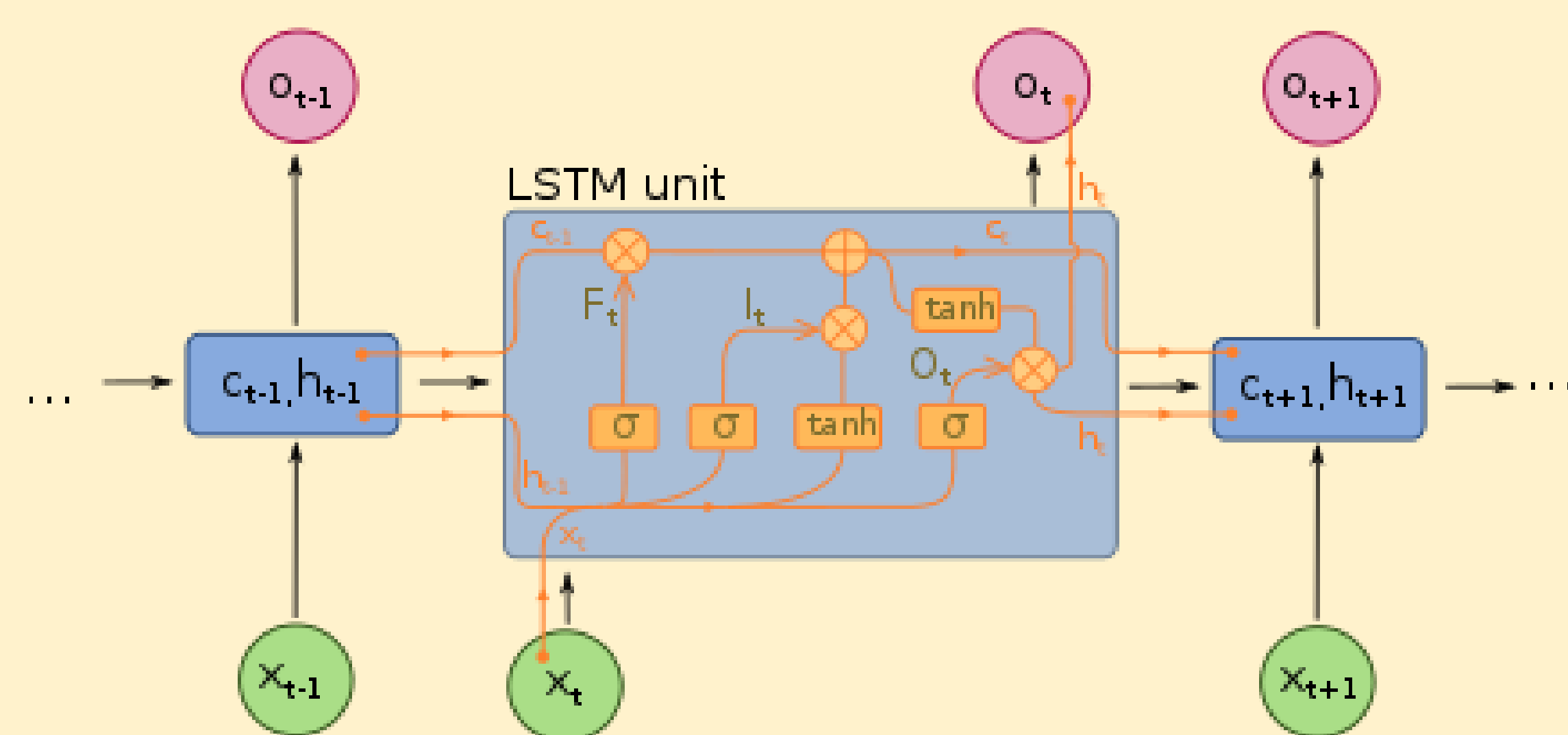
- Data set preparation using the librispeech audio and librispeech text alignment
- Next frame prediction with LSTM
- Peak detection on error curve
- Boundaries spotting

Results



Conclusions

- The poor result of the R-Val still to be improved
- Other speech representation deserves to be explored like wave2vec



- Librosa for MFCCs feaurtes extraction
- Pytorch for the implementation of the model