

Sequential versus Hierarchical Grammar Models for Great Tit Vocal

Sequence Analysis

REU

Kattel, S., Guha, A.

Dr Zhu, K.

Executive Summary

Animal vocalizations provide insight into structured communication and its relation to language evolution. Prior work modeled great tit vocal sequences using coarse call types. We improve this by using syllable-level representation, enabling finer sequence modeling. Our approach achieves high accuracy and significantly outperforming baseline models and demonstrating the importance of representation.



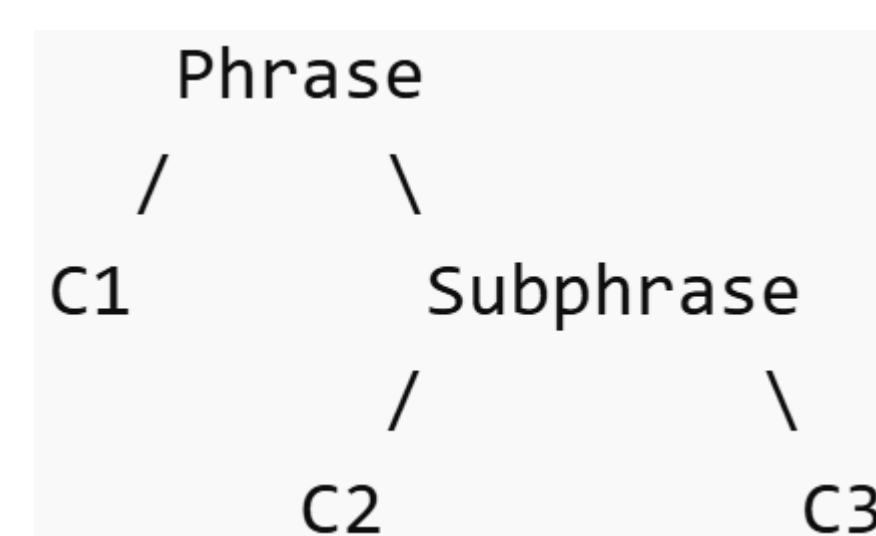
Image source: SongBird Survival

Background

Bird vocalizations exhibit structured patterns, often modeled using sequential approaches. Previous work used **call types**, which are coarse and may miss fine acoustic detail. We instead use **syllables**, providing a more precise representation of vocal structure.

C1 → C2 → C3 → C1 → C4

Sequential: linear, step-by-step transitions between units.



Hierarchical: nested grouping of units into structured levels with dependencies beyond immediate neighbors.

Experimental Setup

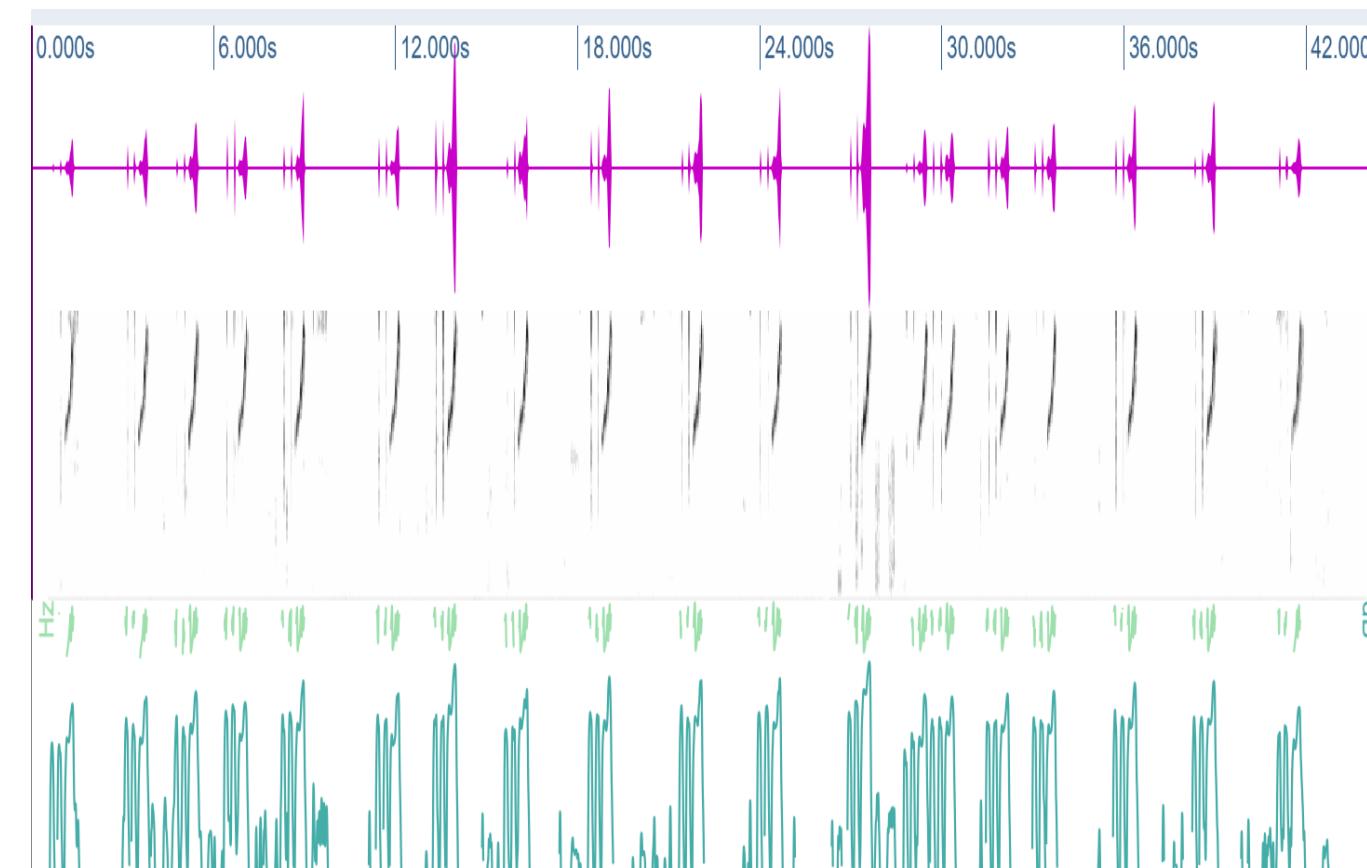


Fig 1. Waveform and spectrogram visualization on Praat++

Audio was segmented into syllables and used to train classification models.

Table 1. Structure of communication in birds

Sl no.	Bird Species	Structure
1	Black-capped Chickadee	Sequential
2	Zebra Finch	Sequential
3	Canary	Sequential
4	Bengalese Finch	Hierarchical
5	European Starling	Hierarchical
6	Mockingbird	Hierarchical

Thresholds were tuned on validation data, and performance was evaluated using accuracy and F1 score.

Experimental Test Plan

Models were trained to classify syllables and evaluated on held-out data. We compared baseline and improved models, analyzing performance using accuracy and F1 to assess both correctness and balance.

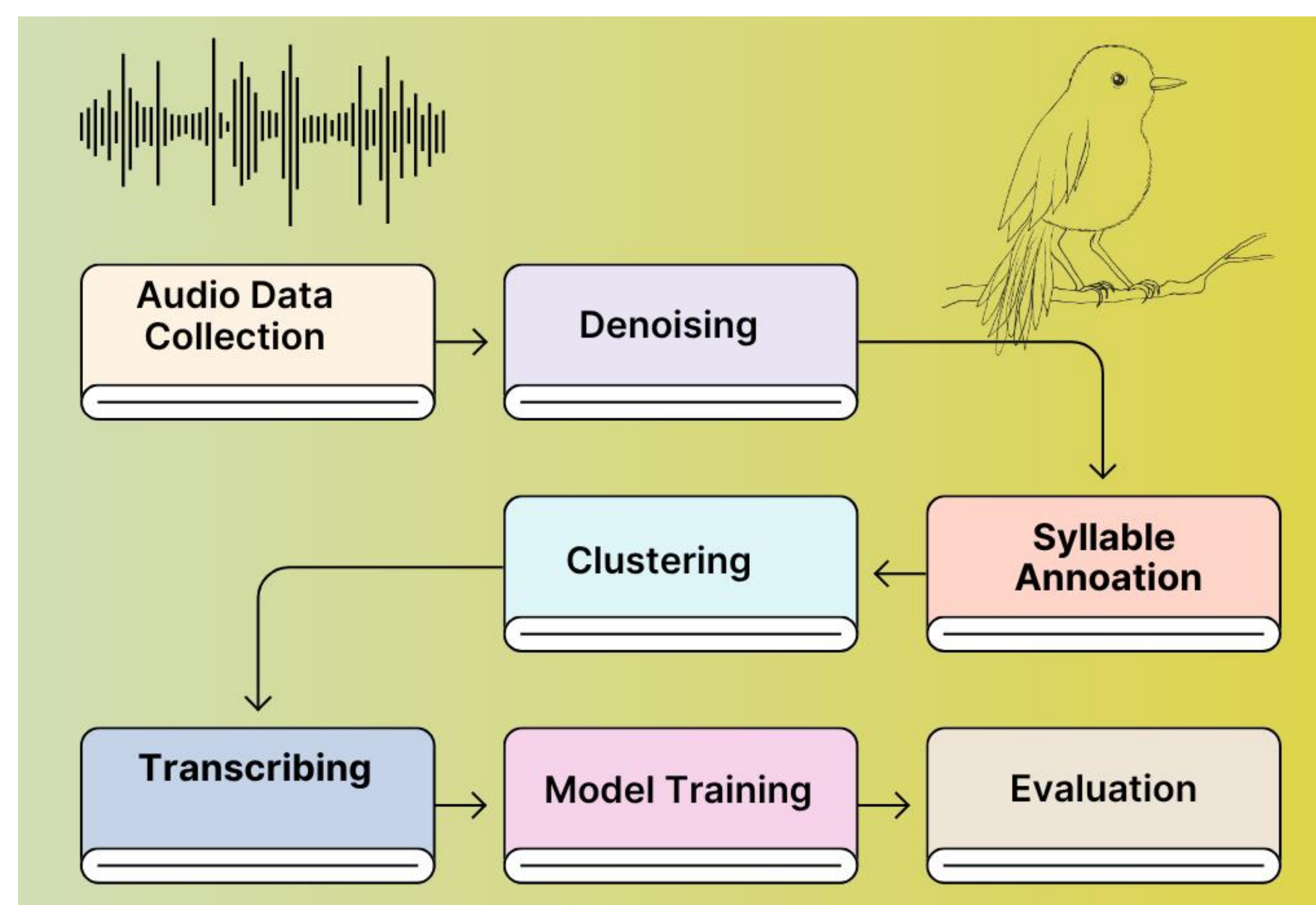


Fig 2. Workflow

Experimental Results

The baseline PANNs model achieved an F1 score of 0.3 on syllable boundary detection, indicating limited performance on our domain-specific task. After fine-tuning on 354 manually annotated audio files, the model improved to an F1 score of 0.77, demonstrating its effectiveness in accurately identifying syllable

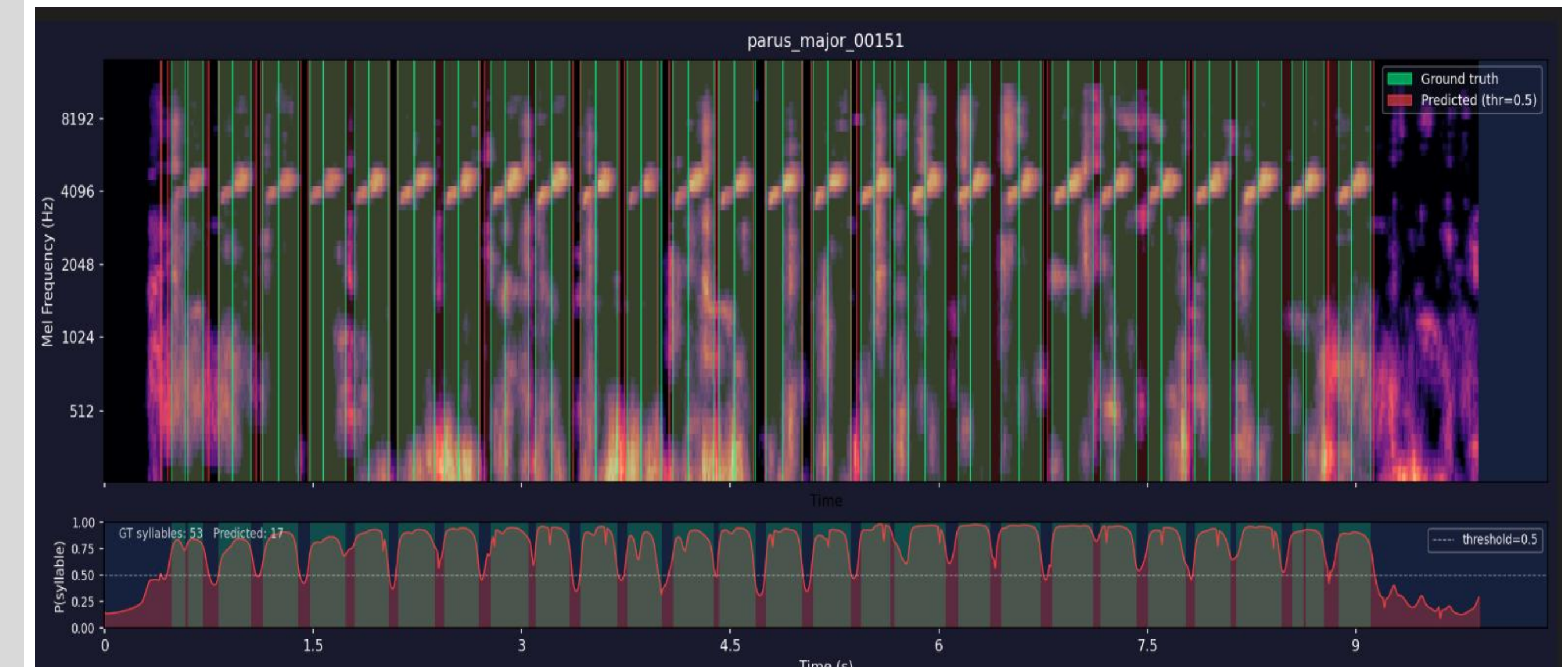


Fig.2 Ground truth vs predicted boundaries

Conclusions

Syllable-level representation improves performance by capturing finer structure. Results show that better representation has a larger impact than model choice. We have successfully done the groundwork for showing that Great Tits follow a sequential model of communication. Future work includes automated segmentation and deeper models.



References

- [1] S. W. Townsend, S. Engesser, S. Stoll, K. Zuberbühler, and B. Bickel, "Compositionality in animals and humans," *PLoS Biology*, vol. 16, no. 8, 2018.
- [2] J. R. Krebs, R. Ashcroft, and M. Webber, "Song repertoires and territory defence in the great tit," *Nature*, vol. 271, pp. 539–542, 1978.
- [3] T. N. Suzuki, D. Wheatcroft, and M. Griesser, "Experimental evidence for compositional syntax in bird calls," *Nature Communications*, vol. 7, 2016.
- [4] T. Q. Gentner et al., "Recursive syntactic pattern learning by songbirds," *Nature*, vol. 440, pp. 1204–1207, 2006.