



HAL
open science

Application to long recordings of babies: Automatic extraction of infant vocalizations from one year home audio recordings

Guillem Bonafos, Clara Bourot, Pierre Pudlo, Jean-Marc Freyermuth, Laurence Reboul, Thierry Legou, Samuel Tronçon, Raoul Desbrières, Jean-Michel Bartoli, Arnaud Rey

► To cite this version:

Guillem Bonafos, Clara Bourot, Pierre Pudlo, Jean-Marc Freyermuth, Laurence Reboul, et al.. Application to long recordings of babies: Automatic extraction of infant vocalizations from one year home audio recordings. Workshop on Infant Language Development (WILD-2024), Jun 2024, Lisbonne, Portugal. hal-04630681

HAL Id: hal-04630681

<https://amu.hal.science/hal-04630681v1>

Submitted on 1 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatic extraction of infant vocalizations from one year home audio recordings

Guillem Bonafos¹, Clara Bourot^{2,6,7}, Pierre Pudlo^{2,7}, Jean-Marc Freyermuth^{2,7}, Laurence Reboul^{2,7},
Thierry Legou^{3,7}, Samuel Tronçon⁴, Raoul Desbrières⁵, Jean-Michel Bartoli⁵, Arnaud Rey^{6,7}

1. ENES, CNRS – Saint Etienne University 3. LPL, CNRS – Aix-Marseille University 5. Saint Joseph's Hospital, Marseille 7. ILCB, CNRS – Aix-Marseille University
2. I2M, CNRS – Aix-Marseille University 4. Résurgences R&D 6. CRPN, CNRS – Aix-Marseille University

Main objectives

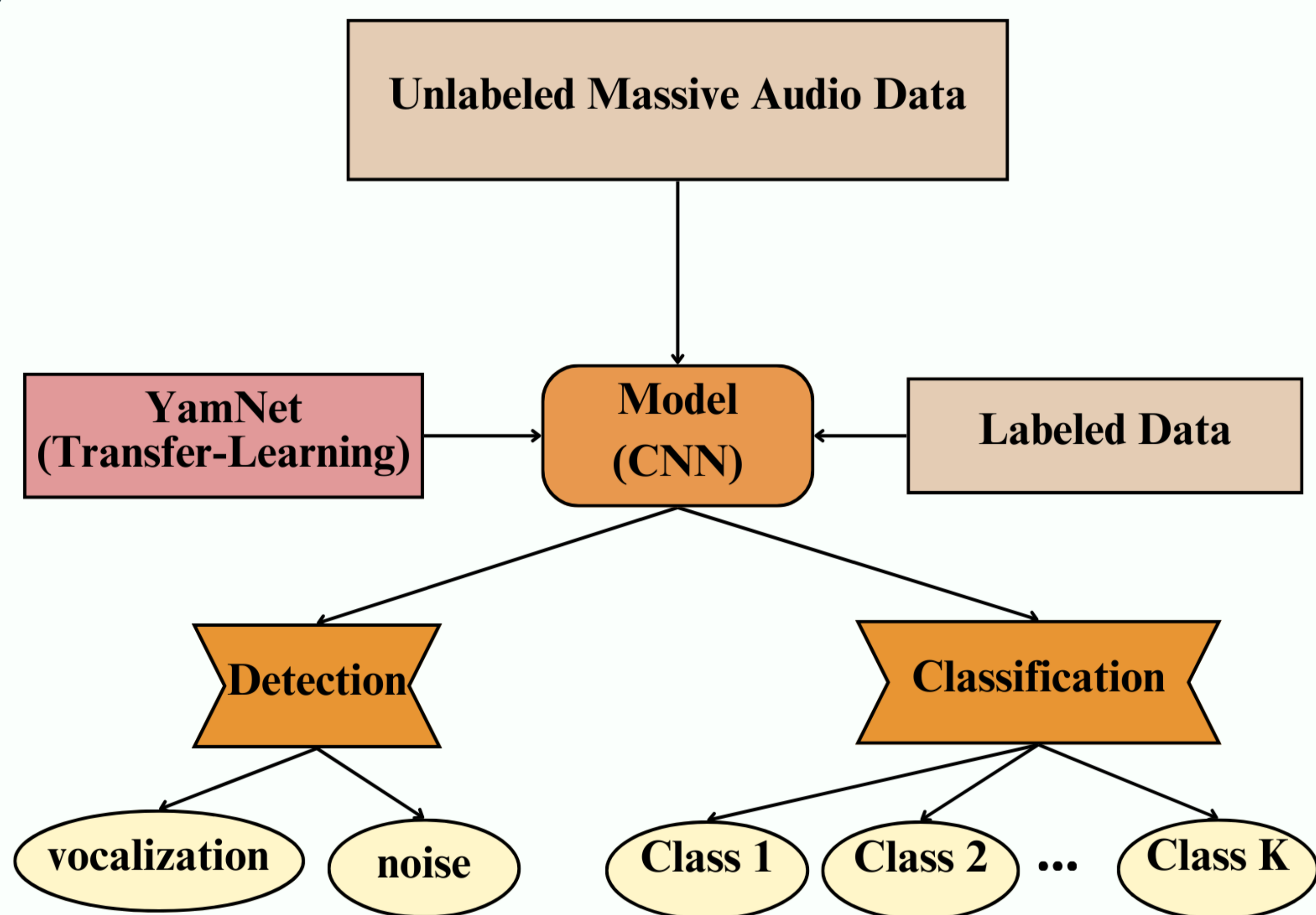
We propose an automatic data processing pipeline based on **deep neural network** that :

- **Extract vocalizations** from **large-scale natural audio recordings**
- **Classify** these **vocalizations**

Advantages

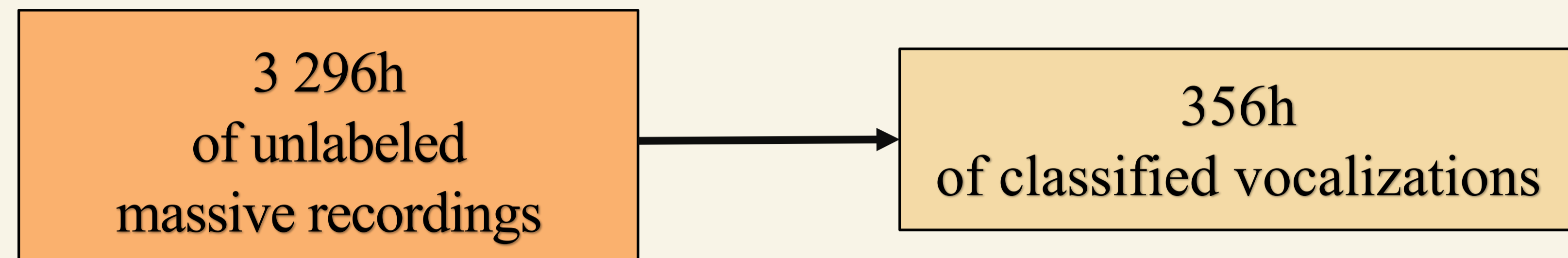
- **Automatic** rather than manual processing
- Do not require a **large amount** of labelled data
- **Robust** against noise

1) Algorithm: [1]



2) Data:

- 13 babies recorded in their first year of life



3) Performance: [1]

Detection:

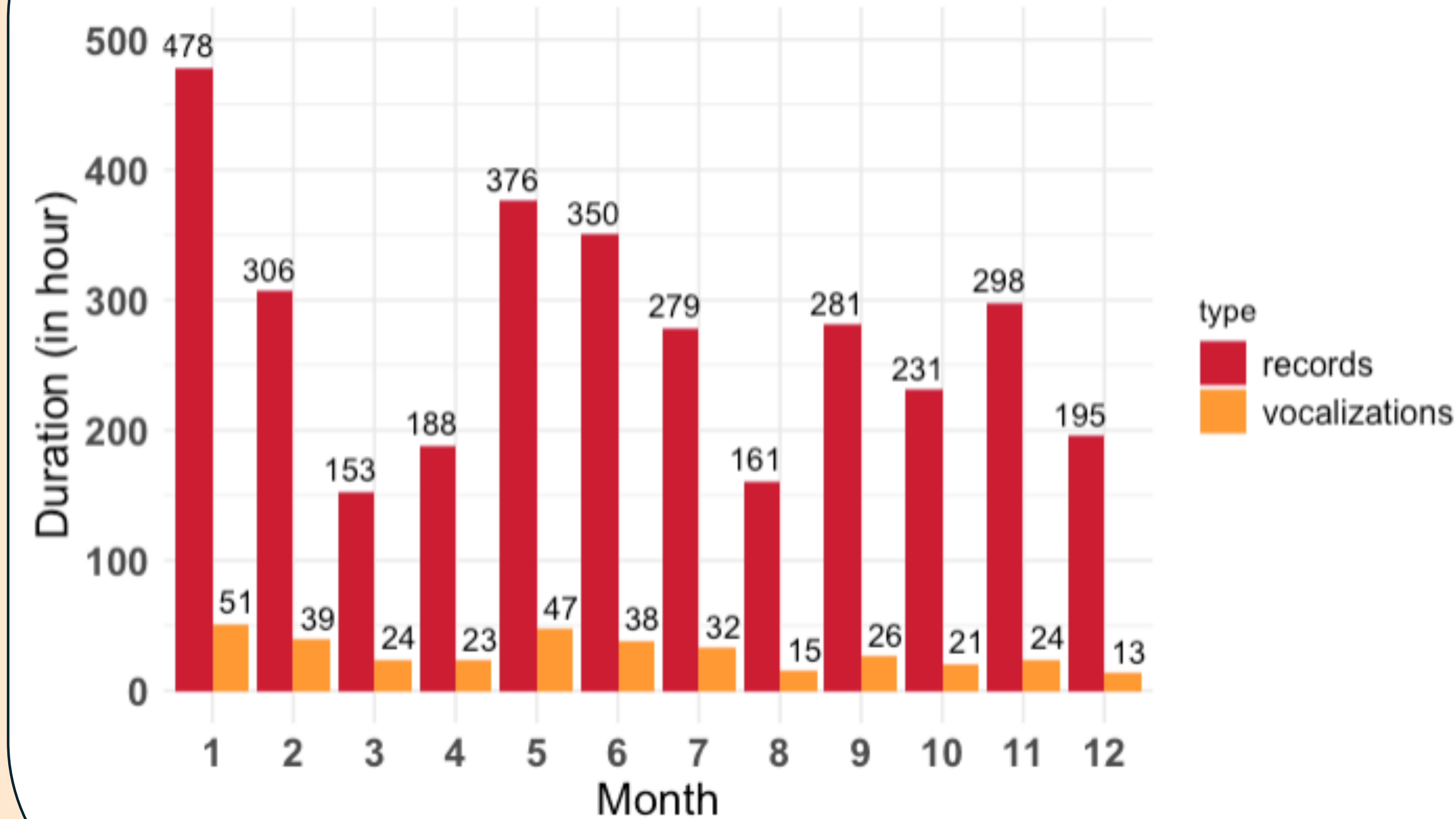
True Class	Predicted Class	
	Vocalization	Noise
Vocalization	100	0
Noise	0	100

Classification:

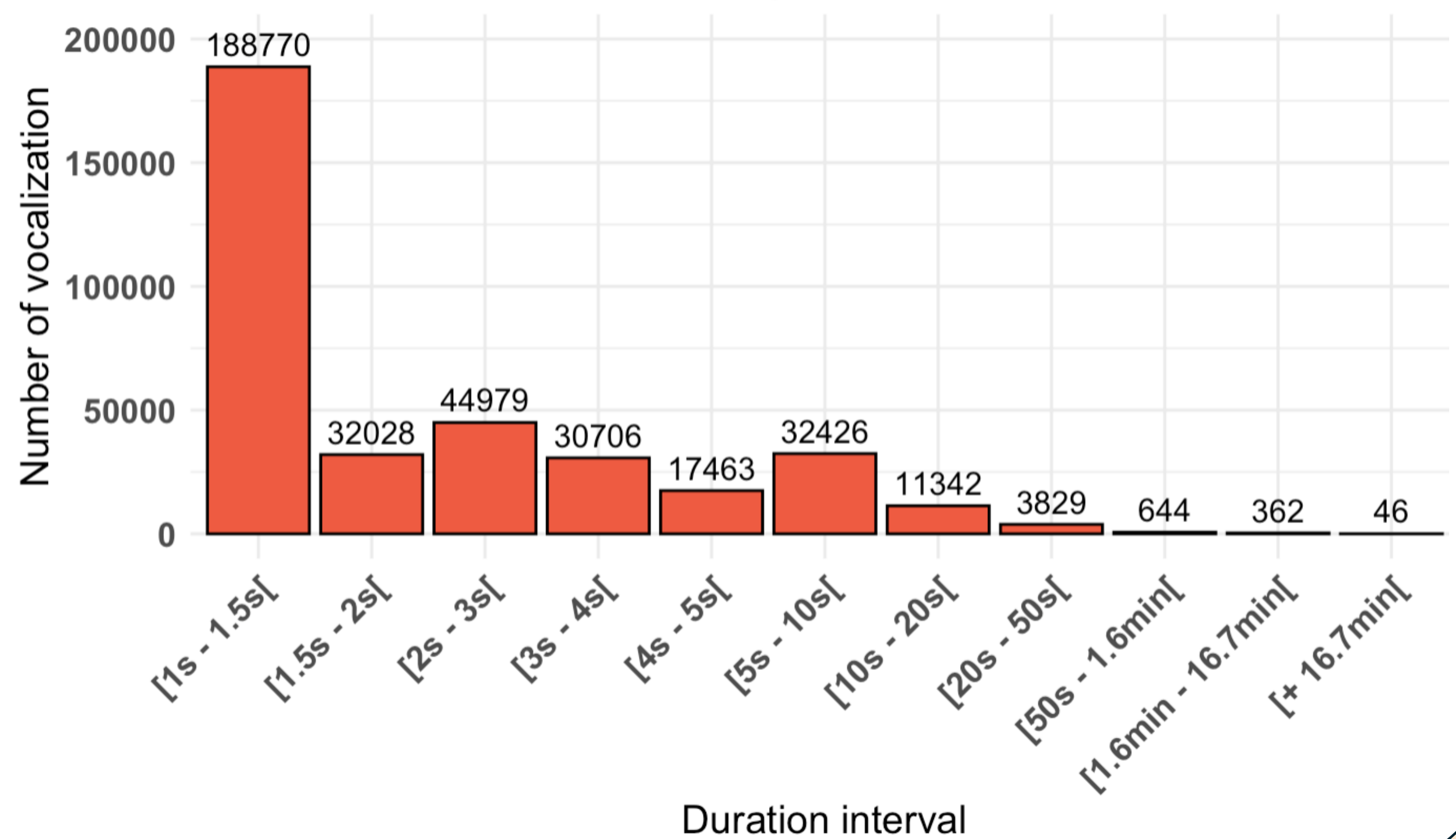
True Class	Predicted Class				
	Canonical	Crying	Junk	Laughing	No-Canonical
Canonical	46	16	11	10	18
Crying	15	44	11	13	17
Junk	19	6	51	9	15
Laughing	13	17	19	36	15
No-Canonical	20	19	24	9	29

4) Descriptive statistics:

Sum of records and vocalizations by month for all babies



Distribution of vocalizations by duration



Conclusion:

- In the **detection task**, the algorithm correctly discriminate between noise and vocalizations

Limitations:

- **Presence** of unlikely **long vocalizations**
- Ongoing improvements are necessary to improve the precision of the **detection task** and **classification task**.

References:

[1] Bonafos et al., (2023) Detecting human and non-human vocal productions in large scale audio recordings. [arXiv:2302.07640](https://arxiv.org/abs/2302.07640)