

Models of laughter dynamics in early child-parent interaction

Stage M2 IAAA

Gabriel Meunier

Aix-Marseille Université

June 10, 2021

Sommaire

Introduction

Données

Méthodes

Expériences

Résultats

Conclusion

Introduction

- ▶ Contexte d'interaction mère / enfant
- ▶ Détection d'événement sonore (rire)
- ▶ Annotation automatique de Providence
- ▶ D'autres application seront possible grâce à ce système
 - ▶ Fréquence et alignement entre rire pendant le développement de l'enfant

Providence et AudioSet

- ▶ Providence
 - ▶ 22 vidéos annotées sur environs 300
 - ▶ 4 enfants sur 6 utilisés
 - ▶ 289 rire augmenté a 840 sur 10h annotées
- ▶ AudioSet
 - ▶ 18678 samples
 - ▶ balanced entre rire et parole

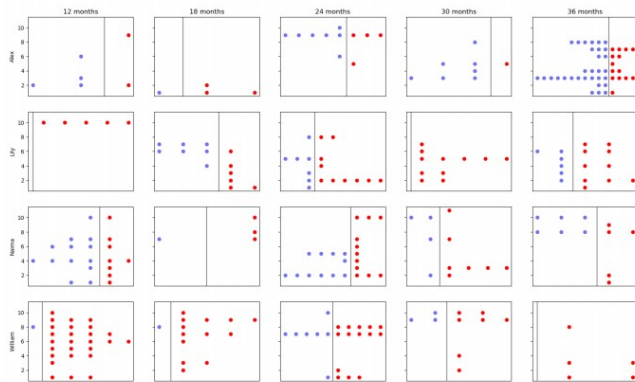


Figure: Occurrence de rire en fonction du temps (rouge/mère, bleu/enfant), (Mazzocconi, 2019)
 L'axe vertical représente le temps de la vidéo découpé en 10 windows de temps équivalent.

Description & Augmentation des données

- ▶ Providence
 - ▶ Rires entre 1 et 4 secondes
 - ▶ Multiples samples de 5 secondes pour chaque rire
- ▶ AudioSet
 - ▶ Rire de durée inconnu dans des samples de 5 secondes
- ▶ Features
 - ▶ MFCCs
 - ▶ Pitches
 - ▶ Embeddings

Méthodes

- ▶ Méthodes classiques

- ▶ SVM
- ▶ KNN
- ▶ RandomForest

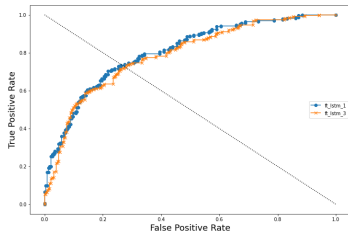
- ▶ Méthodes profondes

- ▶ Simple DNN, CNN, RNN
- ▶ Embedding YAMnet / VGGish / Wav2Vec + RNN

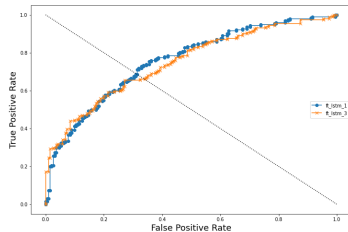
Expériences

- ▶ Choix du modèle
- ▶ Robustesse à la généralisation
- ▶ Optimisation des performances
- ▶ Séparation mère / enfant

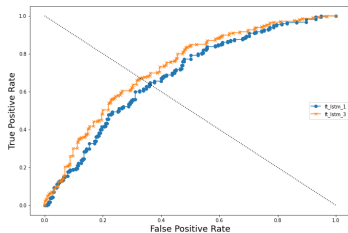
Expériences



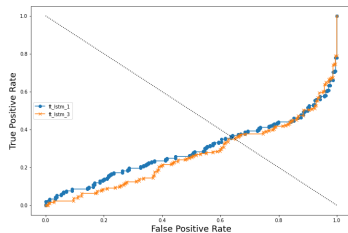
(a) Alex out
AUC : 0.80, 0.79



(b) Lily out
AUC : 0.76, 0.75



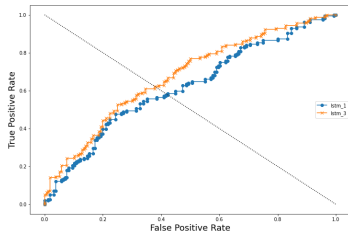
(c) William out
AUC : 0.68, 0.72



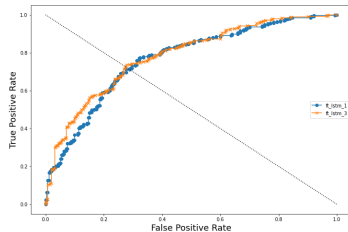
(d) Naima out
AUC : 0.29, 0.26

Figure: ROC curve en fonction de l'enfant out. (AUC 1layer, 3layer)

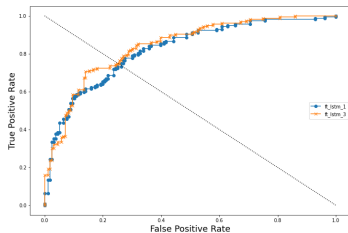
Expériences



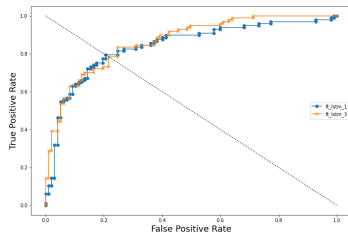
(a) Prov., mères + enfants
AUC : 0.62, 0.67



(b) Audioset + Prov., mères + enfants
AUC : 0.77, 0.78



(c) Audioset + Prov., mères
AUC : 0.82, 0.84



(d) Audioset + Prov., enfants
AUC : 0.84, 0.87

Figure: ROC curve des différentes expériences. (AUC 1-layer, 3-layer)

Résultats

	LSTM 1-layer	LSTM 3-layer
P	0.62	0.67
ASFP	0.77	0.78
ASFP-C	0.84	0.87
ASFP-M	0.82	0.84

Table: AUC des différents modèles.

P (Providence), ASFP (Audioset + finetune Providence), ASFP-C (ASFP version enfant), ASFP-M (ASFP version mère)

	nb frame	% pred child	precision child	% pred mum	precision mum
Alex	3264	1.29%	0.315	1.78%	0.259
Lily	3306	1.09%	0.081	1.97%	0.226
Naima	2364	2.37%	0.20	1.50%	0.308
Violet	2568	0.58%	0.322	1.13%	0.171
William	3174	2.27%	0.014	3.62%	0.136
Moyenne	2935	1.51%	0.178	2.06%	0.206

Table: Résultats d'annotation sur de nouvelles vidéos prise aléatoirement dans la base de donnée Providence et annoté avec ASFP-C/M 3-layer.

Conclusion

- ▶ État actuel
 - ▶ Système fonctionnel fichier audio → fichier d'annotation
 - ▶ Grande augmentation de la rapidité d'annotation
- ▶ Dans le futur
 - ▶ Possibilité d'améliorer le système au fur et à mesure.
 - ▶ Recherche de nouveaux système quand AudioSet sera plus nécessaire
 - ▶ Nouvelle application de recherche