SE&R 2022

https://sites.google.com/view/ser2022/

Automatic Speech Recognition for Spontaneous and Prepared Speech & Speech Emotion Recognition in Portuguese

Shared-Tasks at Propor 2022

Proceedings of the Workshop, Vol. 1

March 21, 2022

About the workshop

The Automatic Speech Recognition for Spontaneous and Prepared Speech & Speech Emotion Recognition in Portuguese (SE&R 2022) took place with the 15th edition of the International Conference on the Computational Processing of Portuguese (PROPOR 2022).

The workshop consisted on two main tracks: Automatic Speech Recognition (ASR) for spontaneous and prepared speech for Portuguese; and Speech Emotion Recognition (SER) in Portuguese. Our main objective in proposing this challenge was to promote research in Portuguese audio processing. Two corpora were proposed for use in the SE&R shared-tasks: CORAA ASR and CORAA SER. CORAA (Corpus of Annotated Audios). A focus is given on the Brazilian Portuguese variant. CORAA ASR corpus contains 389 hours of spontaneous and prepared speech, segmented at utterance level, together with the respective transcription for each utterance. CORAA SER is a 50 minute corpus for emotion recognition. All participants received a corpus training set. The test set (ground truth) was only made available during the workshop.

This edition of the workshop includes six papers describing the datasets and solutions submitted in the SE&R shared-tasks:

- Overview of the Automatic Speech Recognition for Spontaneous and Prepared Speech & Speech Emotion Recognition in Portuguese (S&ER) Shared-tasks at PROPOR 2022
  Arnaldo Candido Junior, Edresson Casanova, Ricardo Marcacini

- Domain Specific Wav2vec 2.0 Fine-tuning For The SE&R 2022 Challenge
  Alef Iury S. Ferreira, Gustavo dos Reis Oliveira

- Pretrained audio neural networks for Speech emotion recognition in Portuguese
  Marcelo Matheus Gauy, Marcelo Finger

- Transfer Learning and Data Augmentation Techniques applied to Speech Emotion Recognition in SE&R 2022
  Caroline Alves, Bruno Carlotto, Bruno Dias, Anátale Garcia, Bruno Gianesi, Renan Izaias, Maria Luiza Morais, Paula Oliveira, Vinícius G. Santos, Rafael Sicoli, Flaviane R. Fernandes Svartman, Sandra Aluisio, Sidney Leal

- Speech Emotion Recognition in Portuguese for SofiaFala: SER SofiaFala
  Alexander Scaranti, Douglas Silva, Fernando Meloni, Alessandra Alaniz

- Transductive Ensemble Learning with Graph Neural Network for Speech Emotion Recognition
  Eliton L. Scardin Perin and Edson Takashi Matsubara

Organizing committee

The SE&R 2022 workshop is promoted by the TaRSila project, which aims to increase speech datasets for Brazilian Portuguese language, looking to achieve state-of-the-art results for the following tasks:

(a) automatic speech recognition (ASR) that automatically transcribes speech;

(b) multi-speaker synthesis (TTS) that generates several voices from different speakers;

(c) speaker identification/verification that selects a speaker from a set of predefined members (speakers seen during the training of the models --- called closed-set scenario --- or in open-set scenario in which the verification occurs with speakers not seen during the training of the models); and

(d) voice cloning that uses a few minute/second voice dataset to train a voice model with synthesis methods, which can read any text in the target voice.

The TaRSila project is part of the Natural Language Processing initiative (NLP2) of the Center for Artificial Intelligence (C4AI) of the University of São Paulo, sponsored by IBM and FAPESP.

The following researchers organized the Shared Task:

- Alessandra Alaniz Macedo, FFCLRP/USP, Brazil (Website Chair)
- Arnaldo Candido Jr., UTFPR, Brazil (Program Chair & Conference Chair)
- Edresson Casanova, ICMC/USP, Brazil (Evaluation Chair)
- Flaviane Romani Fernandes Svartman, FFLCH/USP, Brazil (Program Chair)
- Marcelo Finger, IME/USP, Brazil (Conference Chair)
- Ricardo M. Marcacini, ICMC/USP, Brazil (Publication Chair & Conference Chair)
- Sandra Maria Aluísio, ICMC/USP, Brazil (Website Chair)

Program Committee and External Reviewers

- Bruno Nogueira, FACOM/UFMS, Brazil
- Christopher Shulby, DefinedCrowd, Portugal
- Diego Furtado Silva, UFSCAR, Brazil
- Isabel Maria Martins Trancoso, IST-Lisboa, Portugal
- Nils Murrugarra Llerena, Snapchat, USA
- Rafael Geraldeli Rossi, UFMS, Brazil
- Solange O. Rezende, ICMC/USP, Brazil

Acknowledgements

The SE&R shared-tasks was supported by Center for Artificial Intelligence (C4AI-USP), with support by the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and by the IBM Corporation.