Far-field Speaker Verification Challenge (FFSVC) 2022 : Challenge Evaluation Plan

Xiaoyi Qin[1,2], *Ming Li*[1,2], *Hui Bu*[5],*Shrikanth Narayanan*[4], *Haizhou Li*[3]

[1]Data Science Research Center, Duke Kunshan University, Kunshan, China
[2]School of Computer Science, Wuhan University, Wuhan, China
[3]Department of Electrical & Computer Engineering, National University of Singapore, Singapore
[4]Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, USA
[5]AI Shell Foundation, Beijing, China

ming.li369@duke.edu

1. Introduction

This document is the description of Far-field Speaker Verification Challenge (FFSVC) 2022. The success of FFSVC2020 [1] indicates that more and more researchers are paying attention to the far-field speaker verification task. In this year, the challenge is still focus on the far-field speaker verification scenario and provided a new far-field development and test set collected by real speakers in complex environments with multiple scenarios, e.g., text-dependent, text-independent, cross-channel enroll/test, etc. In addition, in real scenario, speech data is not alway labeled especially for far-field data, which is hard to accurate labeled with close-talking pre-trained model. Therefore, a new focus of this year is cross-language self-supervised/semi-supervised learning, where participants are allowed to generate the pseudo-label for the train/dev set without using speaker label of the FFSVC2020 dataset (in Mandarin) by close-talking model trained by VoxCeleb1&2 (mostly in English) to fine-tune the model.

To this end, this year challenge has two tasks considering the data labeling scenario.

- **Task 1. Fully supervised far-field speaker verification.**
- **Task 2. Semi-supervised far-field speaker verification.**

In contrast to FFSVC2020 tasks, this challenge is focus on the single-channel scenario, which means that both the enrollment and test audio are single-channel data. In addition, considering the real application scenario, a list of trials in this year will more challenge than FFSVC2020. The trial pairs will considering more hard cases, e.g. same gender, cross-domain, cross-channel, cross-time, etc.

The FFSVC2022 is designed to boost the speaker verification research with special focus on far-field scenario under noisy conditions in real scenarios. The objectives of this challenge are to: 1) benchmark the current speech verification technology under this challenging condition, 2) promote the development of new ideas and technologies in speaker verification, 3) provide a real, hard, and exploring challenge to the community that exhibits the far-field characteristics in real scenes.

The new official challenge website has been published in https://ffsvc.github.io. The original FFSVC2020 website (http://2020.ffsvc.org/) will be keep but not updated anymore. The challenge resource and information of FFSVC2020 has been moved to https://ffsvc.github.io/2020. The latest news and challenge information will be update on the new challenge website[1]. If you have any question, please email to ffsvc.challenge@gmail.com and we will reply you as soon as possible.

2. Tasks Description

Task 1 of FFSVC 2022 is the fully supervised far-field speaker verification task that using limited large-scale close-talking databases and FFSVC20 dataset to train a far-field speaker verification system.

Task 2 of FFSVC 2022 is the semi-supervised far-field speaker verification task that participants are allowed to using limited large-scale close-talking databases with speaker label and FFSVC20 dataset without speaker labels to train a far-field speaker verification system.

2.0.1. Trial case

The two tasks adopt the same trial file. In contrast to FFSVC2020, we do not divide the task into two cases: text-dependent and text-independent. However, we will set these test cases in trial files. The final trial cases including but not limited to

- Within same gender. Most trial pairs are selected from same gender and few cross-gender trial pairs are provided.
- Cross-domain. Close-talking or far-field speech segment is chosen as test wav file and enrollment data uses the close-talking speech segment.
- Cross-channel. Telephone data is chosen as enrollment data and tablet/telephone data is chosen as test data.
- Cross-time. Since the recording is not done at once. Each speaker visits three time and each visits has 7-15 days gap. Therefore, enrollment and test data are chosen from different visits data.
- Text mismatch. Contents of enrollment/test data consists of text-dependent and text-independent speech segment.

The trials considering many hard cases in this challenge. The participants are expected to explore more novel and robustness systems.

[1]https://ffsvc.github.io

2.0.2. Training data

We define the task 1&2 as fixed training conditions that the participants can only use special training set to build a speaker verification system. The fixed training set consists of the following:

- VoxCeleb 1&2 [2, 3].
- FFSVC2020 dataset (Train and dev set) [1, 4].

For task 1, participants can only use VoxCeleb 1&2 dataset and the train/dev set of FFSVC2020 dataset with speaker labeled to train the model.

For task 2, in contrast to task1, **participants cannot use the speaker label of FFSVC2020 dataset**. In this task, we encourage the participants adopt the self-supervised or semi-supervised methods to solve the problem of cross-domain unlabeled data, e.g. identity pseudo-label using model pre-trained on the VoxCeleb dataset.

Using any other speech data to the training system is forbidden, while participants are allowed to use the non-speech data to do data augmentation. The self-supervised pre-trained models, such as Wav2Vec [5] and WavLM [6], cannot be used in this challenge.

2.0.3. Development set

In this year, we will publish new development and evaluation sets, which recorded in the same period as FFSVC2020 dataset. A small trials file and wav files with accurate speaker information will be provided for participants as the development set. The development data has the same data distribution as evaluation data. However, the development set is only allowed to tune hyperparameters and test model performance. Any circumstances of training development set is not allowed, eg. using development set to train PLDA or speaker system.

2.0.4. Evaluation set

As mention before, the evaluation set is not completely out-domain data. However, unlike the FFSVC2020 challenge, we introduce more recorded devices and all trial pairs are single-channel speech segments. Evaluation set consists of a large trials file and anonymized audios.

3. Evaluation Protocol

3.1. Evaluation Metrics

In this challenge, we will use two metric to evaluate the system performance. The primary metric we adopt is the Minimum Detection Cost(mDCF). In addition, Equal Error Rate (EER) will be provided to participant as auxiliary metrics.

The mDCF is based on the following detection cost function which is the same function as used in the NIST 2010 SRE. It is a weighted sum of miss and false alarm error probabilities in the form:

$$C_{det} = C_{miss} \times P_{miss} \times P_{tar} + C_{fa} \times P_{fa} \times (1 - P_{tar}) \quad (1)$$

We assume a prior target probability, P_{tar} of 0.01 and equal costs between misses and false alarms. The model parameters are 1.0 for both C_{miss} and C_{fa}.

3.2. Leaderboard platform

This year we move challenge scoring platform to Codalab competition platform. An online leaderboard for each task will be provided and participants can submit maximum one result per day during the evaluation phase. The leaderboard shows the performance of the systems on the fully evaluation set (unlike FFSVC 2020). The challenge leaderboard platforms are available at

- Task 1. https://codalab/competition/xxxx
- Task 2. https://codalab/competition/xxxx

4. Registration and Submission

4.1. Registration

Since the challenge will be held on the Codalab competition platform, please create an account if you do not have one. We kindly request you to associate your account to an institutional e-mail. The organizing committee reserves the right to revoke your participant to the challenge otherwise, please read the evaluation plan carefully. Make sure to set the name of your team in the user's profile, or it will not be visible on the leaderboard.

Participants can register in one or two tasks. If your team participates in multiple tasks, we kindly request you to use the same user account to participate in all tasks.

Please note that any deliberate attempts to bypass the submission limit (for instance, by creating multiple accounts and using them to submit) will lead to automatic disqualification.

4.2. Submission

4.2.1. Score submission

Participants are required to submit at least one valid score file for each participating task to the Codalab platform. The score files must be named as *scores.txt*. The score files should be in UTF-8 format with one line per trial. We will provided a sample in the challenge website.

4.2.2. System description submission

Each registered team is required to submit a technical system description report before the end of the challenge. Please submit this report using the Interspeech 2022 paper template. All reports must be a minimum of 2 pages (including references). Reports must be written in English. The system description does not need to repeat the content of the evaluation plan, such as the introduction of database, evaluation metric, etc. The system description must include the following items:

- A complete description of the system components, including front-end (e.g., speech activity detection, features, normalization, front-end speech enhancement) and back-end (e.g., background models, i-vector/embedding extractor, Probabilistic Linear Discriminant Analysis (PLDA), speaker features fusion) modules along with their configurations (i.e., filterbank configuration, dimensionality and type of the acoustic feature parameters, as well as the acoustic model and the backend model configurations).
- A complete description of the data partitions used to train the various models.
- Performance of the submission systems on the development dataset and the evaluation dataset (calculated by the Codalab platform). Teams are encouraged to quantify the contribution of their major system components that

they believe resulted in significant performance gains.[2]

- Novel ideas, strategies and methods are strongly recommended to be shared.
- A report of the model size, usage of GPUs or CPUs.

The reports should be sent to us as a link to arXiv document, or as a PDF file. In both cases, we will place links to the reports from the challenge website. The report may be used to form all or part of a submission to another conference or workshop. We recommend that you send the report as a link to arXiv document if you intend to do this. The links and PDF files should be sent to ffsvc.challenge@gmail.com.

4.2.3. Paper submission

The organizing committee highly encourages the participating teams to submit a paper to the INTERSPEECH 2022 Far-Field Speaker Verification Challenge special session. However, paper contributions within the scope are also welcome if the authors do not intend to participate in the Challenge itself. In any case, please submit your paper using the standard style info and respecting length limits, and submit to the Interspeech 2022 paper submission system. Important: as topic you should choose only this Special Session (F2SVC 2022). The papers will undergo the normal review process similar to the regular session papers.

5. Schedule

- January 01,2022: Releasing the development data as well as the evaluation plan.
- February 01,2022: Releasing the evaluation data and opening submission of the Codalab evaluation server.
- March 21,2022: Interspeech 2022 paper submission deadline
- March 28,2022: Deadline for results submission
- March 31,2022: Deadline for system description
- Interspeech 2022 special session day: FFSVC2020 special section

6. References

[1] X. Qin, M. Li, H. Bu, W. Rao, R. K. Das, S. Narayanan, and H. Li, "The INTERSPEECH 2020 Far-Field Speaker Verification Challenge," in *Proc. Interspeech*, 2020, pp. 3456–3460.

[2] A. Nagrani, J. S. Chung, and A. Zisserman, "Voxceleb: A Large-Scale Speaker Identification Dataset," in *Proc. Interspeech*, 2017, pp. 2616–2620.

[3] J. S. Chung, A. Nagrani, and A. Zisserman, "Voxceleb2: Deep Speaker Recognition," in *Proc. Interspeech*, 2018.

[4] X. Qin, M. Li, H. Bu, R. K. Das, W. Rao, S. Narayanan, and H. Li, "The FFSVC 2020 Evaluation Plan," *arXiv:2002.00387*, 2020.

[5] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "wav2vec: Unsupervised pre-training for speech recognition," *arXiv:1904.05862*, 2019.

[6] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao, J. Wu, L. Zhou, S. Ren, Y. Qian, Y. Qian, J. Wu, M. Zeng, and F. Wei, "Wavlm: Large-scale self-supervised pre-training for full stack speech processing," *arXiv:2110.13900*, 2021. [Online]. Available: https://arxiv.org/abs/2110.13900

[2] https://www.nist.gov/system/files/documents/2019/07/22/2019_nist_speaker_recognition_challenge_v8.pdf