



2024 Signal Processing Cup

ROBOVOX: FAR-FIELD SPEAKER RECOGNITION BY A MOBILE ROBOT

Official Document of the 2024 Signal Processing Cup
Version 1.1 *Last updated 17 November 2023*

Official Sponsor of the IEEE Signal Processing Cup:



1. Introduction

A speaker recognition system authenticates the identity of claimed users from a speech utterance. For a given speech segment called enrollment and a speech segment from a claimed user, the speaker recognition system will determine automatically whether both segments belong to the same speaker or not. The state-of-the-art speaker recognition systems mainly use Deep Neural Networks (DNN) to extract fixed-length speaker discriminant representations called speaker embeddings. The decision to accept or reject a speaker will be made by comparing speaker embeddings. The DNN-based speaker verification systems perform well in general, but there are some challenges that reduce their performance dramatically. Far-field speaker recognition is among the well-known challenges facing speaker recognition systems. The far-field challenge is intertwined with other variabilities such as noise and reverberation. Two main categories of speaker recognition systems are text-dependent speaker recognition and text-independent speaker recognition. In a text-dependent speaker recognition system, the speaker's voice is recorded from predefined phrases, while, in text-independent speaker recognition, there is no constraint on the content of the spoken dialogue. The task of the IEEE Signal Processing Cup 2024 is text-independent far-field speaker recognition under noise and reverberation for a mobile robot.

2. Competition Details

2.1 Task Description

The Robovox challenge is concerned with doing far-field speaker verification from speech signals recorded by a mobile robot at variable distances in the presence of noise and reverberation. Although there are some benchmarks in this domain such as VoiCes and FFSVC, they don't cover variabilities in the domain of robotics such as the robot's internal noise and the angle between the speaker and the robot. The VoiCes dataset is replayed speech recorded under different acoustical noises. A main drawback of the VoiCes is that it was recorded from played signals whereas our dataset is recorded with people speaking in noisy environments. The FFSVC is another far-field speaker recognition benchmark. However, these benchmarks helped the community significantly, we are introducing a new benchmark for far-field speaker recognition systems in order to address some new aspects. Firstly, our goal is to perform speaker recognition in a real application for the domain of mobile robots. In this domain, there are other variabilities that have not been addressed in previous benchmarks: the robot's internal noise and the angle between the speaker and the robot. Furthermore, the speech signal has been recorded for different distances between the speaker and the robot. In the proposed challenge the following variabilities are present:

- Ambient noise leading to low signal-to-noise ratios (SNR): The speech signal is distorted with noise from fans, air conditioners, heaters, computers, etc.
- Internal robot noises (robot activators): The robot's activator noise reverberates on the audio sensors and degrades the SNR.
- Reverberation: The phenomena of reverberation due to the configuration of the places where the robot is located. The robot is used in different rooms with different surface textures and different room shapes and sizes.
- Distance: The distance between the robot and speakers is not fixed and it is possible for the robot to move during the recognition.
- Babble noise: The potential presence of several speakers speaking simultaneously.
- Angle: The angle between speakers and the robot's microphones In this challenge, two tracks will be proposed:
 - Far-field single-channel tracks: In this task, one channel is used to perform the speaker verification. The main objective is to propose novel robust speaker recognition pipelines to tackle the problem of far-field speaker recognition in the presence of reverberation and noise.

- Far-field multi-channel tracks: In this task, several channels are used to perform speaker verification. The main objective is to develop algorithms that improve the performance of multi-channel speaker verification systems under severe noise and reverberation.

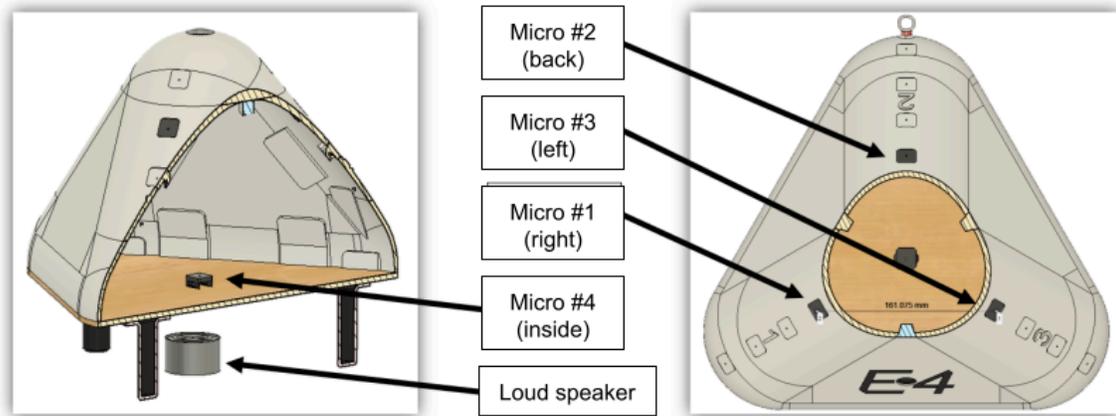


Fig. 1. Robovox (E4): a mobile robot

2.2 Dataset Description

In this challenge, we introduce a novel benchmark that complements previous works and aims at fostering research in farfield single-channel and multi-channel speaker verification. We will propose an evaluation benchmark in which the voice dialogues are recorded by a robot in various acoustic conditions. The Robovox is a French corpus recorded by a mobile robot (E4) in the framework of the ANR project RoboVox. The robot is equipped with a speaker recognition system in noisy environments. There are three microphones on the angles of the robot (Micro #1, Micro #2, Micro #3). The fourth microphone is embedded inside the robot (Micro #4). Another microphone is used as a ground truth microphone (Micro #5). The ground truth microphone is close to the mouth of the speaker. The microphones are depicted in [Figure 1](#). The speech files are recorded from conversations between Robovox and speakers. Robovox utilizes a loudspeaker positioned beneath the robot to articulate its utterances. The dataset includes 78 speakers. The number of conversations between the robot and the speakers is between 24 and 36 which results in 2219 conversations. In each conversation, there are 5 dialogues (speaker turns) on average. Therefore, the total number of recorded dialogues is $\approx 11,000$. The average length of each dialog is 3.6 seconds. Each recording has 8 channels. The channel information is as follows:

- Channel 1 to 3: microphones on the angles of the robot;
- Channel 4: microphone embedded inside the robot;
- Channel 5: ground truth microphone which is close to the speaker;
- Channel 6: Unused channel;
- Channel 7 and channel 8: Are robots dialogues turns.

It is worth noting that having a clean signal recorded by Channel 5, enables us to have the best-expected baseline system and allows us to know the amount of performance degradation for far-field microphones. An example of a recorded signal spectrum is depicted in [Figure 2](#):

The files are recorded from different distances in different acoustical environments with the main following settings:

- 1m, 2m and 3m: Distance of the speaker from the robot: respectively 1, 2, and 3 meters.
- hall, open space, small room (open/close) and medium room (open/close): The sessions are recorded in the different rooms/environments with the door open or close in meeting rooms.
- wall, center, and corner: The robot is placed close to a wall (or window), in the center of the room, or in the corner respectively. Severe reverberation can be spotted.
- calm or noisy: Level of noise in the environment.

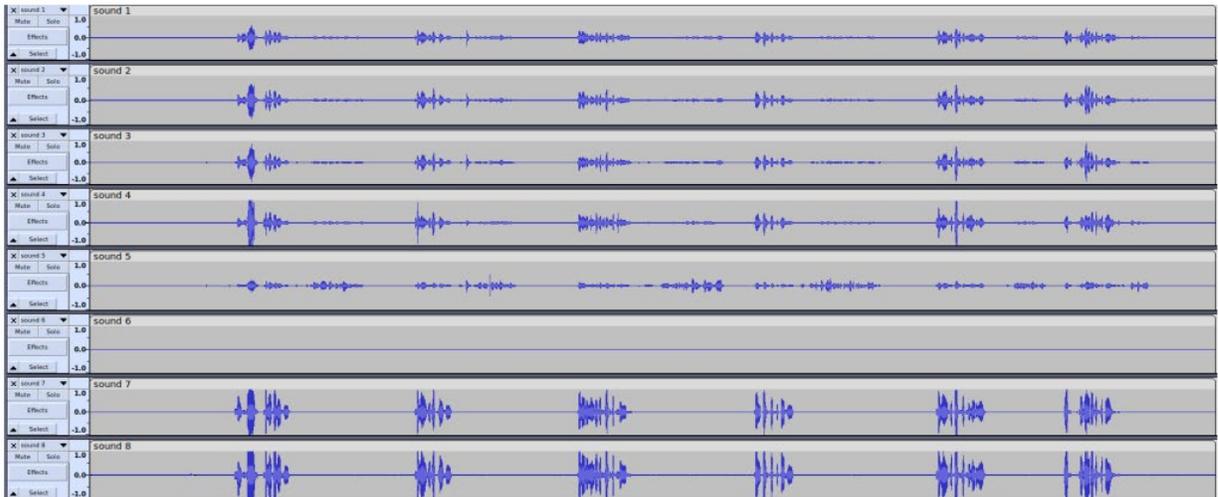


Fig. 2. An example of a recorded speech signal by Robovox

2.3 License

This audio database is made available under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International License. This means that you are free to share (copy, distribute, and transmit the work) and remix (adapt the work), as long as you credit the original authors, do not use this work for commercial purposes, and share any derivative works under a similar license. For more information, please visit the [Creative Commons website](https://creativecommons.org/licenses/by-nc-sa/4.0/).

2.4 Evaluation Protocols

3.4.a. Task1: Far-field single-channel tracks

In the Far-field single-channel track, two channels will be used. The best channel (i.e. channel 5) will be used for enrollment. For each speaker, three dialogues are used as enrollment. If a session is chosen as enrollment the remained dialogues in that session will not be used in the test. The remaining part of the dialogues are used as the test. The most challenging channel (i.e. channel 4) is used for the test.

3.2.2. Task 2: Far-field multi-channel tracks

In this task, all channels can be used for enrollment. Channel 5 is not allowed to be used as the test. Similar to the first task a few dialogues are used in enrollment and the rest of the dialogues (Except dialogues from sessions used in the enrollment) will be used as the test.

3.4.b. Task 2: Far-field multi-channel tracks

In this task, all channels can be used for enrollment. Channel 5 is not allowed to be used as the test. Similar to the first task a few dialogues are used in enrollment and the rest of the dialogues (Except dialogues from sessions used in the enrollment) will be used as the test.

2.5 Data organization

The data is organized as follows:

```
main directory/
  data/
    samples/
    single-channel/
      enrollment/
      test/
      single-channel-trials.trl
    multi-channel/
      enrollment/
      test/
      multi-channel-trials.trl
  docs/
    readme
    robovox-challenge.pdf
```

In the *samples* directory you can find speech sample files from two speakers. You are allowed to use the sample files for training your model or use their information for other tasks such as data simulation or data augmentation. For both single-channel and multi-channel tasks the preprocessing is done on the recorded sessions. The *data/single-channel/test* directory contains the noisy dialogues extracted from channel 4 which is the worst case. The *data/single-channel/enrollment* contains enrollment files for single-channel tasks coming from the ground-truth microphone which has the highest quality among other channels. The *data/multi-channel/enrollment* and *data/multi-channel/test* include the enrollment and test files for multi-channel tasks respectively. The single-channel trials will be found in *single-channel/single-channel-trials.trl* and multi-channel trials are in *multi-channel/multi-channel-trials.trl* file. In order to get the dataset the participants should register on [Codebench](https://www.codabench.org/competitions/1588/) (<https://www.codabench.org/competitions/1588/>) and get the data from the files tab.

3. Evaluation Metrics

The decision cost function (DCF) is used as a primary evaluation metric which is defined as:

$$C_{DET}(\theta) = C_{Miss} \times P_{Miss|Target}(\theta) \times P_{Target} + C_{FA} \times P_{FA|Nontarget}(\theta) \times (1 - P_{Target}) \quad (1)$$

where: ϑ is a decision threshold, C_{Miss} is the cost of false rejection, C_{FA} is the cost of false acceptance, P_{Target} is the prior probability of target speakers. We consider two sets of parameters to calculate $C_{DET}(\theta)$. The parameters are listed in [Table 1](#). The first line shows the parameters using the robot during the day when the probability of having a target speaker is high and the cost of accepting a non-authorized person is less in comparison to night. The last row shows the parameters for using the robot during the night with a low P_{Target} and high C_{FA} .

$C_{DET}(\theta)$	P_{Target}	C_{Miss}	C_{FA}
Day	0.8	1	20
Night	0.01	10	100

Table 1. C_{DET} parameters

The final score will be calculated as the average of C_{DET} for two sets of parameters listed in [Table 1](#).

The Equal Error Rate (EER) will be used as a secondary evaluation metric that is defined as an operating point θ , where $FAR = FRR$. False Acceptance Rate (FAR) is the number of False Acceptance (FA) errors divided by the total number of non-target trials (imposters). And, the False Rejection Rate (FRR) is the number of False Rejection (FR) errors divided by the total number of legitimate trials.

4. Participants submissions

4.1 Results

We provided a scoring and ranking tool on [Codebench](#). The scores will be calculated automatically. The participants should make a valid submission for the single-channel task. A valid submission should contain a score for all trials. A submission with missed trials will not be considered.

! Put the scores in a single text file with this name: yourID.txt. Then put it in a compressed *.zip file and submit it from [CodeBench](#)

The format of a submission is as follows:

```
enroll ID <TAB>test ID <TAB>score
```

where *enroll ID* is the id of an enrollment file, test ID is the id of a test file, and score is the cosine distance between test and enrollment files.

A sample of the result file is shown here:

```
spk_72 93dfe359e7 0.6813565
spk_12 4426dfd7fe 0.7223774
spk_23 bb6235bfce 0.9346804
```

! Note: The multi-channel track will not be considered for ranking the participants. We are proposing this track to foster the research in multi-channel speaker recognition.

4.2 Technical Report

Each group should submit a technical report that includes:

- **System description:** This part should describe the front-end such as hand-made or DNN-based features, the VAD, the speaker embedding extraction architecture, etc.
- **Sample files:** Along with the evaluation protocols recorded sessions from two speakers are provided as sample files. The participants can use these files to adapt their system or use the available information for data augmentation etc. In the case of using these files report the way you exploited them.

! The technical report is obligatory for introducing the winners. Please send your report before final submission date (5 February 2024) in a PDF format to the provided email of the contact person with this subject: yourCodebenchID SPCUP2024 report

5. SPS Student Competition Terms & Conditions

IMPORTANT: All team members must fully read and agree to the terms in this SPS Student Competition Terms and Conditions document in order to be eligible in any SPS Student Competition. By checking the agreement checkbox during team registration, all team members agree to all of the Terms & Conditions mentioned in this document.

IMPORTANT: [Judges and Team Supervisors] The IEEE Conflict of Interest form must be completed before participating in the competition. Full information is included in this document on pages 2 and 3. The Conflict of Interest form can be completed at the following link: <https://www.ieee.org/about/compliance/conflict-of-interest/coiandpob.html>

5.1 Team Formation and Eligibility

5.1.a. Team Composition

- Each team **MUST** be composed of: (i) One faculty member (the Supervisor) and (ii) At least 3 but no more than 10 undergraduates; *Optionally* (iii) At most one graduate student (the Tutor).
 - At least three of the undergraduate team members must be SPS student members at time of team registration.
- Further definitions of each team position are as follows:
 - Faculty (Supervisor): A faculty member or someone with a PhD degree employed by the university. Postdocs and research associates are not considered as faculty members.
 - Graduate Student (Tutor): A graduate student is a student having earned at least a 4-year University degree at the time of submission. *Please note: Tutors are not eligible to receive travel grants or prize money.*
 - Undergraduate: An undergraduate student is a student without a 4-year degree.
- Team members cannot be changed after the team registration deadline.
- At least one undergraduate team member must be in attendance (in-person) of the final round of the competition held at the respective conference (ICASSP or ICIP) to present the team's work.*
- Students receiving the travel grant and prize payments **MUST** be active SPS members at time of team formation. Signal Processing Society membership can be added [here](#).

***Important notice:** Upon registering a team for the competition, the team must commit to at least one undergraduate team member representing the team by attending the physical competition and participating in the final round of the competition at the physical conference. Should a team not be able to participate physically (in-person) in the final round of the competition held at the respective conference (ICASSP or ICIP) for any reason, at any point in the competition, then the team must notify SPS Staff and organizers immediately. This will likely result in the team being ineligible to continue in the competition, therefore forfeiting their position in the competition. Teams must make every effort to attend the final round at the conference; visa issues *may* be an exception. *If all team members are unable to obtain visas, please be prepared to present proof of visa process,*

communication to obtain visa, as well as a visa denial. All eligibility decisions are at the discretion of the SPS Student Services Committee and competition organizers.

Should a team be disqualified or forfeit their finalist position for any reason, the next team selected by the organizers may be contacted to compete in the final round, following the same rules listed above.

5.1.b Team ineligibility (Further clarification)

Specific team **ineligibility**, in addition to the above. Any of these criteria will result in the team being disqualified/ineligible to continue in the competition:

- Teams that are composed with 50% or more of its members being students who have previously participated on a finalist team of another SPS competition within the last calendar year are not eligible.
- Teams with the exact same member composition of a previously placed team in the top 3 of another SPS competition within the last calendar year are not eligible.
- Any team members who have placed in the top three teams of any SPS competition held during the previous conference and/or calendar year, i.e. a member from one of the 3 finalist teams of the 2023 SP Cup or 5-MICC (at ICASSP) will not be eligible to participate in the 2023 VIP Cup or 5-MICC (at ICIP).
- Team members cannot simultaneously participate in more than one competition at the same time.
- Team members cannot participate on more than one team at the same time.

5.2 Final Round Judging Criteria for SPS Competitions

The judging for the final phase of the competition held live at the conference will be based on five equally weighted criteria, unless other specified by the competition organizers in the Official SP Cup Document/Call for Participation. Each of the three finalist teams will be scored on the five criteria and the team with the highest score will place 1st, the team with the second highest score will place 2nd, and the team with the third highest score will place 3rd in the competition.

The five equally weighted criteria are:

1. Innovation of the proposed approach
2. Performance of the first stage competition (by choosing the best submission, score as indicated on the website)
3. Performance of the last submission (second phase held live at the conference) separately on the dataset(s)
4. Quality and clarity of the final report
5. Quality and clarity of the presentation

Each criterion is scored with a 1, 2, or 3; the best team in each criterion will receive 3 points, the second best team will receive 2 points, and the third best team will receive 1 point. The final winning rankings will be based on the highest points awarded from the five criteria during judge deliberations at the end of the competition. Final rankings are ultimately decided by the judges, at their discretion.

5.3 Judge & Team Supervisor Participation and Conflict of Interest

Any judge or team supervisor participating in the competition must sign a Conflict-of-Interest Form agreeing to the following key points. Full information is on the Conflict-of-Interest Form.

Conflict of Interest concerns shall be disclosed and addressed in accordance with IEEE Policies 9.9 A, B, and C and IEEE Policy 4.4.H. - Eligibility and Process Limitations. Judges involved at any stage of the team rankings/scoring process for an SP competition shall be ineligible to judge/vote on the outcome of team rankings for the competition in which the conflict exists. Any real and perceived conflict of interest shall be

avoided. Conflict of interest shall be defined as any relationships, professional or otherwise, that can affect impartiality and objectivity. Such relationships include, but are not limited to the below list. This list also applies

- a. faculty supervisor/student,
- b. faculty supervisor/post-doc,
- c. manager/employee,
- d. shared institutional affiliations,
- e. recent (less than five years) research collaborations or joint authorship,
- f. judge/team supervisor
- g. In the case of a conflict of interest, the judge should neither listen to nor speak in the discussion and should not vote on the team scoring/ranking process.

In our SPS Policies and Procedures (<https://signalprocessingsociety.org/volunteers/policy-and-procedures-manual>)

(Also mentioned above) **The IEEE Conflict of Interest form must be completed before participating in the competition.** The Conflict of Interest form can be completed at the following link: <https://www.ieee.org/about/compliance/conflict-of-interest/coiandpob.html>

Conflict Resolution Process

The Society leadership will create an ad hoc committee to handle each matter requiring conflict resolution.

1. **Composition.** The composition of each ad hoc committee will include area experts. The experts should be chosen based on mediation experience or subject area experience. All members of the ad hoc committee should be non-conflicted, e.g., no prior involvement in the situation, no collegial work relationship, etc. The committee may be augmented with the agreement of all members of the ad hoc committee. The committee will select its own chair.
2. **Process.** During the first meeting of the ad hoc committee, the committee shall create a timeline detailing the conflict resolution process, as well as determine any operational rules for the ad hoc committee's operation (e.g., length of final report; length of statement of dissent, etc.) The individual who brought the conflict matter forward shall be informed of the timeline. All discussions and information presented to the ad hoc committee shall be handled in a confidential manner.

Decisions need not be unanimous; final outcomes may be determined by majority vote of the membership of the ad hoc committee. Dissenting members may include their dissenting opinion as part of the report; the length of such dissent will be determined as part of the committee's operational rules.

After the ad hoc committee has determined its final ruling, the ad hoc committee chair shall be responsible for preparing a short report documenting the committee's findings. The report shall be provided to the individual who brought the conflict matter forward.

3. **Appeal.** If the individual who brought the conflict matter forward feels that the matter has not been adequately resolved by the ad hoc committee at the Society level, the individual may escalate the matter further to TAB or IEEE. The ad hoc committee report shall be shared with TAB and/or IEEE.

5.4 Prizes for Finalists

The three teams with the highest performance in the open competition based on the judging criteria will be selected as finalists and invited to participate in the final competition at ICASSP. The champion team will receive a Grand Prize of \$5,000. The first and the second runner-up teams will receive a prize of \$2,500 and \$1,500, respectively, in addition to travel grants and complimentary conference registrations for up to three team members.

- Up to three student members from each finalist team will be provided limited travel support to attend the conference in-person. In-person attendance of the physical conference is required for reimbursement.

- Travel grant funds are offered on a reimbursement basis of up to \$1,200 for continental travel and \$1,700 for intercontinental travel; there are no exceptions.
- Funds will be issued by way of a bank transfer after the competition via IEEE's NextGen Expense Reimbursement (Concur) tool. *Detailed information on this step will be sent via email to the finalist team members that will be receiving reimbursement.*
- Receipts are required for all expenses and must be uploaded into Concur for reimbursement.
 - Prior to claiming your travel grant award, you must submit receipts of your travel expenses that are equal to or exceed the grant amount in order to receive the full travel grant payment. Travel expenses include: air fare, train tickets, hotel, visa fees, transit, meals, and conference registration. Receipts are required for all items and must equal to or exceed the total reimbursement amount. **The receipts must show** form of payment used, name, date, and amount paid. Hotel reservation confirmations or bookings, invoices, or quotes for airfare are not acceptable receipts unless the proof of payment is also provided with the document.
 - All travel expenses must be submitted through IEEE's NextGen Expense Reimbursement (Concur) tool. *Detailed instructions on how to create an account and properly input receipts will be shared with the (up to) 3 undergraduate members from finalist teams receiving travel support.*
- Complimentary conference registration for up to three team members of finalist teams.
 - These complimentary conference registrations cannot be used to cover any papers accepted by the conference.
 - You must notify Jaqueline Rash, Jaqueline.rash@ieee.org, via email of the three team members who have been chosen to receive complimentary registration. A registration promo code will be provided.
 - VISA: **Once registered for the conference**, each individual will have the opportunity to request an invitation letter through the conference website to be used for the visa application process. This is the **only** way to receive a letter.
- The finalist teams will also be invited to join the Conference Banquet and the SPS Student Job Fair, so that they can meet and talk to SPS leaders and global experts. Please note registration for the Conference Banquet and Student Job Fair is limited and based on availability. You must add these events to your registration. If you are unable to add these events to your registration, then maximum capacity has been reached. There are no exceptions for the Conference Banquet, however, *there may be additional availability for the Student Job Fair. You can inquire by emailing Jaqueline.Rash@ieee.org.*

6. Important Dates

- Challenge announcement: November 2023
- **Team Registration Deadline:** 12 January 2024 [Register here]
- **Final Submission of Team's Work Deadline:** 5 February 2024 [Submit your team's work here]
- Announcement of 3 finalist teams: 14 February 2024
- Final competition in-person at ICASSP 2024: April 14-19, 2024

7. Registration and Important Resources

7.1 Official SP Cup Team Registration

All teams MUST be registered through the official competition registration system in order to be considered as a participating team. Teams also MUST acknowledge, agree to the SPS Student Terms and Conditions, and meet all eligibility requirements at the time of team registration as well as throughout the competition. 

Register your team for the 2024 SP Cup before 12 January 2024 and submit work before 5 February 2024 at the following  link:

7.2 Request Complimentary MATLAB

MathWorks, Inc. continued to support the IEEE SP Cup. Participating students are encouraged to download the complimentary MathWorks Student Competitions Software for use in the competition at the MathWorks' SP Cup webpage:

<https://ww2.mathworks.cn/academia/student-competitions/sp-cup.html>

Please click "Request Software" on the website, fill in the application form, and then submit the form. You will receive an email within 3-5 working days after submission. The email will inform you of the software download, installation, and activation steps if the request is approved.

*** More technical resources such as videos, examples and documentations can be found at the [MathWorks' SP Cup webpage](#).*

Complimentary Software

MathWorks provides complimentary software for this competition. If your team is participating in this competition and needs software, fill out the software request form.

[Request software](#)

8. Contacts

Competition Organizers (technical, competition-specific inquiries):

Mohammad MOHAMMADAMINI

Email: mohammad.mohammadamini@univ-avignon.fr

Tel: +33 7 80 86 21 64

SPS Staff (Terms & Conditions, Travel Grants, Prizes):

Jaqueline Rash, SPS Membership Program and Events Administrator

Email: Jaqueline.Rash@ieee.org

SPS Student Services Committee

Angshul Majumdar, Chair

Email: angshul@iiitd.ac.in

7. Sponsor

We gratefully acknowledge MathWorks, Inc. for their continued support of IEEE Signal Processing Cup. Participating students are encouraged to download the complimentary MathWorks Student Competitions Software for use in the competition.



IEEE Signal Processing Society

Signalprocessingsociety.org



Please note, any work submitted during this competition may be made available to MathWorks for commercial use.