Antoine Deleforge, Diego Di Carlo, Martin Strauss,
Romain Serizel, and Lucio Marcenaro

Audio-Based Search and Rescue With a Drone

Highlights From the IEEE Signal Processing Cup 2019 Student Competition

Increasing interest in unmanned aerial vehicles (UAVs), commonly referred to as *drones*, has occurred in recent years. Search and rescue scenarios where humans in emergency situations need to be quickly found in difficult-to-access areas constitute an important field of application for this technology. Drones have already been used by humanitarian organizations in countries such as Haiti and the Philippines to map areas after a natural disaster using high-resolution embedded cameras, as documented in a recent United Nations report [1]. Although research efforts have focused mostly on developing video-based solutions for this task [2], UAV-embedded audio-based localization has received relatively less attention [3]–[7]. However, UAVs equipped with a microphone array could be of critical help to localize people in emergency situations, especially when video sensors are limited by a lack of visual feedback due to bad lighting conditions (such as at night or in fog) or obstacles limiting the field of view (Figure 1).

This motivated the topic of the sixth edition of the IEEE Signal Processing (SP) Cup: a UAV-embedded sound-source localization (SSL) challenge for search and rescue. The SP Cup is a student competition in which undergraduate students form teams to work on real-life challenges. Each team should include 1) one faculty member (the supervisor), 2) at most, one graduate student (the mentor), and 3) at least three, but no more than ten, undergraduate students. (An undergraduate student is a student without a four-year university degree at the time of submission.) Participation in the SP Cup is open to all teams from around the world that satisfy the aformentioned eligibility criteria. The top three finalist teams were selected to present and compete at the final stage of competition, which was held at the ICASSP 2019 in Brighton, United Kingdom, on 13 May 2019. We will overview the IEEE SP Cup experience, including competition setup, technical approaches, and statistics.

UAVs equipped with a microphone array could be of critical help to localize people in emergency situations.

Drone-embedded SSL

Estimating the direction of a sound source given audio measurements from an array of two or more microphones is a long-standing research topic referred to as *sound source localization* (*SSL*) [8]. The most common approach to this problem is to estimate the sound time difference of arrival (TDOA) in a microphone pair, which can be approximately mapped to an angle of arrival when the source-to-microphones distance is large compared to the intermicrophone distance. For arrays containing more than two microphones and having known geometry, the angles of arrival of different pairs can be combined to estimate the 2D (azimuth and elevation) direction of arrival of the source in the array's coordinate frame. A large number of methods for robustly estimating TDOAs from signal pairs in the presence of noise, reverberation, or interfering sources has been developed, including generalized cross-correlation methods [9] and subspace methods [10], [11]. Alternatively, a number of machine-learning-based SSL methods have recently emerged (see [12] and [13]). However, because acquiring sufficiently large real-world data sets to train SSL models for specific arrays is very costly, most learning-based approaches rely on simulated data sets [13], which do not always generalize well to real-world conditions.

The specific task of UAV-embedded SSL comes with a number of challenges. One major issue is the noise produced by the UAV itself, generically referred

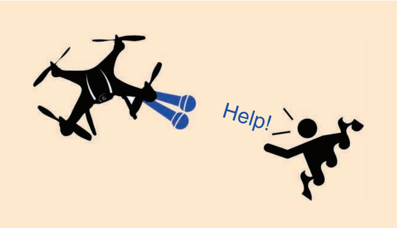

FIGURE 1. Microphones embedded in a UAV may help localize people for search and rescue in disaster areas.

Digital Object Identifier 10.1109/MSP.2019.2924687
Date of publication: 9 September 2019

1053-5888/19©2019IEEE

to as *ego noise* in robotics [14]. Due to the quickly changing speed of motors to stabilize the vehicle in the air or change its position, the noise profile is harmonic but also nonstationary. Additionally, because the microphones are mounted on the drone itself, they are very close to the noise sources, leading to high noise levels. Because of this, the signal-to-noise ratio (SNR) can easily reach –15 dB or lower, making SSL very difficult. Another factor impacting localization performance is wind noise. The wind is produced by the rotating propellers and the UAV movement in the air, and it may also occur naturally in outdoor scenarios. This wind noise has high power and is of low frequency. Thus, it easily overlaps with speech signals, which typically occur in a similar frequency range. Finally, SSL must be performed using relatively short time windows due to the fast movements of the UAV relative to potential sound sources. All of these challenges need to be tackled at the same time and nearly in real time when considering the real-world SSL application of search and rescue. On the bright side, however, UAVs may be equipped with other embedded sensors (such as gyroscopes, motor controllers, inertial measurement units, compasses, and cameras), which may provide useful additional information. In the particular case of the SP Cup, the rotational speeds of each of the drone's propellers at all times were provided along with the audio recordings, and they could optionally be used by the participants for ego-noise estimation.

Estimating the direction of a sound source given audio measurements from an array of two or more microphones is a long-standing research topic referred to as *sound source localization.*

The Drone Egonoise data set

The SP Cup data were built from a subset of the recently released Drone Egonoise (DREGON) data set [7]. The release of parts of the DREGON data set, including ground truth annotations, was delayed for the SP Cup to take place in fair conditions. The entire data set is now publicly available for download at http://dregon.inria.fr. It consists of annotated audio recordings acquired with a specifically designed 3D-printed cube-shaped eight-microphone array rigidly attached under a quadrotor UAV, as shown in Figure 2. The 8SoundsUSB microphones and sound card designed by Sherbrook University, Canada, were used. Specifications can be found at https://sourceforge.net/p/eightsoundsusb/wiki/Main_Page. The quadrotor UAV was the MK-Quadro from MikroKopter (HiSystems GmbH, Moormerland, Germany). The data set includes both static and in-flight recordings with or without the presence of a sound source emulated by a loudspeaker emitting either speech from the TIMIT data set [15] or white noise. Recordings were made inside large rooms with mild reverberation times (lower than 150 ms) and negligible background noise. The synchronized five degrees of freedom coordinates of both the UAV and loudspeaker were obtained using a Vicon motion-capture system, yielding ground-truth source-direction annotation errors of lower than 2° for the entire data set.

Tasks in the 2019 SP Cup

The goal of the competition was for teams to build a system capable of localizing a sound source based on audio recordings made with a microphone array embedded in a UAV. Teams had to use their signal processing expertise to process the audio signals to extract relevant spatial cues to estimate the direction of arrival of a speech source. Key challenges were the large amounts of noise present in the recordings due to the UAV's rotors and wind as well as the dynamics of realistic flights involving fast movements. To help noise estimation, the mean rotational speeds of each of the four propellers were provided for each localization task. The microphone array geometry and coordinate frame were also available.

FIGURE 2. The quadrotor UAV used for the SP Cup, equipped with a 3D-printed eight-microphone array. The circled areas highlight two of the microphones.

The open competition: Static task

For this task, 300 eight-channel audio recordings at 44.1 kHz and of approximately 2 s each were provided in the form of waveform audio file format (WAV) files. All recordings were obtained by adding together a clean recording of a static loudspeaker emitting random utterances from the TIMIT database [15] by an unknown (azimuth and elevation) direction in the UAV microphone array's frame and a recording of UAV noise of the same length in various flight conditions and using various SNRs, from –20 to 5 dB. The goal of this task was to retrieve the azimuth and elevation angles of the static speech source for each of the 300 recordings. A source was considered correctly localized when the great-circle distance between the estimated and ground-truth directions was lower than 10°. One point was given for each correctly localized static file, for a total of 300 points.

The open competition: Flight task

For this task, 36 eight-channel audio recordings at 44.1 kHz lasting precisely 4 s each were provided in the form of WAV files. All recordings were made during flight. In the first 16 recordings, the source was a loudspeaker emitting speech, whereas in the last 20 recordings, the source was a loudspeaker emitting white noise. A white noise source is considered easier to localize because it has a much broader frequency range than speech. The average SNR was approximately –15 dB. Although the source (loudspeaker) was static during flights, the microphone array was moving with the drone. Thus, the (azimuth and elevation) source direction in the array's frame was constantly changing over time. For each of the 4-s recordings, 15 regularly spaced time stamps were defined. The goal of this task was to retrieve the mean azimuth and elevation angles of the source within a 500-ms window centered on each of these time stamps for each of the 32 recordings. Similarly to the static task, one point was given for each correctly localized time stamp, for a total of 540 points.

The open competition—Bonus task: Data collection

On top of the 840 points that could be gained by correctly localizing all of the sources, participating teams were encouraged to send their own audio recordings obtained from one or several microphones embedded in a flying UAV. The recordings had to be made either outdoors or in an environment with mild reverberation, and they weren't to feature sound sources other than the UAV's noise or wind. A report detailing the microphones and UAV used as well as the recording conditions and including pictures of the setup and experiments had to be given with the audio files. Sixty extra points were granted to teams submitting such data. A novel UAV ego-noise data set was created from the teams' contributions to this bonus task, and it is now freely available for research and educational purposes at http://dregon.inria.fr.

The rotational speeds of each of the drone's propellers at all times were provided along with the audio recordings, and they could optionally be used by the participants for ego-noise estimation.

The baseline

A baseline method was provided for the competition. The method, implemented in MATLAB, is based on the open source Multi-channel BSS Locate toolbox (http://bass-db.gforge.inria.fr/bss_locate) and available at https://github.com/Chutlhu/SPCUP19. Although the baseline as provided used the steered-response power phase-transform (SRP-PHAT) method as described in [16], the MBSS Locate toolbox implements 12 different source localization methods, which are detailed in [17] and were also sometimes used by participants. The method chosen as the baseline ranked among the best-performing methods in the single-source localization tasks of the recent IEEE acoustic source localization and tracking challenge [18].

Final competition

The three highest-scoring teams from the open competition stage were selected as finalists invited to compete in the final competition at ICASSP 2019. Teams gave a 5-min presentation of their method followed by 5-min of questions in front of a jury composed of SP Cup organizers and a MathWorks representative. Presentations were marked by the jury according to clarity, content, originality, and answers to the questions. Then, the teams were given previously unseen recordings made with the same UAV as in the open competition, i.e., 20 static speech recordings of roughly 2 s each and one long in-flight speech recording of 20 s. The average SNRs for both tasks were similar to the lowest SNRs encountered during the open competition—approximately –17 dB. The teams had 25 min to run their methods and provide results for these tasks in the same format as in the open stage. Results were evaluated on the spot, and a global score was calculated for each team so that the presentation, the static task, and the flight task each accounted for one-third of the total.

Competition data

The data of both the open and final stages of the 2019 SP Cup as well as corresponding ground-truth result files and MATLAB scripts can be found at http://dregon.inria.fr/datasets/signal-processing-cup-2019/.

2019 SP Cup statistics and results

As in previous years, the SP Cup was run as an online class through the Piazza platform, which allowed for continuous interaction with the teams. In total, 207 students registered for the course, and the number of contributions to the platform was more than 150. An archive of the class is available at https://piazza.com/ieee_sps/other/spcup2019/home. We received complete and valid submissions from 20 eligible teams from 18 different universities in 11 countries across the globe: India, Japan, Brazil, South Korea, New Zealand, China, Germany, Bangladesh, Australia, Poland, and Belgium. The teams had four to 11 members for a total of 132 participants.

Figure 3 summarizes the scores obtained by the 20 participating teams and baseline for all of the open-competition subtasks. Remarkably, 12 teams strictly outperformed the already strong baseline in their overall score (excluding bonus points). As can be seen, near-perfect scores were obtained by the best-performing teams in the static speech and in-flight broadband tasks, with more than 95% of correctly localized sources. In contrast, the in-flight speech task, which formed the heart of the competition and was the closest one to the motivational search and rescue scenario, proved to be extremely challenging. For this task, only one timestamp out of 240 was correctly localized by the baseline. In fact, only nine teams succeeded in localizing more than 5% of the timestamps for this task, while the winning team achieved a stunning 65% score. In addition to the mandatory localization result files, 12 of the teams sent their own UAV recordings for the bonus task, yielding a unique and valuable data set (see the section "The Open Competition—Bonus Task: Data Collection").

Highlights of the technical approaches

In this section, we provide an overview of the methods employed by the top 12 teams that outperformed the baseline. Proposed methods were generally made of at least two components: 1) a preprocessing stage aimed at reducing the noise in the observed signals and 2) an SSL stage. The most popular noise reduction methods used were the multichannel Wiener filter or single channel variations of it (see [19] for a review). These approaches require estimates of the noise statistics, which were obtained using many different techniques. Half of the teams used the motor speeds provided along with the audio files to do so via some form of interpolation among the corresponding individual motor recordings available as development data. Others used voice activity detection to isolate noise-only parts and estimate the statistics on them, or they made use of recursive averaging. In addition to Wiener filtering, several teams used various bandpass filters to reduce the impact of wind noise. Notable alternatives to Wiener filtering included noise reduction methods based on nonnegative matrix factorization or deep neural networks. One team also used spatial filtering to reduce noise in the directions of the four rotors based on the provided UAV geometrical model. Two of the teams developed methods to adaptively remove microphone pairs for which the noise was too important. Many teams combined several of the aforementioned strategies to further reduce the noise.

The goal of the competition was for teams to build a system capable of localizing a sound source based on audio recordings made with a microphone array embedded in a UAV.

For SSL, most of the teams built on the SRP-PHAT method implemented in the baseline. Some others used nonlinear variations of it, beamforming-based methods, or subspace methods. A number of teams used some form of postprocessing on the angular spectra provided by these methods, for instance, by ignoring regions associated to the drone's rotor directions or by clustering local minima. An approach that proved particularly successful for in-flight tasks was to smooth estimated source trajectories. This was done by using Kalman filtering or handcrafted heuristics. Overall,

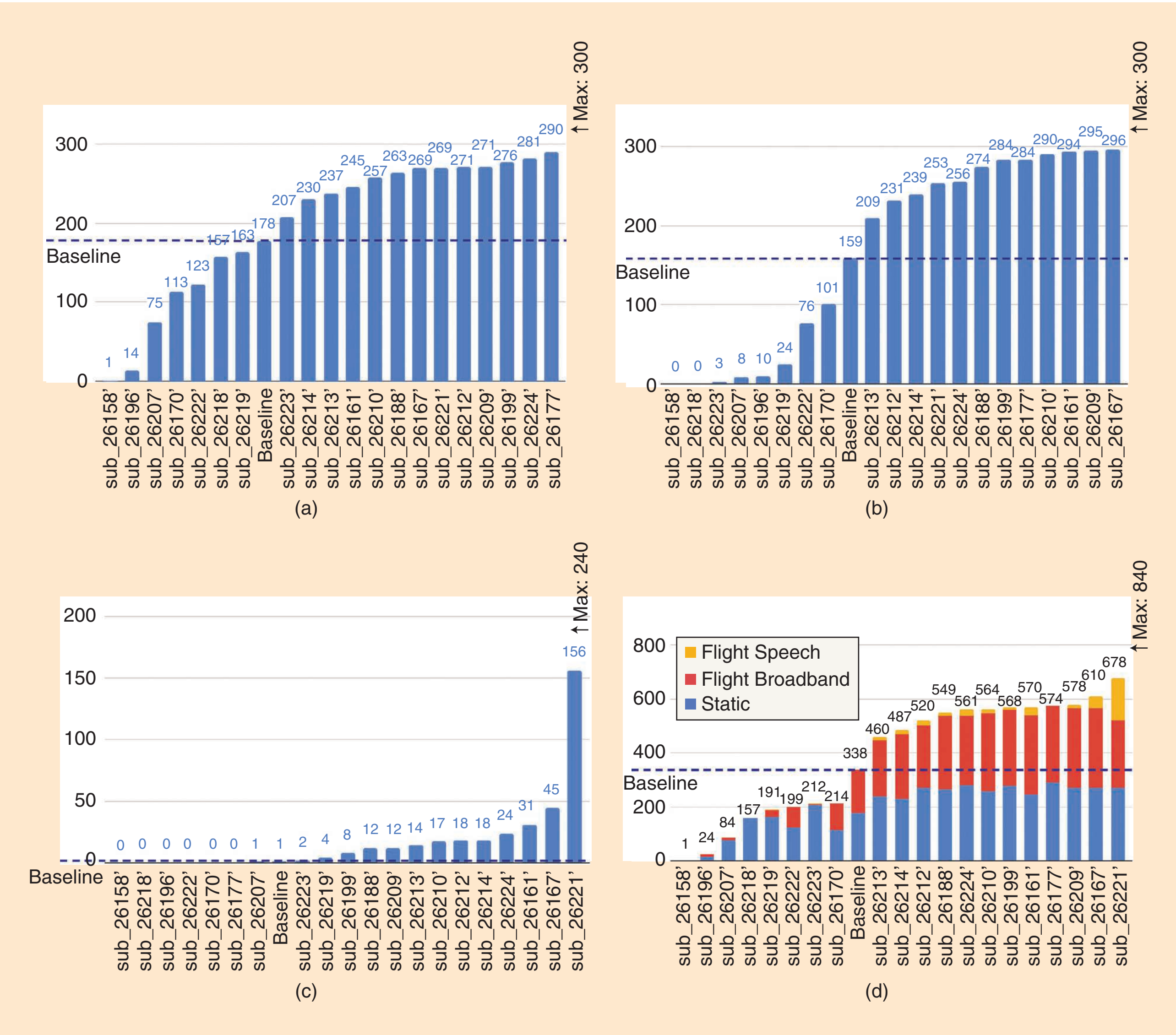

FIGURE 3. The anonymized scores of the 20 teams and the baseline for the three open-competition subtasks: (a) the static speech localization task, (b) in-flight broadband localization task, (c) the in-flight speech localization task, and (d) the total scores.

(a) (b) (c)

FIGURE 4. The members of the three finalist teams after the final competition at ICASSP 2019: (a) first place, Team AGH; (b) second place, Team SHOUT COOEE!; and (c) third place, Team Idea!_SSU.

the finalist teams proved that combining several techniques carefully designed for the task at hand was the only way to achieve good performance on the competition data. This suggests that even better results could be obtained by combining the best ideas from the different competitors.

The winning teams

In the section, we provide details about the three winning teams as well as an overview of some feedback and perspectives received from them. Team members at the final competition are shown in Figure 4.

Team AGH

- *Affiliation*: Akademia Górniczo-Hutnicza, University of Science and Technology, Kraków, Poland
- *Undergraduate students*: Piotr Walas, Mateusz Guzik, and Mieszko Fraś
- *Tutor*: Szymon Woźniak
- *Supervisor*: Jakub Gałka
- *Approach*: The team preprocessed the signals using multichannel Wiener filtering, where the noise covariance matrices were estimated by averaging across several frames as well as across the whole signals. To perform localization, the team combined estimates from the SRP-PHAT baseline and the multiple signal classification based on generalized eigenvalue decomposition [11] methods via *k*-means clustering in the angular-spectrum domain. Angular spectra were presmoothed using a max filter. Finally, a Kalman filter was employed to smooth out estimated trajectories in flight tasks.

Opinions

- "Leading a group of undergrads was a challenging as well as rewarding task. It gave me a perspective on how hard it is to efficiently organize research work in team, even though the team was small in number. During the competition, I especially enjoyed discussing out-of-the-box ideas of undergrads and studying state of the art alongside them. The tricky part of this competition was to figure out how to evaluate accuracy of tested methods, since without ground truth, you never know. On the other hand, the most exciting moments were the announcement of the results of the first phase and incontestably taking part in the final competition at ICASSP. This kind of competition gives an excellent opportunity for undergraduates to try their hands at solving challenging research problem."

  —Szymon Woźniak

- "I chose to join the Signal Processing Cup competition because I searched for a project outside of regular studies that would allow me to develop myself in the field of signal processing. During the work, I got to develop state-of-the-art SSL methods and also had a chance to experience working in a great team. I enjoyed the most the moments when we got some improvements after testing a new idea. Unfortunately, due to the lack of development data, we often had to rely on our intuition in deciding between two solutions, which was the hardest part of the competition. I think those types of events are a great chance for students to get an idea of how the scientific community works and meet like-minded people from around the world."

  —Mateusz Guzik

- "I chose to participate in SP Cup as I saw the opportunity to create a solution which could be potentially used for helping others. During the competition, the most enjoyable and exciting part was studying state-of-the-art algorithms, merging them into one solution, and observing the results. The difficulty of the competition itself was connected to the lack of development data, which made challenging to choose between different solutions. After all, I believe that the most important of taking part in this competition was the knowledge and hands-on experience which we gained."

  —Piotr Walas

Team SHOUT COOEE!

- *Affiliation*: The University of New South Wales, Kensington, Australia
- *Undergraduate students*: Antoni Dimitriadis, Alvin Wong, Ethan Oo, Jingyao Wu, Prasanth Parasu, Qingbei Cheng, Hayden Ooi, and Kevin Yu
- *Supervisor*: Vidhyasaharan Sethu
- *Approach*: The team preprocessed the signals using multichannel Wiener filtering, where the noise covariance matrices were estimated from linear combinations of the provided individual motor recordings, weighted according to the current propeller's

speed. A nonlinear generalized cross-correlation (GCC-NONLIN) method [17] was used to localize the sound source. And, for flight tasks, source trajectories were smoothed using a heuristic method inspired by the Viterbi algorithm.

Opinions

- "I learned a lot from this SP Cup competition, from how directions of arrival can be determined using signal processing techniques to how a Wiener filter can be applied to reduce the noise in recordings. Furthermore, I learned the importance of testing and validation and how it can be utilized to evaluate the effectiveness of strategies as well as determine optimal parameters to produce an algorithm that is accurate and robust. It was an intellectually stimulating and challenging experience. I really enjoyed doing research on various strategies that could be employed in producing more accurate sound source localization results. It was always exciting whenever new strategies developed from our research led to improved performance of our system. I chose to join the competition as I have a passion for signal processing and saw this competition as an opportunity to develop my signal processing skills. Furthermore, I believed I would gain a deeper understanding of how I could apply signal processing methods and techniques to solve practical, real-world problems."

  —Prasanth Parasu

- "The whole experience has been unlike any other that I have been a part of and was very much worth the time spent on the competition. Much was learned during the SP Cup, including the importance of teamwork, clear communication, and (particularly in our team's case) running programs on multiple computers to ensure that we safeguard against unforeseen problems. The UNSW team were all collaborative and supportive of each other, and we have grown closer as a result. The competition gave us the opportunity to challenge ourselves intellectually and gain knowledge and experience that will serve us well in the future. I'd like to thank my team members for being so awesome and particularly our team coordinator, who introduced us to the competition and supported us throughout the whole adventure."

  —Ethan Oo

- "It was great to work in the team SHOUT COOEE! and compete with other brilliant teams all over the world. I love the idea of solving real-world problems. It's challenging and also attractive. Thanks to the SP Cup, I gained a new understanding of speech processing. It provided a good opportunity to learn about the multichannel Wiener filter incorporated with acoustic noise statistics of the drone. We also had chances to research and play around with different DOA estimation algorithms and seek for the best. It's an exciting and unforgettable experience."

  —Jingyao Wu

Teams had to use their signal processing expertise to process the audio signals to extract relevant spatial cues to estimate the direction of arrival of a speech source.

Team Idea!_SSU

- *Affiliation*: Soongsil University, Seoul, South Korea
- *Undergraduate students*: Donggun Lee, Myeonghun Jeong, Minjae Park, Youngjae Lee, and Jinyoung Son
- *Tutor*: Beomhee Jang
- *Supervisor*: Sungbin Im
- *Approach*: The team preprocessed the signals using a combination of single-channel speech enhancement techniques and multichannel Wiener filtering, for which noise statistics were estimated from noise-only segments using voice-activity detection. Wind noise was also reduced by cutting frequencies below 100 Hz. The SSL methods used were SRP-PHAT for the static task and GCC-NONLIN for the flight tasks. To reduce outliers on flight tasks, the team used a two-step procedure: 1) compute a global source direction on a 4-s segment and 2) estimate directions every 250 ms on 1-s segments by limiting the angular search space around the global estimated direction.

Future steps and the upcoming SP Cup

The SP Cup's organizing team hopes that this edition will foster research in the emerging topic of UAV-embedded audition for search and rescue, notably because of its unique data set, which is now publicly available. Participants of the 2019 SP Cup as well as other researchers in this field are encouraged to submit their work to the upcoming special issue of *EURASIP Journal on Audio, Speech, and Music Processing* with the topic "Advances in Audio Signal Processing for Robots and Drones." The submission deadline is 1 December 2019. For details, please visit https://asmp-eurasipjournals.springeropen.com/call-for-papers--advances-in-audio-signal-processing-for-robots-.

The seventh edition of the SP Cup will be held at ICASSP 2020. The theme of the 2020 competition will be announced in September. Teams who are interested in the SP Cup competition may visit https://signalprocessingsociety.org/get-involved/signal-processing-cup. In addition to the SP Cup, the IEEE Signal Processing Society (SPS) also organizes the Video and Image Processing (VIP) Cup. The third edition of the VIP cup will be held at the 2019 IEEE International Conference on Image Processing, in Taipei, Taiwan, on 22–25 September 2019. The theme of this edition is "Activity Recognition From Body Cameras." For details, visit https://signalprocessingsociety.org/get-involved/video-image-processing-cup.

Acknowledgment

The organizers of the 2019 SP Cup would like to express their utmost gratitude to all who made this adventure a reality, including, but not limited to, the participating teams, the local organizers,

and the IEEE SPS Membership Board. In addition, great appreciation goes to MathWorks and its representative Kirthi Devleker, who came to the final competition as a member of the jury. Since its inception, the SP Cup has received generous support from MathWorks, the maker of the popular MATLAB and Simulink platforms. MathWorks kindly provided funding support to the SP Cup, including travel grants and monetary prizes for the finalists. Finally, special thanks go to Pol Mordel, Victor Miguet, Vincent Drevelle, and François Bodin from the Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France), without whom obtaining such valuable UAV-embedded recordings would not have been possible.

Authors

Antoine Deleforge (antoine.deleforge@inria.fr) received his engineering B.Sc. degree in 2008 and his research M.Sc. degree in computer science and mathematics in 2010 from the École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble, France, and the Université Joseph Fourier, Grenoble, France, and his Ph.D. degree in computer sciences and applied mathematics from the University of Grenoble. He is a tenured research scientist with INRIA Nancy-Grand Est, France, with the team MULTISPEECH. He serves as a member of the IEEE Audio and Acoustics Signal Processing Technical Committee (AASP TC) and the IEEE Autonomous System Initiative (ASI). He initiated, coordinated, and chaired this edition of the Signal Processing Cup, endorsed by the AASP TC and ASI. He is a Member of the IEEE.

Diego Di Carlo (diego.di-carlo@inria.fr) is a Ph.D. student with INRIA Rennes-Bretagne Atlantique, France, with the team PANAMA. He was in charge of the competition evaluation and wrote the baseline and evaluation scripts. He is a Student Member of the IEEE.

Martin Strauss (martin.strauss@fau.de) received his B.Sc. degree in sports science in 2013 from the Technical University of Munich and his B.Sc. degree in medical engineering in 2016 from the Friedrich-Alexander University of Erlangen-Nuremberg, Germany, where he is currently a master's student. He is the main author of the Drone Egonoise data set [7], which served as a basis for the competition, and he participated in the early design of the tasks. He is a Student Member of the IEEE.

Romain Serizel (romain.serizel@loria.fr) received his M.Eng. degree in automatic system engineering from the Ecole Nationale Supérieure d'Electricité et de Mécanique, Nancy, France, in 2005, his M.Sc. degree in signal processing from the Université Rennes 1, France, in 2006, and his Ph.D. degree in engineering sciences from the Katholieke Universiteit Leuven, Belgium, in 2011. He is an associate professor at Université de Lorraine, France, doing research with Laboratoire Lorrain de Recherche en Informatique et Ses Applications with the team MULTISPEECH. He contributed to the Signal Processing Cup competition design and evaluation and presided over the final jury. He is a Member of the IEEE.

Lucio Marcenaro (lucio.marcenaro@unige.it) received his degree in electronic engineering and his Ph.D. degree in computer science and electronic engineering from the University of Genova, Italy, in 1999 and 2003, respectively. He is an assistant professor in the Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture, University of Genoa, Italy. He chairs the Student Service Committee of the IEEE Signal Processing Society, supporting the Signal Processing Cup. He is a Member of the IEEE.

References

[1] D. Gilman, "Unmanned aerial vehicles in humanitarian response," United Nations Office for the Coordination of Humanitarian Affairs, 2014. [Online]. Available: https://docs.unocha.org/sites/dms/Documents/Unmanned%20Aerial%20Vehicles%20in%20Humanitarian%20Response%20OCHA%20July%202014.pdf

[2] L. Lopez-Fuentes, J. van de Weijer, M. González-Hidalgo, H. Skinnemoen, and A. D. Bagdanov, "Review on computer vision techniques in emergency situations," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17,069–17,107, 2018. doi: 10.1007/s11042-017-5276-7.

[3] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *Proc. 2012 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 4737–4742.

[4] T. Ohata, K. Nakamura, T. Mizumoto, T. Taiki, and K. Nakadai, "Improvement in outdoor sound source detection using a quadrotor-embedded microphone array," in *Proc. 2014 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 1902–1907.

[5] K. Hoshiba, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, et al., "Design of UAV-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, article no. 2535, 2017. doi: 10.3390/s17112535.

[6] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, "Tracking a moving sound source from a multi-rotor drone," in *Proc. 2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 2511–2516.

[7] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "DREGON: Dataset and methods for UAV-embedded sound source localization," in *Proc. 2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*. doi: 10.1109/IROS.2018.8593581.

[8] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics Autonomous Syst.*, vol. 96, pp. 184–210, Oct. 2017. doi: 10.1016/j.robot.2017.07.011.

[9] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976. doi: 10.1109/TASSP.1976.1162830.

[10] S. Argentieri and P. Danes, "Broadband variations of the MUSIC high-resolution method for sound-source localization in robotics," in *Proc. 2007 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 2009–2014.

[11] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, "Intelligent sound source localization for dynamic environments," in *Proc. 2009 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 664–669.

[12] A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound-source separation and localization on binaural manifolds," *Int. J. Neural Syst.*, vol. 25, no. 1, p. 1440003, 2015. doi: 10.1142/S0129065714400036.

[13] C. Gaultier, S. Kataria, and A. Deleforge, "Vast: The virtual acoustic space traveler dataset," in *Proc. Int. Conf. Latent Variable Analysis and Signal Separation*, 2017, pp. 68–79.

[14] H. W. Löllmann, H. Barfuss, A. Deleforge, S. Meier, and W. Kellermann, "Challenges in acoustic signal enhancement for human-robot communication," in *Proc. 11th ITG Symp. Speech Communication*, 2014, pp. 1–4.

[15] J. S. Garofolo, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium, 1993. [Online]. Available: https://catalog.ldc.upenn.edu/LDC93S1

[16] R. Lebarbenchon, E. Camberlein, D. di Carlo, C. Gaultier, A. Deleforge, and N. Bertin, Evaluation of an open-source implementation of the SRP-PHAT algorithm within the 2018 LOCATA challenge. 2018. [Online]. Available: https://arxiv.org/abs/1812.05901

[17] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Process.*, vol. 92, no. 8, pp. 1950–1960, 2012. doi: 10.1016/j.sigpro.2011.09.032.

[18] H. W. Löllmann, C. Evers, A. Schmidt, H. Mellmann, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge data corpus for acoustic source localization and tracking," in *Proc. 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop*, pp. 410–414.

[19] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 4, pp. 692–730, 2017. doi: 10.1109/TASLP.2016.2647702.

SP