Odyssey 2020 The Speaker and Language Recognition Workshop
1-5 November 2020, Tokyo, Japan

The VOiCES from a Distance Challenge 2019: Analysis of Speaker Verification Results and Remaining Challenges

Mahesh Kumar Nandwana[1], *Michael Lomnitz*[2], *Colleen Richey*[1], *Mitchell McLaren*[1], *Diego Castan*[1], *Luciana Ferrer*[3], *Aaron Lawson*[1]

[1]Speech Technology and Research Laboratory, SRI International, Menlo Park, California, USA
[2]Lab41, In-Q-Tel, Menlo Park, California, USA
[3]Instituto de Investigacón en Ciencias de la Computación, UBA-CONICET, Argentina
mahesh.nandwana@sri.com

Abstract

In early 2019, the VOiCES from a Distance Challenge 2019 was held to foster research in the areas of speaker recognition and automatic speech recognition (ASR), with a special focus on single-channel distant/far-field audio under various noisy conditions. The challenge was based on the VOiCES corpus collected in real reverberant environments. This paper provides details of the challenge and analysis of evaluation results for the speaker recognition task. For the speaker recognition task, a total of 21 international research organizations from academia and industry participated in the challenge and submitted 58 valid systems. We report an in-depth analysis of system performance of the top-performing systems for the task of speaker recognition broken down by multiple factors such as the room acoustics, microphone type, distractor type, and microphone location. We also discuss the remaining challenges in far-field speaker recognition and suggest directions for future research.

1. Introduction

In the speaker recognition community, evaluations and challenges guide a substantial amount of the research. These evaluations provide common data, benchmarks, and performance metrics to the participating teams. This paradigm enables the fair comparison of state-of-the-art performance across the teams working on a specific task, in a manner that can be tracked over time.

One major evaluation in the field is the Speaker Recognition Evaluation (SRE) hosted by the National Institute of Standards and Technology (NIST). Since 1996, NIST has been organizing SREs on either an annual or bi-annual basis. Over the years, SREs have focused on the problem of text-independent speaker verification in a variety of domains, including telephone data; microphone data from a variety of different microphones; different speaking styles (conversational, interviews) and vocal efforts (low, neutral, high); noisy data collected in real environments; multiple bandwidths (8kHz,16kHz); and audio from video (AfV). The most recent evaluation, NIST SRE 2019, focused on the problem of audio-visual speaker recognition. Please refer to [1] for a two-decade detailed overview of the NIST SRE series.

In 2016, SRI International organized the Speakers in the Wild (SITW) challenge based on the SITW dataset [2, 3]. The SITW dataset contains speech samples in English from open-source media representing unconstrained or wild conditions. It also involved single- and multi-speaker recordings at 16kHz. More recently, the VoxCeleb Speaker Recognition Challenge (VoxSRC) also focused on speech obtained in the wild.

These speaker recognition evaluations and datasets resulted in understanding and advancement of research in several domains. However, these evaluations provide very limited insight into the performance of speaker recognition technology operating in the far-field/distant scenario. The understanding and mitigation of reverberation is now even more critical due to the commercial success of digital personal assistants, which typically operate in far-field settings but often assisted with multi-microphone beamforming arrays. To address this gap, SRI International and Lab41 collected the Voices Obscured in Complex Environmental Settings (VOiCES) corpus [4]. VOiCES is a large-scale distant/far-field dataset collected in realistic scenarios that include background noise such as TV, music, or other people talking in the background.

The VOiCES from a Distance Challenge 2019 was held in early 2019 [5, 6] as a part of the Interspeech 2019 special session based on the VOiCES corpus. A total of 21 teams participated in the speaker recognition task for the challenge, suggesting a strong interest in freely available, large-scale datasets for benchmarking technology and driving research toward current and ongoing issues in the distant/far-field speech processing area. The main objectives of this challenge were to: i) benchmark state-of-the-art technology in the area of distant speaker recognition and automatic speech recognition (ASR); (ii) support development of new ideas and technologies in speaker recognition and ASR; (iii) support new research groups entering the field of distant/far-field speech processing; and (iv) provide to the community a new, publicly available dataset that exhibits realistic distance characteristics. In this paper, we present an in-depth analysis of system performance broken down by multiple factors such as room acoustics, microphone type, distractor type, and loudspeaker orientation.

The remainder of this paper is organized as follows: first, the VOiCES corpus is described. Sec. 3 gives an overview of the VOiCES from a Distance Challenge 2019. Next, we discuss the results of the participating teams. In Sec. 5, we provide an in-depth analysis of the results broken down across different parameters. Finally, concluding remarks and remaining challenges are discussed in Sec. 6.

2. VOiCES Corpus

Prior to the Voices Obscured in Complex Environmental Settings (VOiCES) dataset, the existing large-scale corpora for understanding speaker recognition under reverberant conditions

10.21437/Odyssey.2020-24

suffered significant shortcomings. Most prior work either uses software simulations to generate data representing reverberant conditions [7] or uses actual data with very few speakers [8, 9], which results in limited significance in subsequent analysis.

The VOiCES corpus was collected by recording retransmitted audio from high-quality loudspeakers in real rooms, each room having a distinct acoustic profile, capturing natural reverberation using multiple microphones. This dataset has three different background noises, and the source loudspeaker rotation mimics human head movement at predefined intervals during the recordings. It was collected with the intent to push forward the state-of-the-art research in the area of speaker recognition, automatic speech recognition (ASR) [10]; speech activity detection (SAD); and speech enhancement in reverberant environment under noisy conditions. The corpus was released under the Creative Commons-BY 4.0 license, making it accessible for commercial, academic, and government use [4].

For the collection, pre-recorded foreground speech from the LibriSpeech dataset [11] and background noises were played in the rooms and were recorded by up to 20 microphones. The details related to the data-collection protocols and availability can be found at [4].

Table 1: *A summary of the VOiCES datasets across different parameters.*

Parameter	VOiCES
Number of speakers	300
Number of rooms	4
Number of mics	12 (Rm 1&2), 20(Rm 3& 4)
Microphone types	Studio, Lapel, MEMS
Total duration	3800+ hours
Number of audio segments	999,166
Source dataset	LibriSpeech
Speech type	Read
Background noise	Babble, Music, TV
Loudspeaker orientation	0° to 180°
Freely available	Yes

A subset of the VOiCES dataset was released at Interspeech 2018 [4, 12]. The first release included only 200 speakers recorded using 12 microphones in Room 1 and Room 2. The complete VOiCES dataset across different parameters is summarized in Table 1. The entire dataset was released to the public at the VOiCES special session at Interspeech 2019 and is freely available for download[1].

3. VOiCES from a Distance Challenge

The VOiCES from a Distance Challenge 2019 was organized based on the VOiCES corpus. The challenge was designed to foster research in the area of speaker recognition and automatic speech recognition (ASR), with the special focus on single-channel distant/far-field audio, under a variety of noisy and channel conditions. The VOiCES challenge had two tasks: speaker recognition and ASR. Each task had fixed and open training conditions defined by the training data that could be used to train the system. For the remainder of this paper, we only focus on the speaker recognition task of the challenge.

The task for speaker recognition was: given a segment of speech and target speaker enrollment data, automatically determine whether the target speaker is speaking in the segment. The fixed condition training limited the participants to use the Speakers in the Wild (SITW) and VoxCeleb1 and VoxCeleb2 datasets [13, 14]. For the open condition, the teams were allowed to use any propriety and/or public data, including the fixed condition data. More details can be found in the evaluation plan [5].

Table 2: *Details of the enrollment and verification sets for speaker recognition development across different parameters.*

	Enrollment Set	Verification Set
# Speakers	103	189
# Segments	256	15,648
Room ID	Room 1	Room 2
Mic. Type	Studio	Lapel
Mic. ID	01, 03	02, 04, 06, 08–12
Distractors	None	None, Music, TV, Babble
Loudspeaker Orientation	80, 90, 100	0, 60, 90, 120, 180

The development set was created by subsetting the audio files from Room 1 and 2 of the VOiCES corpus. The development set consisted of 15,904 audio segments from 196 speakers. Similarly, the evaluation set was created by subsetting the audio files of the VOiCES corpus to Room 3 and 4, as the associated source data. It consisted of 11,392 audio segments from 100 speakers, disjointed from the development set.

Each audio file contained a single speaker (i.e., single speaker trials only). We ensured that the enroll and test audio segments corresponded to different book chapters from the source corpus (LibriSpeech) to prevent positive bias in performance from trials sourced from the same original session. Tables 2 and 3 provide the details of the enrollment and verification sets for development and evaluation across different parameters.

The evaluation set roughly mirrored the conditions of the development set; however, it contained three major sources of mismatch relative to the development set. First, two enrollment conditions included speakers who were enrolled by either using source data (i.e., no reverberation) or using data from Room 3, the latter being more similar to the development conditions. This element enabled assessing reverberation-mismatch impact. Second, the evaluation set included several unseen microphones relative to the development set. These new microphones were micro-electromechanical systems (MEMS) and boundary microphone types. This subset enabled analyzing the channel mismatch between the development and evaluation sets. Third, the rooms between the development and evaluation sets were disjoint to create an acoustic mismatch between sets. In particular, Room 3 was an L-shaped room with very high reverberation

Table 3: *Details of the enrollment and verification sets for speaker recognition evaluation across different parameters.*

	Enrollment Set	Verification Set
# Speakers	100	96
# Segments	326	11,066
Room ID	Room 3, Source	Room 4
Mic. Type	Studio	Lapel, Boundary, MEMS
Mic. ID	01, 03	04, 06, 08–12, 16–19
Distractors	None	None, TV, Babble
Loudspeaker Orientation	80, 90, 110	0, 30, 60, 90, 120, 150, 180

[1] https://voices18.github.io/

characteristics.

The primary metric for the evaluation was C_{det} with costs of 1 for both errors and a probability of target of 0.01. The C_{det} was normalized by the cost that a naïve system, which always selects the least costly class, would get for the selected parameters. In our case, the normalization factor is given by P_{tar} [15]. The participants were also provided a scoring script that also computed the minimum C_{det}, C_{llr} (secondary metric), equal error rate, and average R_{prec}.

4. Results

In this section, we show the overall evaluation results of all teams. For the speaker recognition task, 21 teams successfully submitted their scores, out of which 4 teams submitted their scores for both the fixed and open conditions. We received 58 submissions, 50 of which were for the fixed training condition, and 8 were for the open training condition. The teams were allowed to submit up to three systems per condition. For the purpose of ranking teams, we picked the best score from each team as their official score for a given condition.

4.1. Overall results

Figure 1a and 1b represents the best actual and corresponding minimum $C_{primary}$ for the fixed and open condition for all teams. The top systems achieve very impressive performance on this challenging dataset.

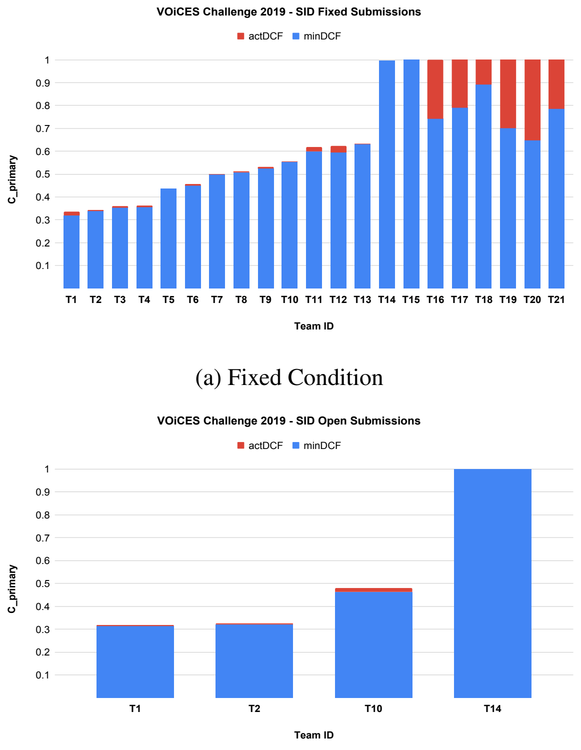
The system calibration was a critical component in the design of speaker recognition systems for this challenge [16]. The majority of the teams had very good calibration performance, with values of minimum C_{det} being very close to the actual C_{det}. This can be attributed to the similarity in the acoustic characteristics of the development and evaluation sets, and most of the teams effectively utilized the development set for the calibration. The six lowest-ranked teams did not successfully apply score calibration, which is reflected by the large difference between the minimum and actual costs.

4.2. Fixed versus open condition

The fixed training condition served the purpose of benchmarking and comparing systems trained with the same data (or a subset thereof). The open training condition provided the means to quantify the gains that could be achieved with an unconstrained amount of data. Figure 2 shows the system performance of four teams that submitted scores for both the fixed and open conditions. For the top two teams, we observe a very limited gain (5% relative) in the open training condition vs. the fixed training condition. Team T10 achieved around 14% relative gain by using the same datasets in fixed and open training conditions. For the open condition, T10 applied weak supervision techniques to increase the total amount of speaker-labeled data by four times compared to the official annotations in VoxCeleb1 and VoxCeleb2.

4.3. Confidence intervals

For all the systems, we also show a 95% confidence interval. The confidence intervals were calculated using a modified version of the joint bootstrapping technique described in [17]. The modification is performed to account for the fact that many models are created for each speaker of interest. Having multiple models per speaker introduces a very strong correlation across trials involving those models. These speakers might even be enrolled with different snippets from the same session. We did not account for the session in this work although it should be done while computing the confidence intervals. To this end, we add another layer of sampling: speakers are sampled first, then models from those speakers, then test signals. The models themselves might be repeated if a speaker was sampled more than once in the first layer of sampling. The trials corresponding to the selected subset of models and test signals are then used to compute the performance metric. We performed the sampling 20 times for each layer to produce 8000 measurements of the metric. The confidence interval that is reported corresponds to the 5 and 95 percentiles of the resulting empirical distribution.



(a) Fixed Condition

(b) Open Condition

Figure 1: Actual and minimum $C_{primary}$ for fixed and open speaker recognition submissions.

Figure 3, shows the confidence intervals for top 13 systems

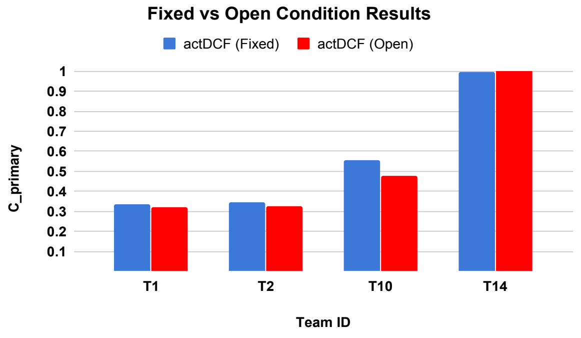


Figure 2: Actual $C_{primary}$ for fixed and open speaker recognition submissions.

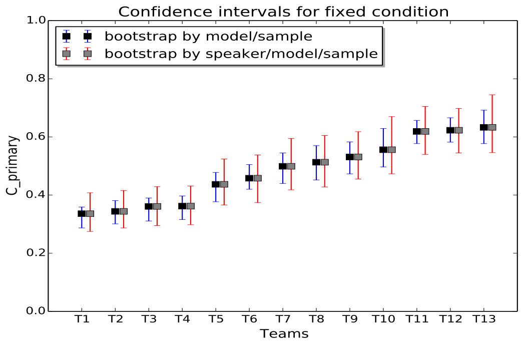


Figure 3: Confidence intervals for fixed condition systems computed with bootstrapping by model/sample and bootstrapping by speaker/model/sample.

with C_{det} value less than 1.0. We observe the effect of speaker sampling on confidence intervals. The confidence intervals are narrower without speaker sampling. Since there is no overlap in the range of confidence intervals between the top four teams and the rest of the teams, it can be concluded that the top four systems are significantly different. Furthermore, the top 4 systems have a very similar range of confidence intervals and perform relatively the same.

5. Analysis of Results

In this section, we present an in-depth analysis of different subconditions. The VOiCES corpus data collection plan was designed very carefully to perform a controlled analysis. This allows us to assess the impact of different factors while keeping other factors constant. The goal of this analysis is to see how current speaker recognition technology fares on different sub-conditions as well as to highlight the remaining challenges.

For this analysis, we split the evaluation trials into different subsets and focus on the top 10 systems from Figure 1a for the fixed condition only.

5.1. Impact of enrollment condition

Figure 4 shows the results for two different enrollment conditions. In the evaluation set, these conditions were defined by either using source data (i.e., no reverberation) or using data from Room 3 for the enrollment of the speaker. The motivation behind this was to access the generalization ability of the speaker recognition systems.

The results show that a change in enrollment condition introduces a mismatch and poses a challenge to the speaker recognition systems. Except for team T6, all other systems resulted in performance degradation. The degradation using Room 3 data ranged between 9% to 42% relative to the source data enrollment. This is useful in cases where the user cooperates in the enrollment process. For example, smart assistants typically operate in a far-field setting, and the user can be enrolled using a mobile application (close-talk) and can be verified using the assistant (far-field) for better performance.

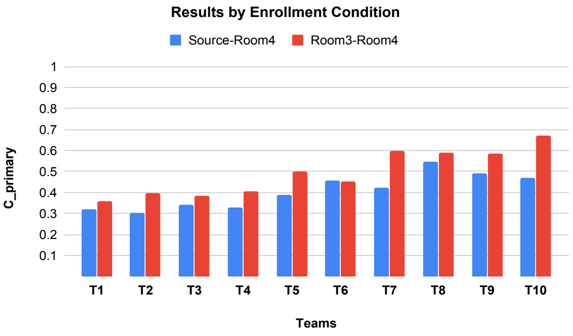


Figure 4: Impact of different enrollment conditions on the speaker recognition performance.

5.2. Impact of background distractor

In Fig. 5, we show the results for different background distractor conditions. The enrollment side includes no distractor, while the verification side includes three commonly occurring different distractor conditions (none, television, and babble). This setup enables an in-depth analysis of the effect of reverberation convoluted with a typical room environment.

We observe that systems are reasonably robust to the effect of TV noise. The performance with TV as a distractor is very close to that of no-distractor condition. The babble noise posed a significant challenge for all the teams, and this resulted in nearly 45% to 50% relative degradation in the performance.

5.3. Impact of microphone type and position

Figure 6 shows the results for different microphone types. In the evaluation, we used a studio mic placed close to the source for enrollment and three different types of microphones (lapel, MEMS, and boundary) for verification placed at different positions in the rooms. From the results, the lapel microphone appears to performs worst among all three mics. However, we noticed that seven lapel mics were placed at different distances, as opposed to three MEMS and one boundary mic. To further investigate, we dissect the results for lapel mics placed at different distances.

Figure 7 shows the results for the lapel microphones placed at seven different distances in Room 4 for the top-five teams. This approach enables studying the impact of microphone position in a room on speaker recognition system performance. We can see that the farther a mic is from the source, the greater the challenge for the systems. This finding is due to the self- and overlap-masking effects of reverberation. Some microphones were partially obstructed, such as under a table, in a wall, or in the ceiling. The performances of these microphones were very

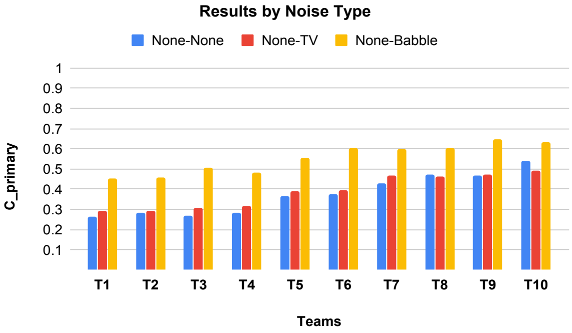


Figure 5: Results for different distractor conditions.

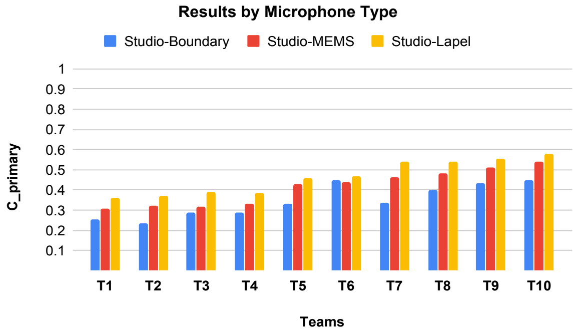


Figure 6: Results for three different microphone types in the evaluation set.

poor due to deteriorating signal quality. Also, this explains the poor performance of the lapel mic in Figure 7, which was likely dominated by obstructed microphones.

6. Conclusions and Remaining Challenges

The VOiCES dataset provides a new dimension for evaluating speaker recognition systems in a manner that reflects real reverberant conditions. Based on the submissions of 21 international research groups, the analysis presented in this paper has highlighted some of the fundamental issues that remain yet unaddressed in the technology, as well as aspects of the database that require further investigation. We summarize these here as future research directions.

The most crucial factor for further study is the significant performance difference observed between the development and evaluation set as described by participating teams in their papers and system descriptions [18, 19, 20, 21, 22, 23, 24]. The performance on the evaluation set was nearly 2-3x times worse than the development set for teams. This becomes even more interesting because the development and evaluation sets were collected in similar conditions. This highlights the fact that the change in acoustic characteristics of a room pose a significant challenge for the speaker recognition systems.

The impact of reverberation is another fundamental issue that is unaddressed by the current speaker recognition technology. The impact of reverberation increases with the increase between the source and the microphone distance. Also, reverberation in the presence of noise hurts the system performance by a significant margin as shown in Figure 5. There is a need to explore novel speaker embeddings architectures capable of handling the long term information (such as the effects of early and late reverberation) and which are robust to multiple conditions as well.

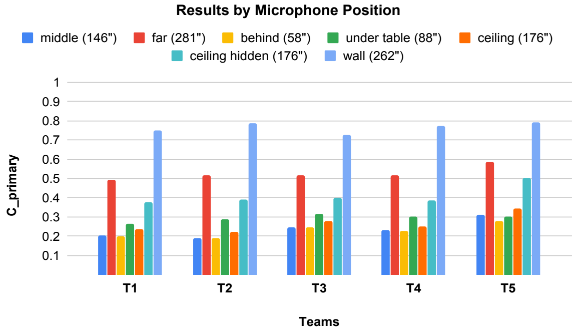


Figure 7: Results for lapel microphones placed at different distances.

Calibration is an important element of the speaker recognition system used/deployed in the real-world. Reverberation occurs very commonly in those situations. The bottom six teams failed to successfully calibrate their systems. A majority of the teams used the development data, which roughly mimicked the conditions of the evaluation set, to train the calibration models. In our previous work [16], we have shown a large degradation in calibration performance when the distance to the microphone is significantly different between calibration and evaluation conditions. In the practical systems, it will be quite challenging to know the distance between the source and microphone apriori and hence calibration methods that dynamically take into account trial conditions can be explored [25].

While the VOiCES challenge focused on single-channel microphone processing, the corpus was collected with multiple channels in the room. This will open new avenues of research in the area of front-end processing such as speech enhancement, beamforming, and dereverberation for the speaker recognition systems. It would be very interesting to see the impact of speech enhancement/dereverberation vs data augmentation on the generalization of speaker recognition systems.

We anticipate that the results and publications that stem from the challenge and corresponding publicly available database will motivate the community to solve the remaining challenges in the far-field/distant speaker recognition arena.

7. References

[1] C. S. Greenberg, L. P. Mason, S. O. Sadjadi, and D. A. Reynolds, "Two decades of speaker recognition evaluation at the national institute of standards and technology," *Computer Speech & Language*, vol. 60, pp. 1–10, 2020.

[2] M. McLaren, L. Ferrer, D. Castan, and A. Lawson, "The 2016 speakers in the wild speaker recognition evaluation," *Proc. Interspeech*, pp. 823–827, 2016.

[3] M. McLaren, L. Ferrer, D. Castan, and A. Lawson, "The speakers in the wild (SITW) speaker recognition database.," *Proc. Interspeech*, pp. 818–822, 2016.

[4] C. Richey, M. A. Barrios, Z. Armstrong, C. Bartels, M. Graciarena, A. Lawson, M. K. Nandwana, et al., "Voices obscured in complex environmental settings (VOiCES) corpus," *Proc. Interspeech*, pp. 1566–1570, 2018.

[5] M. K. Nandwana, J. van Hout, M. McLaren, C. Richey, A. Lawson, and M. A. Barrios, "The VOiCES from a distance challenge 2019 evaluation plan," *arXiv:1902.10828 [eess.AS]*, 2019.

[6] M. K. Nandwana, J. van Hout, C. Richey, M. McLaren, M. A. Barrios, and A. Lawson, "The VOiCES from a distance challenge 2019," *Proc. Interspeech*, pp. 2438–2442, 2019.

[7] Tiago H Falk and Wai-Yip Chan, "Modulation spectral features for robust far-field speaker identification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 90–100, 2010.

[8] Qin Jin, Tanja Schultz, and Alex Waibel, "Far-field speaker recognition," *IEEE Transactions on Audio,*

Speech, and Language Processing, vol. 15, no. 7, pp. 2023–2032, 2007.

[9] D. Garcia-Romero, D. Snyder, S. Watanabe, G. Sell, A. McCree, D. Povey, and S. Khudanpur, "Speaker recognition benchmark using the CHiME-5 corpus," *Proc. Interspeech*, pp. 1506–1510, 2019.

[10] I. Medennikov, Y. Khokhlov, A. Romanenko, I. Sorokin, A. Mitrofanov, V. Bataev, A. Andrusenko, T. Prisyach, M. Korenevskaya, O. Petrov, and A. Zatvornitskiy, "The STC ASR system for the VOiCES from a distance challenge 2019," *Proc. Interspeech*, pp. 2453–2457, 2019.

[11] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5206–5210, 2015.

[12] Mahesh Kumar Nandwana, Julien van Hout, Mitchell McLaren, Allen Stauffer, Colleen Richey, Aaron Lawson, and Martin Graciarena, "Robust speaker recognition from distant speech under real reverberant environments using speaker embeddings," *Proc. Interspeech*, pp. 1106–1110, 2018.

[13] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman, "Voxceleb: A large-scale speaker identification dataset," *Proc. Interspeech*, pp. 2616–2620, 2017.

[14] Joon Son Chung, Arsha Nagrani, and Andrew Zisserman, "Voxceleb2: Deep speaker recognition," *Proc. Interspeech*, pp. 1086–1090, 2018.

[15] Niko Brümmer and Johan Du Preez, "Application-independent evaluation of speaker detection," *Computer Speech & Language*, vol. 20, no. 2-3, pp. 230–275, 2006.

[16] M. K. Nandwana, L. Ferrer, M. McLaren, D. Castan, and A. Lawson, "Analysis of critical metadata factors for the calibration of speaker recognition system," *Proc. Interspeech*, pp. 4325–4329, 2019.

[17] N. Poh and S. Bengio, "Estimating the confidence interval of expected performance curve in biometric authentication using joint bootstrap," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 137–140, 2007.

[18] Sergey Novoselov, Aleksei Gusev, Artem Ivanov, Timur Pekhovsky, Andrey Shulipa, Galina Lavrentyeva, Vladimir Volokhov, and Alexandr Kozlov, "STC Speaker Recognition Systems for the VOiCES from a Distance Challenge," *Proc. Interspeech*, pp. 2443–2447, 2019.

[19] P Matejka, O Plchot, H Zeinali, L Mošner, A Silnova, L Burget, and O Glembek, "Analysis of BUT submission in far-field scenarios of VOiCES 2019 challenge," *Proc. Interspeech*, pp. 2448–2452, 2019.

[20] Arindam Jati, Raghuveer Peri, Monisankha Pal, Tae Jin Park, Naveen Kumar, Ruchir Travadi, Panayiotis Georgiou, and Shrikanth Narayanan, "Multi-task Discriminative Training of Hybrid DNN-TVM Model for Speaker Verification with Noisy and Far-Field Speech," *Proc. Interspeech*, pp. 2463–2467, 2019.

[21] David Snyder, Jesús Villalba, Nanxin Chen, Daniel Povey, Gregory Sell, Najim Dehak, and Sanjeev Khudanpur, "The JHU Speaker Recognition System for the VOiCES 2019 Challenge," *Proc. Interspeech*, pp. 2468–2472, 2019.

[22] Jonathan Huang and Tobias Bocklet, "Intel Far-Field Speaker Recognition System for VOiCES Challenge 2019," *Proc. Interspeech*, pp. 2473–2477, 2019.

[23] Hanwu Sun, Kah Kuan Teh, Ivan Kukanov, and Huy Dat Tran, "The I2Rs Submission to VOiCES Distance Speaker Recognition Challenge 2019," *Proc. Interspeech*, pp. 2478–2482, 2019.

[24] Danwei Cai, Xiaoyi Qin, Weicheng Cai, and Ming Li, "The DKU system for the speaker recognition task of the 2019 VOiCES from a distance challenge," *Proc. Interspeech*, pp. 2493–2497, 2019.

[25] L. Ferrer, M. K. Nandwana, M. McLaren, D. Castan, and A. Lawson, "Toward fail-safe speaker recognition: Trial-Based Calibration with a reject option," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 140–153, 2019.