das-Peak v2.5.0
Performance Report

Release	Date	Description	Author	Reviewer	Approver
2.5.0	24/03/2021	das-Peak 2021Q1	SPC	MSY	MSY

0. What's new?

This new version of das-Peak incorporates these new functionalities:

- Pre-recorded voice detection (Replay-attack) model has been changed to the 2020Q3 anti spoofing model due to high capability detecting smartphone replay attacks in production scenarios.

1. das-Peak microservice

Voice biometrics is a state-of-the-art technology that allows a person to be validated by his/her voice. VERIDAS solution captures the unique physical features of the vocal apparatus and features such as frequency, speed and accents and compiles them together into a virtually **unique voice biometric vector** per person.

The voice biometric vector is a mathematical descriptor obtained from the characteristics of the voice in an audio recording. This mathematical conversion from voice into a biometric vector is irreversible. Therefore, it is not possible to recover a person's voice signal from the calculated biometric vector.

VERIDAS has developed its own speaker verification engine (das-Peak) as a cloud-based solution that can be consumed via APIs.

VERIDAS' voice biometrics group has participated in the **short-duration Speaker Verification Challenge (SdSV) 2020 getting the 3º award** (2º single model), demonstrating best results in the state of the art in Voice Biometrics for **short utterances conditions**. [https://sdsvc.github.io/]

das-Peak calculates the similarity between two audio recordings (in terms of the speakers present in them) using biometric algorithms. das-Peak engine allows to authenticate users voice **without** the need of using a password or predefined phrase (passive recognition) as it is based on **text-independent technology**. This means that the biometric comparison is related to the voice characteristics and not to the content of the sentence. However, the system is flexible to use pre-defined phrases in order to fulfill customer requirements or additional controls.

3. System quality report

3.1. Verification performance (Telephone Channel)

VERIDAS has evaluated its voice biometrics model (das-Peak 2020Q4 version) with an internal Telephone audio database with different duration enrollment audios (10 and 5 seconds) and different duration audios test (10 and 5 seconds) with a total of 26.544 comparisons. This dataset has been recorded in an office environment through telephone calls. The database language is Spanish. In the next table the values of False Positive Rates (FPR) and False Negative Rates (FNR) with different threshold values are showed.

The FPR is the probability to accept a non legit person and the FNR is the probability to reject a legit person. With these values it is possible to choose the desired working point of the voice biometric system.

	Enrollment=5s Verification=5s		Enrollment=10s Verification=5s		Enrollment=10s Verification=10s	
Similarity Threshold	FPR(%)	FNR(%)	FPR(%)	FNR(%)	FPR(%)	FNR(%)
0.5	1	0.36	1	0.08	1	0
0.55	0.74	0.51	0.75	0.17	0.76	0.03
0.6	0.55	0.7	0.54	0.26	0.56	0.05
0.65	0.4	0.94	0.38	0.39	0.41	0.1
0.7	0.27	1.17	0.26	0.47	0.29	0.15
0.75	0.17	1.58	0.17	0.63	0.2	0.18
0.8	0.1	2.03	0.1	0.83	0.1	0.25
0.85	0.05	2.79	0.05	1.13	0.05	0.35
0.9	0.02	4.5	0.02	1.77	0.02	0.45
0.95	0.001	8.7	0.001	3.65	0.01	0.79

This calibration shows different security work points depending on the similarity threshold and the audio voice duration that are compared.

For example, if the use case is 5 seconds to enroll and 5 seconds to verify, with a threshold equal to 0.8, it is obtained FPR=0.1% and FNR=2.03%. In this case, 97.97% of the comparisons of a person's voice and its corresponding voice registration will be considered as the same person, and only 0.1% of the cases comparing voices to different persons will be incorrectly classified as the same person.

3.2. Verification performance (Lossless Audio)

VERIDAS has evaluated its voice biometrics model (das-Peak 2020Q4 version) with Voxceleb 2 test database with different duration enrollment audios (10 and 5 seconds) and different duration audios test (10 and 5 seconds) with a total of 253.188.434 comparisons in different acoustic conditions (street, pubs, TV studio,..).The database language is English. In the next table False Positive Rates (FPR) and False Negative Rates (FNR) with different threshold values are shown. Due to the elevated number of speakers, environments and speaker session variability, the FNR is worse than the telephone channel.

The FPR is the probability to accept a non legit person and the FNR is the probability to reject a legit person. With these values it is possible to choose the desired working point of the voice biometric system.

	Enrollment=5s Verification=5s		Enrollment=10s Verification=5s		Enrollment=10s Verification=10s	
Similarity Threshold	**FPR(%)**	**FNR(%)**	**FPR(%)**	**FNR(%)**	**FPR(%)**	**FNR(%)**
0.5	1	4.28	1	3.05	1	2.13
0.55	0.74	4.94	0.74	3.5	0.74	2.43
0.6	0.54	5.73	0.54	4.11	0.54	2.89
0.65	0.38	6.72	0.38	4.8	0.39	3.28
0.7	0.26	7.99	0.26	5.64	0.27	3.81
0.75	0.17	9.54	0.17	6.75	0.17	4.6
0.8	0.1	11.68	0.1	8.31	0.1	5.57
0.85	0.05	14.61	0.05	10.56	0.05	7.02
0.9	0.02	19.25	0.02	14.2	0.02	9.72

0.95	0.001	27.89	0.001	21.68	0.001	15.59

*Voxceleb 2: http://www.robots.ox.ac.uk/~vgg/publications/2018/Chung18a/chung18a.pdf

This calibration shows different security work points depending on the similarity threshold and the voice duration audios that are compared.

For example, if the use case is 5 seconds to enroll and 5 seconds to verify, with a threshold equal to 0.8, it is obtained FPR=0.1% and FNR=11.68%. In this case, 88.32% of the comparisons of a person's voice and its corresponding voice registration will be considered as the same person, and only 0.1% of the cases comparing voices to different persons will be incorrectly classified as the same person.

3.3. Identification performance

VERIDAS has evaluated its voice biometrics model (das-Peak 2020Q4 version) with an internal database with 256 speakers and 5904 audios for the identification 1:N use case (N=2 to 10). In this database, speakers have been recorded in different sessions with different acoustic conditions (street, pub, train station, room, church,...) through smartphone recording application. The database language is Spanish. A number of 1.000 random identification tasks have been performed for each N (from N=2 to 10). For example, in a 1 to 10 identification process (N=10) the probability to identify the right individual is **97.64 %**. The identification accuracy results for N=2 to 10 can be observed in the following table:

Nº Speakers	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Accuracy (%)	99.77	99.61	99.5	99.36	99.29	99.19	99.09	99.01	98.94

3.4. Voice Authenticity (Anti-spoofing)

The API also includes an anti-spoofing detector model designed to avoid fraudulent access. Given a wav file with voice, the detector estimates the likelihood of the voice authenticity. The performance of the system has been measured for the more important and common spoofing technique: replay attack (a voice capture from a device speaker) through smartphone speakers.

The evaluation has been performed on another internal database with 140094 authentic audios and 109484 spoofing audios, achieving a 98.62% accuracy. The anti-spoofing performance is shown in the following table. Notice the performance of the system is shown for different authenticity thresholds, i.e., 1.00 means authentic and 0.00 means spoof attempt.

Based on the following Table, with an operation point at 0.60, the 0.42% of authentic cases will be rejected, and the 2.6% of replay-attack spoofing attempts will be correctly misclassified as authentic.

Authenticity Threshold	FPR(%)	FNR(%)
0.05	23.97	0.0028
0.1	12.33	0.0056
0.15	8.25	0.0158
0.2	6.18	0.028
0.25	5.06	0.057
0.3	4.33	0.085
0.35	3.82	0.107
0.4	3.42	0.138
0.45	3.12	0.177
0.5	2.91	0.23
0.55	2.74	0.31
0.6	2.60	0.42
0.65	2.50	0.54
0.7	2.43	0.75
0.75	2.35	1.07
0.8	2.29	1.64
0.85	2.26	2.80
0.9	2.21	5.93
0.95	2.15	19.04