ALBAYZIN EVALUATIONS

Iberspeech-Challenges 2024

Supported by

Spanish Thematic Network on Speech Technology (RTTH)

Organized by

ViVoLab & Cátedra RTVE from Universidad de Zaragoza

Universidad San Pablo-CEU

AuDIaS from Universidad Autónoma de Madrid

Universidad del País Vasco

Telefónica Innovación Digital

http://catedrartve.unizar.es/albayzin2024.html

https://iberspeech.tech/albayzin-evaluation-challenge

Open call for Albayzín Evaluations → Five Challenges:

- ✓Speech to Text (S2T) RTVE-UZ

  Automatic transcription of TV shows.

- ✓Speaker Diarization and Identity Assignment (SDIA) RTVE-UZ

  Segmenting broadcast audio documents according to different speakers, linking those segments which originate from the same speaker and identify a closed set of speakers.

- ✓Search on Speech (SoS) U. San Pablo/CEU - UAM

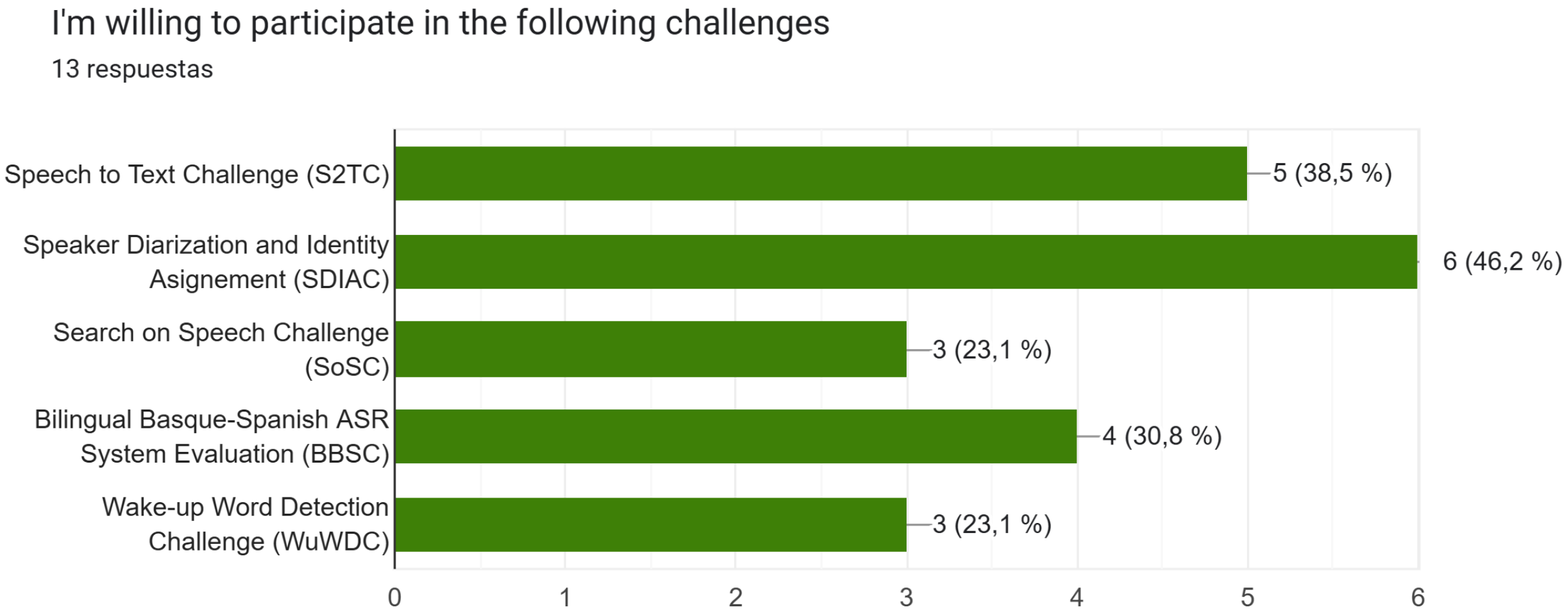
  Searching in audio content a list of terms/queries.

- ✓Bilingual Basque-Spanish ASR (BBS-S2TC) UPV/EHU

  transcribing short segments of speech in Spanish, Basque, or both languages (with code switching).

- ✓Wake-Up Word Detection(WUWDC) Telefónica

  assess the performance of state-of-the-art Keyword Spotting systems


I'm willing to participate in the following challenges
13 respuestas
Speech to Text Challenge (S2TC)
5 (38,5 %)
Speaker Diarization and Identity Asignement (SDIAC)
6 (46,2 %)
Search on Speech Challenge (SoSC)
3 (23,1 %)
Bilingual Basque-Spanish ASR System Evaluation (BBSC)
4 (30,8 %)
Wake-up Word Detection Challenge (WuWDC)
3 (23,1 %)
0
1
2
3
4
5
6

Participation:

12 teams fill the registration form to participate

10 Spanish teams

CIRES21, AhoLab-HiTZ, Ilenia, AUDIAS-UAM, ECASIMM, Vicomtech, AUDIAS, UR, HiTZ-AhoLab, PRHLT

2 International teams

TEAMIV (Intelligent Voice, UK), TPRO (Ireland)

Final participation: 9 teams

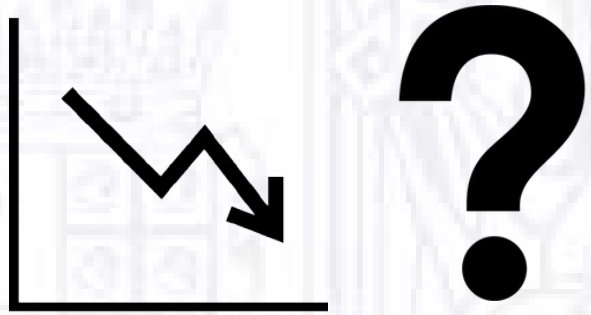
S2TC	SDIAC	BBS-S2TC	SoS	WUWDC
3(-1)	3	3	0	1

Previous evaluations S2TC/SDIAC:

2018: 16 teams (85 systems evaluated)

2020: 12 teams (31 systems evaluated)

2022: 7 teams (20 systems evaluated)



S2TC and SDIAC results https://catedrartve.unizar.es/albayzin2024results.html

17:25-17:50 | S2TC Database and results overview. Eduardo LLeida

TEAM: CIRES21

TEAM: ILENIA

17:50-18:20 | SDIAC Database and results overview. Eduardo Lleida

TEAM: AhoLab-HiTZ

TEAM: AUDIAS-UAM

TEAM: UR

18:20-18:45 | BBS-S2T Database and results overview. Luis Javier Rodríguez

TEAM: HiTZ-AhoLab

TEAM: PRHLT

TEAM: VICOMTECH

18:45-18:55 | WUWDC Database and results overview. Wiliam López

TEAM: AUDIAS-UAM

RTVE2024 Database Description

Vivolab
Aragon Institute for Engineering Resarch (I3A) Universidad de Zaragoza

Virginia Bazán , Carmen Pérez , Pere Vila
Corporación Radiotelevisión Española

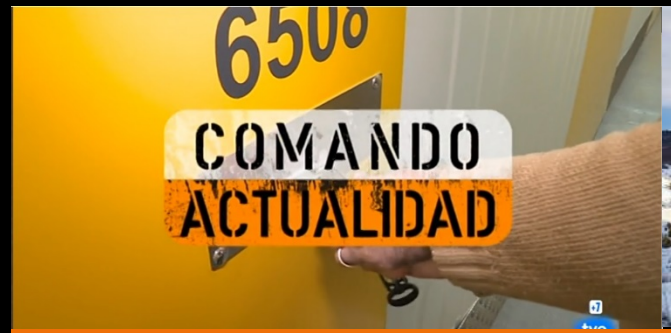
https://catedrartve.unizar.es/rtvedatabase.html

Cátedra RTVE de la Universidad de Zaragoza

July 10th, 2017
The “Corporación Radiotelevisión Española” (RTVE) and the “Universidad de Zaragoza” (UZ) signed an agreement to develop the “Cátedra RTVE de la Universidad de Zaragoza”

The purpose is to boost the technologies associated with the generation of audiovisual metadata.

One of the objectives of the Chair is to launch a set of technological challenges and provide the necessary data to test the technologies.



Databases Description: RTVE2018, RTVE2020, RTVE2022 & RTVE2024

Video and audio quality

- ✓Video files (.mp4): multimodal diarization (2018 & 2020)
  - Format:
    - Default format in the RTVE internet channel “RTVE a la carta”
    - Video:
      - h264 standard codec with pixel format yuv420p, 1024x576 [SAR 1:1,DAR 16:9], 25 fps
      - 1500 kb/s average bit rate
    - Audio:
      - Mpeg Low-Complexity (AAC-LC) codec, sampling frequency 44100Hz stereo
      - Variable Bit rate between 48 y 96 kb/s

- ✓Audio files(.aac)
  - Format:
    - Mpeg Low-Complexity (AAC-LC) codec, sampling frequency 44100Hz stereo
    - Variable Bit rate between 48 y 96 kb/s
    - 2022: some multichannels

Database Content: RTVE2018

http://catedrartve.unizar.es/rtvedatabase.html

Training	Hours	Dev1	Hours	Task	Dev2	Hours	Task	Test	Hours	Task
20H	32:22:37	20H	9:13:13	T2S						
Agrosfera	37:34:32									
Al filo de lo imposible	6:59:54							Al filo de lo imposible	4:10:03	T2S
Asuntos publicos	61:27:00	Asuntos Públicos	8:11:00	T2S						
Arranca en Verde	4:37:35							Arranca en Verde	1:00:30	T2S
Comando actualidad	9:10:28	Comando Actualidad	7:53:13	T2S						
Dicho y Hecho	8:18:00							Dicho y Hecho	1:48:00	T2S
España en comunidad	4:53:27							España en Comunidad	8:09:32	T2S, Diarization, SoS
La mañana	218:12:00	La Mañana	1:30:00	T2S						
La tarde en 24H Economia	4:10:54									
La tarde en 24H Tertulia	17:49:40							La Tarde en 24H Tertulia	8:52:20	T2S,Diarization,Face
La tarde en 24H Entrevista	4:54:03									
La tarde en 24H El tiempo	2:20:12									
Latinoamerica en 24H	12:12:03							Latinoamerica en 24H	4:06:57	T2S, Diarization
Millennium	9:33:01				Millennium	7:42:44	Diarization, T2S, SoS	Millennium	1:52:50	T2S, SoS
Saber y Ganar	26:05:17							Saber y Ganar	2:54:53	T2S
		La noche en 24H	25:44:25	T2S	La noche en 24H	7:26:41	Diarization, T2S, SoS, Face			
	460:40:43		**52:31:51**			**15:09:25**			**41:00:05**	

Human revised manual transcriptions with speakers turns: 108 hours
Aligned transcriptions: 38 hours

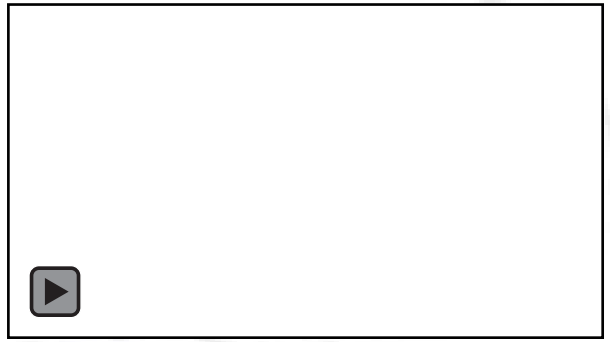
- ✓ Text associated to subtitles (24H channel, year 2017)
  - ✓ More than 3M sentences, 56 M words
  - ✓ Vocabulary of more than 160K unique words

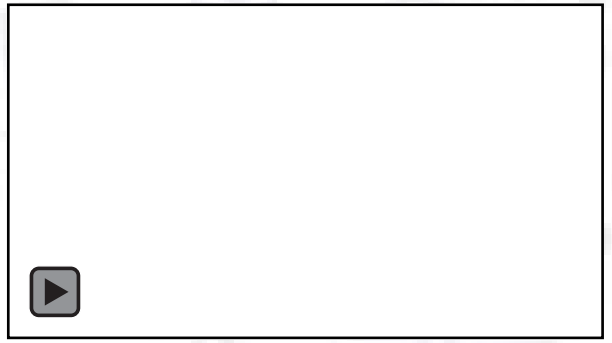
Database Content: RTVE2020

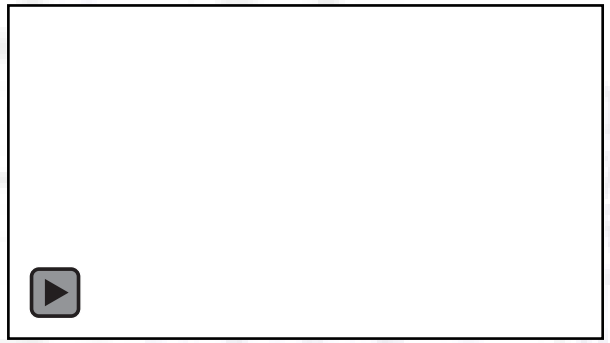
http://catedrartve.unizar.es/rtvedatabase.html

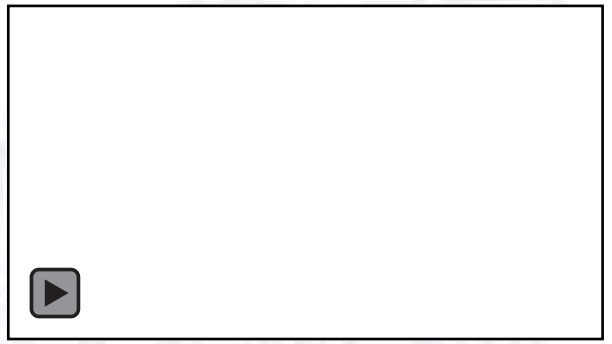
RTVE2020				
Show	Hours	S2T	SD-IA	MD-SD
Millennium	1:56:11	1:56:11	1:56:11	1:56:11
Los desayuno de tve	10:58:34	10:58:34	10:58:34	10:58:34
Comando actualidad	4:01:31	4:01:31	4:01:31	4:01:31
Ese programa del que usted me habla	1:58:36	1:58:36	1:58:36	1:58:36
Neverfilms	0:11:41	0:11:41	0:11:41	0:11:41
Si fueras tu	0:51:14	0:51:14	0:51:14	0:51:14
Bajo la red	0:59:01	0:59:01	0:59:01	0:59:01
Boca norte	1:00:46	1:00:46	1:00:46	1:00:46
Wake-up	0:57:28	0:57:28	0:57:28	0:57:28
Aquí la tierra	10:26:02	10:26:02	10:26:02	10:26:02
Versión española	2:29:12	2:29:12		
Mercado central	8:39:47	8:39:47		
Vaya crack	5:06:00	5:06:00		
Como nos reíamos	2:51:42	2:51:42		
Imprescindibles (2 canales)	3:12:31	3:12:31		
	55:40:16	**55:40:16**	**33:21:04**	**33:21:04**

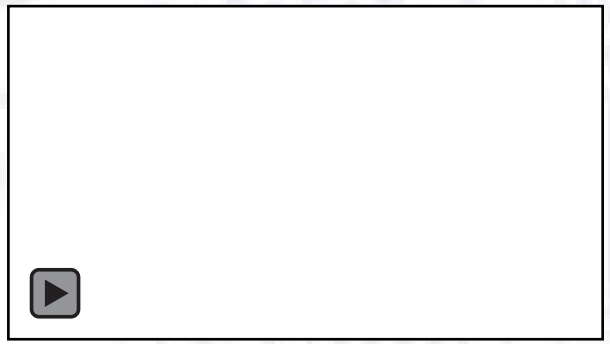
161 characters have been labelled

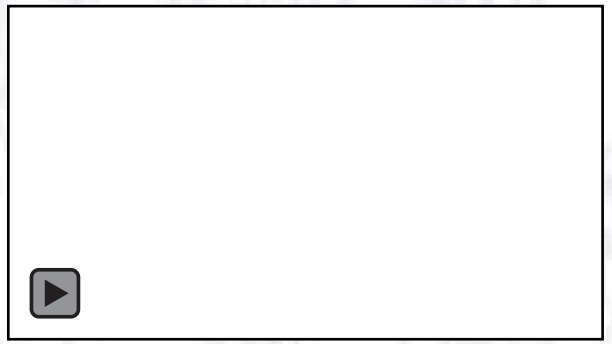










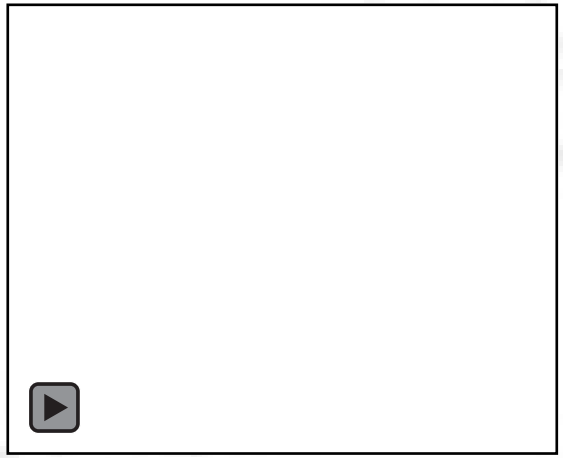


Database Content: RTVE2022

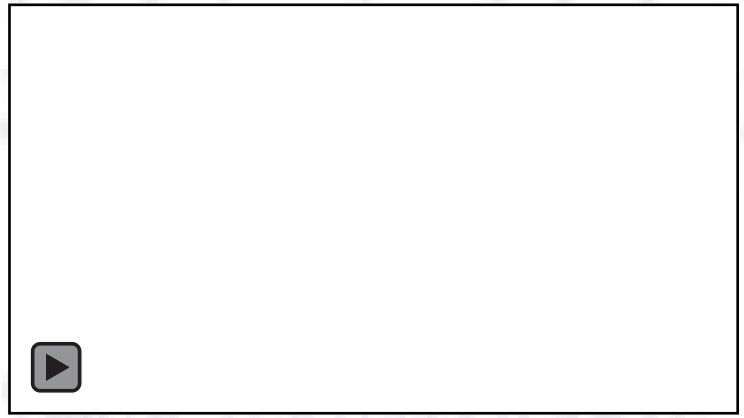
RTVE2022				
Show	Hours	S2T	SD	TaSA
3x4	2:58:17	2:58:17	2:58:17	
A pedir de boca	3:41:38	3:41:38		
Agrosfera	4:15:20	4:15:20	4:15:20	6:38:04
Aquí la Tierra	2:46:44	2:46:44	2:46:44	2:52:35
Ateneo	1:40:09	1:40:09		
Cerámica Popular Española	1:02:35	1:02:35		
Comando Actualidad	3:59:29	3:59:29	3:59:29	
Conversatorios en Casa América	1:58:44	1:58:44		
Corazón	3:00:17	3:00:17	3:00:17	4:33:49
El cazador	3:48:22	3:48:22		
Encuestas con ruido ambiente	2:08:13	2:08:13		
Entrevistas en bruto	3:54:57	3:54:57		
España Directo	4:05:57	4:05:57	4:05:57	
Fiction (Grasa, Yrreal, Riders)	3:53:22	3:53:22	3:53:22	
Informativos UMATIC	0:59:49	0:59:49		
Jara y Sedal	2:29:17	2:29:17		
Noticias Nacional	2:14:32	2:14:32		
Saber y Ganar	4:28:28	4:28:28		
Toros	0:49:57	0:49:57		
21 different shows	**54:16:07**	**54:16:07**	**24:59:26**	**14:04:28**

74 characters have been labelled

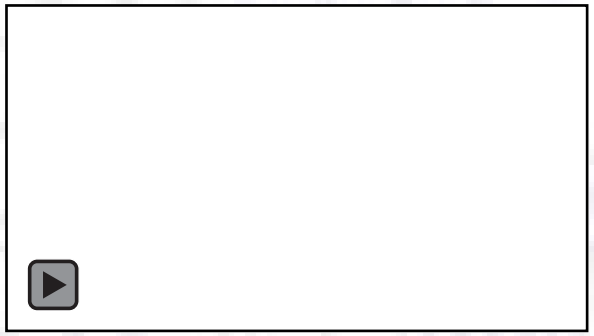
A pedir de boca



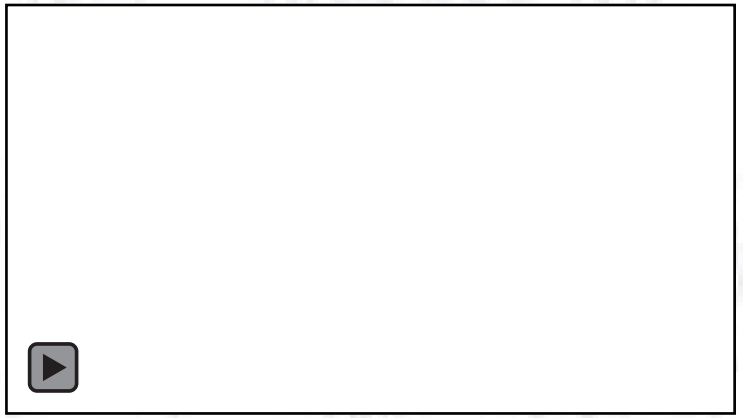
Entrevistas en bruto



Encuestas con ruido ambiente



Fiction (Grasa)



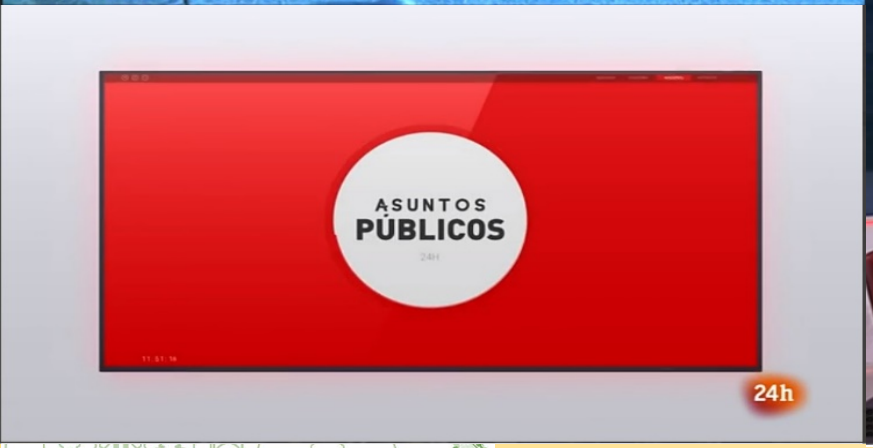
Database Content: RTVE2024

RTVE2024					
Acronym	**Show**	**Hours**	**S2TC**	**SD**	**SDIAC**
AC	APRENDEMOS EN CLAN	1:58:06	1:58:06	1:58:06	
APB	A PEDIR DE BOCA	5:18:14	5:18:14	2:00:26	
AT	AQUI LA TIERRA	2:52:55	2:52:55	2:52:55	2:52:55
CA	COMANDO ACTUALIDAD	3:47:00	3:47:00	0:59:35	0:59:35
CC	CAMINO A CASA	3:43:40	3:43:40	0:59:01	0:59:01
FCNBR	FICCION-BAJO LA RED	0:34:32	0:34:32	0:34:32	0:19:50
FCNGR	FICCION-GRASA	0:39:08	0:39:08	0:39:08	0:39:08
FCNSNS	FICCION-SER O NO SER	0:48:03	0:48:03	0:48:03	
FCNA34	FICCION-AREA 34	0:25:37	0:25:37		
PQ	LA PIJA Y LA QUINQUI	2:04:47	2:04:47	1:05:57	
PL	LA PLAYZ LIST	1:59:57	1:59:57		
LX	LATE XOU CON MARC GIRO	2:07:35	2:07:35		
MA	MAÑANEROS	6:24:50	6:24:50		
RC	RAICES Y CANTARES	2:19:56	2:19:56	0:23:50	0:23:50
SG	SABER Y GANAR	4:27:20	4:27:20		
MP	CONCURSO-MAPI	0:37:02	0:37:02	0:37:02	0:37:02
123	CONCURSO-123	1:26:53	1:26:53		
CD	CANCIONES DEL DESVAN	0:11:34	0:11:34		
AE	AMIGOS DEL ESPACIO	0:25:28	0:25:28	0:25:28	
TO	CANAL TOROS	1:22:04	1:22:04		
DP	DEPORTES	1:54:29	1:54:29	0:29:19	
DC	DIAS DE CINE	0:22:30	0:22:30	0:22:30	
IM	IMPRESCINDIBLES	2:06:14	2:06:14		
CP	CENTROS DEL PODER	1:28:17	1:28:17		
24 different shows		**49:26:11**	**49:26:11**	**14:15:52**	**6:51:21**

33 characters have been labelled

A pedir de boca (Bruto)	
A pedir de boca (emisión)	
Días de cine (bruto)	
Imprescindibles (bruto)	
Aquí la tierra	
Raíces y cantares	

Speech to Text Challenge



Evaluation plan: (http://catedrartve.unizar.es/albayzin2024.html)

Objective:

Evaluate the state of the art in automatic speech recognition for broadcast speech transcription.

Training conditions:

✓ Open condition

Participants are free to use data to train their systems provided that these data are fully documented in the systems description paper.

Test database

49 hours of 24 TV shows covering different genres as live recordings, contests, soap opera, raw material, news, old formats recordings, …

If the file contains music, the lyrics of the song should not be transcribed. They are only transcribed if the singing is a cappella, that is, without music.

Primary metric:

WER (Word Error Rate)

$$WER(\%) = \frac{\#I + \#D + \#S}{N} 100$$

where

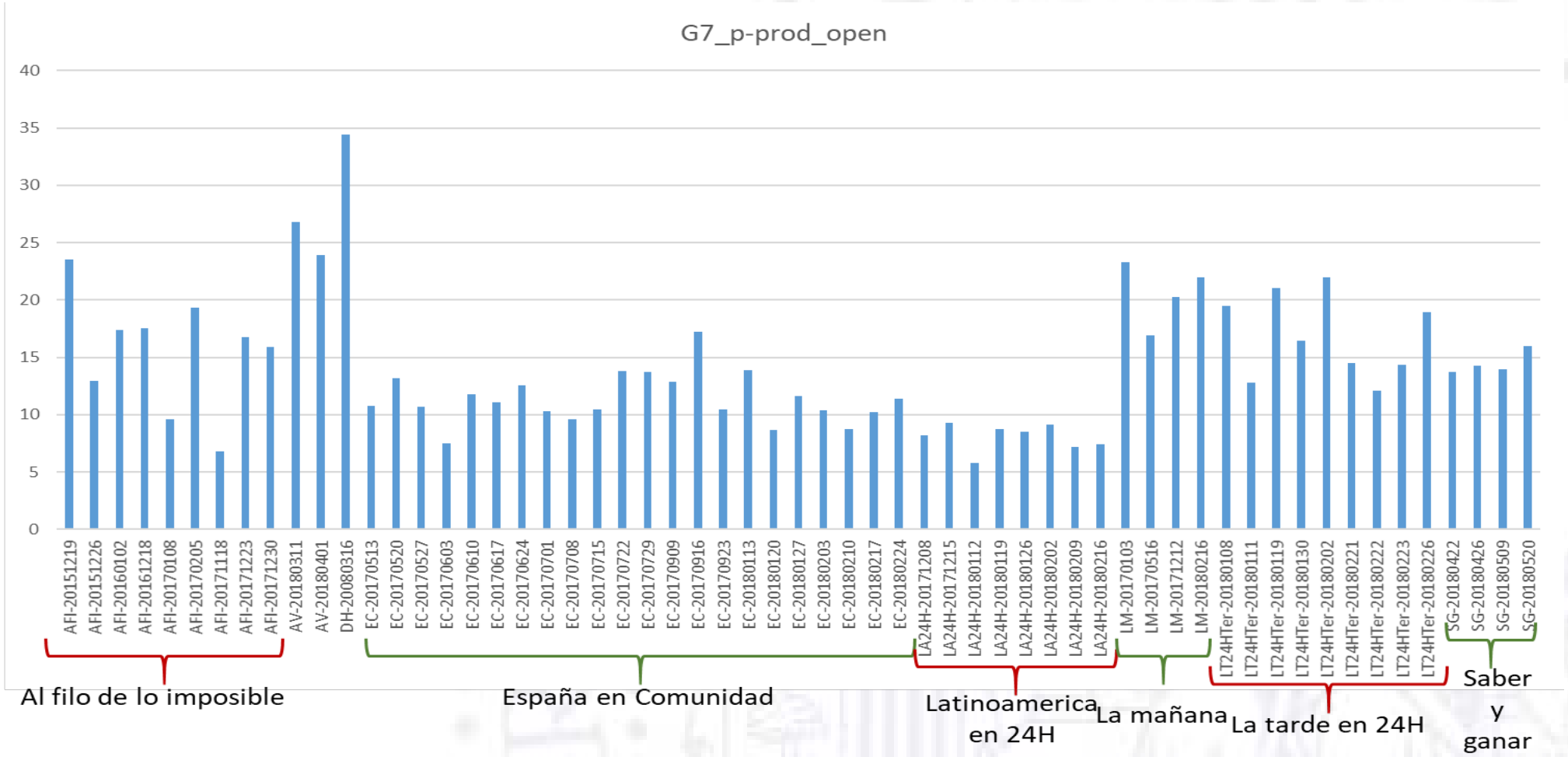
$\#\boldsymbol{I}$ number of insertions, $\#\boldsymbol{D}$ number of deletions, $\#\boldsymbol{S}$ number of substitutions and $\boldsymbol{N}$ total number of words to be recognized

Baseline system

WhisperX with large-v3 model, time alignment with wav2vec2 and pyannote/speaker-diarization-3.1

Previous challenges results

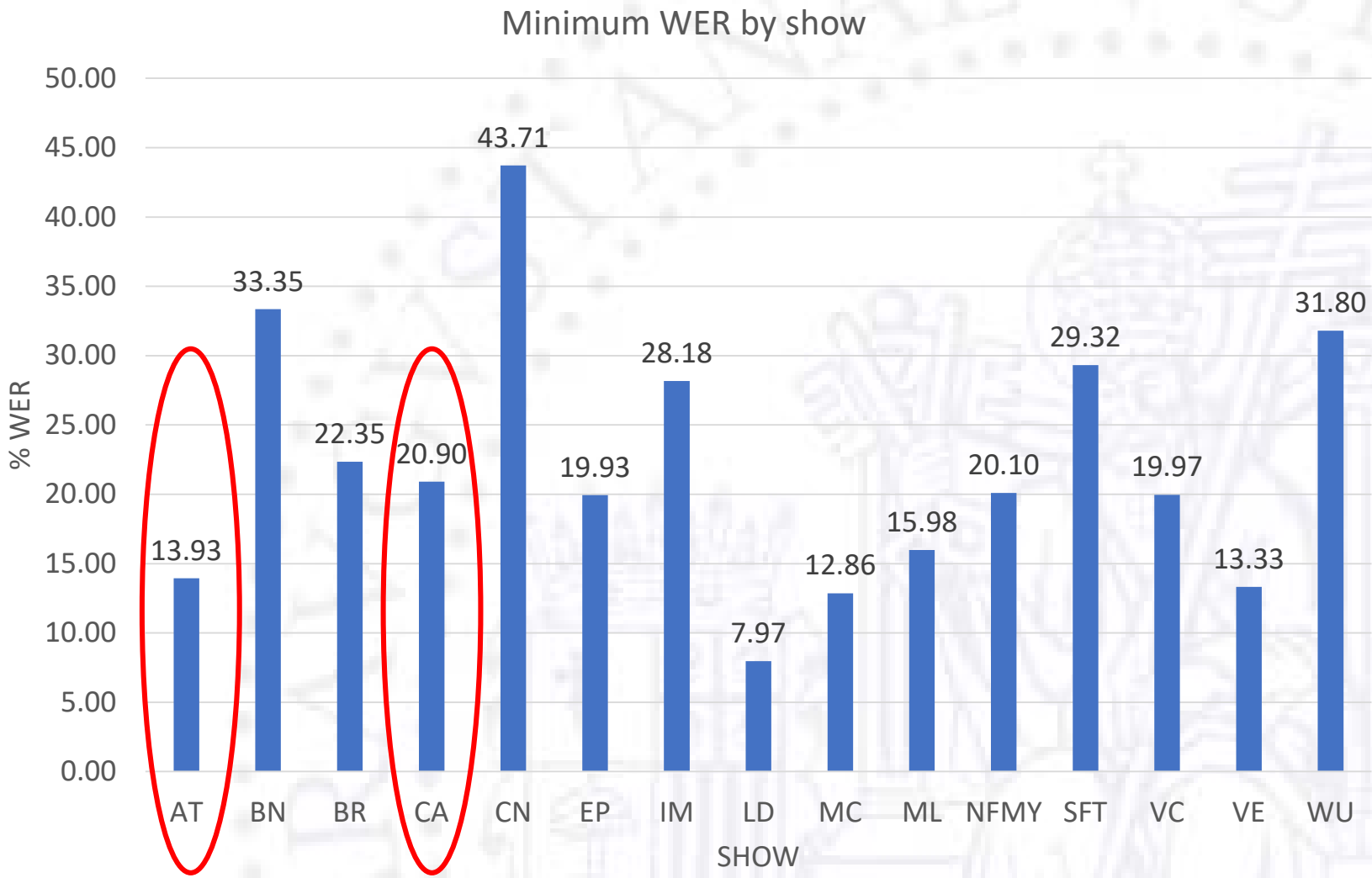
2018:
WER 16,45



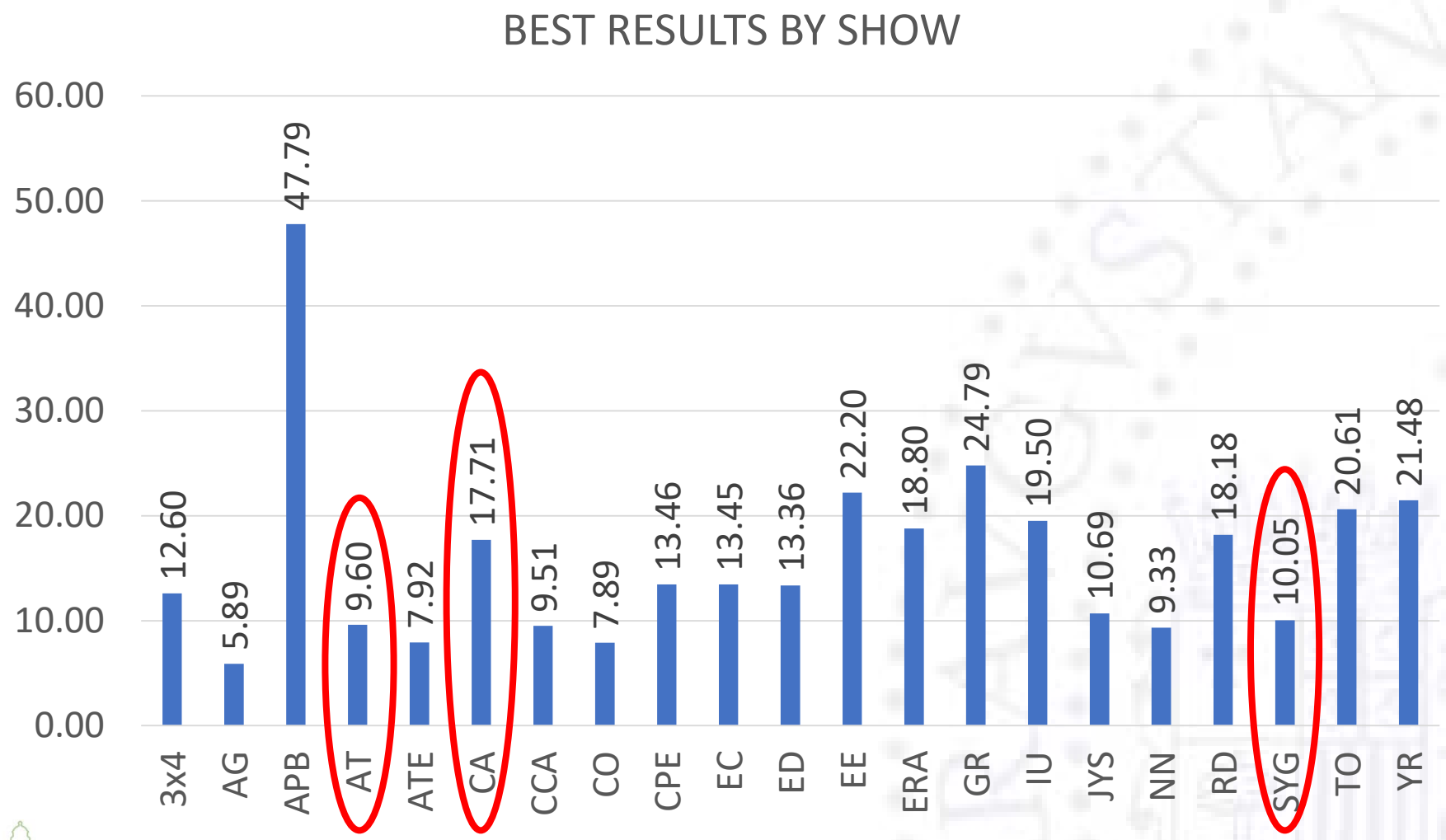
Previous challenges results

2020:
WER 16,04%

Best 2020 system on 2018 test
11,6%



2022 challenge WER: 14,35%

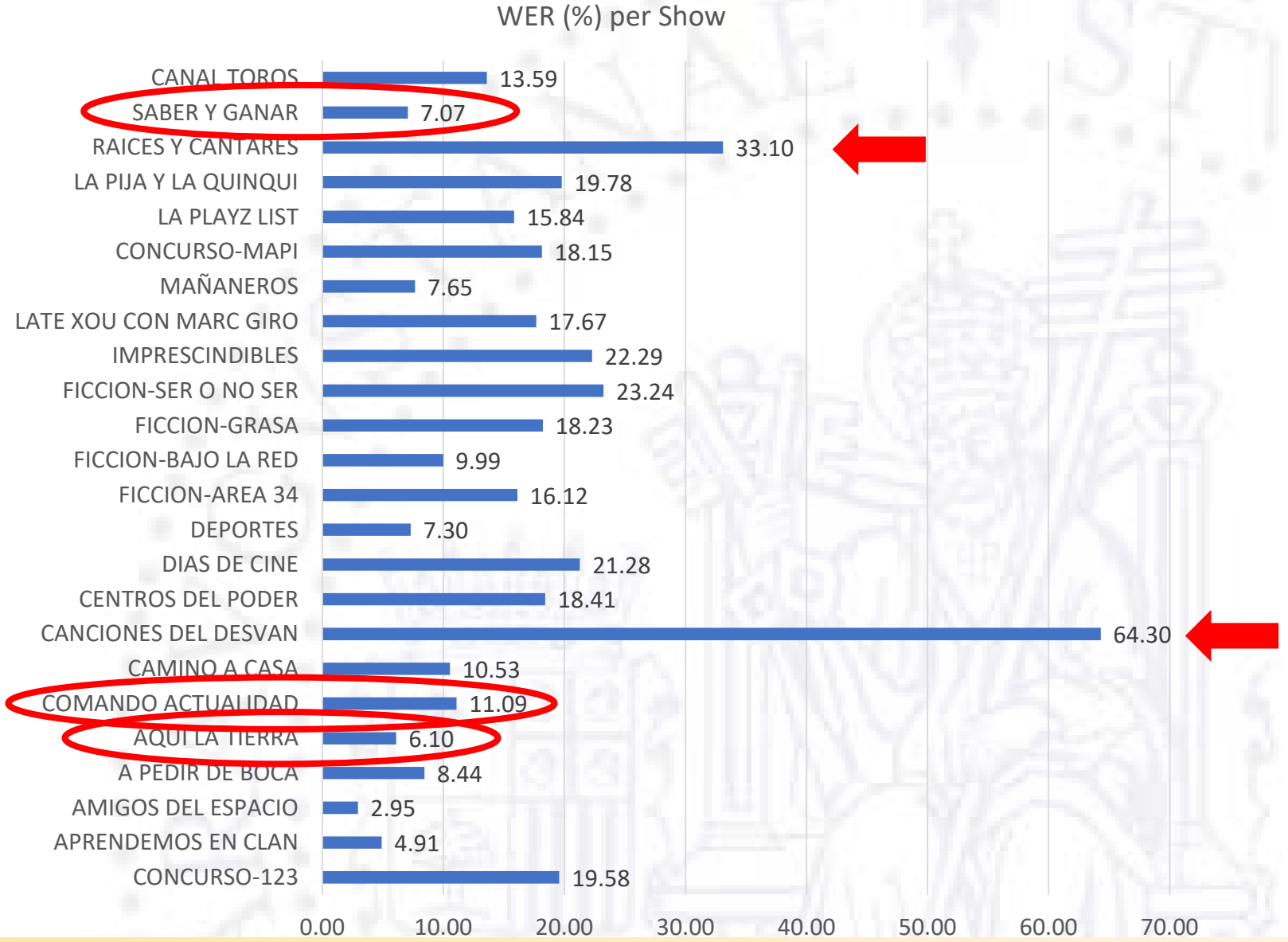


Show	Hours
3x4	2:58:17
A pedir de boca (APB)	3:41:38
Agrosfera (AG)	4:15:20
Aquí la Tierra (AT)	2:46:44
Ateneo (ATE)	1:40:09
Cerámica Popular Española (CPE)	1:02:35
Comando Actualidad (CA)	3:59:29
Conversatorios Casa América (CCA)	1:58:44
Corazón (CO)	3:00:17
El cazador (EC)	3:48:22
Encuestas ruido ambiente (ERA)	2:08:13
Entrevistas en bruto (EE)	3:54:57
España Directo (ED)	4:05:57
Fiction (Grasa-GR, Yrreal-YR, Riders_RD)	3:53:22
Informativos UMATIC (IU)	0:59:49
Jara y Sedal (JYS)	2:29:17
Noticias Nacional (NN)	2:14:32
Saber y Ganar (SYG)	4:28:28
Toros (TO)	0:49:57
21 different shows	**54:16:07**

2024 challenge

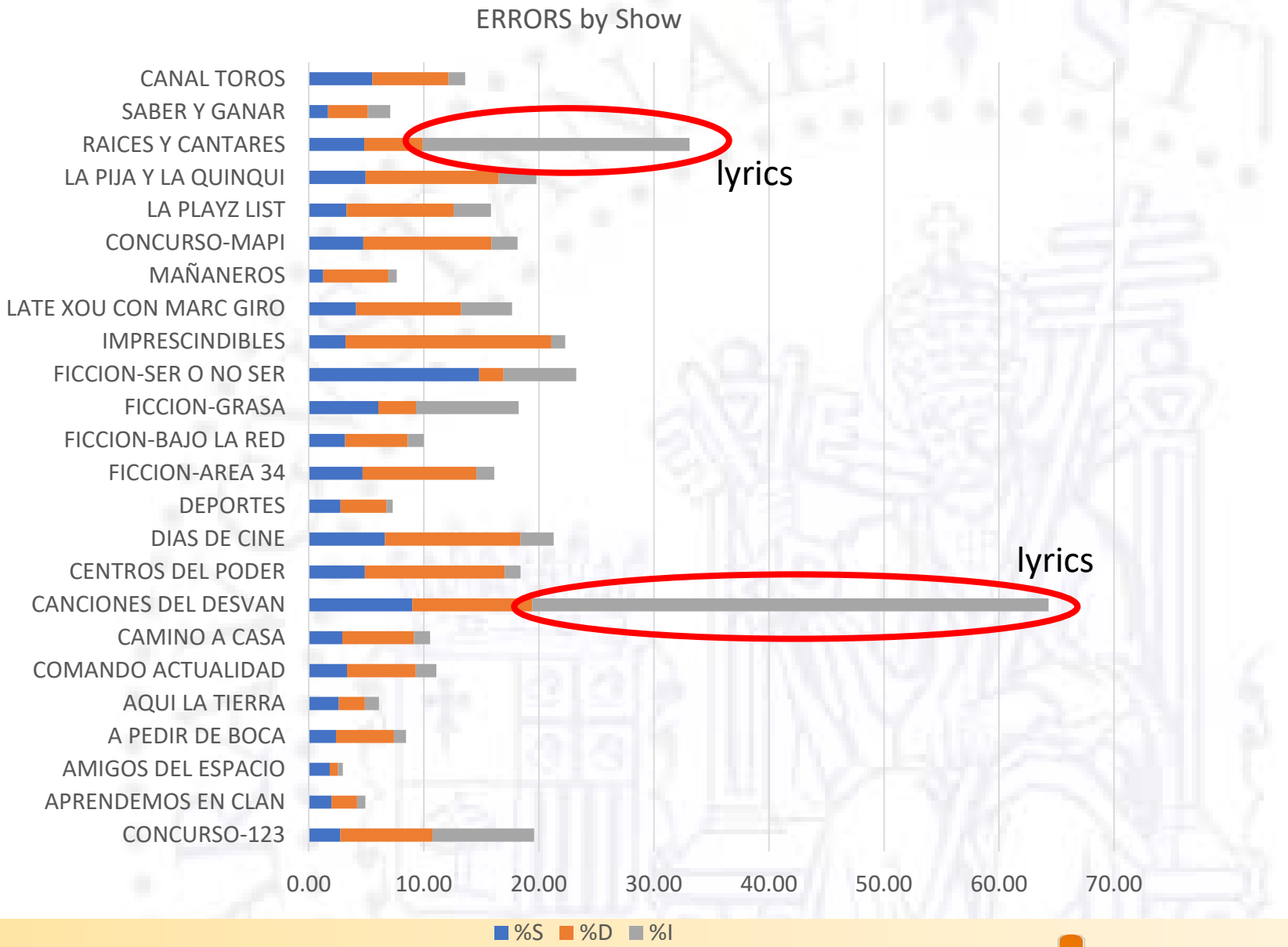
WER	2018	2020	2022	2024
SYG	14,47	-	10,05	7,07
CA	-	20,90	17,71	11,09
AT	-	13,93	9,60	6,10

WER Baseline WhisperX large-v3:
2022: 11,40 %
2024: 12,10 %

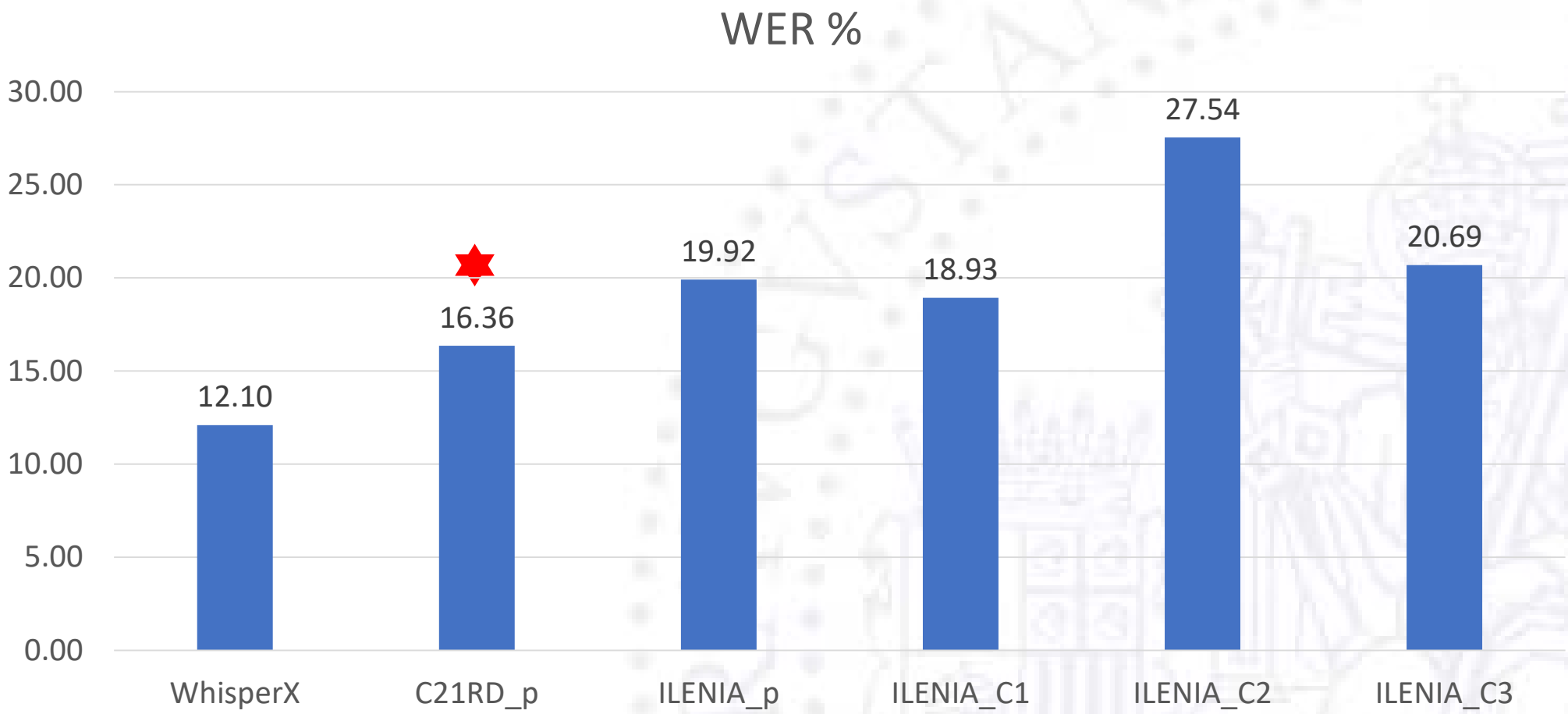


ERRORs Baseline WhisperX large-v3

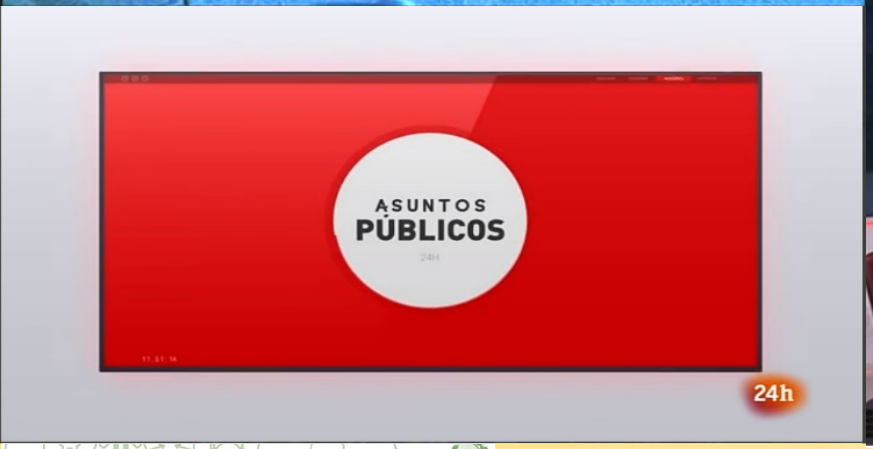
Error	%
Substitutions	3,08
Deletions	**6,57**
Insertions	2,45
WER	12,10



2024 challenge



Speaker Diarization and Identity Assignment Challenge



2024 edition tasks

Evaluate:

- Automatic algorithms for segmenting and clustering speakers in a given audio
- Methods for assigning previously known identities to each speech segment

Evaluation dataset

14 different shows with a total of 14 hours

Baseline:

WhisperX with pyannote/speaker-diarization-3.1

Diarization Error Rate:

Fraction of speaker time not correctly attributed to a specific speaker

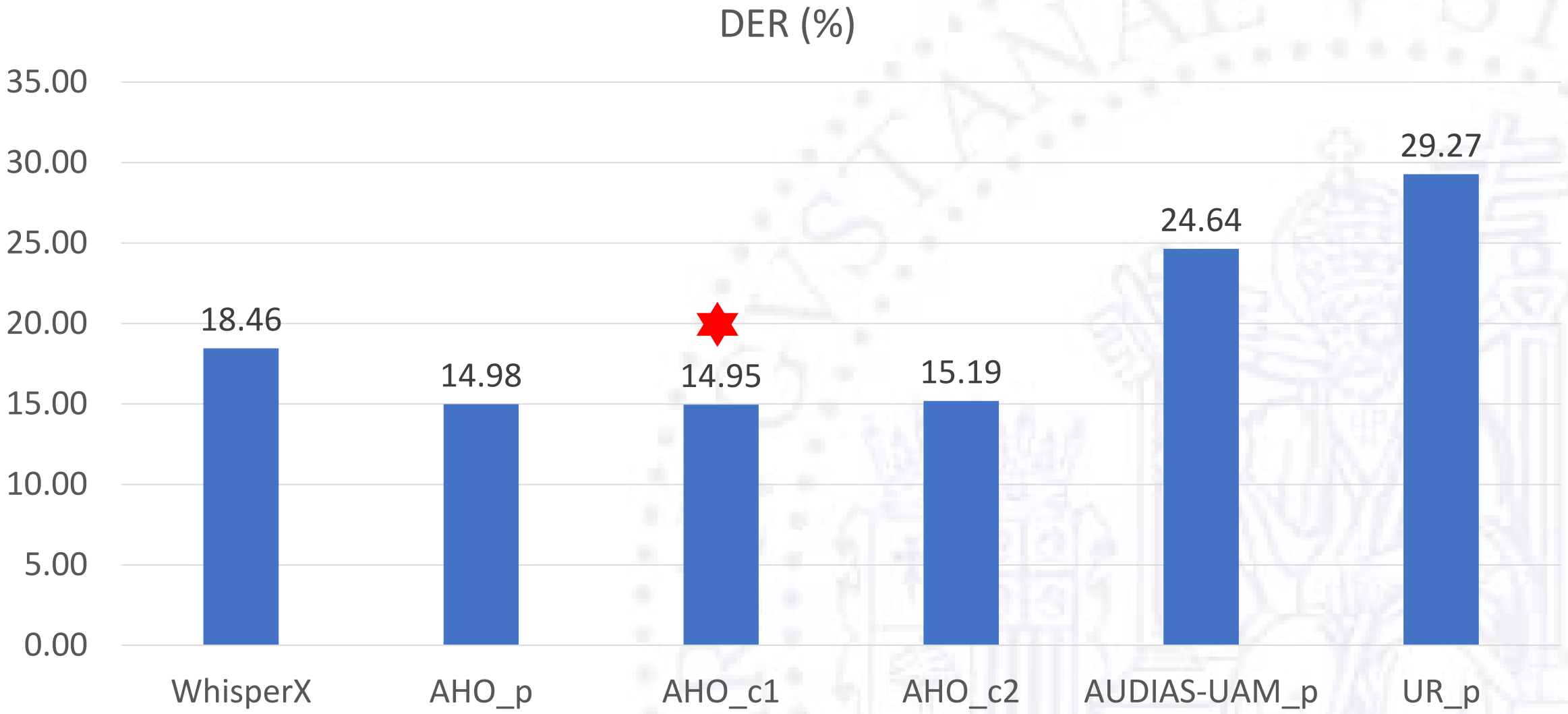
$$DER = \frac{T_{MISS} + T_{FA} + T_{SPK}}{T_{SPEECH}}$$

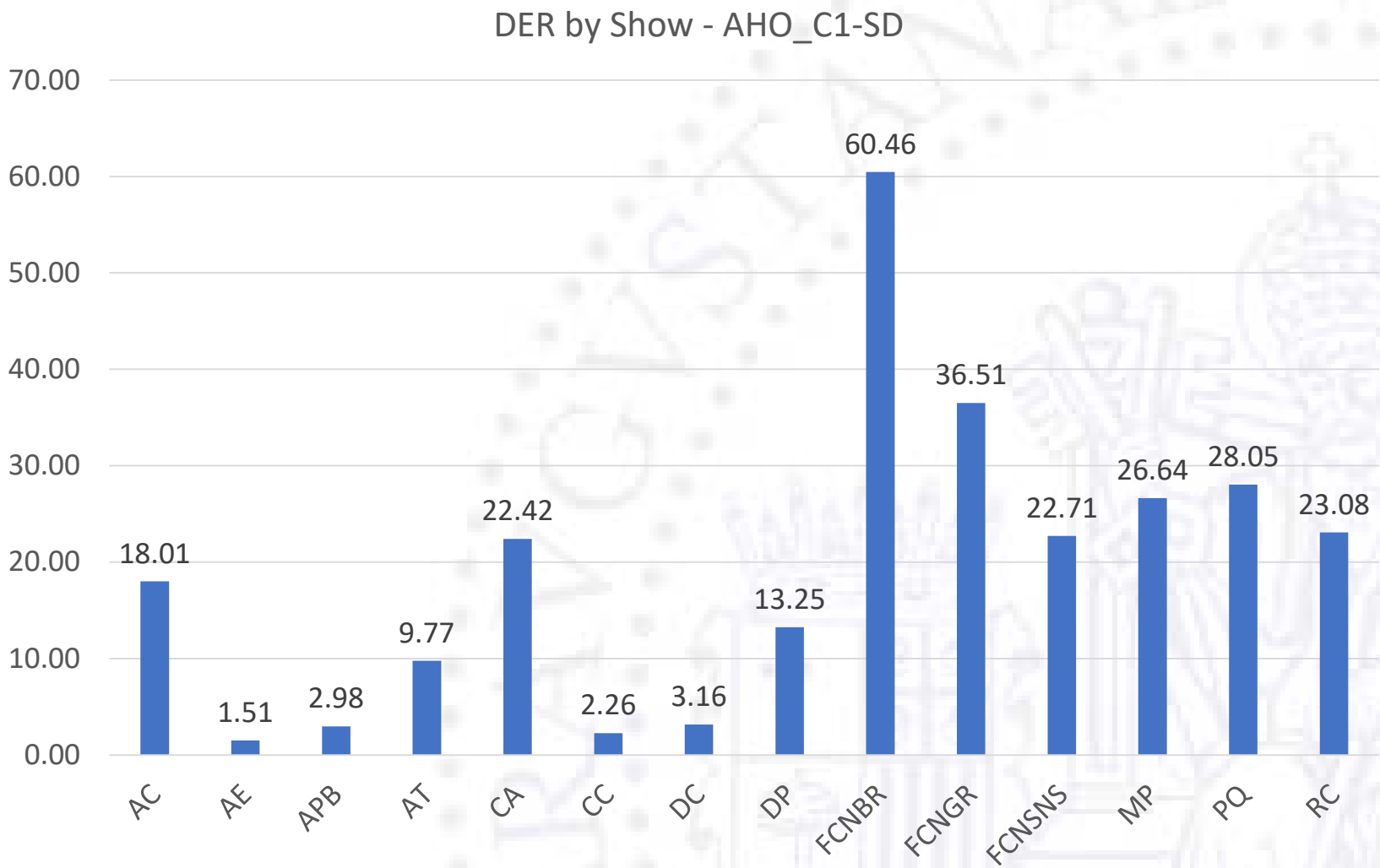
T_{MISS}: Amount of speech considered as non speech

T_{FA}: Amount of non speech considered as speech

T_{SPK}: Amount of speech assigned to a wrong speaker
(contains overlap regions which are evaluated)

250 ms forgiveness collar


DER (%)
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
18.46
14.98
14.95
15.19
24.64
29.27
WhisperX
AHO_p
AHO_c1
AHO_c2
AUDIAS-UAM_p
UR_p


DER by Show - AHO_C1-SD
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
18.01
1.51
2.98
9.77
22.42
2.26
3.16
13.25
60.46
36.51
22.71
26.64
28.05
23.08
AC
AE
APB
AT
CA
CC
DC
DP
FCNBR
FCNGR
FCNSNS
MP
PQ
RC

No participation

The Future of Albayzín Evaluations