Programme Overview

Start	End	22-Oct	23-Oct	24-Oct
8:00	8:30	Toturials		
8:30	9:00		Technical Sessions	Technical Sessions
9:00	9:30			
9:30	10:00			
10:00	10:30		Break	Break
10:30	11:00		Technical Sessions	Technical Sessions
11:00	11:30			
11:30	12:00	Welcome Reception & Opening		
12:00	12:30		Lunch	Lunch
12:30	13:00		WiAPSIPA	
13:00	13:30			
13:30	14:00	Keynote 1	Keynote 2	Keynote 3
14:00	14:30			
14:30	15:00	Technical Sessions	Technical Sessions	Technical Sessions
15:00	15:30			
15:30	16:00			
16:00	16:30	Break	Break	Break
16:30	17:00	Technical Sessions	Technical Sessions	Founders Forum & Closing
17:00	17:30			
17:30	18:00			
18:00	18:30			
18:30	19:00			
19:00	19:30		Banquet	
19:30	20:00			
20:00	20:30			
20:30	21:00			
21:00	21:30			

Onsite registration opens at 10am on 22 Oct.

Best Paper Finalists are given 12-min presentation 3-min Q&A at Conference Awards Sessions I & II

All contributed full papers are given 5-min presentation and 30-min poster interaction (max A0 portrait) in the same room.

All late-breaking reports are given 5-min presentation only.

All authors are required to report to their respective sessions at least 15 mins ahead of the session to upload and test their slides.

Paper Allocations in Technical Sessions

Session	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Active Noise Control I	56	57	72	124	285	289	301	345	511	553	610	626
Active Noise Control II	63	87	134	265	305	310	330	340	458	490	497	513
Advanced Multimedia Applications	113	158	297	364	367	377	382	423	473	516		
Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement	51	184	186	201	234	293	418	439	442	474	501	507
Advanced Signal Processing and Resource Management for Sensing and Communication	144	167	169	219	220	341	400	416				
Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond	58	103	117	157	211	258	275	328	389	398	459	461
Advanced Topics on Music Processing	177	283	298	319	337	349	356	424	441	453	496	517
Advances in Multimodal AI for Multimedia Applications	59	68	96	245	249	252	268	317	354	359	436	
Audio Processing	129	140	174	408	409	434	492	583				
Best Paper Finalists	105	402	69	284	36	355	530	586				
Biomedical Signal Processing and Systems I	26	148	173	214	240	241	336	407	413	415	447	227
Biomedical Signal Processing and Systems II	454	466	531	535	591	630	631	632				
Deep Learning: Algorithm, Implementations, and Applications	49	55	84	108	185	372	385	386	387	422	470	
Emerging Technologies and Applications of Image Processing and Computer Vision	54	143	171	196	269	332	445	518	548	568	572	
Generative AI Models for Vision-Based Applications	91	92	147	242	251	533	536	561	615			
Grand Challenge & Audio Processing	110	138	139	228	401	588	634	642	648	651	652	653
Interactive, Natural, Expressive and Robust Conversation System	42	60	73	78	102	112	152	199	217	321	483	
Machine Learning: Algorithms and Application I	22	47	216	229	274	294	313	361	576	578	609	
Machine Learning: Algorithms and Application II	83	207	232	309	343	352	370	412	460	557	644	
Machine Learning: Algorithms and Application III	35	239	291	311	312	314	411	457	493	520	559	
Machine Learning: Information and Medical Applications	24	260	278	300	303	324	431	558	580	594	595	
Multimedia Security & Forensics	104	135	215	348	379	410	464					
Multimodal AI	23	155	193	299	510	541	550	562				
Music and Singing	71	94	141	307	362	391	421	451	482	498	569	
Privacy and Security in Speech AI	86	133	194	200	417	495	515	538	547	608	623	
Recent Advances in Multimedia Enrichment, Security and Privacy	89	109	136	192	210	218	225	292	320	325	469	
Recent Advances in Multimodal Learning for Computer Vision	52	53	160	205	302	334	363	375	388	399	604	
Research Frontiers in Learned Visual Data Coding and Processing	64	106	161	204	335	383	465	468	532	539	621	
Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)	115	153	206	267	481	505	523	534	563	599	602	
Signal & Information Processing I	46	61	66	76	159	202	244	259	271	318	627	
Signal & Information Processing II	137	322	347	381	393	394	455	480				
Speaker Modeling Beyond Speaker Recognition	41	81	101	131	132	182	189	197	333	590		
Speech and Language Processing I	95	127	142	150	179	213	236	253	262	446	571	
Speech and Language Processing II	97	118	120	154	264	308	331	338	440	587	603	
Speech and Language Processing III	27	130	163	168	191	208	235	237	248	261	414	
Speech and Language Processing IV	29	116	449	478	502	503	514	551	560	577	641	
Speech Communication and Media Generation	256	272	344	384	427	428	450	452	472	487	616	
Speech Recognition and Understanding I	195	282	390	406	426	443	475	524	592	593		
Speech Recognition and Understanding II	151	226	238	280	296	342	392	405	542	556	575	
Speech Synthesis, Enhancement, and Recognition	17	48	98	149	243	304	491	500	554	600		
Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents	126	187	188	203	212	254	276	327	508	509		
Wireless Communications & Networking	39	50	74	128	209	270	287	290	351	456		

22 October

Start	End	Island Ballroom	Lotus I	Lotus II	Hibiscus III	Peony I	Peony II
8:00	8:30	Exhibition Setup	Tutorial 1: Recent Advances in End-to-End Learned Image and Video Coding	Tutorial 2: Deep Speaker Modeling: Theories, Applications and Practice	Tutorial 3: From Detection to Direction: An Overview of Sound Event Localization and Detection	Tutorial 4: Adaptive Sensor Networks in Digital Health	Conference Subcomm Meeting
8:30	9:00						
9:00	9:30						
9:30	10:00						
10:00	10:30	Registration					
10:30	11:00						
11:00	11:30						
11:30	12:00	Welcome Reception & Opening Ceremony					
12:00	12:30						
12:30	13:00						
13:00	13:30						
13:30	14:00	Keynote 1 by Emanuël Habets					
14:00	14:30						
14:30	15:00	Perspective 1: Voice Privacy & Security	Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement	Interactive, Natural, Expressive and Robust Conversation System	Generative AI Models for Vision-Based Applications	Multimedia Security & Forensics	Wireless Communications & Networking
15:00	15:30						
15:30	16:00						
16:00	16:30	Break					
16:30	17:00	Perspective 2: Utilization of the Foundation Models and the Future	Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents	Speaker Modeling Beyond Speaker Recognition	Three-Minute Thesis (3MT) Competition	OC Meeting	Advanced Multimedia Applications
17:00	17:30						
17:30	18:00						
18:00	18:30						
18:30	19:00						
19:00	19:30		APSIPA BoG Meeting				
19:30	20:00						
20:00	20:30						
20:30	21:00						
21:00	21:30						

23 October

Start	End	Island Ballroom	Lotus I	Lotus II	Hibiscus III	Peony I	Peony II
8:00	8:30	Registration					
8:30	9:00	Active Noise Cancellation Workshop Keynote	Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond	Speech and Language Processing I	Research Frontiers in Learned Visual Data Coding and Processing	Multimodal AI	Biomedical Signal Processing and Systems I
9:00	9:30						
9:30	10:00						
10:00	10:30	Break					
10:30	11:00	Active Noise Cancellation Panel Discussion	Advanced Topics on Music Processing	Speech and Language Processing II	Emerging Technologies and Applications of Image Processing and Computer Vision	Biomedical Signal Processing and Systems II	Machine Learning: Algorithms and Application I
11:00	11:30						
11:30	12:00						
12:00	12:30	Lunch					
12:30	13:00	Women in APSIPA Forum	TC Meeting	TC Meeting	TC Meeting	TC Meeting	TC Meeting
13:00	13:30						
13:30	14:00	Keynote 2 by Jane Wang					
14:00	14:30						
14:30	15:00	Perspective 3: Neural Speech Assessment and Its Application	Active Noise Control I	Speech and Language Processing III	Machine Learning: Information and Medical Applications	Audio Processing	Signal & Information Processing I
15:00	15:30						
15:30	16:00						
16:00	16:30	Break					
16:30	17:00	Education Forum	Active Noise Control II	Speech and Language Processing IV	Recent Advances in Multimedia Enrichment, Security and Privacy	OC Meeting	Advances in Multimodal AI for Multimedia Applications
17:00	17:30						
17:30	18:00						
18:00	18:30						
18:30	19:00						
19:00	19:30	Banquet					
19:30	20:00						
20:00	20:30						
20:30	21:00						
21:00	21:30						

24 October

Start	End	Island Ballroom	Lotus I	Lotus II	Hibiscus III	Peony I	Peony II
8:00	8:30	Registration					
8:30	9:00	Sadaoki Furui Prize Papers and Conference Award Session I	Music and Singing	Speech Communication and Media Generation	Deep Learning: Algorithm, Implementations, and Applications	Advanced Signal Processing and Resource Management for Sensing and Communication	Machine Learning: Algorithms and Application II
9:00	9:30						
9:30	10:00						
10:00	10:30	Break					
10:30	11:00	Conference Award Session II	Grand Challenge & Audio Processing	Speech Recognition and Understanding I	Recent Advances in Multimodal Learning for Computer Vision	Signal & Information Processing II	Machine Learning: Algorithms and Application III
11:00	11:30						
11:30	12:00						
12:00	12:30	Lunch					
12:30	13:00		TC Meeting	TC Meeting	TC Meeting	TC Meeting	TC Meeting
13:00	13:30						
13:30	14:00	Keyntoe 3 by Xiao-Ping (Steven) Zhang					
14:00	14:30						
14:30	15:00	Perspective 4: Leveraging LLMs in Interdisciplinary Study of Speech and Language	Speech Synthesis, Enhancement, and Recognition	Speech Recognition and Understanding II	Privacy and Security in Speech AI	Late Breaking Reports	Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)
15:00	15:30						
15:30	16:00						
16:00	16:30	Break					
16:30	17:00	APSIPA General Assembly, Founders' Forum & Closing Ceremony					
17:00	17:30						
17:30	18:00						

Date	Start	End	Island Ballroom	Chair
	8:00	10:00		
	10:00	10:30	Set Up	
	10:30	11:30		
22	11:30	13:30	Welcome Reception & Opening Ceremony	Woon Seng Gan
Oct	13:30	14:30	Keynote 1 by Emanuël Habets	Shoji Makino
2025	14:30	16:00	Perspective 1: Voice Privacy & Security	Liping Chen
Wed	16:00	16:30	Break	
	16:30	18:00	Perspective 2: Utilization of the Foundation Models and the Future	Isao Echizen
	18:00	19:00		
	19:00	21:30		
	8:30	10:00	Active Noise Cancellation Workshop Keynote	Woon Seng Gan
	10:00	10:30	Break	
	10:30	12:00	Active Noise Cancellation Panel Discussion	Woon Seng Gan
23	12:00	13:30	Women in APSIPA Forum (12.20-13.30)	Bonnie Law
Oct	13:30	14:30	Keynote 2 by Jane Wang	Jay Kuo
2025	14:30	16:00	Perspective 3: Neural Speech Assessment and Its Application	Hsin-Min Wang
Thu	16:00	16:30	Break	
	16:30	18:00	Education Forum	Anthony Kuh
	18:00	19:00	Turnover for Banquet	
	19:00	21:30	Banquet	
	8:30	10:00	Sadaoki Furui Prize Papers and Conference Award Session I	Sang Hoon Lee
	10:00	10:30	Break	
24	10:30	12:00	Conference Award Session II	Yan Wu
Oct	12:00	13:30	Lunch	
2025	13:30	14:30	Keyntoe 3 by Xiao-Ping (Steven) Zhang	Haizhou Li
Fri	14:30	16:00	Perspective 4: Leveraging LLMs in Interdisciplinary Study of Speech and Language	Tan Lee
	16:00	16:30	Break	
	16:30	18:00	APSIPA General Assembly, Founders' Forum & Closing Ceremony	Woon Seng Gan

Date	Start	End	Lotus I	Chair	Co-Chair
	8:00	10:00			
	10:00	10:30	Tutorial 1: Recent Advances in End-to-End Learned Image and Video Coding		
	10:30	11:30			
22	11:30	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement	Shoji Makino	Kouei Yamaoka
Wed	16:00	16:30	Break		
	16:30	18:00	Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents	Haizhou Li	Madhusudan Singh
	18:00	19:00	*Turnover for BoG Meeting*		
	19:00	21:30	BOG Meeting	Woon Seng Gan	
	8:30	10:00	Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond	Nobutaka Ono	Keisuke Imoto
	10:00	10:30	Break		
	10:30	12:00	Advanced Topics on Music Processing	Tetsuro Kitahara	Eita Nakamura
23	12:00	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Active Noise Control I	Bai Mingsian	Shi Dong Yuan
Thu	16:00	16:30	Break		
	16:30	18:00	Active Noise Control II	Thushara Abhayapala	Jiancheng Tao
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Music and Singing	Soham Gangopadhyay	Tomohiko Nakamura
	10:00	10:30	Break		
24	10:30	12:00	Grand Challenge & Audio Processing	Ramesh R	Yang Xiao
Oct	12:00	13:30			
2025	13:30	14:30			
Fri	14:30	16:00	Speech Synthesis, Enhancement, and Recognition	Dyah A. M. G. Wisnu	Saisha Suresh Bore
	16:00	16:30	Break		
	16:30	18:00			

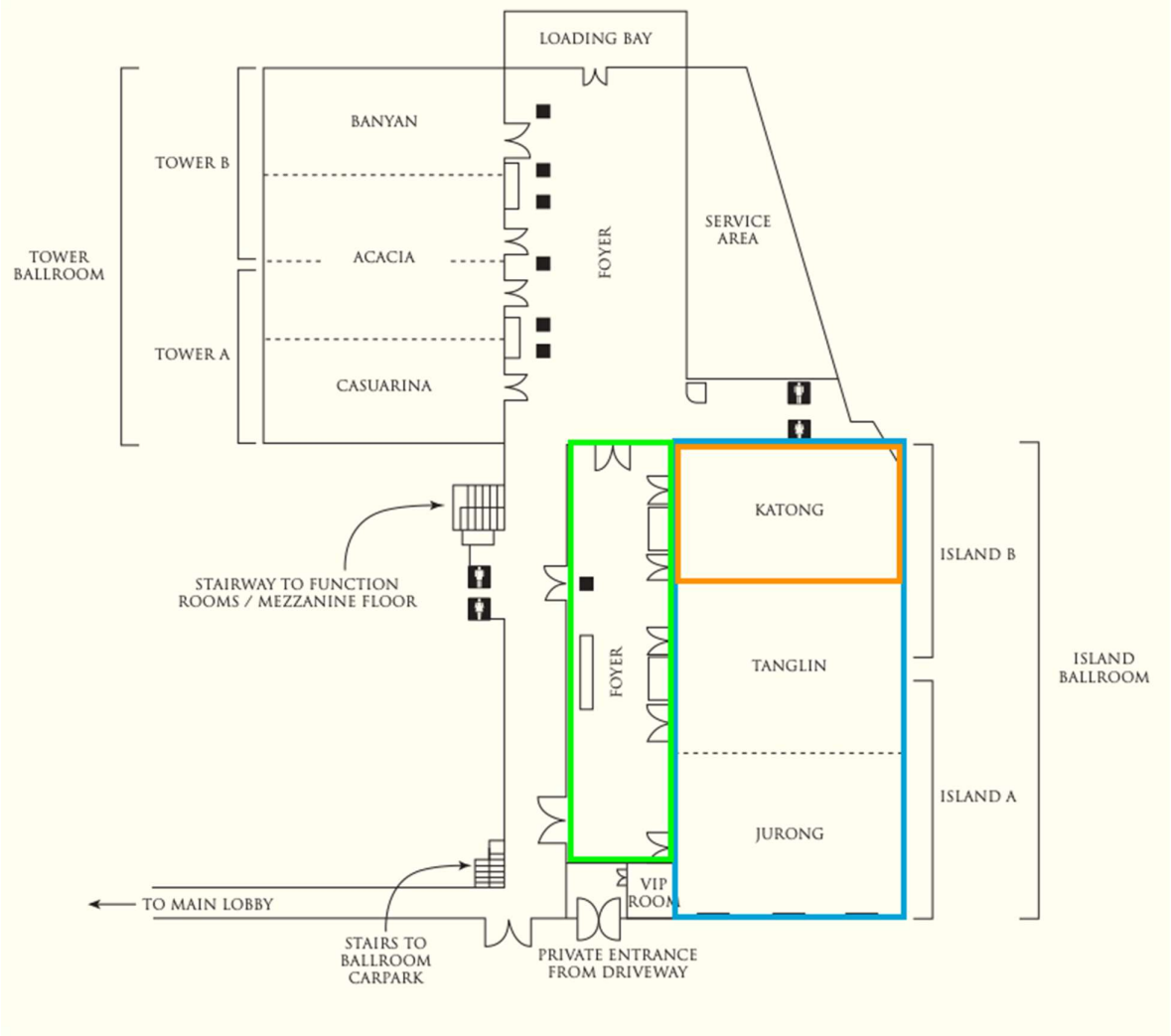
Date	Start	End	Lotus II	Chair	Co-Chair
	8:00	10:00			
	10:00	10:30	Tutorial 2: Deep Speaker Modeling: Theories, Applications and Practice		
	10:30	11:30			
22	11:30	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Interactive, Natural, Expressive and Robust Conversation System	Tetsuji Ogawa	Tzu-Quan Lin
Wed	16:00	16:30	Break		
	16:30	18:00	Speaker Modeling Beyond Speaker Recognition	Shuai Wang	Yanmin Qian
	18:00	19:00	*Turnover for BoG Meeting*		
	19:00	21:30	BOG Meeting	Woon Seng Gan	
	8:30	10:00	Speech and Language Processing I	Yang Ai	Eliathamby Ambikairajah
	10:00	10:30	Break		
	10:30	12:00	Speech and Language Processing II	Ashish Panda	Bagus Tris Atmaja
23	12:00	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Speech and Language Processing III	Liping Chen	Ivan Kukanov
Thu	16:00	16:30	Break		
	16:30	18:00	Speech and Language Processing IV	Xin Wang	Madhusudan Singh
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Speech Communication and Media Generation	Wen-Chin Huang	Kengo Ohta
	10:00	10:30	Break		
24	10:30	12:00	Speech Recognition and Understanding I	Sayaka Shiota	Virender kadyan
Oct	12:00	13:30			
2025	13:30	14:30			
Fri	14:30	16:00	Speech Recognition and Understanding II	Yu Tsao	Huijing Zhan
	16:00	16:30	Break		
	16:30	18:00			

Date	Start	End	Hibiscus III	Chair	Co-Chair
	8:00	10:00			
	10:00	10:30	Tutorial 3: From Detection to Direction: An Overview of Sound Event Localization and Detection		
	10:30	11:30			
22	11:30	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Generative AI Models for Vision-Based Applications	Po-Chyi Su	Hsiao-Chi Li
Wed	16:00	16:30	Break		
	16:30	18:00	Three-Minute Thesis (3MT) Competition	Woon Seng Gan	
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Research Frontiers in Learned Visual Data Coding and Processing	Wen-Hsiao Peng	Xiem Hoang
	10:00	10:30	Break		
	10:30	12:00	Emerging Technologies and Applications of Image Processing and Computer Vision	Chang-Su Kim	Keunsoo Ko
23	12:00	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Machine Learning: Information and Medical Applications	Kazushi Ikeda	Ting Dang
Thu	16:00	16:30	Break		
	16:30	18:00	Recent Advances in Multimedia Enrichment, Security and Privacy	Masaki Kawamura	Shoko Imaizumi
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Deep Learning: Algorithm, Implementations, and Applications	Shoji Makino	Shogo MURAMATSU
	10:00	10:30	Break		
24	10:30	12:00	Recent Advances in Multimodal Learning for Computer Vision	Akshita Abrol	Stephen Karungaru
Oct	12:00	13:30			
2025	13:30	14:30			
Fri	14:30	16:00	Privacy and Security in Speech AI	Tomoki Toda	Yiming Wang
	16:00	16:30	Break		
	16:30	18:00			

Date	Start	End	Peony I	Chair	Co-Chair
	8:00	10:00			
	10:00	10:30	Tutorial 4: Adaptive Sensor Networks in Digital Health		
	10:30	11:30			
22	11:30	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Multimedia Security & Forensics	Hong Huy Nguyen	Mehmet Sinan YILDIRIM
Wed	16:00	16:30	Break		
	16:30	18:00	OC Meeting		
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Multimodal AI	Chi-Chun (Jeremy) Lee	Keunsoo Ko
	10:00	10:30	Break		
	10:30	12:00	Biomedical Signal Processing and Systems II	Ingon Chanpornpakdi	Kazushi Ikeda
23	12:00	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Audio Processing	Jihui (Aimee) Zhang	Huawei Zhang
Thu	16:00	16:30	Break		
	16:30	18:00	OC Meeting		
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Advanced Signal Processing and Resource Management for Sensing and Communication	Koichi Adachi	Kazunori Hayashi
	10:00	10:30	Break		
24	10:30	12:00	Signal & Information Processing II	Candy Olivia Mawalim	Candy Olivia Mawalim
Oct	12:00	13:30			
2025	13:30	14:30			
Fri	14:30	16:00	Late Breaking Reports (638 & 654-6667)	Ying-Ren Chien	Jihui Aimee Zhang
	16:00	16:30	Break		
	16:30	18:00			

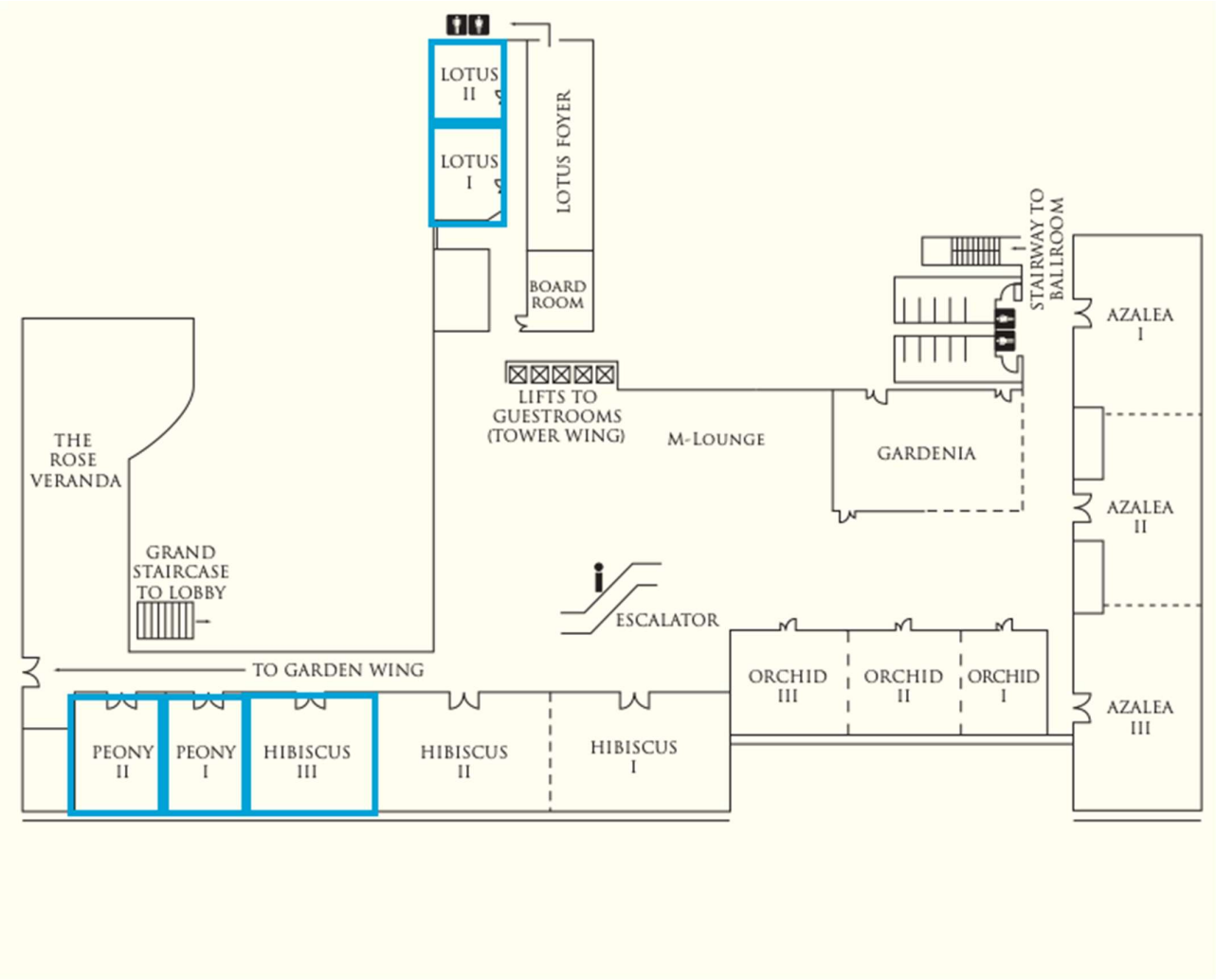
Date	Start	End	Peony II	Chair	Co-Chair
	8:00	10:00			
	10:00	10:30	Conference Subcomm Meeting		
	10:30	11:30			
22	11:30	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Wireless Communications & Networking	Kouji Hirata	Shu-Yu Lin
Wed	16:00	16:30	Break		
	16:30	18:00	Advanced Multimedia Applications	Zhi Hu	Dayan Perera
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Biomedical Signal Processing and Systems I	Jetsada Arnin	Wen Zhang
	10:00	10:30	Break		
	10:30	12:00	Machine Learning: Algorithms and Application I	Jia-Ching Wang	Guo-Shiang Lin
23	12:00	13:30			
Oct	13:30	14:30			
2025	14:30	16:00	Signal & Information Processing I	Dipanita Chakraborty	S. Nimishan
Thu	16:00	16:30	Break		
	16:30	18:00	Advances in Multimodal AI for Multimedia Applications	Phat Nguyen	Chen-Kuo Chiang
	18:00	19:00			
	19:00	21:30			
	8:30	10:00	Machine Learning: Algorithms and Application II	Junya Hara	Yuichi Tanaka
	10:00	10:30	Break		
24	10:30	12:00	Machine Learning: Algorithms and Application III	Koichi Shinoda	Chih-Yi Chiu
Oct	12:00	13:30			
2025	13:30	14:30			
Fri	14:30	16:00	Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)	Yang Xiao	Jisheng Bai
	16:00	16:30	Break		
	16:30	18:00			

Floorplan for Island Ballroom



Registration: Foyer of Island Ballroom
Plenary Sessions: Island Ballroom
Banquet: Island Ballroom
Exhibition: Katong

Floorplan for Mezzanine Floor



Parallel Sessions:

- **Lotus I**
- **Lotus II**
- **Hibiscus III**
- **Peony I**
- **Peony II**

2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

22-24 October 2025, Singapore

Technical Program

Day 1 | Day 2 | Day 3

Day 1: Wednesday, 22 Oct 2025 - Overview	
08:00-10:00	D1-0800_IB **Exhibition Setup** Location: Island Ballroom
08:00-11:30	D1-0800_L1 **Tutorial 1: Recent Advances in End-to-End Learned Image and Video Coding** Location: Lotus I
08:00-11:30	D1-0800_L2 **Tutorial 2: Deep Speaker Modeling: Theories, Applications and Practice** Location: Lotus II
08:00-11:30	D1-0800_H3 **Tutorial 3: From Detection to Direction: An Overview of Sound Event Localization and Detection** Location: Hibiscus III
08:00-11:30	D1-0800_P1 **Tutorial 4: Adaptive Sensor Networks in Digital Health** Location: Peony I
10:00-11:30	D1-1000_IB **Registration** Location: Island Ballroom
11:30-13:30	D1-1130_IB **Welcome Reception & Opening Ceremony** Location: Island Ballroom
13:30-14:30	D1-1330_IB **Keynote 1 by Emanuël Habets** Location: Island Ballroom
14:30-16:00	D1-1430_IB **Perspective 1: Voice Privacy & Security** Location: Island Ballroom
14:30-16:00	D1-1430_L1 **Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement** Location: Lotus I
14:30-16:00	D1-1430_L2 **Interactive, Natural, Expressive and Robust Conversation System** Location: Lotus II

14:30-16:00	D1-1430_H3 **Generative AI Models for Vision-Based Applications** Location: Hibiscus III
14:30-16:00	D1-1430_P1 **Multimedia Security & Forensics** Location: Peony I
14:30-16:00	D1-1430_P2 **Wireless Communications & Networking** Location: Peony II
16:00-16:30	**Break**
16:30-18:00	D1-1630_IB **Perspective 2: Utilization of the Foundation Models and the Future** Location: Island Ballroom
16:30-18:00	D1-1630_L1 **Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents** Location: Lotus I
16:30-18:00	D1-1630_L2 **Speaker Modeling Beyond Speaker Recognition** Location: Lotus II
16:30-18:00	D1-1630_H3 **Three-Minute Thesis (3MT) Competition** Location: Hibiscus III
16:30-18:00	D1-1630_P2 **Advanced Multimedia Applications** Location: Peony II

Day 1: Wednesday, 22 Oct 2025 - With Papers	
08:00-10:00	D1-0800_IB **Exhibition Setup** Location: Island Ballroom
08:00-11:30	D1-0800_L1 **Tutorial 1: Recent Advances in End-to-End Learned Image and Video Coding** Location: Lotus I
08:00-11:30	D1-0800_L2 **Tutorial 2: Deep Speaker Modeling: Theories, Applications and Practice** Location: Lotus II
08:00-11:30	D1-0800_H3 **Tutorial 3: From Detection to Direction: An Overview of Sound Event Localization and Detection** Location: Hibiscus III
08:00-11:30	D1-0800_P1 **Tutorial 4: Adaptive Sensor Networks in Digital Health** Location: Peony I
10:00-11:30	D1-1000_IB **Registration** Location: Island Ballroom
11:30-13:30	D1-1130_IB **Welcome Reception & Opening Ceremony** Location: Island Ballroom
13:30-14:30	D1-1330_IB **Keynote 1 by Emanuël Habets** Location: Island Ballroom

14:30-16:00	**D1-1430_IB Perspective 1: Voice Privacy & Security** Location: Island Ballroom **D1-1430_IB.1** 635 Advancing Speech Quality Assessment Through Scientific Challenges and Open-source Activities *Wen-Chin Huang* **D1-1430_IB.2** 637 Progress and Challenges in DNN-based Objective Quality Assessment of Synthesized Speech *Erica Cooper* **D1-1430_IB.3** 643 Foundation Models as Guardrails: LLM- and VLM-Based Approaches to Safety and Alignment *Huy Nguyen, Pride Kavumba, Tomoya Kurosawa, Koki Wataoka* **D1-1430_IB.4** 645 Speaker Privacy and Security in the Big Data Era: Protection and Defense against Deepfake *Liping Chen, Kong Aik Lee, Zhen-Hua Ling, Xin Wang, Rohan Kumar Das, Tomoki Toda, Haizhou Li* **D1-1430_IB.5** 646 Non-Intrusive Intelligibility Prediction for Hearing Aids: Recent Advances, Trends, and Challenges *Ryandhimas Zezario* **D1-1430_IB.6** 647 From Evaluation to Optimization: Neural Speech Assessment for Downstream Applications *Yu Tsao* **D1-1430_IB.7** 649 Normalization through Fine-tuning: Understanding Wav2vec2.0 Embeddings for Phonetic Analysis *Yiming Wang, Jiahong Yuan* **D1-1430_IB.8** 668 Enabling Internationalization of Affective Speech Technology using LLMs *Bo-Hao Su, Shinji Watanabe, Chi-Chun Lee*

14:30-16:00	**D1-1430_L1 Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement** Location: Lotus I **D1-1430_L1.1** 51 MVDR beamforming for underdetermined sound source separation using iterative PSD estimation in beamspace *Jin Xuan Teh, Yusuke Hioka* **D1-1430_L1.2** 184 Exploring Dual-Mode Training for Real-time Target Speaker Extraction *Li Li, Shogo Seki* **D1-1430_L1.3** 186 Switching Constant Separating Vector for Moving Source Extraction with Geometric Constraints *Changda Chen, Yichen Yang, Yuehao Zhao, Shoji Makino, Jingdong Chen* **D1-1430_L1.4** 201 Neural Network-Assisted Joint DOA Estimation and Beamforming with First-Order Reflection Modeling *Yichen Yang, Chao Pan, Qiang Gao, Jacob Benesty, Shoji Makino, Jingdong Chen* **D1-1430_L1.5** 234 Speaker Localization in Classroom Environments Using GCC-PHAT Features and Mamba State Space Models with Ad-hoc Microphone Arrays *Rashed Iqbal, Christian Ritz, Jack Yang, Sarah Howard* **D1-1430_L1.6** 293 Joint Separation and Tracking of Moving Sources With Distributed Microphone Arrays Based on Time-Varying Inertial Spatial Models *Ryunosuke Nihei, Yoshiaki Bando, Aditya Arie Nugraha, Diego Di Carlo, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii* **D1-1430_L1.7** 418 Visually-Informed Multichannel Sound Source Separation Based on 3D Gaussian Primitives *Haruki Asano, Ryunosuke Nihei, Yoshiaki Bando, Aditya Arie Nugraha, Diego Di Carlo, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii* **D1-1430_L1.8** 439 Joint Optimization of Sampling Rate Offsets and Demixing Filters Using Auxiliary Function Method *Hayato Takeuchi, Takao Kawamura, Nobutaka Ono, Shoko Araki* **D1-1430_L1.9** 442 First Demonstration of Acoustic Scene Classification Based on Trained Sound-to-Light Conversion *Shun Kotsugi, Takao Kawamura, Nobutaka Ono* **D1-1430_L1.10** 474 Auxiliary-Function-Based Decentralized Independent Vector Analysis for Distributed Microphone Arrays *Kouei Yamaoka, Katsuhiro Morita, Norihiro Takamune, Hiroshi Saruwatari* **D1-1430_L1.11** 501 Interactive Spatial Audio Rendering on Mobile Devices: A Two-Stage User Interface with Adaptive HRTF Selection and Real-Time Room Acoustics Simulation *Shravan Raghunath, Kanishk AL, Sailesh S, Rishabh Gupta, Saurav Gupta, Ramesh R* **D1-1430_L1.12** 507 Are Identical Sounds Present in Distributed Recordings to Serve as Spatio-Temporal Anchors? A Case Study Using the SINS Database *Takao Kawamura, Nobutaka Ono*

14:30-16:00	**D1-1430_L2 Interactive, Natural, Expressive and Robust Conversation System** Location: Lotus II **D1-1430_L2.1** 42 Language Adaptation Wake Word Spotting via Latent Space from Pre-trained Speech Models *Shifu Xiong, Hengshun Zhou, Kai Shen, Shi Cheng, Hang Chen, Genshun Wan, Kewei Li, Jun Du, Lirong Dai* **D1-1430_L2.2** 60 Identifying Speaker Information in Feed-Forward Layers of Self-Supervised Speech Transformers *Tzu-Quan Lin, Hsi-Chun Cheng, Hung-yi Lee, Hao Tang* **D1-1430_L2.3** 73 Multi-task Pretraining for Enhancing Interpretable L2 Pronunciation Assessment *Jiun-Ting Li, Bi-Cheng Yan, Yi-Cheng Wang, Berlin Chen* **D1-1430_L2.4** 78 End-to-End Integration of Speech Emotion Recognition and Voice Activity Detection with a Self-Supervised Model for Noise Robustness *Natsuo Yamashita, Masaaki Yamamoto, Yohei Kawaguchi* **D1-1430_L2.5** 102 SCSMT: A Multilingual Children's Speech Corpus for Singapore's Mother Tongues *Bowen Zhang, Nur Afiqah Abdul Latiff, Rong Tong, Donny Soh, Ian McLoughlin* **D1-1430_L2.6** 112 Reducing Orthographic Dependency on Paired Data by Probabilistic Integration via Syllabogram for Japanese Dialogue Speech Recognition *Ryu Takeda, Kazunori Komatani* **D1-1430_L2.7** 152 Expressive Prompting: Improving Emotion Intensity and Speaker Consistency in Zero-Shot TTS *Haoyu Wang, Chunyu Qiang, Tianrui Wang, Cheng Gong, Yu Jiang, Yuheng Lu, Chen Zhang, Longbiao Wang, Jianwu Dang* **D1-1430_L2.8** 199 Constructing an In-the-Wild Spoken Dialogue Dataset Based on YouTube Dialogue Videos *Yuki Sato, Sanae Yamashita, Shinnosuke Takamichi, Ryuichiro Higashinaka* **D1-1430_L2.9** 217 Emotional Text-To-Speech Based on Mutual-Information-Guided Emotion-Timbre Disentanglement *Jianing Yang, Sheng Li, Takahiro Shinozaki, Yuki Saito, Hiroshi Saruwatari* **D1-1430_L2.10** 321 Conversation Context-aware Direct Preference Optimization for Style-Controlled Speech Synthesis *Atsushi Kojima, Yusuke Fujita, Hao Shi, Tomoya Mizumoto, Mengjie Zhao, Yui Sudo* **D1-1430_L2.11** 483 A Hybrid Attention Mechanism to Improve Tacotron 2 Performance for Indonesian Text-to-Speech Synthesis *Angela Catherina, Bima Prihasto, Boby Mugi Pratama, Li-Wei Kang, Jia-Ching Wang*

14:30-16:00	**D1-1430_H3 Generative AI Models for Vision-Based Applications** Location: Hibiscus III **D1-1430_H3.1** 91 TRUST: Token-dRiven Ultrasound Style Transfer for Cross-Device Adaptation *Nhat-Tuong Do-Tran, Ngoc-Hoang-Lam Le, Ian Chiu, Po-Tsun Paul Kuo, Ching-Chun Huang* **D1-1430_H3.2** 92 Two-Stage Transformer-based Deep Hyperspectral and Multispectral Image Fusion Network for Hyperspectral Image Super-Resolution *Wo-Yen Li, Chia-Ming Lee, Chih-Chung Hsu, Volodymyr Khylenko, Li-Wei Kang* **D1-1430_H3.3** 147 Pedestrian Detection based on Visible Guided Occlusion Handling *Lien-Chieh Huang, Ching-Te Chiu, Yung-Cheng Su* **D1-1430_H3.4** 242 Spatial-Frequency Guided Moiré Removal with Multi-Stage Feature Fusion *Chen Lo, Chia-Hung Yeh* **D1-1430_H3.5** 251 Registration of Infrared and Visible Images Using Style Transfer-Based Semantic Segmentation *Si-Ting Lin, Chih-Hung Han, Chieh-Ling Lee, Po-Chyi Su, Feng-Tsun Chien, Min-Kuan Chang* **D1-1430_H3.6** 533 Peransformer: Improving Low-informed Expressive Performance Rendering with Score-aware Discriminator *Xian He, Wei Zeng, Ye Wang* **D1-1430_H3.7** 536 Prompt-Based Vertebral Segmentation Using a Generative AI Approach in OVCF Spinal Radiographs *Po-Kai Su, Pei-Rong Jiang, Kai-Xuan Xu, Meng-Lei Su, Jiann-Her Lin, Hsin-Han Chiang, Hsiao-Chi Li* **D1-1430_H3.8** 561 A Dual-Stream Diffusion Model with Physically-Based Rendering for Single Image Reflection Removal *Cheng-Wei Hsu, Ming-Sui Lee* **D1-1430_H3.9** 615 DYNAMIC FACIAL EXPRESSION RECOGNITION IN THE WILD USING MAMBA-STYLE SELECTIVE SSM AND FACIAL ATTENTION MECHANISM *Yudhistira Arditya Pratama, Theophilus Ezra Nugroho Pandin, Yi-Zeng Hsieh*
14:30-16:00	**D1-1430_P1 Multimedia Security & Forensics** Location: Peony I **D1-1430_P1.1** 104 The Potential of LLMs for Generating Malicious Domain Names *Lim Kit Michael Ye, Kaijian Zheng, Ngai Fong Law, Jianping Li* **D1-1430_P1.2** 135 Reducing Implicit Class Imbalance in Unlabeled Datasets Using Text-Specified Sensitive Attributes *Kosei Suyama, Kazuaki Nakamura* **D1-1430_P1.3** 215 DRASP: A Dual-Resolution Attentive Statistics Pooling Framework for Automatic MOS Prediction *Cheng-Yeh Yang, Kuan-Tang Huang, Chien-Chun Wang, Hung-Shin Lee, Hsin-Min Wang, Berlin Chen* **D1-1430_P1.4** 348 Multimodal Large Language Model for Deepfake Video Detection and Description *Haoran Sun, Chen Cai, Kong Aik Lee, Lap-Pui Chau, Yi Wang* **D1-1430_P1.5** 379 Biometric Identification Using Default Mode Network Features Extracted from Eyes-Open Resting-State EEG Data *Parvathy Remesh, Jijomon Chettuthara Moncy, Vinod Achutavarrier Prasad* **D1-1430_P1.6** 410 Backdoor Poisoning Attack Against Face Spoofing Attack Detection Methods *Shota Iwamatsu, Koichi Ito, Takafumi Aoki* **D1-1430_P1.7** 464 Access Control for Diffusion Models by Random Masking the Covariance of Initial Noise Distribution *Temma Tanaka, Kazuaki Nakamura*

14:30-16:00	**D1-1430_P2 Wireless Communications & Networking** Location: Peony II **D1-1430_P2.1** 39 Low-Complexity Sparse Channel Estimation for Reconfigurable Intelligent Surface-Aided MIMO *Wei-Lin Chiang, Shu-Yu Lin, Jung-Chun Chi, Yuan-Hao Huang* **D1-1430_P2.2** 50 BFIS: Efficient Unknown Protocol Feature Extraction Method For Satellite Communication Systems *Xianwen Ling, Kun Zhang, Rong Tong, Dianying Chen* **D1-1430_P2.3** 74 Outdoor Experiment of Deep Joint Source-Channel Coding Using FFT-Enabled Convolutional Neural Network for Image Transmission *Tomoka Mori, Hiroshi Tatsukawa, Yuji Kawai, Yoshinori Shinohara, Hiroki Ikeda, Daisuke Hisano* **D1-1430_P2.4** 128 On LSTM-Based Behavioral Modeling of Radio-Frequency Power Amplifiers with a Small Training Dataset *Ryoki Yamaguchi, Satoshi Miyata, Suehiro Shimauchi, Eiji Mochida, Seiji Fujiwara* **D1-1430_P2.5** 209 DL-based Optical Fibre Fault Detection for Healthcare Telesurgery Communication System *Khushi Shah, Lakshit Pathak, Akshita Abrol, Kanak Jain, Rajesh Gupta, Parishi Shah, Sudeep Tanwar, Umesh Bodkhe, Tong Rong* **D1-1430_P2.6** 270 Overcoming Imperfect Detection Limitations: Deep Learning-Based Calibration Strategy for Rotating Interferometric Arrays *Zhaohang Zhang, Chunzhe Wang, Zhen Huang, Yafeng Zhan* **D1-1430_P2.7** 287 A Regional Clustering Method Based on Propagation Similarity for Modeling Cumulative Interference from Large Numbers of Terminals *Tatsuro Hidaka, Osamu Takyu, Kei Inage, Takeo Fujii, Kohei Yoshida, Masayuki Ariyoshi* **D1-1430_P2.8** 290 Radio Frequency Fingerprinting-Based Device Identification Using Deep Metric Learning *Dinh Tuan Anh, Bui Tung Lam, Pham An Duy, Pham Minh Tuan, Tran Vinh Co, Nguyen Huu Tinh, Huynh Cong Bang* **D1-1430_P2.9** 351 GNSS Spoofing Detection Based on LSTM-TNN-CVAE Network *Chaowen Tang, Tian Qin* **D1-1430_P2.10** 456 Enhancing Speech Quality in Scintillating Satellite Communications: A Rician Fading Modeling Approach *Teh Kah Kuan, Sun Hanwu, Tran Huy Dat*
16:00-16:30	**Break**
16:30-18:00	**D1-1630_IB Perspective 2: Utilization of the Foundation Models and the Future** Location: Island Ballroom

16:30-18:00	**D1-1630_L1 Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents** Location: Lotus I **D1-1630_L1.1** 126 NE-PADD: Leveraging Named Entity Knowledge for Robust Partial Audio Deepfake Detection via Attention Aggregation *Huhong Xian, Rui Liu, Berrak Sisman, Haizhou Li* **D1-1630_L1.2** 187 Robust Localization of Partially Fake Speech: Metrics and Out-of-Domain Evaluation *Hieu-Thi Luong, Inbal Rimon, Haim Permuter, Kong Aik Lee, Eng Siong Chng* **D1-1630_L1.3** 188 Mixture of Low-Rank Adapter Experts in Generalizable Audio Deepfake Detection *Janne Laakkonen, Ivan Kukanov, Ville Hautamäki* **D1-1630_L1.4** 203 Continual Audio Deepfake Detection via Universal Adversarial Perturbation *Wangjie Li, Lin Li, Qingyang Hong* **D1-1630_L1.5** 212 Exploring Source Features with Deep Residual Neural Networks for Replay Attack Detection *Suresh Veesa, Badugu Vamsi Krishna, Madhusudan Singh* **D1-1630_L1.6** 254 A Preliminary Study on Sectional Voice Anonymization and Detection *Shaoqi Tang, Zeyan Liu, Liping Chen, Kong Aik Lee, Tomoki Toda, Zhenhua Ling* **D1-1630_L1.7** 276 ArcticEcho: A Novel Speaker-Controlled Voice Cloning Dataset for Modern Deepfake Detection Benchmarking *Soham Gangopadhyay, Inderpreet Singh, Prateek Pandya, Ashish Mani, Sumit Goswami* **D1-1630_L1.8** 327 Variational Regularization for End-to-end Speech Deepfake Detection *Siqing Qin, Kong Aik Lee, Man-Wai Mak, Pasquale Lisena, Massimiliano Todisco* **D1-1630_L1.9** 508 A Wavelet tour of Audio Deepfake Detection *Arth Shah, Aniket Pandey, Manav Giakwad, Hemant Patil* **D1-1630_L1.10** 509 Fusion of Modulation Spectrogram and SSL with Multi-head Attention for Fake Speech Detection *Rishith Sadashiv T N, Abhishek Bedge, Saisha Suresh Bore, Jagabandhu Mishra, Mrinmoy Bhattacharjee, S R Mahadeva Prasanna*

16:30-18:00	**D1-1630_L2 Speaker Modeling Beyond Speaker Recognition** Location: Lotus II **D1-1630_L2.1** 41 SpkAugTSE: A Simple and Efficient Approach to Address Target Confusion in End-to-End Speaker Extraction *Zhenghai You, Zhenyu Zhou, Lantian Li, Dong Wang* **D1-1630_L2.2** 81 Interpolating Speaker Identities in Embedding Space for Data Expansion *Tianchi Liu, Ruijie Tao, Qiongqiong Wang, Yidi Jiang, Hardik B. Sailor, Ke Zhang, Jingru Lin, Haizhou Li* **D1-1630_L2.3** 101 MDD: a Mask Diffusion Detector to Protect Speaker Verification Systems from Adversarial Perturbations *Yibo Bai, Sizhou Chen, Michele Panariello, Xiao-Lei Zhang, Massimiliano Todisco, Nicholas Evans* **D1-1630_L2.4** 131 Fusing Multi-layer Features of the Pre-trained Model With Grouped Cross Attention for Spoofing Speech Detection *Yu Guan, Wu Guo, Jie Zhang, Zhijun Zhang* **D1-1630_L2.5** 132 Fusing Blocked Deep Features of Pre-Trained Models for Short-Duration Speaker Verification *Zhi jun Zhang, Wu Guo, Jie Zhang, Yu Guan* **D1-1630_L2.6** 182 Multi-level Adversarial Training with Data Augmentation for Robust Speaker Verification *Xiaolei Zhang, Zhihua Fang, Liang He* **D1-1630_L2.7** 189 Analysis of Speaker Verification Performance Trade-offs with Neural Audio Codec Transmission *Nirmalya Mallick Thakur, Jia Qi Yip, Eng Siong Chng* **D1-1630_L2.8** 197 Estimating Speaker's Seating Position from Monaural Speech in a Simulated Vehicle Interior Sound Field *Masataka Kaneko, Wen-Chin Huang, Tomoki Toda* **D1-1630_L2.9** 333 TS-VAD+: Modularized Target-Speaker Voice Activity Detection for Robust Speaker Diarization *Tran The Anh, Azmat Adnan, Wu Yihao, Chng Eng Siong* **D1-1630_L2.10** 590 Are Multimodal Foundation Models All That Is Needed for Emofake Detection? *Mohd Mujtaba Akhtar, Girish, Orchid Chetia Phukan, Swarup Ranjan Behera, Pailla Balakrishna Reddy, Ananda Chandra Nayak, Sanjib Kumar Nayak, Arun Balaji Buduru*
16:30-18:00	**D1-1630_H3 Three-Minute Thesis (3MT) Competition** Location: Hibiscus III

16:30-18:00	**D1-1630_P2 Advanced Multimedia Applications** Location: Peony II **D1-1630_P2.1** 113 Ensemble Methods for Estimating the Localization of Coronary Stenosis from CT Images Using 3D CNN Models *Minori Kondo, Masaki Aono, Kazuki Shimizu, Masashi Hashimoto, Takeshi Miyaji, Kei Nomura* **D1-1630_P2.2** 158 Tiered Assessment for DSP Education: Exploring Students' Motivation and Performance *Eliathamby Ambikairajah, Tharmakulasingam Sirojan, Vidhyasaharan Sethu* **D1-1630_P2.3** 297 An Investigation of Parameter Scheduling for Image Restoration in Optical Analog Circuits *Taisei Kato, Ryo Hayakawa, Soma Furusawa, Kazunori Hayashi, Youji Iiguni* **D1-1630_P2.4** 364 Robust Cloud Removal from Optical Satellite Images Using Synthetic Aperture Radar and Multimodal Embedding Prior *Taishin Miura, Shunsuke Ono, Ryo Matsuoka* **D1-1630_P2.5** 367 Reflection and Noise Separation from Polarized Images via Joint Nonnegative Matrix Factorization and Plug-and-Play Denoising *Maharu Oda, Ryo Matsuoka* **D1-1630_P2.6** 377 Gated probabilistic diffusion for temporal action segmentation *Yun LI, Hanmin Li, Kin-Man Lam* **D1-1630_P2.7** 382 Theory of Spherical VR model for Landscape Representation *Hiroyuki Nishimoto, Toru Takahashi, Masakazu Yoshida* **D1-1630_P2.8** 423 HyTver: A Novel Loss Function for Longitudinal Multiple Sclerosis Lesion Segmentation. *Dayan Perera, Ting Fung Fung, Vishnu Monn Baskaran* **D1-1630_P2.9** 473 KH-FUNSD: A Hierarchical and Fine-Grained Layout Analysis Dataset for Low-Resource Khmer Business Document *Nimol Thuon, Jun Du* **D1-1630_P2.10** 516 Effective Speckle Noise Reduction Using Transformed Bayesian Likelihood with Wiener-Based and Sketch-Based Geometric Priors *Ming-Hsun Mo, Pin-Wen Huang, Jian-Jiun Ding*

2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

22-24 October 2025, Singapore

Technical Program

Day 1 | Day 2 | Day 3

Day 2: Thursday, 23 Oct 2025 - Overview	
08:00-08:30	D2-0800_IB **Registration** Location: Island Ballroom
08:30-10:00	D2-0830_IB **Active Noise Cancellation Workshop Keynote** Location: Island Ballroom
08:30-10:00	D2-0830_L1 **Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond** Location: Lotus I
08:30-10:00	D2-0830_L2 **Speech and Language Processing I** Location: Lotus II
08:30-10:00	D2-0830_H3 **Research Frontiers in Learned Visual Data Coding and Processing** Location: Hibiscus III
08:30-10:00	D2-0830_P1 **Multimodal AI** Location: Peony I
08:30-10:00	D2-0830_P2 **Biomedical Signal Processing and Systems I** Location: Peony II
10:00-10:30	**Break**
10:30-12:00	D2-1030_IB **Active Noise Cancellation Panel Discussion** Location: Island Ballroom
10:30-12:00	D2-1030_L1 **Advanced Topics on Music Processing** Location: Lotus I
10:30-12:00	D2-1030_L2 **Speech and Language Processing II** Location: Lotus II

10:30-12:00	D2-1030_H3 **Emerging Technologies and Applications of Image Processing and Computer Vision** Location: Hibiscus III
10:30-12:00	D2-1030_P1 **Biomedical Signal Processing and Systems II** Location: Peony I
10:30-12:00	D2-1030_P2 **Machine Learning: Algorithms and Application I** Location: Peony II
12:00-12:30	**Lunch**
12:30-13:30	D2-1230_IB **Women in APSIPA Forum** Location: Island Ballroom
13:30-14:30	D2-1330_IB **Keynote 2 by Jane Wang** Location: Island Ballroom
14:30-16:00	D2-1430_IB **Perspective 3: Neural Speech Assessment and Its Application** Location: Island Ballroom
14:30-16:00	D2-1430_L1 **Active Noise Control I** Location: Lotus I
14:30-16:00	D2-1430_L2 **Speech and Language Processing III** Location: Lotus II
14:30-16:00	D2-1430_H3 **Machine Learning: Information and Medical Applications** Location: Hibiscus III
14:30-16:00	D2-1430_P1 **Audio Processing** Location: Peony I
14:30-16:00	D2-1430_P2 **Signal & Information Processing I** Location: Peony II
16:00-16:30	**Break**
16:30-18:00	D2-1630_IB **Education Forum** Location: Island Ballroom
16:30-18:00	D2-1630_L1 **Active Noise Control II** Location: Lotus I
16:30-18:00	D2-1630_L2 **Speech and Language Processing IV** Location: Lotus II
16:30-18:00	D2-1630_H3 **Recent Advances in Multimedia Enrichment, Security and Privacy** Location: Hibiscus III
16:30-18:00	D2-1630_P2 **Advances in Multimodal AI for Multimedia Applications** Location: Peony II
19:00-21:30	D2-1900_IB **Banquet** Location: Island Ballroom

Day 2: Thursday, 23 Oct 2025 - With Papers

08:00-08:30	**D2-0800_IB Registration** Location: Island Ballroom
08:30-10:00	**D2-0830_IB Active Noise Cancellation Workshop Keynote** Location: Island Ballroom
08:30-10:00	**D2-0830_L1 Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond** Location: Lotus I **D2-0830_L1.1** 58 Evaluation of auditory and tactile perception for augmented sound-image enhancement using pre-virtual-leading hypersonic signals *Ryota Imanaka, Yuting Geng, Masato Nakayama, Takanobu Nishiura* **D2-0830_L1.2** 103 Improvement in Variance Estimation in Variable-Step-Size Shared-Error NLMS Algorithm for Acoustic Echo and Noise Canceller *Kenta Iwai* **D2-0830_L1.3** 117 Hierarchical Sparse Sound Field Reconstruction with Spherical and Linear Microphone Arrays *Shunxi Xu, Craig T. Jin* **D2-0830_L1.4** 157 Robust Superdirective Beamforming Using a Uniform Circular Array with Directional Microphones *Weilong Huang, Longfei Felix Yan, Emanuël A.P. Habets* **D2-0830_L1.5** 211 Towards Robust Stereo 3-D SELD: A Study of Perceptual Features and Data Augmentation *Jun Wei Yeow, Ee-Leng Tan, Santi Peksi, Woon-Seng Gan, Huang Qirui* **D2-0830_L1.6** 258 Pre-training Autoencoder for Acoustic Event Classification via Blinky *Xiaoyang Liu, Yuma Kinoshita* **D2-0830_L1.7** 275 Sound source enhancement using power spectral density estimation in beamspace for a dual unmanned aerial vehicle system *Mingxue Song, Jin Xuan Teh, Yusuke Hioka, Benjamin Yen, Hiroshi Saruwatari* **D2-0830_L1.8** 328 Three-Dimensional Gradient-Based Tracking of Multiple Sound Sources *Shaoheng Xu, Wei-Ting Lai, Yile (Angela) Zhang, Jihui (Aimee) Zhang, Amy Bastine, Prasanga Samarasinghe, Thushara Abhayapala* **D2-0830_L1.9** 389 Retrieval-Augmented Difference Captioning to Explain Unsupervised Anomalous Sound Detection *Ryoya Ogura, Tomoya Nishida, Yohei Kawaguchi* **D2-0830_L1.10** 398 An Evaluation of Supervised Virtual Microphone Estimators in Reverberant Sound Fields *Kimihiro Hattori, Wen-Chin Huang, Kazuya Takeda, Tomoki Toda* **D2-0830_L1.11** 459 Human-CLAP: Human-perception-based contrastive language–audio pretraining *Taisei Takano, Yuki Okamoto, Yusuke Kanamori, Yuki Saito, Ryotaro Nagase, Hiroshi Saruwatari* **D2-0830_L1.12** 461 Training Acoustic Scene Classification Models Robust to Asynchrony in Distributed Microphone Arrays *Takao Kawamura, Nobutaka Ono*

08:30-10:00	**D2-0830_L2 Speech and Language Processing I** Location: Lotus II **D2-0830_L2.1** 95 Neural Speech Separation with Parallel Amplitude and Phase Spectrum Estimation *Fei Liu, Yang Ai, Zhen-Hua Ling* **D2-0830_L2.2** 127 Single-Channel Speech Enhancement in Spherical-Mapped Short-Time Spectral Domain *Yu Morinaga, Naoto Kotake, Iori Hashimoto, Suehiro Shimauchi, Shigeaki Aoki* **D2-0830_L2.3** 142 Introducing Self-Supervised Learning Models for Spoken Query-Spoken Term Detection *Masato Nagase, Kazunori Kojima, Shi-wook Lee, Yoshiaki Itoh* **D2-0830_L2.4** 150 Characterization of Speech Similarity Between Australian Aboriginal and High-Resource Languages: A Case Study on Dharawal *Ting Dang, Trini Manoj Jeyaseelan, Eliathamby Ambikairajah, Vidhyasaharan Sethu* **D2-0830_L2.5** 179 Segment Transformer: AI-Generated Music Detection via Music Structural Analysis *Yumin Kim, Seonghyeon Go* **D2-0830_L2.6** 213 Dialect Identification Using Resource-Efficient Fine-Tuning Approaches *Zirui Lin, Haris Gulzar, Monnika Roslianna Busto, Akiko Masaki, Takeharu Eda, Kazuhiro Nakadai* **D2-0830_L2.7** 236 A High-Quality and Low-Complexity Streamable Neural Speech Codec with Knowledge Distillation *En-Wei Zhang, Hui-Peng Du, Xiao-Hang Jiang, Yang Ai, Zhen-Hua Ling* **D2-0830_L2.8** 253 Effectiveness of streaming ASR for real-time laughter and screaming detection *Mizuki Kurasawa, Yoshiko Arimoto* **D2-0830_L2.9** 262 Mitigating Data Imbalance in Automated Speaking Assessment *Fong-Chun Tsai, Kuan-Tang Huang, Bi-Cheng Yan, Tien-Hong Lo, Berlin Chen* **D2-0830_L2.10** 446 An Information-Theoretic Approach to Data Selection for Generative Topic Modeling *Michael Santoso, Bhone Tay Zar Kyaw, Valentinus Roby Hananto, Victor Kryssanov* **D2-0830_L2.11** 571 Collective Learning-based Optimal Transport GAN with Multi-Level Fine-Grained and Global Discriminators for Voice Conversion *Sandipan Dhar, Md. Tousin Akhter, Nanda Dulal Jana, Swagatam Das, Monorama Swain, Saurav Chowdhury*

08:30-10:00	**D2-0830_H3 Research Frontiers in Learned Visual Data Coding and Processing** Location: Hibiscus III **D2-0830_H3.1** 64 Neural Implicit Representations for Object-centric Machine Vision Tasks *Yeoneui Kim, Je-Won Kang* **D2-0830_H3.2** 106 GoP-to-Frame Encoder Adaptation for Learned Video Compression *Xiaohan Pan, Runsen Feng, Henan Wang, Yixin Gao, Zhibo Chen* **D2-0830_H3.3** 161 Efficient Adversarial Attack and Training on Learned Image Compression *Jun Kurihara, Heming Sun* **D2-0830_H3.4** 204 Accelerating VVC Inter-Frame Coding: A Lightweight CNN for Fast QTMT Partitioning *Jui-Chen Luo, Jiann-Jone Chen, Tien-Ying Kuo, Yi-Fan Wu, Zhang Kai-Jie* **D2-0830_H3.5** 335 Multimodal Speech Analysis for Early Detection of Mild Cognitive Impairment: A Scalable Approach *Muhammad Bilal, Waleed Abdulla, Gary Cheung, Lynette Tippett, Reza Shahamiri* **D2-0830_H3.6** 383 Boundary-Enhanced Attention Network for Breast Mass Segmentation *Rong Chen, Karungaru Stephen, Kenji Terada, Linhuang Wang* **D2-0830_H3.7** 465 Scale and Rotation Estimation of Similarity-Transformed Images via Cross-Correlation Maximization Based on Auxiliary Function Method *Shinji Yamashita, Yuma Kinoshita, Hitoshi Kiya* **D2-0830_H3.8** 468 Strong Eye Closure Detection in Children with Profound Intellectual and Multiple Disabilities Using Robust Temporal Difference Features *Kaito Kosaki, Teppei Nakano, Mari Wakabayashi, Tomomi Sato, Tetsuji Ogawa* **D2-0830_H3.9** 532 A Rate-Quality Model for Learned Video Coding *Sang NguyenQuang, Cheng-Wei Chen, Xiem HoangVan, Wen-Hsiao Peng* **D2-0830_H3.10** 539 Low-Light RAW Image Enhancement with Additive Parameterization and State Space Model *Shugo Yamashita, Masaaki Ikehara* **D2-0830_H3.11** 621 Synthesizing and Restoring Weather-corrupted Images with Conditional Diffusion Models *Youngho Go, Sung-Hak Lee*
08:30-10:00	**D2-0830_P1 Multimodal AI** Location: Peony I **D2-0830_P1.1** 23 VoxRep: Enhancing 3D Spatial Understanding in 2D Vision-Language Models via Voxel Representation *Alan (Gia Tuan Dao) Dao, Norapat Buppodom* **D2-0830_P1.2** 155 Active Multi-Object Tracking for 3D Reconstruction with Hierarchical Reinforcement Learning *Heng Li, Cheng Cai* **D2-0830_P1.3** 193 Multimodal Sentiment Analysis with Missing Modality: A Knowledge-Transfer Approach *Weide Liu, Huijing Zhan* **D2-0830_P1.4** 299 Modeling Spatiotemporal Multimodal Data With Kernel Graph Regression Models And Copulas *Jeffrey Wu, Gareth Peters* **D2-0830_P1.5** 510 CopeCap: A lightweight image captioning model with collaborative prompt learning *Xiwei Yu, Guoshun He* **D2-0830_P1.6** 541 Lyric-Aware Karaoke Background Video Selection Using Large Language Models and Moment Retrieval *Tomoki Ariga, Jun Taniguchi, Yosuke Higuchi, Sayaka Toma, Kunihiro Abe, Rie Shigyo, Tetsuji Ogawa* **D2-0830_P1.7** 550 Audio-Visual Speech Recognition Based on Cross-Lingual Transfer Learning *Fumiya Kondo, Tamura Satoshi* **D2-0830_P1.8** 562 Exploring Machine Learning and Language Models for Multimodal Depression Detection *Javier Si Zhao Hong, Timothy Zoe Delaya, Sherwyn Chan Yin Kit, Pai Chet Ng, Xiaoxiao Miao*

08:30-10:00	**D2-0830_P2 Biomedical Signal Processing and Systems I** Location: Peony II **D2-0830_P2.1** 26 Predicting Problematic Internet Use in Children Using Feature-Rich Structured Data with Ensemble Machine Learning and Bayesian Optimisation *Niteesh K R, Pooja T S* **D2-0830_P2.2** 148 Phonocardiogram Signal Analysis for Myocardial Infarction Level Prediction Using Deep Learning Model *Ira Puspasari, Tati L.R. Mengko, Agung W. Setiawan, Miftah Pramudyo, Nobuo Watanabe, Trio Adiono* **D2-0830_P2.3** 173 Prediction of Maximum and Minimum Postprandial Blood Glucose Levels in People with Diabetes *Kotaro Nagayama, Shota Kato, Kana Eguchi, Masahide Hamaguchi, Hiroyuki Tominaga, Youji Hamaguchi, Michiaki Fukui, Manabu Kano* **D2-0830_P2.4** 214 Towards Telepathic Communication: A Multi-Band EEG Model for Imaginary Speech Decoding *Yifan Zhang, Yuting Ding, Fei Chen* **D2-0830_P2.5** 240 Tiny-VRN: A Lightweight Variational Residual Network for EEG-Based Emotion Recognition *Sivaraj Nimishan, Selvarajah Thuseethan, Shanmuganathan Vasanthapriyan, Roshan G. Ragel* **D2-0830_P2.6** 241 A Comparison of Solicited and Longitudinal Cough Sounds for Tuberculosis Detection *Aprianto Dwi Prasetyo, Bagus Tris Atmaja, Dhany Arifianto, Sakriani Sakti* **D2-0830_P2.7** 336 Detecting Defecation Premonition from the Acoustic Activity of Bowel Sounds *Shota Miyagawa, Toshitaka Yamakawa, Masayuki Tanabe, Kazushi Ikeda* **D2-0830_P2.8** 407 EegCNR: A Novel Feature for Attention Estimation from EEG *Asif M S, Sagila Gangadharan K, Achutavarrier Prasad Vinod* **D2-0830_P2.9** 413 Lower Limb Calf Muscle Segmentation from Diffusion-Weighted Magnetic Resonance Images Using Deep Learning *Eshan Pandey, Xiaomeng Wang, Julian Gan, Ying-Hwey Nai, Derek Hausenloy, Pek Lan Khong, Forest Su Lim Tan, Thiruneepan Selvakulasingam, Ryan Fraser Kirwan, Cheryl Pei Ling Lian* **D2-0830_P2.10** 415 Principal Component Regularization in Iterative Inversion of DBIM for Ultrasound Tomography *Nguyen Thi Thu, Tran Quang-Huy, Luong Thi Theu, Duc-Tan Tran* **D2-0830_P2.11** 447 Reasoning Visualization for Critical Care EEG Classification with Prototypical Part Networks *Takuma Bingo, Hajime Yano, Taichiro Ashizaki, Kazuma Koda, Masaya Togo, Riki Matsumoto, Tetsuya Takiguchi* **D2-0830_P2.12** 227 Plant Species-Specific Anomaly Detection Based on Electrophysiological Signals *Andy Desman Lo, Elvin Nur Furqon, Junaidul Islam, Isack Farady, Kahlil Muchtar, Ronnie Concepcion II, Chih-Yang Lin*
10:00-10:30	**Break**
10:30-12:00	**D2-1030_IB Active Noise Cancellation Panel Discussion** Location: Island Ballroom

10:30-12:00	**D2-1030_L1 Advanced Topics on Music Processing** Location: Lotus I **D2-1030_L1.1** 177 Drum-to-Vocal Percussion Sound Conversion and Its Evaluation Methodology *Rinka Nobukawa, Makito Kitamura, Tomohiko Nakamura, Shinnosuke Takamichi, Hiroshi Saruwatari* **D2-1030_L1.2** 283 How do Deaf and Hard of Hearing people listen to Music Instruments? Subjective Evaluation and Acoustic Features *Rumi Hiraga, Yuhki Shiraishi, Keiichi Yasu* **D2-1030_L1.3** 298 Quality Assessment of DNN–Based Algorithms for Music Boundary Detection *Aneeka Azmat, Li Su, ChengHsin Hsu* **D2-1030_L1.4** 319 Note-level Nonchord-tone Identification with Graph Neural Networks *Yui Uehara, Satoshi Tojo* **D2-1030_L1.5** 337 Evaluation Score Prediction for Japanese Songs Based on Melody Fitness to Lyrics *Sosuke Nishimura, Eita Nakamura* **D2-1030_L1.6** 349 A Comparative Study of Statistical Features and Deep Learning for Orchestral Texture Classification *Zih-Syuan Lin, Jun-You Wang, Li Su* **D2-1030_L1.7** 356 Efficient Transformer-Based Piano Transcription With Sparse Attention Mechanisms *Weixing Wei, Kazuyoshi Yoshii* **D2-1030_L1.8** 424 Transformer-Based Unpaired Piano Accompaniment Style Transfer *Hsin Ai, Yi-Hsuan Yang* **D2-1030_L1.9** 441 Designing a Music Difficulty Measure for Controllable Automatic Piano Rearrangement *Hikari Miyaji, Keito Sawada, Wen-Chin Huang, Tomoki Toda* **D2-1030_L1.10** 453 Vocal onset detection and pitch segmentation in medieval choral music guided by original notational sources *Samuel Bellows, Sarabeth Mullins, Brian Katz* **D2-1030_L1.11** 496 MORTM: MoE-Optimized Rhythmic Transformer Model for Symbolic MIDI Generation *Takaaki Nagoshi, Tetsuro Kitahara* **D2-1030_L1.12** 517 TAPA-ICL: Taxonomy-Aware Prompt Augmentation for In-Context Learning in Music Understanding *Jiahao Zhao, Yunjia Li, Kazuyoshi Yoshii*

10:30-12:00	**D2-1030_L2 Speech and Language Processing II** Location: Lotus II **D2-1030_L2.1** 97 Beyond Binary Detection: Multi-Etiology Dysarthria Classification with Pre-trained Speech Models *Zihan Zhong, Qianli Wang, Satwinder Singh, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, Seyed Reza Shahamiri* **D2-1030_L2.2** 118 A Dual-Path Speaker-Independent Acoustic-to-Articulatory Inversion Model Based On Content and Speaker Information Disentanglement *Qiang Fang* **D2-1030_L2.3** 120 Dementia Prediction From Speech Signal Using Optimized Prosodic Features *Bagus Tris Atmaja, Sakriani Sakti* **D2-1030_L2.4** 154 Speech Emotion Recognition via Entropy-Aware Score Selection *ChenYi Chua, JunKai Wong, Chengxin Chen, Xiaoxiao Miao* **D2-1030_L2.5** 264 Improving Exemplar-Based Electrolaryngeal Speech Voice Conversion via Robust Content Representations *Fo-Rui Li, Hsin-Te Hwang, Ming-Chi Yen, Men-Tung Lo, Yu Tsao, Hsin-Min Wang* **D2-1030_L2.6** 308 An Efficient Transfer Learning Method Based on Adapter with Local Attributes for Speech Emotion Recognition *Haoyu Song, Mcloughlin Ian, Qing Gu, Nan Jiang, Yan Song* **D2-1030_L2.7** 331 ASRQ-VC: ASR-Guided Speech Content Quantization for High-Fidelity Voice Conversion *Songting Liu, Deheng Ye, Wei Yang, Haoyang Li, Eng Siong Chng* **D2-1030_L2.8** 338 PUNSER: Large-Scale Pre-trained and Unified Model for Practical Speech Emotion Recognition *Yu Hayashizaki, Takashi Nose, Sumiharu Kobayashi, Satoru Fukayama, Akinori Ito* **D2-1030_L2.9** 440 Investigation of the effectiveness of converted speech auditory feedback in low-latency real-time voice conversion *Kiseki Niwa, Kazuhiro Kobayashi, Tomoki Toda* **D2-1030_L2.10** 587 Study on Signal Processing Techniques in Protecting Voice Personae Against Speech Synthesis Systems *Nopparut Li, Candy Olivia Mawalim, Masashi Unoki* **D2-1030_L2.11** 603 MixedG2P-T5: G2P-free Speech Synthesis for Mixed-script texts using Speech Self-Supervised Learning and Language Model *Joonyong Park, Daisuke Saito, Nobuaki Minematsu*

10:30-12:00	**D2-1030_H3 Emerging Technologies and Applications of Image Processing and Computer Vision** Location: Hibiscus III **D2-1030_H3.1** 54 You Only Touch Once: One-Touch System for Personalized 3D Music Video Generation *Kyungjune Lee, Youngjin Shin, Jungwoo Huh, Sanghoon Lee* **D2-1030_H3.2** 143 Single-Image Pupil Localization via Implicit 3D Eye Reconstruction *Taejun Roh, Yejin Cho, Duong Hai Nguyen, Chul Lee* **D2-1030_H3.3** 171 Flow-Guided Consistent Video Depth Estimation for Cross-Dataset Generalization *Jaeseok Jang, Chang-Su Kim* **D2-1030_H3.4** 196 DCB: An Efficient Approach for Building Long-Range Dependencies in CNNs *Tianxiang Lan, Mingyi He, Yuchao Dai* **D2-1030_H3.5** 269 A User-Guided and Local Motion-Adaptive Framework for Virtual Product Placement in Video *Tianwen Zhang, Ju-Won Seo, Kang-Min Kim, Keunsoo Ko* **D2-1030_H3.6** 332 Shallow yet Perceptual Decoding for Neural Image Compression through Minimal Nonlinearity *JaeKyung Ryu, Nam Ik Cho* **D2-1030_H3.7** 445 SyncScore: A Framework for Synchronization Scoring in Group Sports via Human Pose Estimation *Khai Pin Ang, Iven Zi Yin Low, Yumun Hooi, Yuen Peng Loh* **D2-1030_H3.8** 518 Data Augmentation-Driven Segmentation of Ovarian Tumor Ultrasound Images using Vision Mamba *Thanh-Phuc Dao, Huyen-Trang To, Hoang-Son Bui, Thi-Lan Le* **D2-1030_H3.9** 548 Optimizing JPEG Decoder for Bitstream-Corrupted Image Restoration *Shumin Jiang, Hao Qin, Tianyi Liu, Yi Wang* **D2-1030_H3.10** 568 Semantic Scene Completion from a Single Depth Image with Coarse-Grained Segmentation *Jiun Yen Ching, Lai-Kuan Wong, Wai Lee Kung* **D2-1030_H3.11** 572 Pixel-weighted Domain Adaptation for Agricultural Segmentation *Shunta Kimura, Handie Shao, Shogo Matsumoto, Daiki Yamada, Toshihiro Kitajima, Hideki Nakayama*
10:30-12:00	**D2-1030_P1 Biomedical Signal Processing and Systems II** Location: Peony I **D2-1030_P1.1** 454 Freeze and Learn using KAN for Infant Cry Classification *Arth Shah, Vishnu Vardhan, Hemant Patil* **D2-1030_P1.2** 466 Investigation of Enhancement Strategies for Recurrent Spiking Neural Network based Brain-Machine Interface Decoding *Wilson Tansil, Nur Ahmadi, Timothy Constandinou, Dessi Puji Lestari* **D2-1030_P1.3** 531 Detecting Deceptive Responses Due to Psychological Bias by the Probability Density Function of EEG Content Rate Dynamics During NEO-FFI Answering *Yuto Ashikawa, Yosuke Kurihara* **D2-1030_P1.4** 535 A Comparative Analysis of Statistical, Regional CNN, and Sequential Transformer Approaches for Alzheimer's Disease Classification *Trí Huynh, Xuan Hoc Pham, Nhu Nguyen, Thi Thu Nguyen, Huong Ha, Lua Ngo* **D2-1030_P1.5** 591 Beyond Speech and More: Investigating the Emergent Ability of Speech Pre-Trained Models for Classifying Physiological Time-Series Signals *Orchid Chetia Phukan, Swarup Ranjan Behera, Girish, Mohd Mujtaba Akhtar, Arun Balaji Buduru, Rajesh Sharma* **D2-1030_P1.6** 630 Channel Selection Guided by Layer-wise Relevance Propagation for CNN-Based EEG Classification of Major Depressive Disorder *Woo-Seok Ahn, Seung-Hwan Lee, Han-Jeong Hwang* **D2-1030_P1.7** 631 Development of HRV-Based Biomarkers for Predicting Blood Glucose Levels *Ju-An Park, Jun-Seok Lee, Na-Ri Kim, Han-Jeong Hwang* **D2-1030_P1.8** 632 Development of 3D Textile Electrodes for Electrocardiography Measurement *Sang-Ho Lee, In-Su Park, Han-Jeong Hwang*

10:30-12:00	**D2-1030_P2 Machine Learning: Algorithms and Application I** Location: Peony II **D2-1030_P2.1** 22 Kernel Ridge Regression for Efficient Learning of High-Capacity Hopfield Networks *Akira Tamamori* **D2-1030_P2.2** 47 Enhanced Sliding Discrete Fourier Transform (eSDFT) with Error-Bound Control for Real-Time Parallel Processing *Jetsada Arnin, Danial Kahani, Bernard A. Conway* **D2-1030_P2.3** 216 Sparse-Coded Time-Delay DMD with Control for Nonlinear State-Space Modeling on Graphs *Ryuto Ito, Hiromu Kanauchi, Hiroyasu Yasuda, Masaaki Nagahara, Shogo Muramatsu* **D2-1030_P2.4** 229 Nonnegative Matrix Factorization Using Dirichlet-Distribution-Based Regularization *Haru Ogawa, Daichi Kitamura, Shoma Ayano* **D2-1030_P2.5** 274 Significance of co-occurring biomarkers in localization of epileptic seizure onset zone *Nawara Mahmood Broti, Masaki Iwasaki, Yumie Ono* **D2-1030_P2.6** 294 Reinforcement Learning in Portfolio Management: A Survey of Methods and Trends *Silan Hu, Yulin Huang, Arjun Agarwal, Tanya Warrier, Yuwen Wang, Haozhe Ma, Zhengding Luo* **D2-1030_P2.7** 313 Large Sparse Covariance Matrix Estimation via Dual Proximal Gradient Method *Fengpei Li, Ziping Zhao* **D2-1030_P2.8** 361 An improved method for Image Shadow Removal by Combining Deterministic and Stochastic Models *Hongjun Sheng, Lanqing Guo, Xinggan Peng, Zhiping Lin, Bihan Wen* **D2-1030_P2.9** 576 Knowledge-Infused Topic Model for Empathetic Dialogue Response *Po-Chuan Chen, Jen-Tzung Chien* **D2-1030_P2.10** 578 Cross-Patient Seizure Onset Zone Classification by Patient-Dependent Weight *Xuyang ZHAO, Hidenori Sugano, Toshihisa Tanaka* **D2-1030_P2.11** 609 NOCTUA: A High-Efficiency Reconfigurable NoC-based Transformer Universal Accelerator *Kun-Chih Chen, Pin-Ching Shen, Bo-Chun Chen*
12:00-12:30	**Lunch**
12:30-13:30	**D2-1230_IB Women in APSIPA Forum** Location: Island Ballroom
13:30-14:30	**D2-1330_IB Keynote 2 by Jane Wang** Location: Island Ballroom
14:30-16:00	**D2-1430_IB Perspective 3: Neural Speech Assessment and Its Application** Location: Island Ballroom

14:30-16:00	**D2-1430_L1 Active Noise Control I** Location: Lotus I **D2-1430_L1.1** 56 Design of speech leakage-suppressed audio-spot based on auditory masking area control with active masker cancellation using parametric array loudspeakers *Tomoki Hashida, Yuting Geng, Masato Nakayama, Takanobu Nishiura* **D2-1430_L1.2** 57 Multichannel feedforward active noise control system with optical laser microphone in reverberant environments *Maoto Mizutani, Kenta Iwai, Masato Nakayama, Takanobu Nishiura, Yoshiharu Soeta* **D2-1430_L1.3** 72 Frequency-domain online modeling of multiple secondary paths without auxiliary noise for active noise control *Siyuan Lian, Xiaofeng Zeng, Ruquan Sun, Jing Lu* **D2-1430_L1.4** 124 Applying Model-Agnostic Meta-Learning with Iterative Dichotomiser 3 for Alternating-Switching Active Noise Control Systems *Xiaoyi Shen, Dongyuan Shi, Woon-Seng Gan, Jun Yang* **D2-1430_L1.5** 285 A Robust Proactive Communication Strategy for Distributed Active Noise Control Systems *Junwei Ji, Dongyuan Shi, Zhengding Luo, Boxiang Wang, Ziyi Yang, Haowen Li, Woon-Seng Gan* **D2-1430_L1.6** 289 Directional Selective Fixed-Filter Active Noise Control Based on Convolutional Neural Network in Reverberant Environments *Boxiang Wang, Zhengding Luo, Haowen Li, Dongyuan Shi, Junwei Ji, Ziyi Yang, Woon-Seng Gan* **D2-1430_L1.7** 301 An Online Secondary Path Modeling Technique in a Hybrid Active Noise Control System *Harold Alexis Lao, Cheng-Yuan Chang* **D2-1430_L1.8** 345 A Diffusion Remote Microphone Technique for Distributed Active Noise Control *Tianyou Li, Sipei Zhao, Haowen Li, Xiaofeng Zeng, Ruquan Sun, Jing Lu* **D2-1430_L1.9** 511 An Integrated Active Noise Control and Crosstalk Cancellation System Designed Under a Generalized Model-Matching Framework *Michael Edy, Chih Yen Wang, Ching En Huang, You Siang Chen, Mingsian R. Bai* **D2-1430_L1.10** 553 Improvement of Noise Reduction in a Panel Combined with Multiple Loudspeakers Using Active Noise Control *Tatsuya Murao* **D2-1430_L1.11** 610 Selective Fixed Filter Sub-band Active Noise Control System Based on Reference Signal Power Estimation *Shota Toyooka, Ryo Matsuura, Kenta Iwai, Yoshinobu Kajikawa* **D2-1430_L1.12** 626 Performance analysis of active noise control over a spatial region *Jihui (Aimee) Zhang, Thushara Abhayapala, Naoki Murata, Prasanga Samarasinghe, Yu Maeno, Yuki Mitsufuji*

14:30-16:00	**D2-1430_L2 Speech and Language Processing III** Location: Lotus II **D2-1430_L2.1** 27 I^2TTS: Image-indicated Immersive Text-to-speech Synthesis with Spatial Perception *Jiawei Zhang, Tian-Hao Zhang, Jun Wang, Jiaran Gao, Ruijie Tao, Xinyuan Qian, Xu-Cheng Yin* **D2-1430_L2.2** 130 Chain-of-Thought Distillation for ASR Error Correction with Multimodal Large Language Models *Shaomeng Yang, Jiaming Luo, Jinran Wang, Rongfeng Su, Yongjie Zhou, Lan Wang, Nan Yan* **D2-1430_L2.3** 163 Direction-guided Spatial Attention for Multichannel Speech Enhancement *Shuai Nie, Yaran Chen, Shan Liang, Jiaming Xu, Runyu Shi* **D2-1430_L2.4** 168 A Study of Japanese Mixed Emotional Speech Synthesis Based on an End-to-End Emotional Speech Synthesis Model *Issei Sakata, Tetsuo Kosaka* **D2-1430_L2.5** 191 EFTTS: Zero-Shot Emotional Speech Synthesis via Conditional Flow Matching and Self-Supervised Representations *Haoyu Wang, Jiale Chen, Jiaxun Li, Sizhe Shan, Yuehai Wang* **D2-1430_L2.6** 208 Improving Speech-to-Speech Translation for Low-Resource Languages via Transfer Learning *Rui Zhou, Akinori Ito, Takashi Nose* **D2-1430_L2.7** 235 DialoSpeech: Dual-Speaker Dialogue Generation with LLM and Flow Matching *Hanke Xie, Dake Guo, Chengyou Wang, Yue Li, Wenjie Tian, Xinfa Zhu, Xinsheng Wang, Xiulin Li, Guanqiong Miao, Bo Liu, Lei Xie* **D2-1430_L2.8** 237 VICNet: FaderNet-Based Voice Impression Conversion with Affective Dimensional Representation *Takuya Takahashi, Saki Kugimoto, Toru Nakashika* **D2-1430_L2.9** 248 Strategic Re-weighting of U-Net Components in Diffusion Models for Enhanced Speech Enhancement without Retraining *Yuehai Zhang, Yang Li, Yuehao Zhao, Shoji Makino* **D2-1430_L2.10** 261 Fast and Speaker-Independent Utterance Selection for ASR-Free CALL Systems of Minority Languages *Takaki Koshikawa, Akinori Ito, Takashi Nose* **D2-1430_L2.11** 414 Speech-Content-Driven Highlighting of Translated Lecture Slides for Foreign Language Lecture Understanding *Naoki Muto, Chee Siang Leow, Junichi Hoshino, Takehito Utsuro, Hiromitsu Nishizaki*

14:30-16:00	**D2-1430_H3 Machine Learning: Information and Medical Applications** Location: Hibiscus III **D2-1430_H3.1** 24 Class Incremental Learning using Continual Backpropagation on Honey Botanic Origin Classification with Hyperspectral Imaging *Guyang Zhang, Iman Ardekani, Waleed Abdulla* **D2-1430_H3.2** 260 Multi-strategy improved electric eel foraging optimisation algorithm for UAV path planning *Zexin Zhang, Chengbiao Fu, Hongwei Guo, Anhong Tian* **D2-1430_H3.3** 278 A Deep Reinforcement Learning Approach to Roundabout Traffic Signal Control *Cheng-Yu Chen, Daniil Buryakov, Valentinus Roby Hananto, Victor Kryssanov* **D2-1430_H3.4** 300 A preliminary study on machine learning to predict circuit exchange in pediatric patients with ECMO *Tatsuya Hasegawa, Toshiyuki Nakanishi, Koichi Fujiwara* **D2-1430_H3.5** 303 HasRL Robot: A Heterogeneous Asynchronous Reinforcement Learning System for High-Dimensional Bipedal Control *Jingyang Mai, Zechen Guo, Zhengding Luo, Haozhe Ma* **D2-1430_H3.6** 324 A Psychological Strategy Annotation Method Using Multiple LLMs with a Chain of Thought Based on Deductive Reasoning *Jinran Wang, Jiaming Luo, Shaomeng Yang, Yongjie Zhou, Xuefang Zhang, Rongfeng Su, Nan Yan, Lan Wang* **D2-1430_H3.7** 431 Outlier Removal in MEG Data for Imagined Speech Classification *Koki Nose, Hajime Yano, Tetsuya Takiguchi, Seiji Nakagawa* **D2-1430_H3.8** 558 Performance Evaluation of CHIRPS and ETCCDI Indices for Extreme Rainfall Risk Mapping in Thailand Using XGBoost *Vinitar Khettar, Nuntikorn Kitratporn, Sawarin Lerk-u-suke, Jirabhorn Chaiwongsai, Phaisarn Jeefoo, Chanika Sukawattanavijit* **D2-1430_H3.9** 580 Riverbed Estimation Using Locally-Structured Unitary Network *Seiyu Hitomi, Hiroyasu Yasuda, Kiyoshi Hayasaka, Shogo Muramatsu* **D2-1430_H3.10** 594 Contrastive Learning of Temporal and Event-Based Behavioral Views for Universal User Embeddings *Yuuki Tachioka* **D2-1430_H3.11** 595 Market Forecasting Using LSTM-ARIMA Model with MACD Decomposition *Teng-Chih Yu, Jian-Jiun Ding*
14:30-16:00	**D2-1430_P1 Audio Processing** Location: Peony I **D2-1430_P1.1** 129 Anomalous Sound Detection Based on Derivative Features of Short-Time Holomorphic Fourier Transform *Iori Hashimoto, Yu Morinaga, Suehiro Shimauchi, Shigeaki Aoki* **D2-1430_P1.2** 140 Elastic Additive Angular Margin Loss Integrated with Mixup for Anomalous Sound Detection *Yihao Zhao, Yichen Yang, Xiao Zhang, Shoji Makino* **D2-1430_P1.3** 174 A Distilled Low-Latency Neural Vocoder with Explicit Amplitude and Phase Prediction *Hui-Peng Du, Yang Ai, Zhen-Hua Ling* **D2-1430_P1.4** 408 Directional Filtering of Sound Fields for Emphasizing Specific Directions of Arrival and Its Applications *Ryo Murakami, Natsuki Ueno* **D2-1430_P1.5** 409 Sound Field Estimation Method Robust to Microphone Position and Directivity Errors *Takumi Koga, Natsuki Ueno* **D2-1430_P1.6** 434 Anomalous Sound Detection Using Time-Frequency Derivative of Instantaneous Phase Features *Tran-Quang-Tuan Vo, Quoc-Huy Nguyen, Masashi Unoki* **D2-1430_P1.7** 492 Few-Step Diffusion-Based Voice Conversion Using Consistency Trajectory Models *Ryuichi Hatakeyama, Toru Nakashika, Takuya Takahashi* **D2-1430_P1.8** 583 Spatial Audio Signal Enhancement: A Multi-output MVDR Method in The Spherical Harmonic-domain *Huawei Zhang, Jihui (Aimee) Zhang, Huiyuan (June) Sun, Prasanga Samarasinghe*

14:30-16:00	**D2-1430_P2 Signal & Information Processing I** Location: Peony II **D2-1430_P2.1** 46 Generalized Student's t Sparse Kernel Learning for Robust Signal Processing *Long Pan, Libiao Peng, Xifeng Li, Dongjie Bi, Yongle Xie* **D2-1430_P2.2** 61 A Hierarchical Attention Model for Local and Global Feature Integration in RCS Classification *Yida Wu, Caiyun Wang, Jianing Wang, Xiaofei Li, Ying Nan* **D2-1430_P2.3** 66 A Sliding-Window Range–Bearing Scan STAP for Underwater Active Sonar Target Detection *Weisi Hua, Yixin Yang, Yuxuan Chen, Xianghao Hou* **D2-1430_P2.4** 76 TH-LDV: Transformer-based Hybrid method for Signal Detection in Laser Doppler Velocimetry *Yue Wang, Ruifeng Li, Changsong Liu, Liangrui Peng, Ning Ding, Gang Yao* **D2-1430_P2.5** 159 Estimating Dynamic Graph Flows with Kernel Models and Hadamard-Structured Riemannian Constraints *Duc Thien Nguyen, Konstantinos Slavakis, Dimitris Pados* **D2-1430_P2.6** 202 Period Estimation for Time-Varying Graph Signals and Its Application to Graph Wiener Filter *Tsutahiro Fukuhara, Junya Hara, Hiroshi Higashi, Yuichi Tanaka* **D2-1430_P2.7** 244 Computationally Efficient Sparse Signal Recovery by Deep Unfolded-Periodic Sketched ISTA *Tatsuki Tokumura, Ayano Nakai-Kasai, Tadashi Wadayama* **D2-1430_P2.8** 259 Fisher Information-based Metrics for Representation Learning *Do Nguyen Dang Thi, Le Quoc Anh, Tran Trong Duy, Le Vu Ha, Nguyen Linh Trung* **D2-1430_P2.9** 271 Wave Direction Estimation Based on Local Gradient Techniques from Satellite Imagery for Coastal Dynamics Monitoring *Woramet Simrum, Paweena Kanokhong, Chakapat Chokchaisiri, Somrudee Deepaisarn, Kittipisut Chansri, Chanyut Lisawat, Waranrach Viriyavit, Akkharawoot Takhom, Phutphalla Kong, Didin Agustian Permadi, Sharifah Hafizah Syed Ariffin, Surasak Boonkla, Kasorn Galajit, Jessada Karnjana* **D2-1430_P2.10** 318 HIQA-DB: A Benchmark Dataset for Image Quality Assessment in Hospital Surveillance *Yujin Han, Taewan Kim* **D2-1430_P2.11** 627 Semantic Neural View Synthesis for Key Content Preservation in Horizontal-to-Vertical Video Conversion *Dipanita Chakraborty, Minoru Okada, Kosin Chamnongthai*
16:00-16:30	**Break**
16:30-18:00	**D2-1630_IB Education Forum** Location: Island Ballroom

16:30-18:00	**D2-1630_L1 Active Noise Control II** Location: Lotus I **D2-1630_L1.1** 63 Electro-acoustic component placement optimization for helicopter cabin ANC systems *Yuhang Yang, Liquan Shi, Ningyuan Liang, Guoyong Jin* **D2-1630_L1.2** 87 Spatial-Correlation-Based Error Weighting Method for Efficient Application of Filtered Reference Algorithm in Multichannel Active Noise Control *Meiling Hu, Jing Lu, Qingyu Ma* **D2-1630_L1.3** 134 An Alternating Mode Strategy for Adaptive Sound Field Control and Acoustic Path Tracking *Junqing Zhang, Jingli Xie, Dongyuan Shi, Wen Zhang, Jingdong Chen, Jacob Benesty* **D2-1630_L1.4** 265 DOA Estimation with Lightweight Network on LLM-Aided Simulated Acoustic Scenes *Haowen Li, Zhengding Luo, Dongyuan Shi, Boxiang Wang, Junwei Ji, Ziyi Yang, Woon-Seng Gan* **D2-1630_L1.5** 305 Co-forecasting of Time-varying Spatial-frequency Map for Selective Fixed-Filter Multichannel ANC based on Dynamic Factor Graph *Xiruo Su, Bin Wu* **D2-1630_L1.6** 310 Unsupervised Spectrogram Enhancement Algorithm Based on Bi-LSTM *Hanwen Zhang, Xiruo Su, Zhijuan Zhu, Bin Wu, Lingyun Ye* **D2-1630_L1.7** 330 Continual Learning-Based Selective Fixed-filter Active Noise Control *Jingsong Xiao, Qirui Huang* **D2-1630_L1.8** 340 Meta-Learned Regional Initialization of Control Filters for Headphone Active Noise Control *Ziyi Yang, Zhengding Luo, Dongyuan Shi, Junwei Ji, Boxiang Wang, Haowen Li, Qirui Huang, Woon-Seng Gan* **D2-1630_L1.9** 458 RAMDC: Room-Aware Multi-Device Clustering for Large Scale Teleconferencing *Yile Zhang, Weiting Lai, Amy Bastine, Xingyu Chen, Lachlan Birnie, Thushara Abhayapala, Prasanga Samarasinghe* **D2-1630_L1.10** 490 Multi-channel ANC with Adaptive Kernel Assisted On-line Secondary Path Modeling *Hucheng Wang, Tao Liu, Junqing Zhang, Wen Zhang* **D2-1630_L1.11** 497 A Laplace Distribution-Based Variable Step-Size FxlogLMS Algorithm for Active Impulsive Noise Control *Aoi Haneda, Yosuke Sugiura, Tetsuya Shimamura* **D2-1630_L1.12** 513 Research Progress on Active Control of Road Noise in Vehicles *Wangxiaoxu Chen, Jiancheng Tao, Shuping Wang, Kai Chen, Haishan Zou, Xiaojun Qiu*

16:30-18:00	**D2-1630_L2 Speech and Language Processing IV** Location: Lotus II **D2-1630_L2.1** 29 Leveraging Language Information for Target Language Extraction *Mehmet Sinan Yildirim, Ruijie Tao, Wupeng Wang, Junyi Ao, Haizhou Li* **D2-1630_L2.2** 116 VietLyrics: A Large-Scale Dataset and Models for Vietnamese Automatic Lyrics Transcription *Nguyen Quoc Anh, Bernard Cheng, Kelvin Soh* **D2-1630_L2.3** 449 Autofocus Neural Beamformer Based on Steering Vector Estimation *Reiya Marukawa, Takeshi Yamada* **D2-1630_L2.4** 478 Estimating User Sentiment at Sub-exchange Granularity from Exchange-level Annotations *Daichi Yukizawa, Kazunori Komatani, Ryu Takeda, Kenta Yamamoto* **D2-1630_L2.5** 502 DAU-KDAH Dysarthic Multi-Lingual and Multimodal Speech Corpora for Indic Languages *Arth Shah, Hiya Chaudhari, Kavya Kumar, Arushi Srivastava, Priya Damdar, Ravindrakumar Purohit, Dharmendra Vaghera, Bhavna Singh, Aparna Walanj, Abhishek Srivastava, Hemant Patil* **D2-1630_L2.6** 503 Gamma-VAE-VC: Voice conversion based on VAE assuming gamma distribution for both latent variables and observation *Nanako Imaichi, Takuya Takahashi, Toru Nakashika* **D2-1630_L2.7** 514 Zero-shot Context Biasing with Trie-based Decoding using Synthetic Multi-Pronunciation *Changsong Liu, Yizhou Peng, Eng Siong Chng* **D2-1630_L2.8** 551 Dimension 414 and Minimal Embedding Dimensions for Phonetic Feature Encoding in WavLM *Narthana Sivalingam, Uthayasanker Thayasivam* **D2-1630_L2.9** 560 Directional Hybrid Optimization of HRTFs for Low-Order Spherical Harmonics Binaural Rendering *Rui Zhang, Yuxuan Ke, Qunping Ni, Ge Yao, Xiaodong Li, Chengshi Zheng* **D2-1630_L2.10** 577 Speech Enhancement Network With Windowed Cross Attention Using Noise-Reference Microphone *Kota Suzuki, Yosuke Sugiura, Tetsuya Shimamura* **D2-1630_L2.11** 641 BAANI: A 296M-Parameter Neural Vocoder for End-to-End Punjabi Speech Synthesis *Siddharth Kumar, Nisarg Trivedi, Ravindrakumar Purohit, Hemant Patil*

16:30-18:00	**D2-1630_H3 Recent Advances in Multimedia Enrichment, Security and Privacy** Location: Hibiscus III **D2-1630_H3.1** 89 Reversible Data Hiding in EtC Images with Flexible Access Privileges *Yusaku Kato, Shoko Imaizumi* **D2-1630_H3.2** 109 Robust Ownership Verification of DNN Models Against JPEG Compression via Probability-Controlled Adversarial Attacks *Teruki Sano, Minoru Kuribayashi, Masao Sakai, Shuji Ishobe, Eisuke Koizumi, Zhang Zhang* **D2-1630_H3.3** 136 Detoxification of Poisoned Recognition Models by Fine-tuning with Out-of-Distribution Samples *Junsuke Takano, Kazuaki Nakamura* **D2-1630_H3.4** 192 Layer-Wise Weight Statistics for Node Classification and Defense of Federated Large Language Models *Alexander Berns, Reon Akai, Minoru Kuribayashi, Rémi Cogranne* **D2-1630_H3.5** 210 Robustness evaluation against fine-tuning in associative watermarking method for CNN *Keiichi Mori, Masaki Kawamura* **D2-1630_H3.6** 218 Lossless Image Processing for OpenEXR Images with Flexible Functions *Anna Yamaguchi, Shoko Imaizumi* **D2-1630_H3.7** 225 Proposal of a Random Encoding Layer Compatible with Arbitrary Message Lengths for DiffuseTrace *Ou Egami, Masaki Kawamura* **D2-1630_H3.8** 292 Automatic Dependent Surveillance-Broadcast Preamble Classification for Spoofing Detection *Darren Kah Hou Quek, Guang Hua, Zhiping Lin* **D2-1630_H3.9** 320 Model Extraction Attack and Its Countermeasure for Denoising Diffusion Implicit Models *Hayato Shoji, Kazuaki Nakamura* **D2-1630_H3.10** 325 Content-Aware Dominant Color Extraction and Its Application to MItiple-key-Color Image Retrieval *Mei Hashimoto, Michiharu Niimi* **D2-1630_H3.11** 469 Privacy-Preserving Image Retrieval Scheme Using Combined Features in Cloud Computing *Jing Liang, Yuxuan Wang, Tingting Song, Ce Zheng, Peiya Li*

16:30-18:00	**D2-1630_P2 Advances in Multimodal AI for Multimedia Applications** Location: Peony II **D2-1630_P2.1** 59 Efficient Generative Adversarial Networks for Color Document Image Enhancement and Binarization Using Multi-scale Feature Extraction *Rui-Yang Ju, KokSheik Wong, Jen-Shiun Chiang* **D2-1630_P2.2** 68 Leveraging Large Language Models in Visual Speech Recognition: Model Scaling, Context-Aware Decoding, and Iterative Polishing *Zehua Liu, Xiaolou Li, Li Guo, Lantian Li, Dong Wang* **D2-1630_P2.3** 96 Computationally-efficient Call Classification of New Zealand Birds using Texture-based Features *Yonghui Tao, Mathis Quere, Yusuke Hioka, Stephen Marsland* **D2-1630_P2.4** 245 Incorporating Semantic Visual Content into Click-Through Rate Prediction for Video Advertisements *Yoshiaki Tanabe, Shuntaro Masuda, Gakumatsu Ryu, Naoto Tanji, Hiroyuki Seshime, Ling Xiao, Toshihiko Yamasaki* **D2-1630_P2.5** 249 From Blurry to Brilliant Detection: YOLO-Based Aerial Object Detection with Super Resolution *Ragib Amin Nihal, Benjamin Yen, Takeshi Ashizawa, Katsutoshi Itoyama, Kazuhiro Nakadai* **D2-1630_P2.6** 252 ATJO: Adaptive three-dimensional joint optimization for remote sensing video super-resolution *Tian Qin, Lijing Bu, Zhengpeng Zhang, Mingjun Deng, Yin Yang, Jingxue Wang, Xinyu Lan, Wenjuan Peng, Yang Hu* **D2-1630_P2.7** 268 Block-level Lagrange multiplier adaptation based on distortion propagation factors *Hongwei Guo, Yipeng Liu, Lei Luo, Chengbiao Fu, Ce Zhu* **D2-1630_P2.8** 317 Distributed Compressed Video Sensing with Enhanced Boundary Handling Based on Extended Convolutional Sparse Representation *Ibuki Muta, Yoshimitsu Kuroki* **D2-1630_P2.9** 354 Joint Modeling of Big Five and HEXACO for Multimodal Apparent Personality-trait Recognition *Ryo Masumura, Shota Orihashi, Mana Ihori, Tomohiro Tanaka, Naoki Makishima, Taiga Yamane, Naotaka Kawata, Satoshi Suzuki, Taichi Katayama* **D2-1630_P2.10** 359 Foreground-Background Segmentation Based Surveillance Video Coding *Jiyong Yu, Luheng Jia, Yifan Zang, Zhaoyang Yu, Shuyuan Zhu, Li Song, Kebin Jia* **D2-1630_P2.11** 436 Rain Removal via VAE-Enhanced Transformer with Hierarchical Feature Integration *Yaya Huang, Litong Liu, KokSheik Wong*
19:00-21:30	**D2-1900_IB Banquet** Location: Island Ballroom

2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

22-24 October 2025, Singapore

Technical Program

Day 1 | Day 2 | Day 3

Day 3: Friday, 24 Oct 2025 - Overview	
08:00-08:30	D3-0800_IB **Registration** Location: Island Ballroom
08:30-10:00	D3-0830_IB **Sadaoki Furui Prize Papers and Conference Award Session I** Location: Island Ballroom
08:30-10:00	D3-0830_L1 **Music and Singing** Location: Lotus I
08:30-10:00	D3-0830_L2 **Speech Communication and Media Generation** Location: Lotus II
08:30-10:00	D3-0830_H3 **Deep Learning: Algorithm, Implementations, and Applications** Location: Hibiscus III
08:30-10:00	D3-0830_P1 **Advanced Signal Processing and Resource Management for Sensing and Communication** Location: Peony I
08:30-10:00	D3-0830_P2 **Machine Learning: Algorithms and Application II** Location: Peony II
10:00-10:30	**Break**
10:30-12:00	D3-1030_IB **Conference Award Session II** Location: Island Ballroom
10:30-12:00	D3-1030_L1 **Grand Challenge & Audio Processing** Location: Lotus I
10:30-12:00	D3-1030_L2 **Speech Recognition and Understanding I** Location: Lotus II

10:30-12:00	D3-1030_H3 **Recent Advances in Multimodal Learning for Computer Vision** Location: Hibiscus III
10:30-12:00	D3-1030_P1 **Signal & Information Processing II** Location: Peony I
10:30-12:00	D3-1030_P2 **Machine Learning: Algorithms and Application III** Location: Peony II
12:00-13:30	**Lunch**
13:30-14:30	D3-1330_IB **Keyntoe 3 by Xiao-Ping (Steven) Zhang** Location: Island Ballroom
14:30-16:00	D3-1430_IB **Perspective 4: Leveraging LLMs in Interdisciplinary Study of Speech and Language** Location: Island Ballroom
14:30-16:00	D3-1430_L1 **Speech Synthesis, Enhancement, and Recognition** Location: Lotus I
14:30-16:00	D3-1430_L2 **Speech Recognition and Understanding II** Location: Lotus II
14:30-16:00	D3-1430_H3 **Privacy and Security in Speech AI** Location: Hibiscus III
14:30-16:00	D3-1430_P1 **Late Breaking Reports** Location: Peony I
14:30-16:00	D3-1430_P2 **Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)** Location: Peony II
16:00-16:30	**Break**
16:30-18:00	D3-1630_IB **APSIPA General Assembly, Founders' Forum & Closing Ceremony** Location: Island Ballroom

Day 3: Friday, 24 Oct 2025 - With Papers	
08:00-08:30	D3-0800_IB **Registration** Location: Island Ballroom
08:30-10:00	D3-0830_IB **Sadaoki Furui Prize Papers and Conference Award Session I** Location: Island Ballroom **D3-0830_IB.1** 105 Integrating Semantic Knowledge for Enhanced Weakly-Supervised Group Activity Recognition *Muhammad Adi Nugroho, Jinyoung Park, Yeeun Seong, Changick Kim* **D3-0830_IB.2** 402 Directed Graph Dynamic Mode Decomposition for Nonlinear State-Space Modeling *Hiromu Kanauchi, Ryuto Ito, Hiroyasu Yasuda, Masaaki Nagahara, Yuichi Tanaka, Shogo Muramatsu*

08:30-10:00	**D3-0830_L1 Music and Singing** Location: Lotus I **D3-0830_L1.1** 71 Unified Timbre Transfer: A Compact Model for Real-Time Multi-Instrument Sound Morphing *Anders R. Bargum, Naotake Masuda, Bogdan Teleaga, Andrew Fyfe, Cumhur Erkut* **D3-0830_L1.2** 94 REAL-WORLD MUSIC PLAGIARISM DETECTION WITH MUSIC SEGMENT TRANSCRIPTION SYSTEM *Seonghyeon Go* **D3-0830_L1.3** 141 Attention-based Adaptive Structured Patchout Spectrogram Transformer for Music Classification *Yuan Liu, Shoji Makino, Lingqing Liu, Yichen Yang* **D3-0830_L1.4** 307 Accuracy Improvement of Automatic Chord Recognition with Source Separation Preprocessing *Ayumu Mitoma, Ken'ichi Furuya* **D3-0830_L1.5** 362 Effects of Music Training Experience on the Production of English Rhythm by Chinese Learners *Ying Chen, Chenyu Li, Ruizhe Wang, Yujia Zhang* **D3-0830_L1.6** 391 Hierarchical Symbolic Music Generation with Variational Autoencoder-based Bar-wise Feature Sequences *Keito Sawada, Wen-Chin Huang, Tomoki Toda* **D3-0830_L1.7** 421 Singing MIDI Transcription with Music Language Models: Formulation and Comparison *Yu Sugimoto, Jun-You Wang, Li Su, Eita Nakamura* **D3-0830_L1.8** 451 Data-Efficient Music Captioning via Contrastive and Semantic Alignment *Leekyung Kim, Jonghun Park* **D3-0830_L1.9** 482 GAN-Enhanced InpaintNet for Music Inpainting on Limited Data *Koumei Naemura, Boyu Cao, Ryotaro Nagase, Ryoichi Takashima, Yoichi Yamashita* **D3-0830_L1.10** 498 An Analysis of Singing Accuracy towards Quantifying the Melodic Singability *Minami Kawahara, Tetsuro Kitahara* **D3-0830_L1.11** 569 Guitar Tone Morphing by Diffusion-based Model *Kuan-Yu Chen, Kuan-Lin Chen, Yu-Chieh Yu, Jian-Jiun Ding*

08:30-10:00	**D3-0830_L2 Speech Communication and Media Generation** Location: Lotus II **D3-0830_L2.1** 256 Active Learning for Text-to-Speech Synthesis with Informative Sample Collection *Kentaro Seki, Shinnosuke Takamichi, Takaaki Saeki, Hiroshi Saruwatari* **D3-0830_L2.2** 272 Semi-Supervised End-to-End Speech-to-Text Translation with Joint Text-to-Text and Speech-to-Text Decoding *Tomohiro Tanaka, Ryo Masumura, Naoki Makishima, Mana Ihori, Shota Orihashi, Satoshi Suzuki, Taiga Yamane* **D3-0830_L2.3** 344 UTRo-NAST: Non-Autoregressive Speech Translation via Understanding, Translation, and Reordering *Yu-Chen Kuan, Kuan-Yu Chen* **D3-0830_L2.4** 384 Laughing Across Borders: A Culturally-Aware Joke Generator for Asian Regions *Ashley Fang Cai Xian, Chen Ting Ng, Ashley Kok Siu Cheng, Wah Yang Tan, Mohan Raj Chanthran, Lay-Ki Soon, Meisin Lee* **D3-0830_L2.5** 427 Synthesizing Vowel-Like Tones with Pitch Circularity *Kaori Hashimoto, Takao Kawamura, Nobutaka Ono* **D3-0830_L2.6** 428 Error Correction Using LLMs for Sentence Estimation from Ambiguous Inputs via Wearable Keyboards *Matsuri Iwasaki, Masanobu Abe, Sunao Hara* **D3-0830_L2.7** 450 A Robust End to End Spoken Grammar Assessment System *Sunil Kumar Kopparapu, Chitralekha Bhat, Ashish Panda* **D3-0830_L2.8** 452 LAPS-Diff: A Diffusion-Based Framework for Hindi Singing Voice Synthesis With Language Aware Prosody-Style Guided Learning *Sandipan Dhar, Mayank Gupta, Preeti Rao* **D3-0830_L2.9** 472 End-to-end multi-channel speaker extraction and binaural speech synthesis *Cheng Chi, Xiaoyu Li, Yuxuan Ke, Qunping Ni, Ge Yao, Xiaodong Li, Chengshi Zheng* **D3-0830_L2.10** 487 Improving Listening Head Generation Performance Using Speech Representations from Self-Supervised Learning *Tamon Mikawa, Yasuhisa Fujii, Yukoh Wakabayashi, Kengo Ohta, Ryota Nishimura, Norihide Kitaoka* **D3-0830_L2.11** 616 ULF-TTS: An Uncluttered Hybrid TTS System using Language and Flow Matching Models *Jae Hyun Park, Young Sik Eom, SeungJae Choi, Allison Shindell, Min-Gwan Seo, Gyeong-hoon Lee*

08:30-10:00	**D3-0830_H3 Deep Learning: Algorithm, Implementations, and Applications** Location: Hibiscus III **D3-0830_H3.1** 49 Honey Adulteration Detection via Robust Diffusion Classifier and Hyperspectral Imaging *Weihao Tang, Guyang Zhang, Waleed Abdulla* **D3-0830_H3.2** 55 Semantic-Fast-SAM: Efficient Semantic Segmenter *Byunghyun Kim* **D3-0830_H3.3** 84 Micro-expression Recognition Using VideoMamba and Regional Selective Mixup *Yu-Chen Lin, Yi-Jing Chen, Chih-Chang Yu, Hsu-Yung Cheng* **D3-0830_H3.4** 108 100x Monolingual Data Augmentation Using LLMs to Build a Parallel Corpus for Machine Translation *Hitoshi Ito, Naoto Shirai, Kazutaka Kinugawa, Hideya Mino, Yoshihiko Kawai* **D3-0830_H3.5** 185 Neural Analog Filter for Sampling-Frequency-Independent Convolutional Layer *Kanami Imamura, Tomohiko Nakamura, Kohei Yatabe, Hiroshi Saruwatari* **D3-0830_H3.6** 372 Enhancing Technical Documents Retrieval for RAG *Songjiang Lai, Tsun-Hin Cheung, Ka-Chun Fung, Kaiwen Xue, Kwan-Ho Lin, Yan-Ming Choi, Vincent Ng, Kin-Man Lam* **D3-0830_H3.7** 385 Lightweight Zero-Shot Keyword Spotting via Multi-Granular Knowledge Distillation *Yun-Ting Sun, Lo-Ya Li, Tien-Hong Lo, Jeih-Weih Hung, Shih-Chieh Huang, Berlin Chen* **D3-0830_H3.8** 386 Monomial Matrix Relocation on the Loss Function Level-Set of Feedforward Neural Networks *Ozgur Soysal, Arda Ozdemir, Yiğit Yıldırım, Orhan Arikan* **D3-0830_H3.9** 387 Low-Rank Compression of Neural Network Weights by Null-Space Encouragement *Arda Ozdemir, Ozgur Soysal, Ege Doğanay, Yiğit Yıldırım, Orhan Arikan* **D3-0830_H3.10** 422 Sign-MExD: An Expert-Infused Diffusion Model for Sign Language Production *Jiayu Shen, Kalin Stefanov, Vee Yee Chong, Lay-Ki Soon, KokSheik Wong* **D3-0830_H3.11** 470 FEKF: Flow-based Extended Kalman Filter *Pham Hai Anh, Tran Trong Duy, Do Hai Son, Karim Abed-Meraim, Nguyen Linh Trung*
08:30-10:00	**D3-0830_P1 Advanced Signal Processing and Resource Management for Sensing and Communication** Location: Peony I **D3-0830_P1.1** 144 A reinforcement learning-based approach to cooperative multi-UAV task allocation *Naohiro Kubota, Hideyoshi Miura, Tomotaka Kimura, Kouji Hirata* **D3-0830_P1.2** 167 Trajectory Design of UAVs-Assisted Edge Computing Systems for Efficient Data Collection from Animal Herds *Nao Maeda, Tomotaka Kimura, Kouji Hirata* **D3-0830_P1.3** 169 Priority-based RCSA method considering required frequency slot width in multi-core fiber networks *Funa Fukui, Yutaka Fukuchi, Kouji Hirata* **D3-0830_P1.4** 219 Retraining-Free Blockage Prediction for Millimeter-Wave Communications Based on Minor Components of Angular Power Profiles *Xiaoqing Tong, Kohei Mitani, Kazunori Hayashi, Koji Yamamoto, Takuto Arai, Shuki Wai, Tatsuhiko Iwakuni, Daisei Uchida* **D3-0830_P1.5** 220 Modified Resource Allocation Algorithm based on Co-channel Interference Prediction in Local 5G Environmentts *Takeru Nanjo, Osamu Takyu* **D3-0830_P1.6** 341 Implicit Interference Status Notification Through Time & Frequency Resource Selection in LoRaWAN *Yuto Hayasaka, Koichi Adachi* **D3-0830_P1.7** 400 Wireless Environment Estimation with Directional Antennas using Radio Environment Database for Wireless Information and Power Transfer in Smart Factories *Kohei Yuzawa, Takeo Fujii, Yoshiaki Narusue, Zhengdong Lin, Yu Kagaya* **D3-0830_P1.8** 416 Data-Driven Tuning of Neural Network Aided Least Squares for UWB-TDoA Indoor Positioning *Ryoichi Kawaguchi, Shinsuke Ibi, Hisato Iwai*

08:30-10:00	**D3-0830_P2 Machine Learning: Algorithms and Application II** Location: Peony II **D3-0830_P2.1** 83 Equivalence of Graph Signal Processing Using a Hermitian Graph Laplacian and Its Corresponding Graph Laplacian with Duplicated Nodes *Akira Tanaka* **D3-0830_P2.2** 207 Skeleton-sequence-based Early Action Recognition by Using Graph Convolutional Neural Networks and Knowledge Distillation Techniques *Wen-Nung Lie, Kien Truc Le, Veasna Vann, Jui-Chiu Chiang, Ngoc Dung Bui* **D3-0830_P2.3** 232 A State-Dependent Model for Identification of Time-varying Directed Graphs *Yuzhe Li, Hangjing Zhang, H. Vicky Zhao* **D3-0830_P2.4** 309 Unrolled Multimodal Signal Restoration with Signed Twofold Graph Learning *Haruki Yokota, Yuichi Tanaka, Higashi Hiroshi* **D3-0830_P2.5** 343 Efficient Sparse Matrix Acceleration for Deep Learning via Two-Step Bitmap Tensor Architecture *Jia-Hong Weng, Tu Wei-Chen* **D3-0830_P2.6** 352 Distance-based Laplacian Algebra for Effective Subgraph Filter Learning *Purui Zhang, Feng Ji, Yanan Zhao, Wee Peng Tay, Bihan Wen* **D3-0830_P2.7** 370 Evaluation of Low-Resource and High-Efficiency Deep Learning Accelerator for Clinical Dental Diagnosis *Lin Yuan-Jin, Chang Yu-Jen, Liang Chin-Hao, Wei Sung-Tsun, Weng Jia-Hong, Chen Shih-Lun, Tu Wei-Chen* **D3-0830_P2.8** 412 Quantization Index Modulation-based Reversible Data Hiding in Compressed Neural Network *Jun Hirano, Jonethe Tan Yang, Fathin Acyuta Makarim, Daham Jayasinghe, KokSheik Wong* **D3-0830_P2.9** 460 Dense Vector Retrieval in Data Federation *Amorntip Prayoonwong, Yang-Chun Hsu, Xin-Jie Ye, Po-Kai Lu, Chih-Hang Wang, Chih-Yi Chiu* **D3-0830_P2.10** 557 Organ Detection Based on Vision-Language Model For Abdominal CT Images *Jun-Hong Ou, Bo-Xian Wang, Yu-Hong Zheng, Sufal K. Chhabra, Guo-SHiang Lin, Shen-Lei Yan, Chen-Kuo Chiang* **D3-0830_P2.11** 644 Locally-Structured Unitary Network *Shogo Muramatsu*
10:00-10:30	**Break**
10:30-12:00	**D3-1030_IB Conference Award Session II** Location: Island Ballroom **D3-1030_IB.1** 69 Digital-Optical Hybrid Computation for Deep Unfolding-Aided MIMO Signal Detection *Takumi Nishiyama, Lantian Wei, Tadashi Wadayama* **D3-1030_IB.2** 284 Uncolorable Examples: Preventing Unauthorized AI Colorization via Perception-Aware Chroma-Restrictive Perturbation *Yuki Nii, Futa Waseda, Ching-Chun Chang, Isao Echizen* **D3-1030_IB.3** 36 Voice Conversion Augmentation for Speaker Recognition on Defective Datasets *Ruijie Tao, Zhan Shi, Yidi Jiang, Tianchi Liu, Haizhou Li* **D3-1030_IB.4** 355 Explainable Disentanglement on Discrete Speech Representations for Noise-Robust ASR *Shreyas Gopal, Ashutosh Anshul, Haoyang Li, Yue Heng Yeo, Hexin Liu, Eng Siong Chng* **D3-1030_IB.5** 530 Rethinking Robust ASR Strategies: Can Textual In-Context Learning Improve Acoustic Robustness? *Benita Angela Titalim, Faisal Mehmood, Sakriani Sakti* **D3-1030_IB.6** 586 Noro: Noise-Robust One-shot Voice Conversion with Hidden Speaker Representation Learning *Haorui He, Yucheng Song, Yuancheng Wang, Haoyang Li, Xueyao Zhang, Li Wang, Gongping Huang, Eng Siong Chng, Zhizheng Wu*

10:30-12:00	**D3-1030_L1 Grand Challenge & Audio Processing** Location: Lotus I **D3-1030_L1.1** 110 DG-SED: Domain Generalization for Sound Event Detection with Heterogeneous Training Data *Yang Xiao, Han Yin, Jisheng Bai, Rohan Kumar Das* **D3-1030_L1.2** 138 Accelerated Convolutive Transfer Function-Based Multichannel NMF Using Iterative Source Steering *Xuemai Xie, Xianrui Wang, Liyuan Zhang, Yichen Yang, Shoji Makino* **D3-1030_L1.3** 139 DySiME: Dynamic Single-Source Multichannel Enhancement Using Time-Varying Directional Cues *Hao Liang, Yichen Yang, Xiao Zhang, Shoji Makino, Jingdong Chen* **D3-1030_L1.4** 228 Demixing Filter Estimation for Bleeding-Sound Reduction of a Vocal Microphone *Soushi Taninomiya, Daichi Kitamura, Norihiro Takamune, Kouei Yamaoka, Hiroshi Saruwatari, Yu Takahashi, Kazunobu Kondo, Hayato Yamakawa* **D3-1030_L1.5** 401 Prior-Guided Source Separation with Direct Update of Back-projected Demixing Vectors *Kukuru Koiso, Taishi Nakashima, Nobutaka Ono* **D3-1030_L1.6** 588 Meta-Learning with Pretrained Audio Representations Enables One-Shot Acoustic Signal Classification *Haoxiang Wu, Zhengqiao Zhao, Jingdong Chen, Jacob Benesty* **D3-1030_L1.7** 634 Self-Rotation-Robust Online-Independent Vector Analysis with Sound Field Interpolation on Circular Microphone Array *Taishi Nakashima, Yukoh Wakabayashi, Nobutaka Ono* **D3-1030_L1.8** 642 A Streamable Neural Audio Codec with Residual Scalar-Vector Quantization for Real-Time Communication *Xiao-Hang Jiang, Yang Ai, Rui-Chen Zheng, Zhen-Hua Ling* **D3-1030_L1.9** 648 A Semi-Supervised Acoustic Scene Classification Network Based on Multi-Modal Information Fusion *Junkang Yang, Hongqing Liu, Liming Shi, Lu Gan, Hiromitsu Nishizaki, Chee Siang Leow* **D3-1030_L1.10** 651 ASCMamba: Multimodal Time-Frequency Mamba for Acoustic Scene Classification *Bochao Sun, Dong Wang, Zhanlong Yang, Jun Yang, Han Yin* **D3-1030_L1.11** 652 Evaluation of Low-Frequency Restriction, Pitch-Shift Augmentation, and Average Pooling for Acoustic Scene Classification under Unseen-City Conditions *Takao Kawamura, Masayuki Sera, Nobutaka Ono* **D3-1030_L1.12** 653 The APSIPA ASC 2025 Grand Challenge on City and Time-Aware Semi-supervised Acoustic Scene Classification: Summary and Results *Jisheng Bai, Mou Wang, Haohe Liu, Bin Xiang, Yin Liu, Jianfeng Chen, Dongyuan Shi, Mark Plumbley, Susanto Rahardja, Woon-Seng Gan*

10:30-12:00	**D3-1030_L2 Speech Recognition and Understanding I** Location: Lotus II **D3-1030_L2.1** 195 Phoneme-grapheme Dictionary-based Prompting for Robust Proper Noun Recognition in Japanese ASR *Ryuga Sugano, Hiroaki Sato, Asahi Sakuma, Tadashi Kumano, Yoshihiko Kawai, Shinji Watanabe* **D3-1030_L2.2** 282 LLM-Driven Hypothesis Set Refinement for Enhanced ASR Post-Processing *Chen-Han Wu, Kuan-Yu Chen* **D3-1030_L2.3** 390 Real-time VAD-less speech recognition by fine-tuning SSL model with data containing tagged non-speech segments *Jotaro Emoto, Ryota Nishimura, Kengo Ohta, Norihide Kitaoka* **D3-1030_L2.4** 406 Improving Automatic Speech Recognition Model for Super-Elderly Voice Using Speech Synthesis Model *Ryota Uematsu, Chee Siang Leow, Norihide Kitaoka, Hiromitsu Nishizaki* **D3-1030_L2.5** 426 Improving Code-Switching Speech Recognition with TTS Data Augmentation *Yue Heng Yeo, Yu Chen Hu, Shreyas Gopal, Yi Zhou Peng, He Xin Liu, Eng Siong Chng* **D3-1030_L2.6** 443 PQSR: A Speech Corpus of Polar Questions and Spontaneous Responses in Standard Chinese with Complex Intentions Annotated *Yingyi Luo, Yue Huang, Qingke Sun, Shuwen Chen* **D3-1030_L2.7** 475 Toward Natural System Repair: An Analysis of Human Other-Initiated Self-Repair Patterns in Japanese Casual Conversations *Kazuya Tsubokura, Yurie Iribe, Norihide Kitaoka* **D3-1030_L2.8** 524 Self-Supervised Learning for Classification of Normal vs. Dysarthric Speech *Hiya Chaudhari, Kavya Kumar, Hemant Patil* **D3-1030_L2.9** 592 Investigating Polyglot Speech Foundation Models for Learning Collective Emotion from Crowds *Orchid Chetia Phukan, Girish, Mohd Mujtaba Akhtar, Panchal Nayak, Priyabrata Mallick, Swarup Ranjan Behera, Parabattina Bhagath, Pailla Balakrishna Reddy, Arun Balaji Buduru* **D3-1030_L2.10** 593 Rethinking Cross-Corpus Speech Emotion Recognition Benchmarking: Are Paralinguistic Pre-Trained Representations Sufficient? *Orchid Chetia Phukan, Mohd Mujtaba Akhtar, Girish, Swarup Ranjan Behera, Parabattina Bhagath, Pailla Balakrishna Reddy, Arun Balaji Buduru*

10:30-12:00	**D3-1030_H3 Recent Advances in Multimodal Learning for Computer Vision** Location: Hibiscus III **D3-1030_H3.1** 52 Allegory of the Cave: Breakdown of Illusions in Multimodal Perception with Neural Radiance Fields *Axel Paivansalo, Ching-Chun Chang, Hanrui Wang, Futa Waseda, Isao Echizen* **D3-1030_H3.2** 53 Overlapped Coffee Beans Detection and Localization Using a Low-Cost 3D Monocular Point Cloud Clustering Method *Isack Farady, Alifya Febriana, Chih-Yang Lin* **D3-1030_H3.3** 160 Interpretable Video-Text Alignment (VTA) for Cross-Modal Retrieval *Tsung-Shan Yang, Yun-Cheng Wang, Chengwei Wei, Suya You, C.-C. Jay Kuo* **D3-1030_H3.4** 205 Sequence Modeling and Generative Model Driven Non-Rigid 3D Reconstruction *Yuxin He, Hui Deng, Mingyi He, Yuchao Dai* **D3-1030_H3.5** 302 Robust Audio-Visual Speech Recognition in Noisy Clinical Environments *Akshita Abrol, Ridwan Arefeen, Haotong Yu, Alexi George, Kelvin Li, Zhengkui Wang, Rong Tong* **D3-1030_H3.6** 334 Integrating Visual XAI and LLMs for Interpretable Medical Image Analysis *Chern Hong Lim, Xin Hui Lor* **D3-1030_H3.7** 363 InternVL-VPR: Hierarchical Zero-Shot Visual Place Recognition with VLM-driven Re-Ranking *Zhi Hu, Liang Liao, Weisi Lin* **D3-1030_H3.8** 375 Token Compression Meets Compact Vision Transformers: A Survey and Comparative Evaluation for Edge AI *Phat Nguyen, Ngai-Man Cheung* **D3-1030_H3.9** 388 Adapting Vision-Language Models for Information Extraction from Bilingual Medical Invoices *Ha Do, Anh Dung DO, Thanh-Ha DO* **D3-1030_H3.10** 399 Zero-shot Artistic Text Recognition with Multimodal Language Models *Tien Do, Thuyen Tran, Duy-Dinh Le, Thanh Duc Ngo* **D3-1030_H3.11** 604 Attention Based Deep Reference Frame Enhancement for VVC Inter Prediction *Linchen Xu, Zhikai Liu, Fan Liang*
10:30-12:00	**D3-1030_P1 Signal & Information Processing II** Location: Peony I **D3-1030_P1.1** 137 Low-Complexity Total Variation-based Signal Reconstruction with Adaptive Gradient Descent for Compressive sensing *Pei-Cheng Yeh, Chieh-Li Wang, Yuan-Hao Huang* **D3-1030_P1.2** 322 Robust Initialization Strategies for Hankel Structured Low-Rank Approximation via Variable Projection *Natsuki Yoshino, Akira Tanaka* **D3-1030_P1.3** 347 High-Resolution ISAR Imaging for High-Speed Targets via Joint Intra-Pulse and Inter-Pulse Translational Motion Compensation *Jiabao Wang, Shuai Shao, Jiaqi Wei* **D3-1030_P1.4** 381 Sparse Echo Reconstruction of Micro-motion Targets Under the Joint Constraints of Low-rank and Periodic Consistency *Mingming Jin, Jun Wang, Shaoming Wei, Peng Lei* **D3-1030_P1.5** 393 Distributed Extended Object Tracking with Adaptive Networks *Kaidi Yang, Wei Xia* **D3-1030_P1.6** 394 Extended Object Tracking: A DNN-aided approach *Runhe Gan, Wei Xia* **D3-1030_P1.7** 455 Non-negative Learned ISTA with Reflected-ReLU-Augmented L1 Regularization *Haruki Esaki, Towa Yasui, Seisuke Kyochi* **D3-1030_P1.8** 480 Phoneme-Specific Challenges to Intelligibility in Hearing Impairment Under Noisy Condition *Denawati Junia, Candy Olivia Mawalim*

10:30-12:00	**D3-1030_P2 Machine Learning: Algorithms and Application III** Location: Peony II **D3-1030_P2.1** 35 Audio-Visual Fusion Framework for Low-Resource Language Speech Recognition Based on Progressive Down-sampling and Grouped Multi-Heads Attention Mechanism *ChongChong Yu, Xiaolong Xu, Zhaopeng Qian, Kejing Xiao, Yuchen Tan* **D3-1030_P2.2** 239 A Data-Driven Control Framework Using Deep Reinforcement Learning for Autonomous Driving *Mei-Lin Huang, Ching-Hung Lee, Cheng-Ting Huang, Hsin-Han Chiang* **D3-1030_P2.3** 291 Recipe Diffusion: Cross-Frame Attention and Region-Aware Diffusion for Coherent Visual Recipe Instruction Generation *Weiyi Xia, Satoru Fujita* **D3-1030_P2.4** 311 Improving Few-Shot Classification via Feature-Aligned AI-Generated Images *Yu-Wen Tung, Mei-Chen Yeh* **D3-1030_P2.5** 312 Rotation Invariant Automatic Rigging for 3D Human Scan Data *Yiqing Li, Satoru Fujita* **D3-1030_P2.6** 314 SinDiffPhase: High-Quality Phase Estimation with Ultra-Fast Single-Step Diffusion *Yifei Ni, Andong Li, Lingling Dai, Erwei Yin, Qunping Ni, Chengshi Zheng* **D3-1030_P2.7** 411 MapCVAE: Probabilistic Prediction of Diverse Pedestrian Behaviors on General Roads *Konosuke Kobayashi, Satoru Fujita* **D3-1030_P2.8** 457 Herald: Democratizing Compositional Reasoning for Visual Tasks without Any Training *Guan Yuan Tan, Arghya Pal, Sailaja Rajanala, Raphael C.-W. Phan, Chee-Ming Ting* **D3-1030_P2.9** 493 Canopy to Canopy: Evaluating Model Generalization In 3D Tropical Forest Semantic Segmentation *Sue Han Lee, Brenda Ru Yi Sim, Chung Siung Choo, Yuen Peng Loh* **D3-1030_P2.10** 520 LSTM-Transformer Hybrid Network for UAV-Bird Classification Using Radar Track Information *Anning Jiang, Dianfeng Qiao, Shun Liu, Yan Liang* **D3-1030_P2.11** 559 A Unified Framework for Interpretable and Uncertainty-Aware Battery State of Health Estimation Using Deep Neural Networks *Elias Isaac Huai-En Lim, Nicholas Heng Loong Wong*
12:00-13:30	**Lunch**
13:30-14:30	**D3-1330_IB Keyntoe 3 by Xiao-Ping (Steven) Zhang** Location: Island Ballroom
14:30-16:00	**D3-1430_IB Perspective 4: Leveraging LLMs in Interdisciplinary Study of Speech and Language** Location: Island Ballroom

14:30-16:00	**D3-1430_L1 Speech Synthesis, Enhancement, and Recognition** Location: Lotus I **D3-1430_L1.1** 17 End-to-End Simultaneous Dysarthric Speech Reconstruction with Frame-Level Adaptor and Multiple Wait-k Knowledge Distillation *Minghui Wu, Haitao Tang, Jiahuan Fan, Ruizhi Liao, Yanyong Zhang* **D3-1430_L1.2** 48 Disfluency Disentanglement Enhancement in Spoken-Text-Style Transfer for Spontaneous Speech Synthesis *Yuuto Nakata, Daiki Yoshioka, Wen-Chin Huang, Tomoki Toda* **D3-1430_L1.3** 98 DARS: Dysarthria-Aware Rhythm-Style Synthesis for ASR Enhancement *Minghui Wu, Xueling Liu, Jiahuan Fan, Haitao Tang, Yanyong Zhang, Yue Zhang* **D3-1430_L1.4** 149 Emotion-Rich Cross-Speaker TTS via Contrastive Prosody Enhancement *Jen-Tzung Chien, Bryan Gautama Ngo* **D3-1430_L1.5** 243 Face-conditioned Large-scale Text-to-Speech via Speaker Embedding Prediction from Facial Images *Umi Okamoto, Sei Ueno, Akinobu Lee* **D3-1430_L1.6** 304 Few-shot Speaker Adaptation for Text-to-Speech Synthesis Using Non-Target Speaker Corpora for Glossectomy Patients *Masayori Okamura, Masanobu Abe, Sunao Hara* **D3-1430_L1.7** 491 Personalized Bone-Conduction Bandwidth Extension with Speaker Characteristics *Pan Xu, Zhongyu Zhang, Zhonghua Fu* **D3-1430_L1.8** 500 Time-Aligned Laughter Sound Event Recognition for Conversational Laughter Analysis and Synthesis *Hiroki Mori* **D3-1430_L1.9** 554 PALGAN: A Joint Optimization-Based Preprocessing Method for Speech Restoration in Parametric Array Loudspeakers *Wenyao Ma, Jun Yang* **D3-1430_L1.10** 600 Beyond One-Shot Dubbing: Leveraging N-Best Translation and Prompted Paraphrasing with Synchrony-Aware Re-Ranking *Jan Saragih, Faisal Mehmood, Sakriani Sakti*

14:30-16:00	**D3-1430_L2 Speech Recognition and Understanding II** Location: Lotus II **D3-1430_L2.1** 151 Probabilistic Language-Aware Speech Recognition *Jen-Tzung Chien, Willianto Sulaiman, Chung-Hsuan Wang* **D3-1430_L2.2** 226 LLMs-Integrated Automatic Hate Speech Recognition Using Controllable Text Generation Models *Ryutaro Oshima, Yuya Hosoda, Yoji Iiguni* **D3-1430_L2.3** 238 Multi-stage Speech Enhancement with Cascaded SNR Domain Shifts *Xiaoran Li, Zilu Guo, Jun Du* **D3-1430_L2.4** 280 Autoencoder-Driven Latent Representation Learning for Language-Agnostic Disordered Speech Classification using a Universal Feature set *Puneet Bawa, Virender Kadyan, Shareef Babu Kalluri* **D3-1430_L2.5** 296 FH-RestoreASR: Frequency-Hopping Robust Air Traffic Control Speech Restoration and Recognition *Youngeun Kwon, Yeri Byun, Hyunsung Cho, Jongwon Choi* **D3-1430_L2.6** 342 Speech Intelligibility Assessment with Uncertainty-Aware Whisper Embeddings and sLSTM *Ryandhimas Zezario, Dyah Wisnu, Hsin-Min Wang, Yu Tsao* **D3-1430_L2.7** 392 GCI detection and glottal wave estimation based on TV-CAR speech analysis *Keiichi Funaki* **D3-1430_L2.8** 405 HIPA-MoE: A Parameter-Efficient Fine-Tuning Architecture with Hierarchical Adapter-based Mixture-of-Experts for Multilingual ASR *Xun Lu, Xuyang Wang, Gaofeng Cheng, Lin Zheng, Pengyuan Zhang* **D3-1430_L2.9** 542 Mild Cognitive Impairment Detection via Linear Discriminant Analysis of Picture Description Speech Features: A Cross Corpus Comparison *Yan-Lin Lai, Erh-Yun Chang, Yi-Wen Liu, Jung Lung Hsu, Hui-Chuan Hsu* **D3-1430_L2.10** 556 Parameter-Efficient Fine-Tuning of Foundation Models for CLP Speech Classification *Susmita Bhattacharjee, Jagabandhu Mishra, H.S. Shekhawat, S R Mahadeva Prasanna* **D3-1430_L2.11** 575 Language Awareness in Code-Switching Speech Recognition *Jen-Tzung Chien, Bobbi Aditya*

14:30-16:00	**D3-1430_H3 Privacy and Security in Speech AI** Location: Hibiscus III **D3-1430_H3.1** 86 Voice Privacy Protection with Adversarial Examples Using Anchor Speaker Embedding *Shunya Ishikawa, Yuki Katsumata, Toru Nakashika* **D3-1430_H3.2** 133 Investigation of perception inconsistency in speaker embedding for asynchronous voice anonymization *Rui Wang, Liping Chen, Kong Aik Lee, Zhengpeng Zha, Zhenhua Ling* **D3-1430_H3.3** 194 SegReConcat: A Data Augmentation Method for Voice Anonymization Attack *Ridwan Arefeen, Xiaoxiao Miao, Rong Tong, Aik Beng Ng, Simon See* **D3-1430_H3.4** 200 An Enhanced Probabilistic Approach for Singfake Generation *Arth Shah, Aniket Pandey, Satyam R. Tiwari, Hemant Patil* **D3-1430_H3.5** 417 Neural Semi-fragile Watermarking for Proactive Deepfake Speech Detection *Dohyun Yoon, Tomoki Toda* **D3-1430_H3.6** 495 Investigating Self-supervised Learning-Based Front-End for Multi-Channel Replay Attack Detection *Takuo Yamaguchi, Sayaka Shiota, Naohiro Tawara* **D3-1430_H3.7** 515 Transferability of Adversarial Examples across Speaker Embedding Models for Voice Privacy Protection *Kotaro Nakamura, Takuya Takahashi, Toru Nakashika* **D3-1430_H3.8** 538 Voice Privacy Preservation with Multiple Random Orthogonal Secret Keys: Attack Resistance Analysis *Kohei Tanaka, Hitoshi Kiya, Sayaka Shiota* **D3-1430_H3.9** 547 CycleSiFiNF-VC: Controllable Non-Parallel Voice Conversion by Neural Formant Manipulation with Improved Cycle-Consistency Loss *Sumiharu Kobayashi, Takashi Nose, Akinori Ito* **D3-1430_H3.10** 608 Recoverable Audio Adversarial Examples for Voice Protection in One-shot Voice Conversion *Chenshuai Shu, Tianpeng Zheng, Yanxiang Chen* **D3-1430_H3.11** 623 Reference-free Adversarial Sex Obfuscation in Speech *Yangyang Qu, Michele Panariello, Massimiliano Todisco, Nicholas Evans*

14:30-16:00	**D3-1430_P1 Late Breaking Reports** Location: Peony I **D3-1430_P1.1** 638 DCMix:Distance Based ClassMix for Unsupervised Domain Adaptation *Chen-Kuo Chiang, Yue-Lin Yang, Huai En Lee* **D3-1430_P1.2** 654 BiGaitNet: Deep CNN-Based Classification of Parkinson's Disease Gait Abnormalities Using a Smart Insole Robust to Fewer Plantar Sensors *Eun-Seo Park, Xianghong Liu, Chang-Hee Han* **D3-1430_P1.3** 655 EEG Modulation Associated with Music-Induced Emotional Responses *Mayu Goto, Motoko Iwashita, Ingon Chanpornpakdi, Makiko Ishikawa, Kenji Ishida, Toshihisa Tanaka* **D3-1430_P1.4** 656 Nonlinear System Identification Approach under Noisy Input Signals and Impulse Observed Noise by Kernel Adaptive Filtering Algorithm *Ying-Ren Chien, En-Ting Lin* **D3-1430_P1.5** 657 Low-Resource Atayal Speech Recognition into Atayal and Chinese Texts *Po-Cheng Chan, Ching-Ting Hsin, Di Tam Luu, Chi-Tao Chen, Jia Ching Wang* **D3-1430_P1.6** 658 Retinal Artery–Vein Segmentation via Attention-Guided W-Net and GAN-Based Boundary Refinement *Jing-Ming Guo, De-Yu Guu, Yih-Ping Luh, Yi-Chong Zeng* **D3-1430_P1.7** 659 Study on Optimal Online Modeling of the Feedback and Secondary Paths in ANC Systems *Fuma Tsujiwaki, Shota Toyooka, Kenta Iwai, Yoshinobu Kajikawa* **D3-1430_P1.8** 660 Local Contrast Enhancement in LDR Images via Adaptive Distribution of Clipped-histogram Excess *Seong-hyun Jin, Dong-Min Son, Young-Ho Go, Sung-Hak Lee* **D3-1430_P1.9** 661 A Lightweight and Reversible Audio Watermarking Scheme Based on Integer Wavelet Transform *Xuping HUANG, Akinori Ito* **D3-1430_P1.10** 662 A Novel Hybrid Active Noise Control System with Remote Microphone based Virtual Sensing Algorithm *Shota Toyooka, Yoshinobu Kajikawa* **D3-1430_P1.11** 663 Enhancing Speech Quality in Scintillating Satellite Communications: A Rician Fading Modeling Approach (final version) *Teh Kah Kuan* **D3-1430_P1.12** 664 Variance-driven U-Net Training and Chroma-scale-based Multi-exposure Image Fusion *Changwoo Son, Youngho Go, Seunghwan Lee, Sunghak Lee* **D3-1430_P1.13** 665 A New Generation with Variational Autoencoder and Music Transformer for 8-bit MIDI Music *Qian-Yi Zhuang, Shih-Ming Wang, Mau-Yi Tian, Te-Jen Su, Mong-Fong Horng* **D3-1430_P1.14** 666 Toward Physics-Guided LLM Libraries for SINDy: Framework and Rheology Example *Shota Kato, Takeshi Sato, Souta Miyamoto* **D3-1430_P1.15** 667 Joint Design of Low Sidelobe Radar Waveform and Filter with Hardware Platform Verification *Haoqian Rong, Shaojie Wang, Zining Zhao, Jiawei Zhang*

14:30-16:00	**D3-1430_P2 Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)** Location: Peony II **D3-1430_P2.1** 115 Exploring Audio-Visual Fusion Methods in Foundation Model-Based Deception Detection *Jiaxiang Meng, Hardik B. Sailor, Qiongqiong Wang, Tianchi Liu, Kong Aik Lee, Xingmei Wang* **D3-1430_P2.2** 153 Emot-CM-BERT: Adaptive Attention and Class-Aware Cross-Modal Learning for Emotion Recognition from Audio and Text *Shintami Chusnul Hidayati, James Rafferty Lee, Kevin Davi Samuel* **D3-1430_P2.3** 206 DP-GS: Depth-prior & Perception-guided Gaussian Splatting for Sparse-view Novel View Synthesis *Bowen Gao, Zhicheng Lu, Mingyi He, Yuchao Dai* **D3-1430_P2.4** 267 Efficient Video to Audio Mapper with Visual Scene Detection *Mingjing Yi, Yuxi Wang, Ming Li* **D3-1430_P2.5** 481 Adversarial Learning for Duration Prediction in Indonesian Text-to-Speech: Modification to Stochastic and Deterministic Predictors *Yoga Tiara Wiguna, Bima Prihasto, Boby Mugi Pratama, Chia-Hung Yeh, Jia-Ching Wang* **D3-1430_P2.6** 505 Narrativity-Aware Video Summarization Based on Vision and Language Foundation Models *Shumpei Saito, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii* **D3-1430_P2.7** 523 RawTFNet: A Lightweight CNN Architecture for Speech Anti-spoofing *Yang Xiao, Ting Dang, Rohan Kumar Das* **D3-1430_P2.8** 534 Dynamic Fusion Multimodal Network for SpeechWellness Detection *Wenqiang Sun, Han Yin, Jisheng Bai, Jianfeng Chen* **D3-1430_P2.9** 563 AIGuard: Anomaly Detection in Surveillance Videos with YOLOv8 *Rungpilin Anantathanavit, Supakorn Suthirat, Po-Chyi Su* **D3-1430_P2.10** 599 Ensemble Confidence Calibration for Sound Event Detection in Open-environment *Yuanjian Chen, Han Yin* **D3-1430_P2.11** 602 Enhancing Stereo Sound Event Detection with BiMamba and Pretrained PSELDnet *Wenmiao Gao, Han Yin*
16:00-16:30	**Break**
16:30-18:00	**D3-1630_IB APSIPA General Assembly, Founders' Forum & Closing Ceremony** Location: Island Ballroom