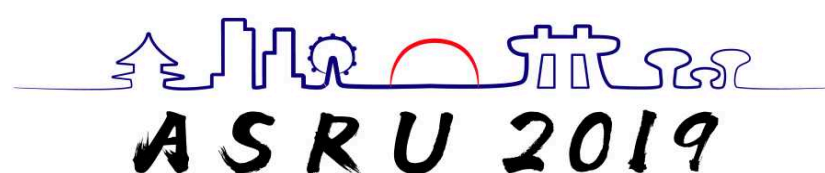# ASRU 2019 POSTER SESSION 3

**Monday 16th Dec 2019 Time: 10:30 am - 12:00 noon**

**LOCATION: EVENTS CENTRE**

**SESSION CHAIR: Hemant Patil**

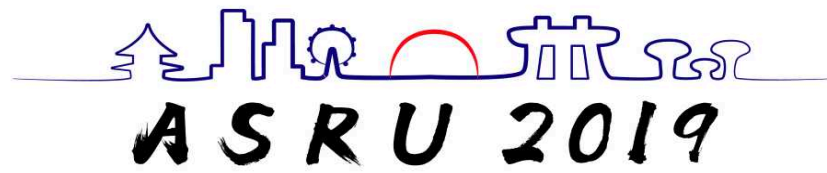| Poster Board | Room | ID | Paper Num | Paper Title |
|---|---|---|---|---|
| P1 | L2 | ASR-2.1 | 1156 | TRAINING LANGUAGE MODELS FOR LONG-SPAN CROSS-SENTENCE EVALUATION |
| P2 | L2 | ASR-2.2 | 1178 | TRANSFORMER ASR WITH CONTEXTUAL BLOCK PROCESSING |
| P3 | L2 | ASR-2.3 | 1185 | A DENSITY RATIO APPROACH TO LANGUAGE MODEL FUSION IN END-TO-END AUTOMATIC SPEECH RECOGNITION |
| P4 | L2 | ASR-2.4 | 1192 | IMPROVING GRAPHEME-TO-PHONEME CONVERSION BY INVESTIGATING COPYING MECHANISM IN RECURRENT ARCHITECTURES |
| P5 | L2 | ASR-2.5 | 1193 | A COMPARATIVE STUDY ON TRANSFORMER VS RNN IN SPEECH APPLICATIONS |
| P6 | L2 | ASR-2.6 | 1194 | FROM SENONES TO CHENONES: TIED CONTEXT-DEPENDENT GRAPHEMES FOR HYBRID SPEECH RECOGNITION |
| P7 | L2 | ASR-2.7 | 1201 | ATTENTION-BASED SPEECH RECOGNITION USING GAZE INFORMATION |
| P8 | L2 | ASR-2.8 | 1203 | LISTENING WHILE SPEAKING AND VISUALIZING: IMPROVING ASR THROUGH MULTIMODAL CHAIN |
| P9 | L2 | ASR-2.9 | 1215 | EMBEDDINGS FOR DNN SPEAKER ADAPTIVE TRAINING |
| P10 | L2 | ASR-2.10 | 1219 | LANGUAGE MODEL BOOTSTRAPPING USING NEURAL MACHINE TRANSLATION FOR CONVERSATIONAL SPEECH RECOGNITION |
| P11 | L2 | ASR-2.11 | 1224 | SPEAKER AND LANGUAGE AWARE TRAINING FOR END-TO-END ASR |
| P12 | L2 | ASR-2.12 | 1225 | DATA AUGMENTATION BASED ON VOWEL STRETCH FOR IMPROVING CHILDREN'S SPEECH RECOGNITION |
| P13 | L2 | ASR-2.13 | 1234 | MIXED BANDWIDTH ACOUSTIC MODELING LEVERAGING KNOWLEDGE DISTILLATION |
| P14 | L2 | ASR-2.14 | 1256 | ON TEMPORAL CONTEXT INFORMATION FOR HYBRID BLSTM-BASED PHONEME RECOGNITION |
| P15 | L2 | ASR-2.15 | 1268 | EXPLORING MODEL UNITS AND TRAINING STRATEGIES FOR END-TO-END SPEECH RECOGNITION |
| P16 | L2 | ASR-2.16 | 1269 | QUERY-BY-EXAMPLE ON-DEVICE KEYWORD SPOTTING |
| P17 | L2 | ASR-2.17 | 1276 | SMALL-FOOTPRINT KEYWORD SPOTTING WITH GRAPH CONVOLUTIONAL NETWORK |
| P18 | L2 | ASR-2.18 | 1327 | SIMPLIFIED LSTMS FOR SPEECH RECOGNITION |
| P19 | L2 | ASR-2.19 | 1395 | GENERALIZED LARGE-CONTEXT LANGUAGE MODELS BASED ON FORWARD-BACKWARD HIERARCHICAL RECURRENT ENCODER-DECODER MODELS |
| P20 | L2 | ASR-2.20 | 1409 | END-TO-END TRAINING OF A LARGE VOCABULARY END-TO-END SPEECH RECOGNITION SYSTEM |
| P21 | A | S2S.1 | 1004 | MULTILINGUAL END-TO-END SPEECH TRANSLATION |
| P22 | A | S2S.2 | 1218 | NEURAL MACHINE TRANSLATION WITH ACOUSTIC EMBEDDING |
| P23 | A | S2S.3 | 1360 | ONE-TO-MANY MULTILINGUAL END-TO-END SPEECH TRANSLATION |
| P24 | A | S2S.4 | 1390 | SPEECH-TO-SPEECH TRANSLATION BETWEEN UNTRANSCRIBED UNKNOWN LANGUAGES |
| P25 | B | SDR.1 | 1055 | ENHANCED BERT-BASED RANKING MODELS FOR SPOKEN DOCUMENT RETRIEVAL |
| P26 | B | SDR.2 | 1065 | VIRTUAL ADVERSARIAL TRAINING FOR DS-CNN BASED SMALL-FOOTPRINT KEYWORD SPOTTING |
| P27 | B | SDR.3 | 1091 | VERIFYING DEEP KEYWORD SPOTTING DETECTION WITH ACOUSTIC WORD EMBEDDINGS |
| P28 | B | SDR.4 | 1124 | MULTILINGUAL BOTTLENECK FEATURES FOR QUERY BY EXAMPLE SPOKEN TERM DETECTION |
| P29 | B | SDR.5 | 1133 | ADDITIONAL SHARED DECODER ON SIAMESE MULTI-VIEW ENCODERS FOR LEARNING ACOUSTIC WORD EMBEDDINGS |
| P30 | B | SDR.6 | 1257 | EFFICIENT FREE KEYWORD DETECTION BASED ON CNN AND END-TO-END CONTINUOUS DP-MATCHING |

# ASRU 2019 POSTER SESSION 4

**Tuesday 17th Dec 2019 Time: 10:30 am - 12:00 noon**
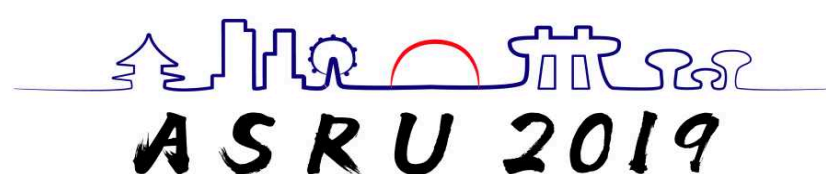
## LOCATION: EVENTS CENTRE

### SESSION CHAIR: Yifan Gong

| Poster Board | Room | ID | Paper Num | Paper Title |
|---|---|---|---|---|
| P1 | A | NEW.1 | 1059 | IMPROVING SPEECH ENHANCEMENT WITH PHONETIC EMBEDDING FEATURES |
| P2 | A | NEW.2 | 1098 | DETECTING DECEPTION IN POLITICAL DEBATES USING ACOUSTIC AND TEXTUAL FEATURES |
| P3 | A | NEW.3 | 1205 | END-TO-END OVERLAPPED SPEECH DETECTION AND SPEAKER COUNTING WITH RAW WAVEFORM |
| P4 | A | NEW.4 | 1244 | TIME DOMAIN AUDIO VISUAL SPEECH SEPARATION |
| P5 | A | NEW.5 | 1283 | SPEECH REVEALS FUTURE RISK OF DEVELOPING DEMENTIA: PREDICTIVE DEMENTIA SCREENING FROM BIOGRAPHIC INTERVIEWS |
| P6 | A | NEW.6 | 1284 | IMPROVING FUNDAMENTAL FREQUENCY GENERATION IN EMG-TO-SPEECH CONVERSION USING A QUANTIZATION APPROACH |
| P7 | A | NEW.7 | 1363 | TOWARDS REAL-TIME MISPRONUNCIATION DETECTION IN KIDS' SPEECH |
| P8 | B | SLR-2.1 | 1114 | INCORPORATING PRIOR KNOWLEDGE INTO SPEAKER DIARIZATION AND LINKING FOR IDENTIFYING COMMON SPEAKER |
| P9 | B | SLR-2.2 | 1195 | LOGISTIC SIMILARITY METRIC LEARNING VIA AFFINITY MATRIX FOR TEXT-INDEPENDENT SPEAKER VERIFICATION |
| P10 | B | SLR-2.3 | 1304 | LOW-RESOURCE DOMAIN ADAPTATION FOR SPEAKER RECOGNITION USING CYCLE-GANS |
| P11 | B | SLR-2.4 | 1319 | CNN WITH PHONETIC ATTENTION FOR TEXT-INDEPENDENT SPEAKER VERIFICATION |
| P12 | B | SLR-2.5 | 1320 | PROBING THE INFORMATION ENCODED IN X-VECTORS |
| P13 | B | SLR-2.6 | 1326 | IN-THE-WILD END-TO-END DETECTION OF SPEECH AFFECTING DISEASES |
| P14 | B | SLR-2.7 | 1330 | OPTIMIZING NEURAL NETWORK EMBEDDINGS USING PAIR-WISE LOSS FOR TEXT-INDEPENDENT SPEAKER VERIFICATION |
| P15 | B | SLR-2.8 | 1334 | TOWARDS CONTROLLING FALSE ALARM --- MISS TRADE-OFF IN PERCEPTUAL SPEAKER COMPARISON VIA NON-NEUTRAL LISTENING TASK FRAMING |
| P16 | B | SLR-2.9 | 1348 | DOVER: A METHOD FOR COMBINING DIARIZATION OUTPUTS |
| P17 | L2 | SLU.1 | 1014 | USING VERY DEEP CONVOLUTIONAL NEURAL NETWORKS TO AUTOMATICALLY DETECT PLAGIARIZED SPOKEN RESPONSES |
| P18 | L2 | SLU.2 | 1016 | SPOKEN MULTIPLE-CHOICE QUESTION ANSWERING USING MULTIMODAL CONVOLUTIONAL NEURAL NETWORKS |
| P19 | L2 | SLU.3 | 1043 | TRANSFER LEARNING FOR CONTEXT-AWARE SPOKEN LANGUAGE UNDERSTANDING |
| P20 | L2 | SLU.4 | 1191 | EMOCEPTION: AN INCEPTION INSPIRED EFFICIENT SPEECH EMOTION RECOGNITION NETWORK |
| P21 | L2 | SLU.5 | 1209 | A COMPARATIVE STUDY ON END-TO-END SPEECH TO TEXT TRANSLATION |
| P22 | L2 | SLU.6 | 1249 | JOINT LEARNING OF WORD AND LABEL EMBEDDINGS FOR SEQUENCE LABELLING IN SPOKEN LANGUAGE UNDERSTANDING |
| P23 | L2 | SLU.7 | 1264 | MARKOV RECURRENT NEURAL NETWORK LANGUAGE MODEL |
| P24 | L2 | SLU.8 | 1266 | TOPIC-AWARE POINTER-GENERATOR NETWORKS FOR SUMMARIZING SPOKEN CONVERSATIONS |
| P25 | L2 | SLU.9 | 1279 | SLU FOR VOICE COMMAND IN SMART HOME: COMPARISON OF PIPELINE AND END-TO-END APPROACHES |
| P26 | L2 | SLU.10 | 1313 | SCALABLE NEURAL DIALOGUE STATE TRACKING |
| P27 | L2 | SLU.11 | 1351 | HIERARCHICAL TRANSFORMERS FOR LONG DOCUMENT CLASSIFICATION |
| P28 | L2 | SLU.12 | 1352 | ADAPTING PRETRAINED TRANSFORMER TO LATTICES FOR SPOKEN LANGUAGE UNDERSTANDING |
| P29 | L2 | SLU.13 | 1355 | SPATIO-TEMPORAL CONTEXT MODELLING FOR SPEECH EMOTION CLASSIFICATION |
| P30 | L2 | SLU.14 | 1387 | PARAPHRASE GENERATION BASED ON VAE AND POINTER-GENERATOR NETWORKS |

**ASRU 2019 DEMO SESSION**
**LOCATION: EVENTS CENTER LEVEL 1**

**Tuesday 17th Dec 2019 Time: 16:00-17:30**
**SESSION CHAIR: Kong Aik Lee**

| Poster Board | Room | ID | Paper Title |
|---|---|---|---|
| P1 | A | DEMO-1.1 | TOWARDS AUTOMATIC SPEECH EVALUATION FOR MULTILINGUAL SOCIETIES: PROTOTYPE SYSTEM FOR SINGAPORE CHILDREN LEARNING MALAY |
| P3 | A | DEMO-1.2 | A SOUND TRACKING ROBOT ON BIO-INSPIRED ALGORITHM |
| P5 | A | DEMO-1.3 | SIMMC: SITUATED INTERACTIVE MULTI-MODAL CONVERSATIONAL DATA COLLECTION AND EVALUATION PLATFORM |
| P7 | A | DEMO-1.4 | MUSIGPRO: AUTOMATIC LEADERBOARD OF SINGERS USING REFERENCE INDEPENDENT EVALUATION |

## ASRU 2019 POSTER SESSIONS
## LOCATION: EVENTS CENTRE

## Wed. 18th Dec 2019 Time: 10:30-12:00 noon
## SESSION CHAIR: ROHAN KUMAR DAS

| Poster Board | Room | ID | Paper Num | Paper Title |
|---|---|---|---|---|
| P1 | L2 | ASR-3.1 | 1285 | SEMI-SUPERVISED TRAINING AND DATA AUGMENTATION FOR ADAPTATION OF AUTOMATIC BROADCAST NEWS CAPTIONING SYSTEMS |
| P2 | L2 | ASR-3.2 | 1286 | ONLINE BATCH NORMALIZATION ADAPTATION FOR AUTOMATIC SPEECH RECOGNITION |
| P3 | L2 | ASR-3.3 | 1288 | SPEAKER ADAPTIVE TRAINING USING MODEL AGNOSTIC META-LEARNING |
| P4 | L2 | ASR-3.4 | 1294 | A COMPARISON OF END-TO-END MODELS FOR LONG-FORM SPEECH RECOGNITION |
| P5 | L2 | ASR-3.5 | 1298 | ACOUSTIC MODEL ADAPTATION FROM RAW WAVEFORMS WITH SINCNET |
| P6 | L2 | ASR-3.6 | 1305 | RECURRENT NEURAL NETWORK TRANSDUCER FOR AUDIO-VISUAL SPEECH RECOGNITION |
| P7 | L2 | ASR-3.7 | 1312 | EXPLICIT ALIGNMENT OF TEXT AND SPEECH ENCODINGS FOR ATTENTION-BASED END-TO-END SPEECH RECOGNITION |
| P8 | L2 | ASR-3.8 | 1329 | RECOGNIZING LONG-FORM SPEECH USING STREAMING END-TO-END MODELS |
| P9 | L2 | ASR-3.9 | 1335 | LEVERAGING LANGUAGE ID IN MULTILINGUAL END-TO-END SPEECH RECOGNITION |
| P10 | L2 | ASR-3.10 | 1336 | STREAMING END-TO-END SPEECH RECOGNITION WITH JOINT CTC-ATTENTION BASED MODELS |
| P11 | L2 | ASR-3.11 | 1344 | MONOTONIC RECURRENT NEURAL NETWORK TRANSDUCER AND DECODING STRATEGIES |
| P12 | L2 | ASR-3.12 | 1368 | CHARACTER-AWARE ATTENTION-BASED END-TO-END SPEECH RECOGNITION |
| P13 | L2 | ASR-3.13 | 1370 | ATTENTION BASED ON-DEVICE STREAMING SPEECH RECOGNITION WITH LARGE SPEECH CORPUS |
| P14 | L2 | ASR-3.14 | 1375 | ZERO-SHOT CODE-SWITCHING ASR AND TTS WITH MULTILINGUAL MACHINE SPEECH CHAIN |
| P15 | L2 | ASR-3.15 | 1376 | END-TO-END CODE-SWITCHING ASR FOR LOW-RESOURCED LANGUAGE PAIRS |
| P16 | L2 | ASR-3.16 | 1408 | UNSUPERVISED ADAPTATION OF ACOUSTIC MODELS FOR ASR USING UTTERANCE-LEVEL EMBEDDINGS FROM SQUEEZE AND EXCITATION NETWORKS |
| P17 | L2 | ASR-3.17 | 1410 | POWER-LAW NONLINEARITY WITH MAXIMALLY UNIFORM DISTRIBUTION CRITERION FOR IMPROVED NEURAL NETWORK TRAINING IN AUTOMATIC SPEECH RECOGNITION |
| P18 | L2 | ASR-3.18 | 1317 | SPEECH RECOGNITION WITH AUGMENTED SYNTHESIZED SPEECH |
| P19 | A | MLP.1 | 1152 | SECOND LANGUAGE TRANSFER LEARNING IN HUMANS AND MACHINES USING IMAGE SUPERVISION |
| P20 | A | MLP.2 | 1377 | ZERO-SHOT PRONUNCIATION LEXICONS FOR CROSS-LANGUAGE ACOUSTIC MODEL TRANSFER |
| P21 | A | SDS.1 | 1070 | ROBUST BELIEF STATE SPACE REPRESENTATION FOR STATISTICAL DIALOGUE MANAGERS USING DEEP AUTOENCODERS |
| P22 | A | SDS.2 | 1097 | IMPROVING SPEECH-BASED END-OF-TURN DETECTION VIA CROSS-MODAL   REPRESENTATION LEARNING WITH PUNCTUATED TEXT DATA |
| P23 | A | SDS.3 | 1271 | DIALOGUE ENVIRONMENTS ARE DIFFERENT FROM GAMES: INVESTIGATING VARIANTS OF DEEP Q-NETWORKS FOR DIALOGUE POLICY |
| P24 | A | SDS.4 | 1325 | EFFICIENT SEMI-SUPERVISED LEARNING FOR NATURAL LANGUAGE UNDERSTANDING BY OPTIMIZING DIVERSITY |
| P25 | B | SS-1.1 | 1143 | DEVELOPMENT OF VOICE SPOOFING DETECTION SYSTEMS FOR 2019 EDITION OF AUTOMATIC SPEAKER VERIFICATION AND COUNTERMEASURES CHALLENGE |
| P26 | B | SS-1.2 | 1281 | SPOOF DETECTION USING TIME-DELAY SHALLOW NEURAL NETWORK AND FEATURE SWITCHING |
| P27 | B | SS-1.3 | 1364 | LONG RANGE ACOUSTIC AND DEEP FEATURES PERSPECTIVE ON ASVSPOOF 2019 |
| P28 | B | SS-2.1 | 1101 | MGB-5: ARABIC DIALECT IDENTIFICATION ACROSS 17 DIALECTS AND MOROCCAN SPEECH RECOGNITION |
| P29 | B | SS-2.2 | 1333 | RDI-CU SYSTEM FOR THE 2019 ARABIC MULTI-GENRE BROADCAST CHALLENGE |