



HAL
open science

Short Report on the IEEE ICASSP 2018 Conference

Joseph J Mariani

► **To cite this version:**

Joseph J Mariani. Short Report on the IEEE ICASSP 2018 Conference. Laboratoire Interdisciplinaire des Sciences du Numérique. 2018. hal-04413200

HAL Id: hal-04413200

<https://hal.science/hal-04413200>

Submitted on 23 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Short Report on ICASSP 2018
Cagliari (Canada), April 16-20, 2018

J. Mariani
LIMSI-CNRS

- 43rd ICASSP (1975)
- Speech processing (incl. HLT), Signal Processing, Audio, Music, some image, video
- Pascale Fung (HKUST) co-Technical Chair
- 1 Day Tutorials, 4 days conference
- 2830 submissions (2nd largest)
- 1406 accepted (about 50%)
- Ranked 1st conference on Speech Processing in Google Scholar (H-Index 67) re. Interspeech (47), ACL (67), NAACL (47), LREC (43), COLING (35), CVPR (158), NIPS (101)
- 77 Oral sessions, 110 poster sessions, 16 Special sessions (upon 42 proposed)
- **17 sessions in parallel (7 Oral, 10 Poster)!**
- Forum “AI and Signal Processing”
- 10-15% of IEEE SPS society attend ICASSP
- **Positive financial result: will reimburse Student Early Registration!**
- 40% US, 27% Europe/Africa, 27% Asia, 6% Others
- 80% male, 9% female, 11% unknown
- **50% academia -50% industry!**
- **US: 65% industry!**
- Many papers from Google, Microsoft, Facebook, Apple...
- In the future
 - o **IEEE Student membership = 1 \$!!!**
 - o **IEEE Resource Center (Video, Papers, Data, Code,...)!**
 - o App for the IEEE society
- **Interactive data entry from audience in real time through Conference App during Opening session**
- Award Mari Ostendorf (SPS Society and IEEE J.F. Flanagan awards)
- Keynote J. Hirschberg (Columbia Univ.)
 - o Detecting Deceptive Speech
- Keynote A. Acero (IEEE Award, Apple)
 - o ASR story
 - o Apple achievements (SIRI)
 - o Icassp attendance stable for years (about 2000) re. NIPS (from 1000 to 8000 in Long Island 2017)
 - o iPhone X
- Keynote Y. LeCun (FAIR)
 - o Neural Network revenge after NN winter
 - o Paper on Convolutional Neural Networks (CNN) rejected at CVPR 2009, many papers on CNN at CVPR 2 years later
 - o “Practice first, Theory second (Telecommunications, Planes, Cars, etc.): Same for NN. Make it work, then find out why it works”
 - o “In real world, you cannot go faster than real time. In virtual world, you can.”
 - o GPU. Now many NN chips.
 - o Most papers cited by Y. Le Cun with ArXiv numbering
 - o **Facebook: 2 Billion Images per day, 2 Billion Users. All images are processed (face recognition, images of interest, close caption)!**
 - o “Domination doesn’t reflect intelligence, but a high degree of testosterone”
 - o French program on AI
- Keynote Luc Vincent (Lyft) on Autonomous Cars
- Problems visas (Turkey, Viet Nam): no show

- Problems audio in rooms (questions from the stage!)
- Problems room (very long/very wide)-(empty/crowded)
- Papers: 33% North America (30% US, 3% Canada), 32% Asia (16% China (incl. Hong Kong)!, 16% others (incl. Japan, Korea, Taiwan, India, Singapore), 29% Europe (7% France, 22% others), 6% Others (incl. Australia).
- 46 Sessions on Neural Networks or Machine Learning (incl. one track specifically on “Machine Learning for Signal Processing” (22 sessions), 6 sessions specifically on Deep Learning, but excluding 3 sessions specifically on Big data): 23% of 203 sessions
- 531 papers on Neural Networks or Machine Learning: about 40% of 1406 papers!
- Papers on (D)NN / ML
 - o End2End Processing
 - o GAN (Generative Adversarial Networks)
 - o Teacher-student learning
 - o Zero-shot ASR/SLU
 - o Vision as an interlingua
 - o Spoken Language Understanding without Automatic Speech Recognition (for a limited task)
 - o Computing required size of NN classifier
 - o Transferring information between NN
 - o Interpreting DNN output layer activations
 - o Understanding Recurrent Neural State
 - o Fooling end2end speaker verification through adversarial examples
- Other sessions/papers
 - o Special Session Signal Processing for Human Behavior
 - o Self-aware agents
 - o ASR systems performance Prediction (LNE+LIUM)
 - o Silent Speech / Lip Reading (data available from Univ. Nottingham)
 - o Human identification through short trivial information (coughs, laughs...)
 - More difficult to mimic!
 - o Dog barks understanding
 - o Bird vocalization identification
 - o Age estimation from speech
 - o Speaker recognition from mixed datasets “OK, Google” (150 MUtts)/ “Hey, Google” (1.2 Mutts)!
 - o Geographic adaptation of LM (Apple)
 - o Recognition of 6 Indian languages with a single model
 - o Predicting Image Memorability (Technicolor)
 - o Remote microphone (Orange)
- Forum AI and SP
 - o Not enough female researchers at ICASSP (P. Fung) Applause.
- Demo
 - o Arabic dialects recognition (QCRI)
- Exhibition
 - o 27 exhibitors (10 from China (Alibaba, Didi, Tencent,...)!)
 - o Including 9 Language Resources booths, 6 from China! (LDC, Appen, Defined Crowd, Speech Ocean, Huiting Data, Magic Data, Databaker, Datatang, Speech Ocean)