


Machine learning applied to speech recognition

Alice Coucke

Head of Machine Learning Research

SONOS

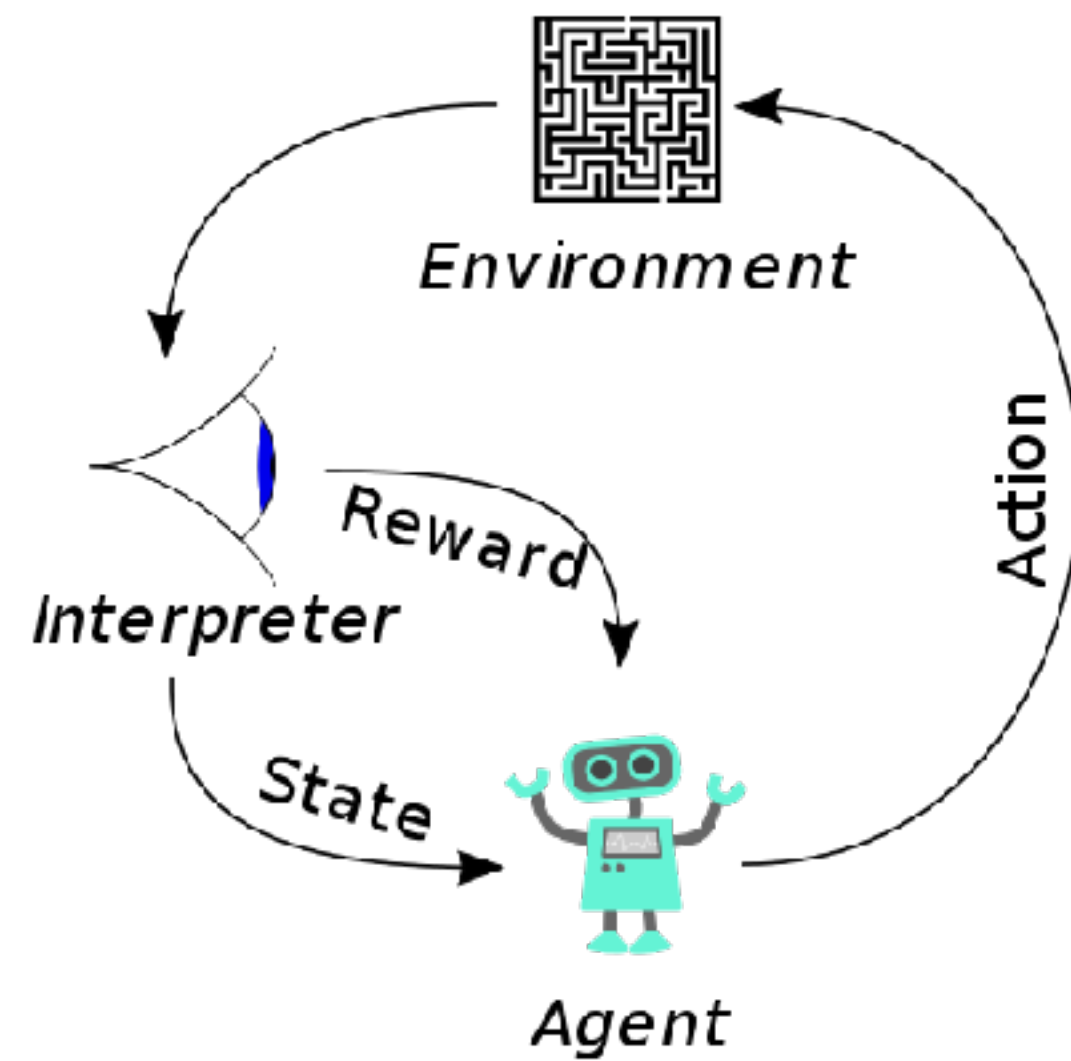
Outline:

- 
- 1. Recent advances in applied ML**
 - 2. From physics to machine learning**
 - 3. Working at Sonos**

Recent advances in applied Machine Learning

Reinforcement learning

Learning goal-oriented behavior within simulated environments



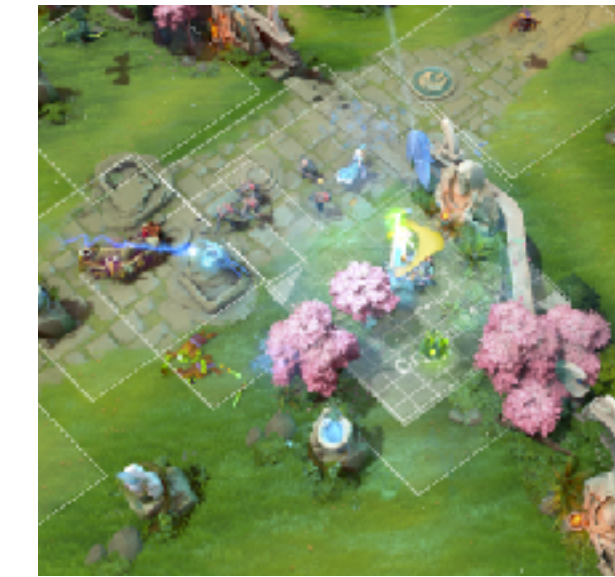
Go

AlphaGo (Deepmind, 2016)



Starcraft II

AlphaStar (Deepmind)



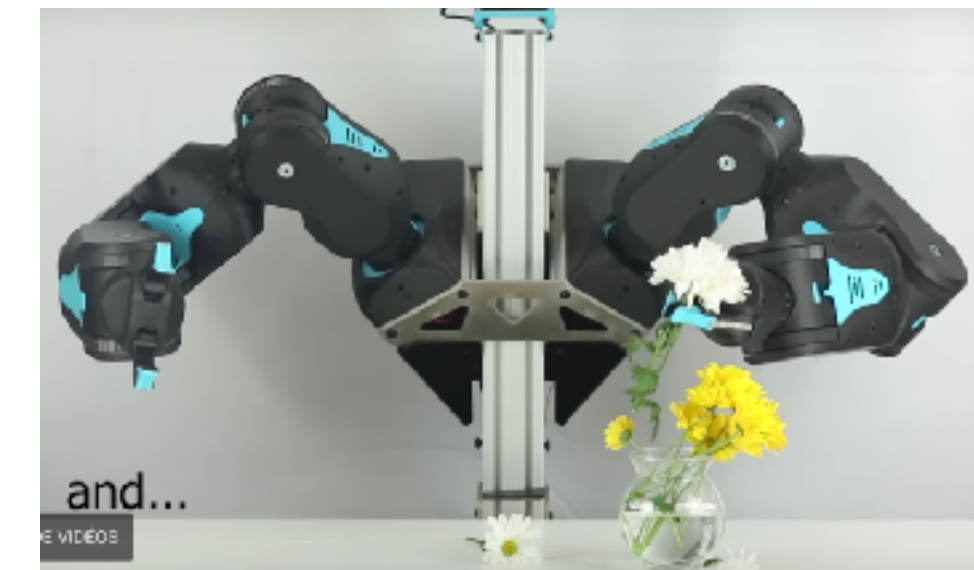
Dota 2

OpenAI Five (OpenAI)



Play-driven learning for robots

(Google Brain)

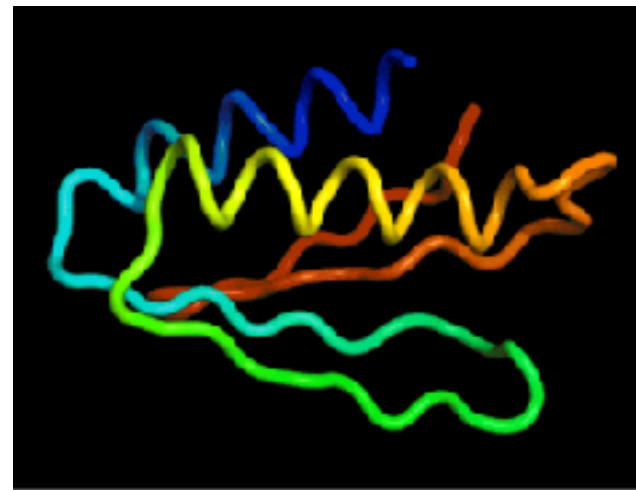


Sim-to-real dexterity learning

Project BLUE (UC Berkeley)

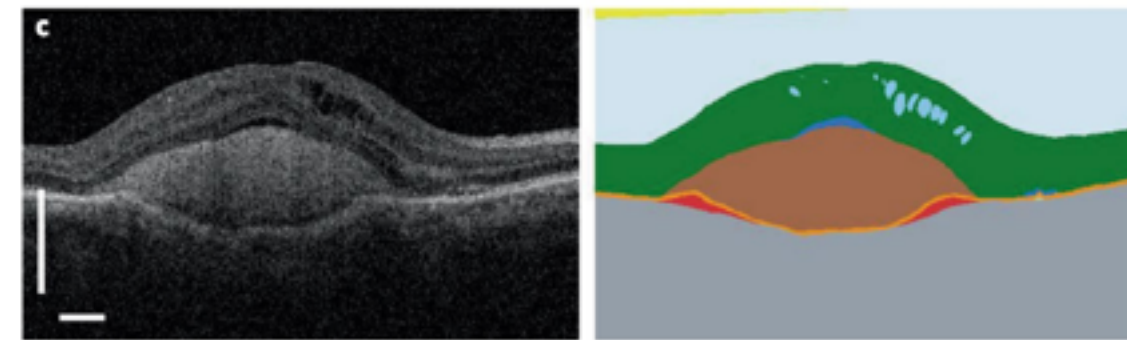
Machine Learning for Life Sciences

Deep learning applied to biology and medicine



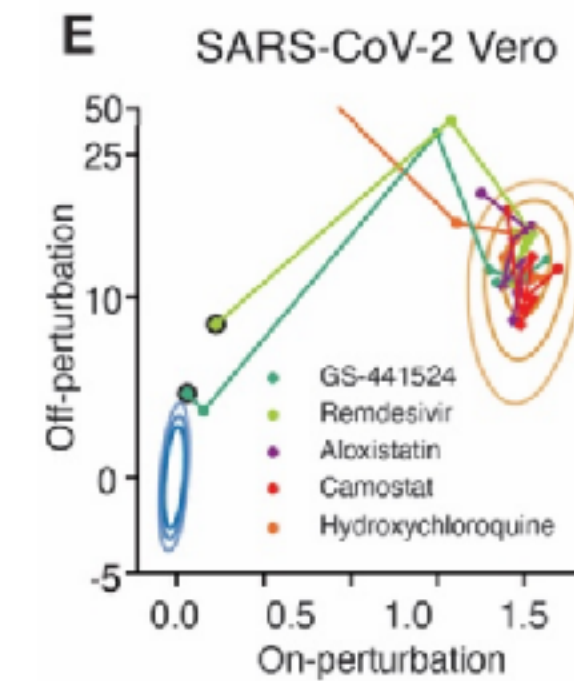
Protein folding & structure prediction

AlphaFold (Deepmind)



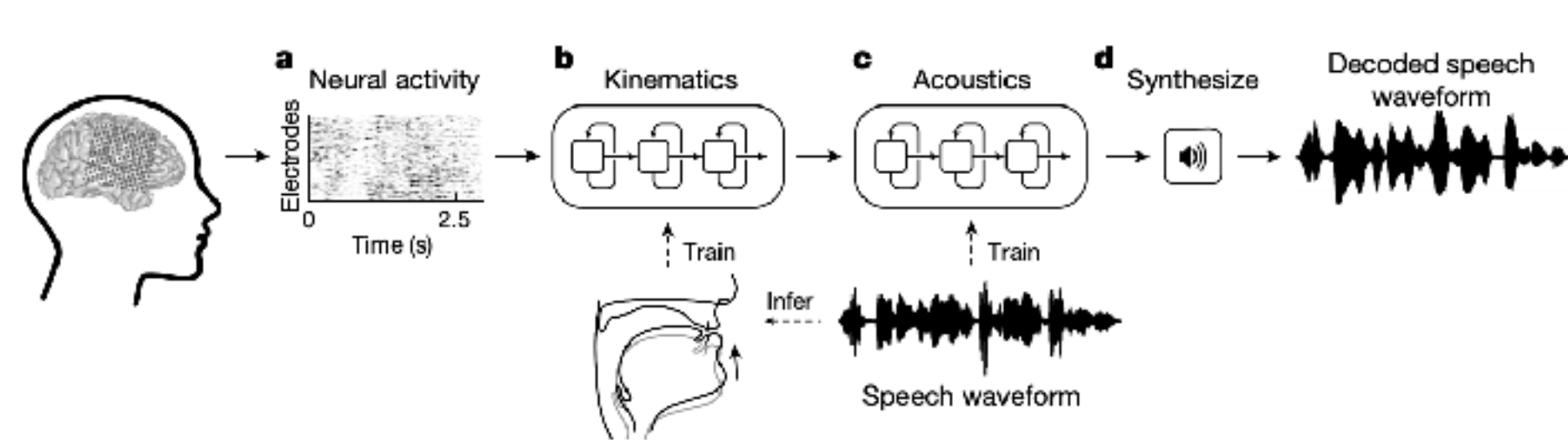
Eye disease diagnosis

(NHS, UCL, Deepmind)



Drug discovery

(Recursion, MIT, Harvard)



Reconstruct speech from neural activity

(UCSF)



Limb control restoration

(Batelle, Ohio State Univ)

Computer vision

High-level understanding of digital images or videos



What is on the jersey of the player on the right?

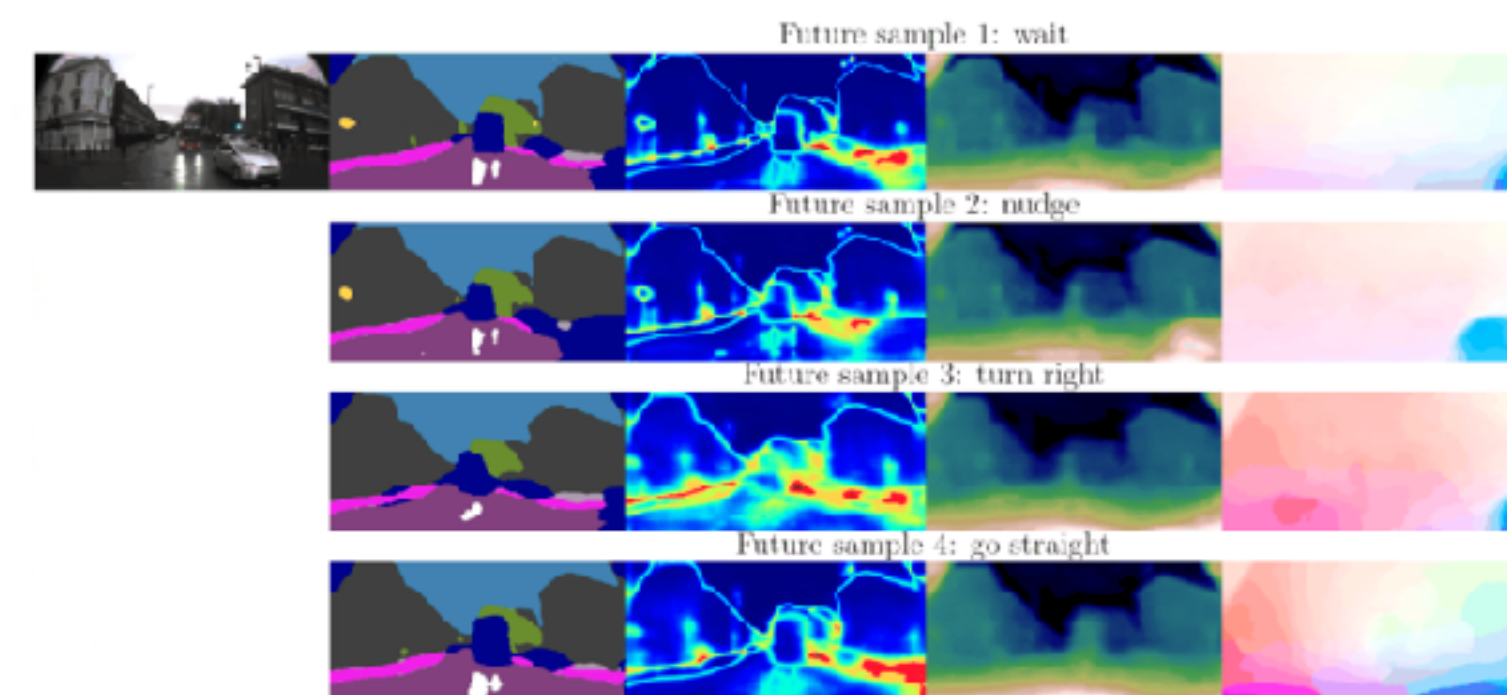


Visual Question Answering on everyday images

(Facebook Research, Georgia Tech)



GANs for artificial video dubbing (Synthesis)



Future state prediction in dynamic scenes

(Wayve, Univ of Cambridge)

From physics to machine learning and back

A surge of interest from the physics community



NeurIPS: workshop on « machine learning and the physical sciences »

- Application of machine and deep learning to physical sciences
- Generative models
- Likelihood-free inference
- Variational inference
- Simulation-based models
- Implicit models
- Probabilistic models
- Model interpretability
- Approximate Bayesian computation
- Strategies for incorporating prior scientific knowledge into machine learning algorithms
- Experimental design
- Any other area related to the subject of the workshop

The jamming transition as a paradigm to understand the loss landscape of deep neural networks

Mario Geiger,^{1,*} Stefano Spigler,^{1,*} Stéphane d'Ascoli,^{2,3} Levent Sagun,^{2,1} Marco Baity-Jesi,⁴ Giulio Biroli,^{2,3} and Matthieu Wyart¹

¹*Institute of Physics, EPFL, CH-1015 Lausanne, Switzerland*

²*Institut de Physique Théorique, Université Paris-Saclay, CEA, CNRS, F-91191 Gif-sur-Yvette, France*

³*Laboratoire de Physique Statistique, École Normale Supérieure, PSL Research University, F-75005 Paris, France*

⁴*Department of Chemistry, Columbia University, 10027 New York, USA*

(Dated: June 18, 2019)

Modelling the influence of data structure on learning in neural networks

Sebastian Goldt¹, Marc Mézard²,
Florent Krzakala² and Lenka Zdeborová¹

¹*Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, France*

²*Laboratoire de Physique de l'École Normale Supérieure, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, Paris, France*

Capacity-resolution trade-off in the optimal learning of multiple low-dimensional manifolds by attractor neural networks

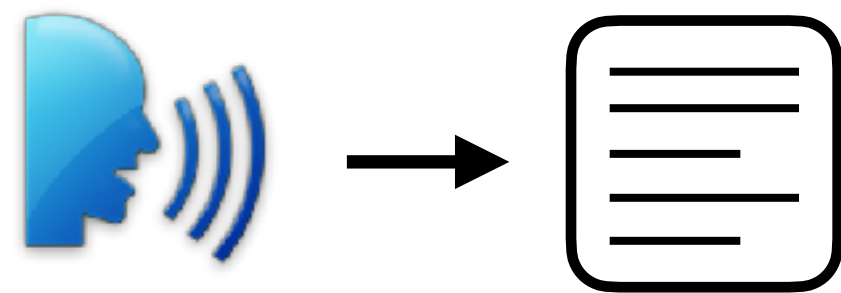
Aldo Battista and Rémi Monasson

*Laboratory of Physics of the École Normale Supérieure,
CNRS UMR 8028 & PSL Research, Paris, France*

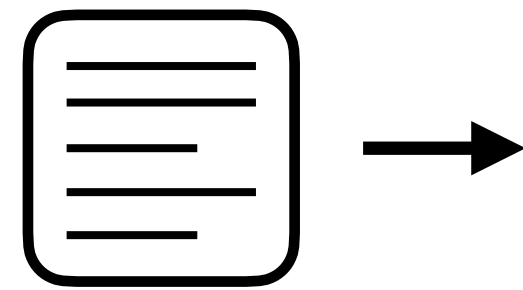
(Dated: October 15, 2019)

Speech and language

Understand and analyze human speech

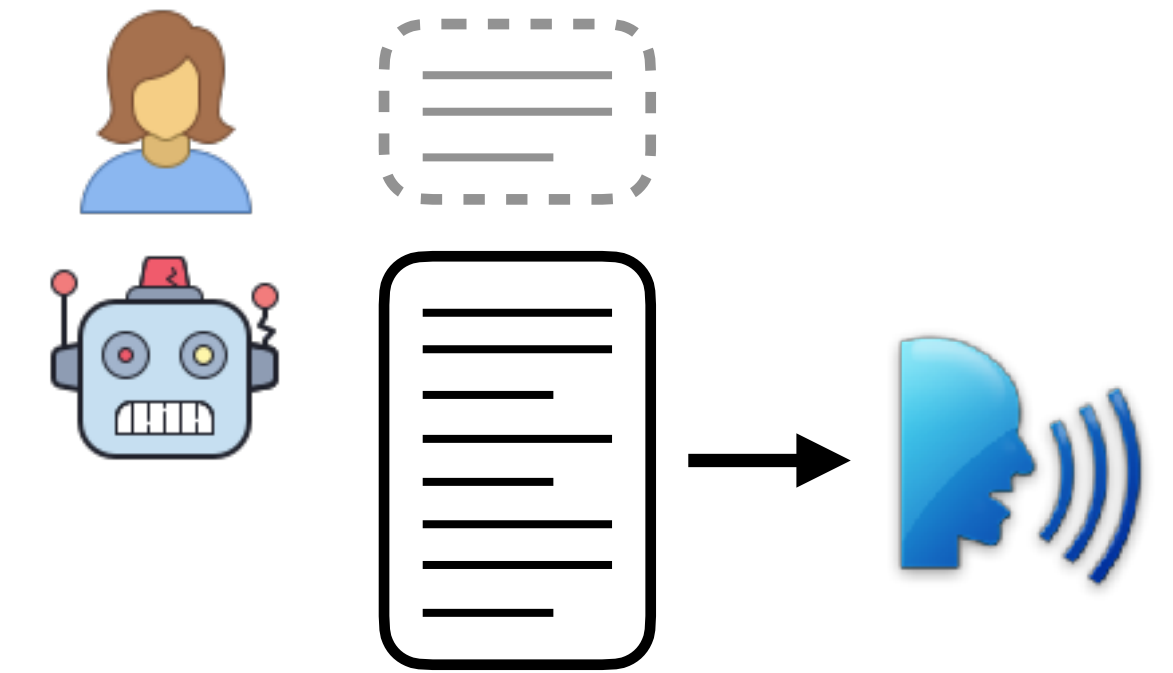
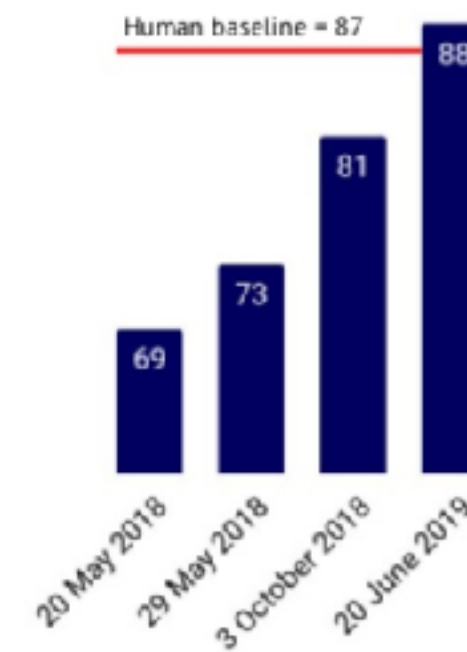


Speech transcription
Human Parity (Microsoft)



```
{  
  intent: FindWeather,  
  entities: {  
    datetime: 11/19/2020,  
    location: Paris  
  }  
}
```

Spoken language understanding
(Super)GLUE benchmarks (Google, Facebook, IBM, Stanford ...)



Text & speech generation
*GPT-2 (Open AI)
Bert (Google)
XLNet (CMU) ...*



Neural machine translation
Unsupervised MT (Facebook)



Voice activity detection
Detect speech from audio



Speaker identification
Recognize unique speakers



Sentiment analysis
Detect emotions in text and speech

Fairness, ethics, and explainability

We, as scientists, have a say in the future of AI



From physics to machine learning

My background

From physics to machine learning



2012: M2 ICFP Theoretical physics

2013-2016: PhD in statistical physics @LPTENS



Feb 2017: senior ML scientist @ Snips

2019: head of ML research @ Snips

Today: head of ML research @ Sonos

A few takeaways

(please go ask other people too)

- PhD? Postdoc?
- Working at a startup company
- Physicists and machine learning

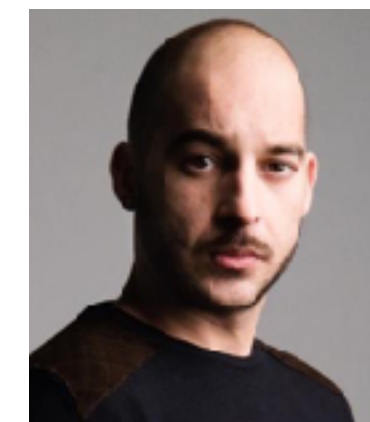
**Physicists
@ Snips/Sonos**



Me :)



Raffaele Tavarone
ML Director
Acoustic modeling



Francesco Caltagirone
ML Director
Spoken Language
Understanding



Alaa Saade
Sr ML Scientist
Now: DeepMind



Stéphane d'Ascoli
ML research intern
Now: PhD ENS & FAIR

Sonos Voice Experience

Sonos

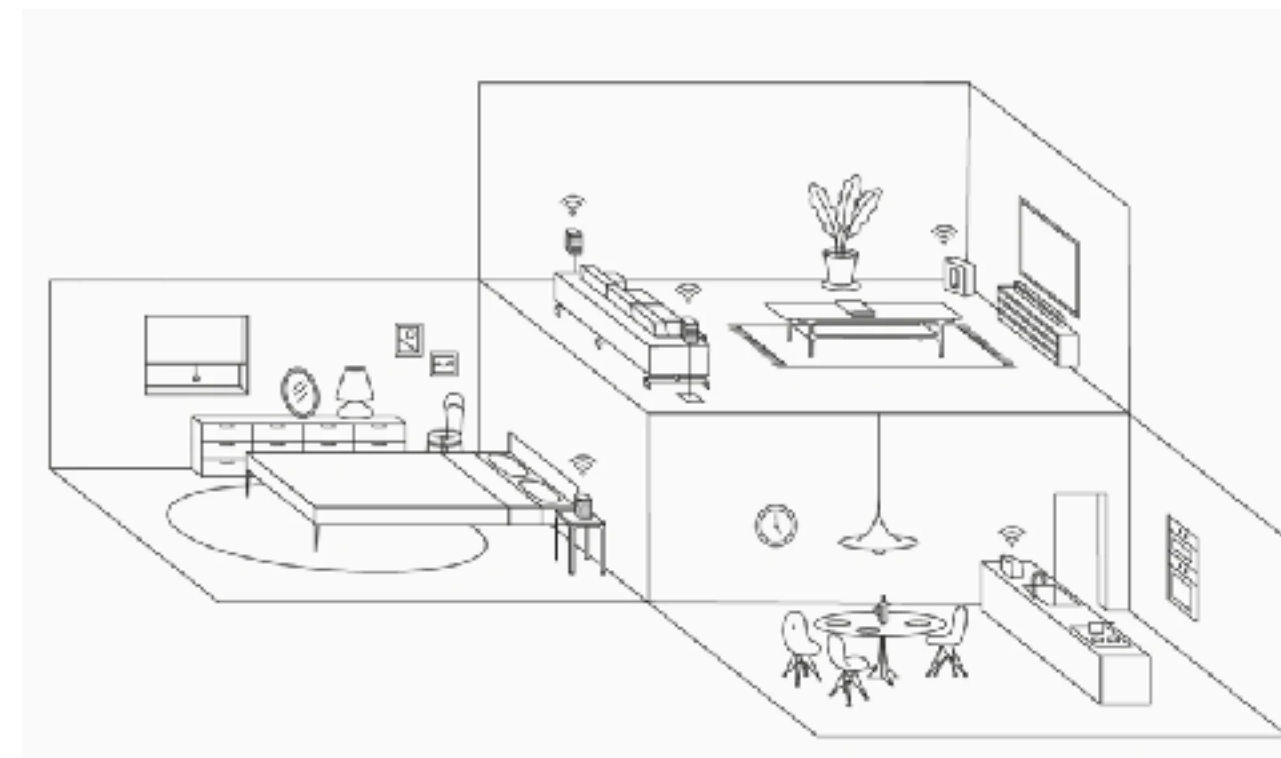
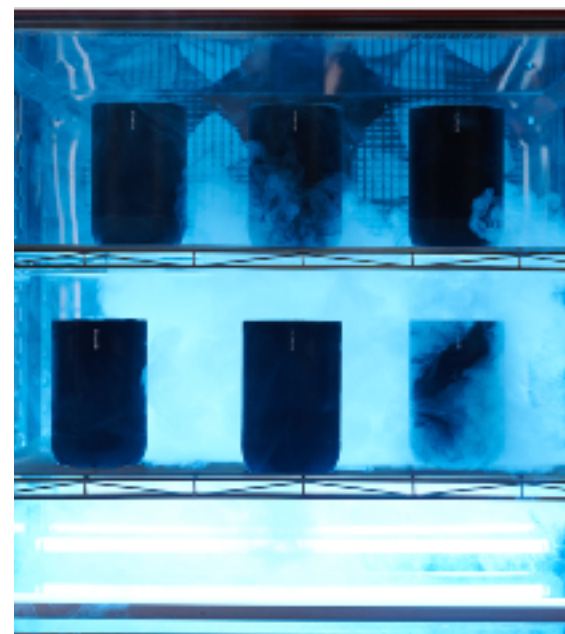
Home sound system

Sonos

- ◆ Smart speakers in multi room
- ◆ Hardware & software (~400 people)
- ◆ 1.5k employees in 6 countries (US, France, Netherlands, UK, Australia, China)

Sonos Voice Experience

- ◆ Ex Snips, joined Sonos 2019
- ◆ 50 employees: ML, Linguistics, SW engineering
- ◆ Audio signal proc, speech, NLP, embedded systems

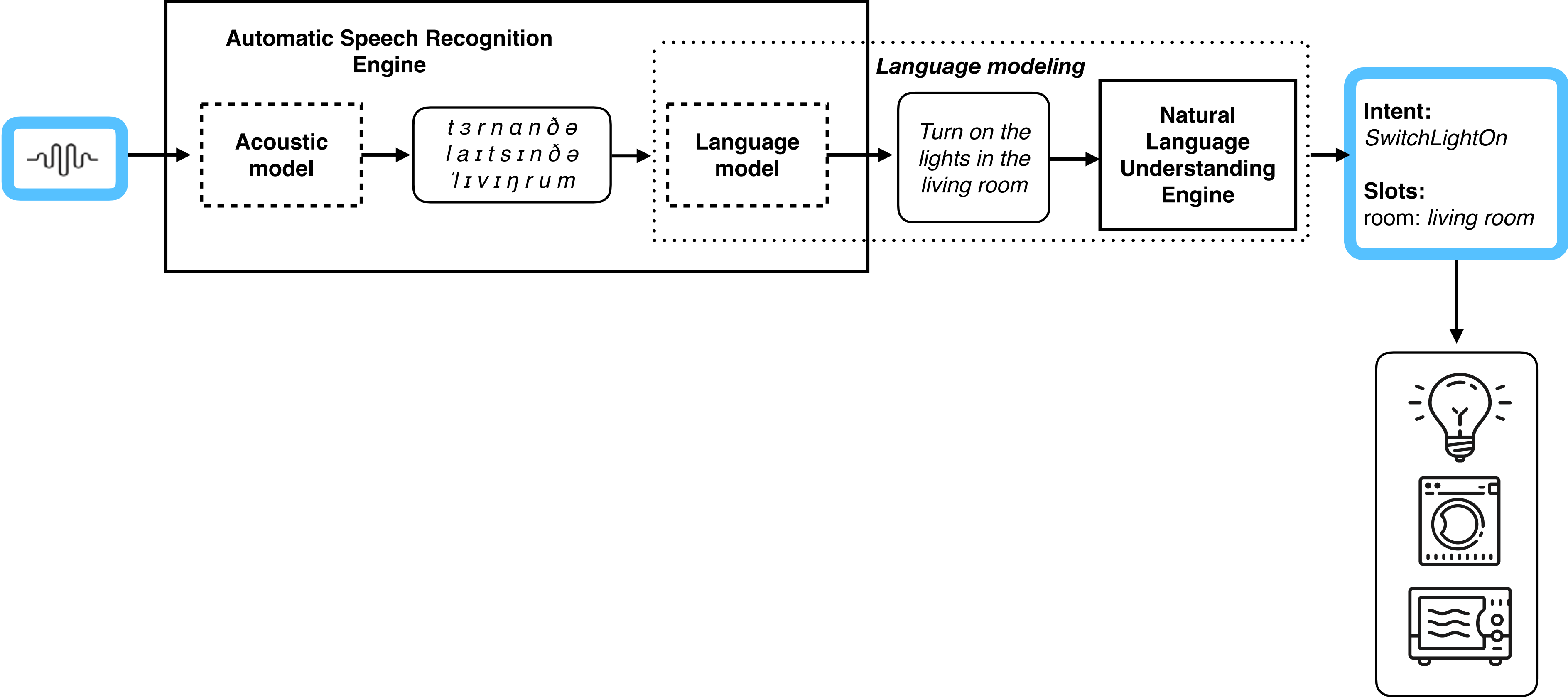


Brand new tech blog:

<https://tech-blog.sonos.com/>

Machine Learning at Sonos Voice Experience

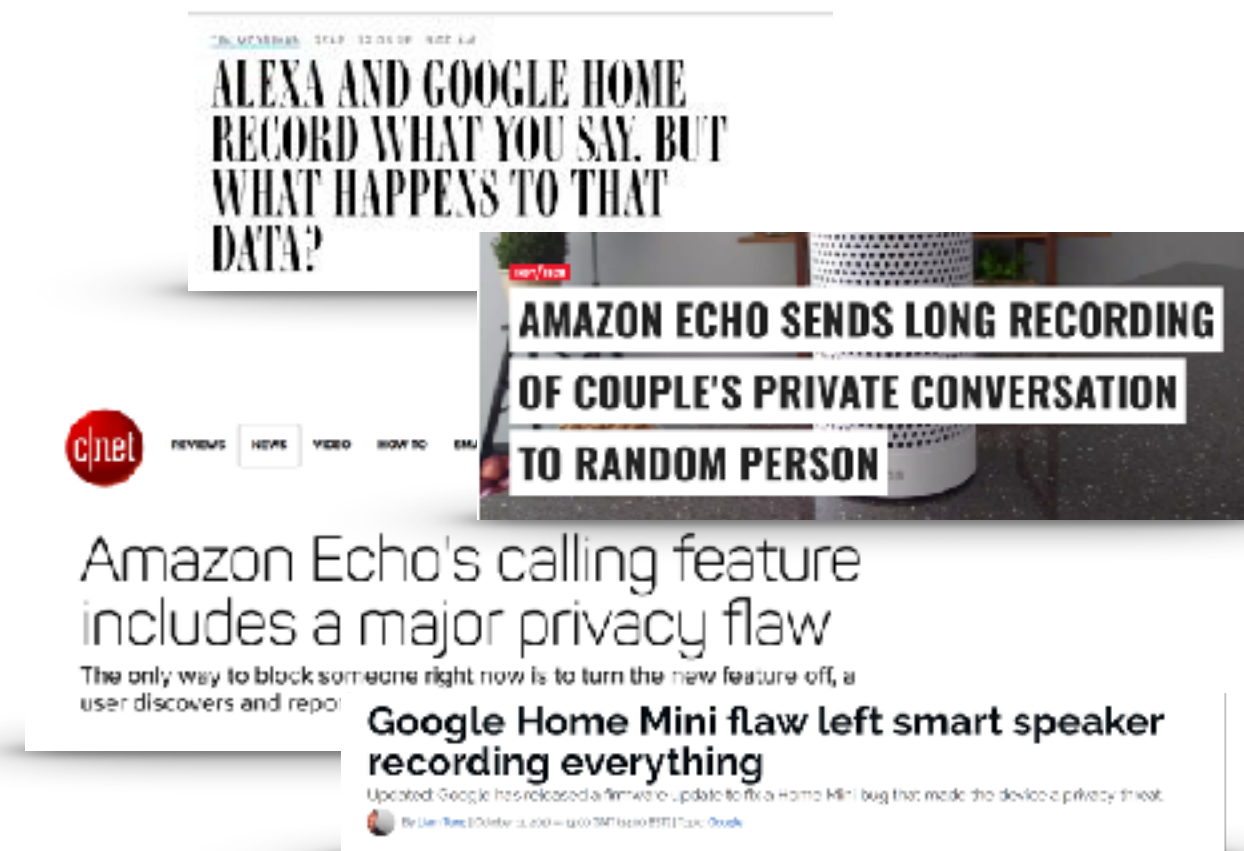
Spoken Language understanding for voice assistants



Our approach

Our own voice in a vast ecosystem

Privacy by design



Resource constrained ML: small data & hardware



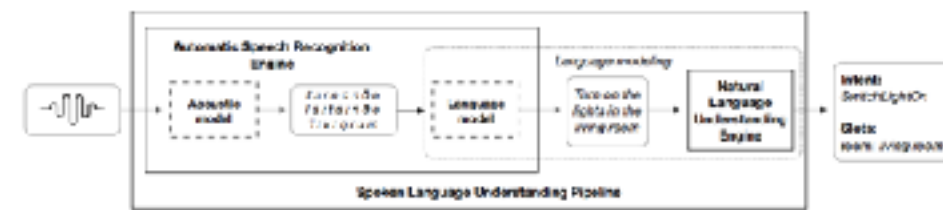
- ✓ A new **popular trend** in the ML community (low resource ML, transfer learning, miniaturization, etc)
- ✓ Numerous conferences and workshops on the topic
- ✓ Towards a **safer, greener and more private** conversational AI

Research activity

Publishing research in industry

**Snips voice platform:
an embedded spoken language
understanding system for private-
by-design voice interfaces**

arXiv

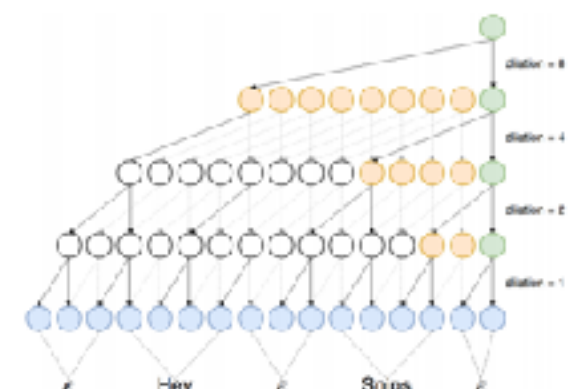


ICML 2018
Workshop PiMLAI



**Efficient keyword spotting using
dilated convolutions and gating**

arXiv

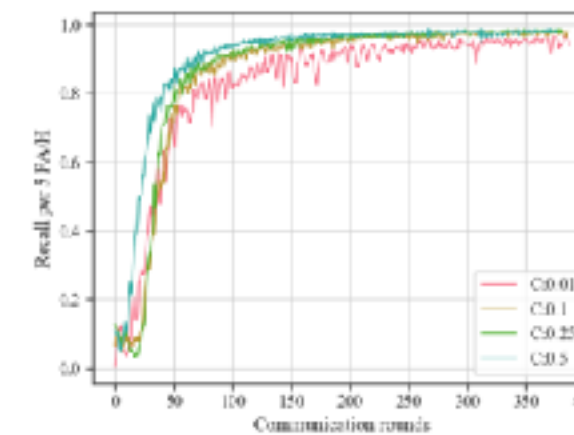


ICASSP 2019
Main track



**Federated learning for keyword
spotting**

arXiv



ICASSP 2019
Main track



**Spoken language understanding on
the edge**

arXiv

Quantity	Close field		Far field	
	Snips	Google	Snips	Google
Intent classification (F1)	0.92	0.89	0.84	0.86
Perfect parsing (%)	0.84	0.79	0.72	0.73

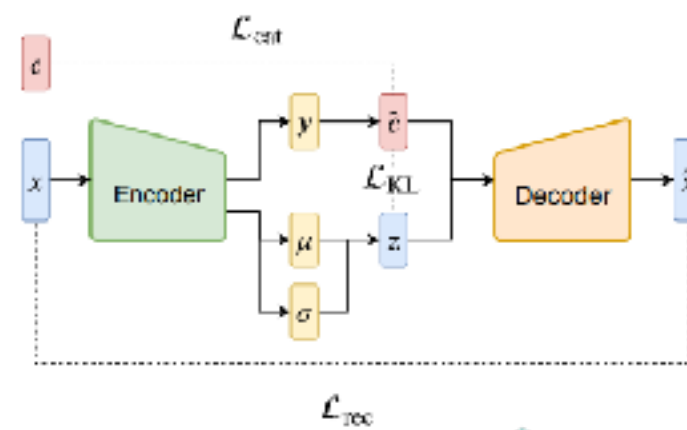
Table 3. End-to-end generalization performance on the "Smartlights" assistant: comparison with Google's Dialogflow cloud service on a 5-fold cross-validation experiment.

NeurIPS 2019
Workshop EMC2



**Conditioned text generation with
transfer for closed-domain dialogue
systems**

arXiv

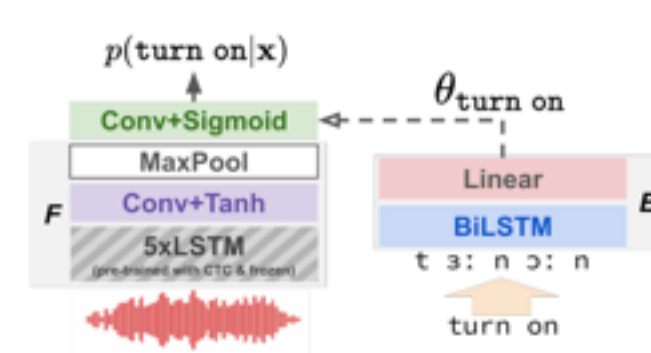


SLSP 2020
Main track



**Predicting detection filters for small
footprint open-vocabulary keyword
spotting**

arXiv

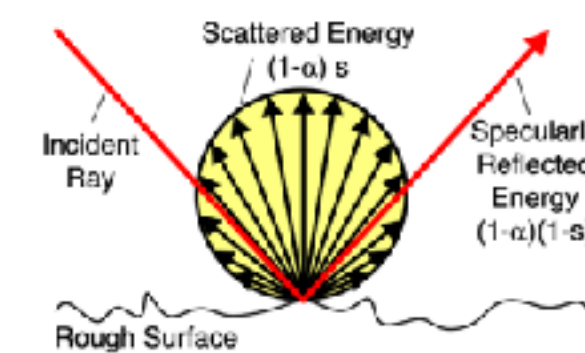


Interspeech 2020
Main track



**A study on more realistic room
simulation for far-field keyword**

arXiv

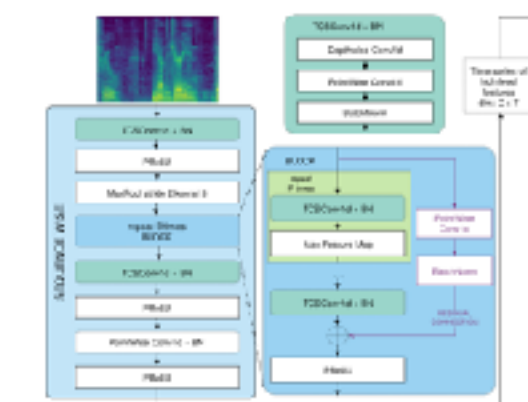


APSIPA 2020
Main track



**Small footprint Text-Independent
Speaker Verification for Embedded
Systems**

arXiv



ArXiv preprint
2020

Thank you for your attention!

Questions?

SONOS