



**SAILLABS**

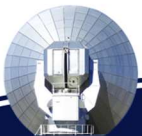


**T E C H N O L O G Y**

**Christoph Prinz / Automatic Speech Recognition Research Progress Hits the Road**

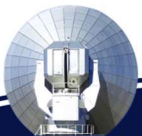
# Traditional Automatic Speech Recognition Paradigms

- Statistical Models
  - Model building: Parameters estimated from recorded audio and transcriptions
  - Recognition: most likely words with respect to observed acoustic speech signal and syntax of language
  - Most likely sequence is  $\max P(W | A)$ 
    - $W$  is word sequence
    - $A$  is acoustic speech signal
  - **Acoustic Models (AM): assign probabilities to acoustic information using Hidden Markov Models (HMMs). HMM model variance in temporal and spectral dimension.**
  - Language Models (LM): assign probabilities to sequence of words using n-gram counts represented in Weighted Finite State Transducers (WFST)



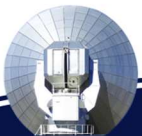
# Traditional Automatic Speech Recognition Paradigms

- Statistical Models
  - Model building: Parameters estimated from recorded audio and transcriptions
  - Recognition: most likely words with respect to observed acoustic speech signal and syntax of language
  - Most likely sequence is  $\max P(W|A)$ 
    - $W$  is word sequence
    - $A$  is acoustic speech signal
  - **Acoustic Models (AM): assign probabilities to acoustic information using Deep Neural Networks (DNNs). DNNs have multiple hidden layers, can enable composition of features from lower layers, which gives them a huge learning capacity and thus the potential of modeling complex patterns of speech (deep learning).**
  - Language Models (LM): assign probabilities to sequence of words using n-gram counts represented in Weighted Finite State Transducers (WFST)



# Comparison & Short History of Deep Neural Networks (DNNs)

- Compared to HMMs, DNNs are observed to reduce word-error-rates (WER) on average by 25%. i.e. a HMM system having 80% correct (20% WER) will have 85% correct (15% WER) using DNNs. This is the most substantial single ASR technology steps in the last 10 years.
- DNNs theory is old, DNNs started appearing in the field 2010 (ICASSP & Interspeech Conference 2010/2014)
- Used by all major commercial ASR systems e.g., Microsoft Cortana, Xbox, Skype Translator, Google Now, Apple Siri, Baidu and iFlyTek voice search, and a range of Nuance speech products, etc.



Danke - Thank you - Merci - Gracias - Terima kasih - Dziękuję - Спасибо - تشكر - Takk - Tack - Grazie - Gràcies - 感謝 - Ευχαριστώ - תודה

# SAILLABS

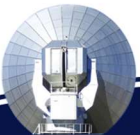
TECHNOLOGY

Your name here \ Name of your PP

Mariannengasse 14 · 1090 Vienna · Austria

Ph +43-1-580 95- 0 · Fax +43-1-580 95-580

[www.sail-labs.com](http://www.sail-labs.com)



SAILLABS