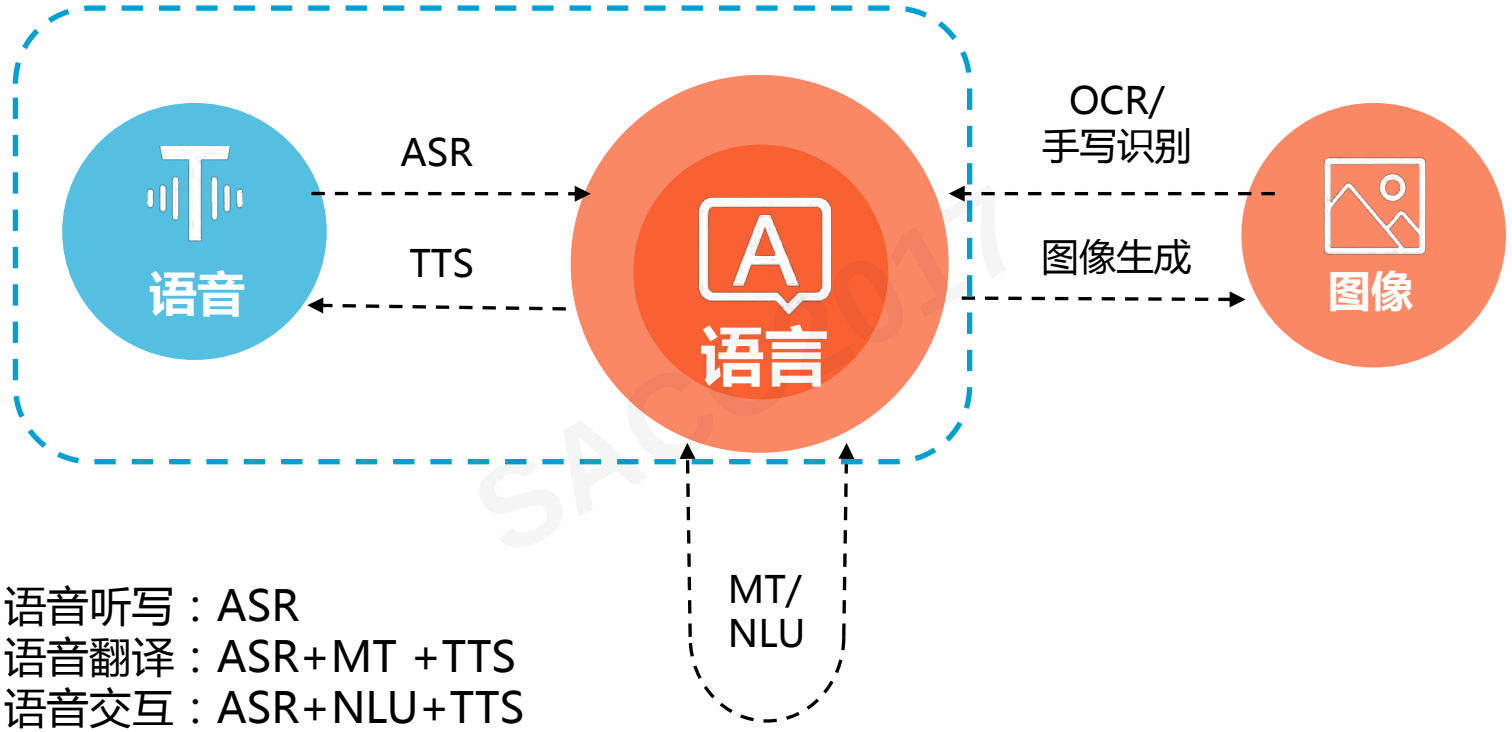
搜狗智能语音之路

搜狗语音交互技术中心 | 陈伟

从移动互联网时代迈向智能时代

语言是人工智能的核心

语言是思想和知识的载体



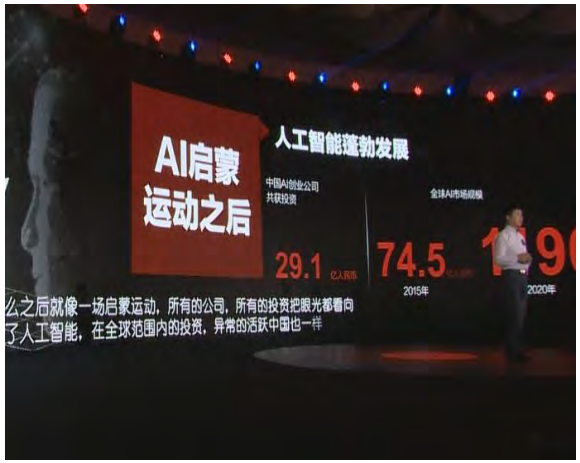
语音听写：ASR
语音翻译：ASR+MT +TTS
语音交互：ASR+NLU+TTS

语音听写技术已走向实用

语音识别可以更好提升输入/记录的效率

语音听写技术已走向实用

语音识别可以更好提升输入/记录的效率



语音听写技术落地法院庭审

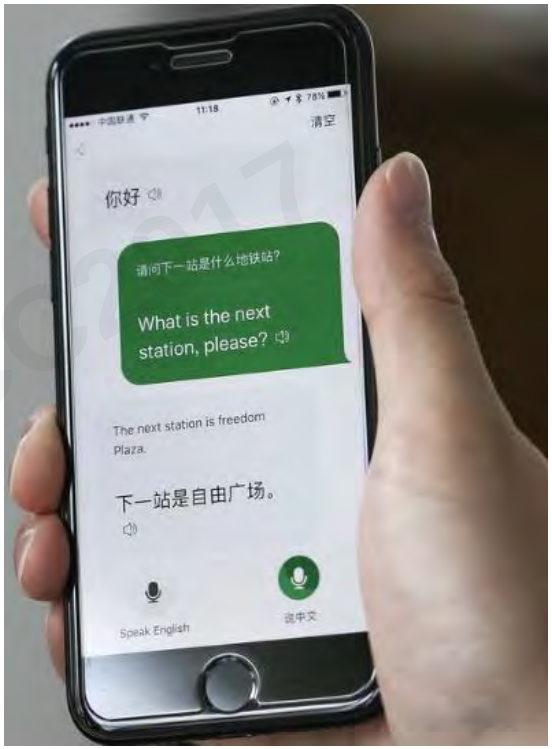
截止到8月15日

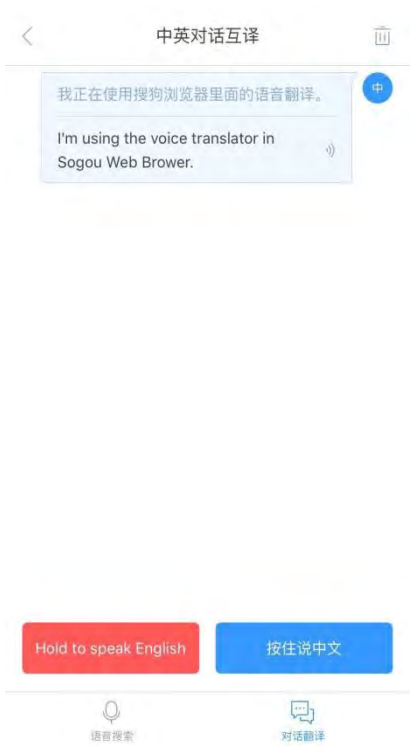
17个 省份	50家 法院
357场次 庭审	194场次 庭审直播

语音翻译技术逐渐可用

更好服务跨语言的交流

应用场景	
出国旅行	国际交流
演讲同传	视频字幕





搜狗语音同传技术

语音断句

语音识别

1.大家早上好
2.我叫李磊
3.很高兴认识大家

文本断句

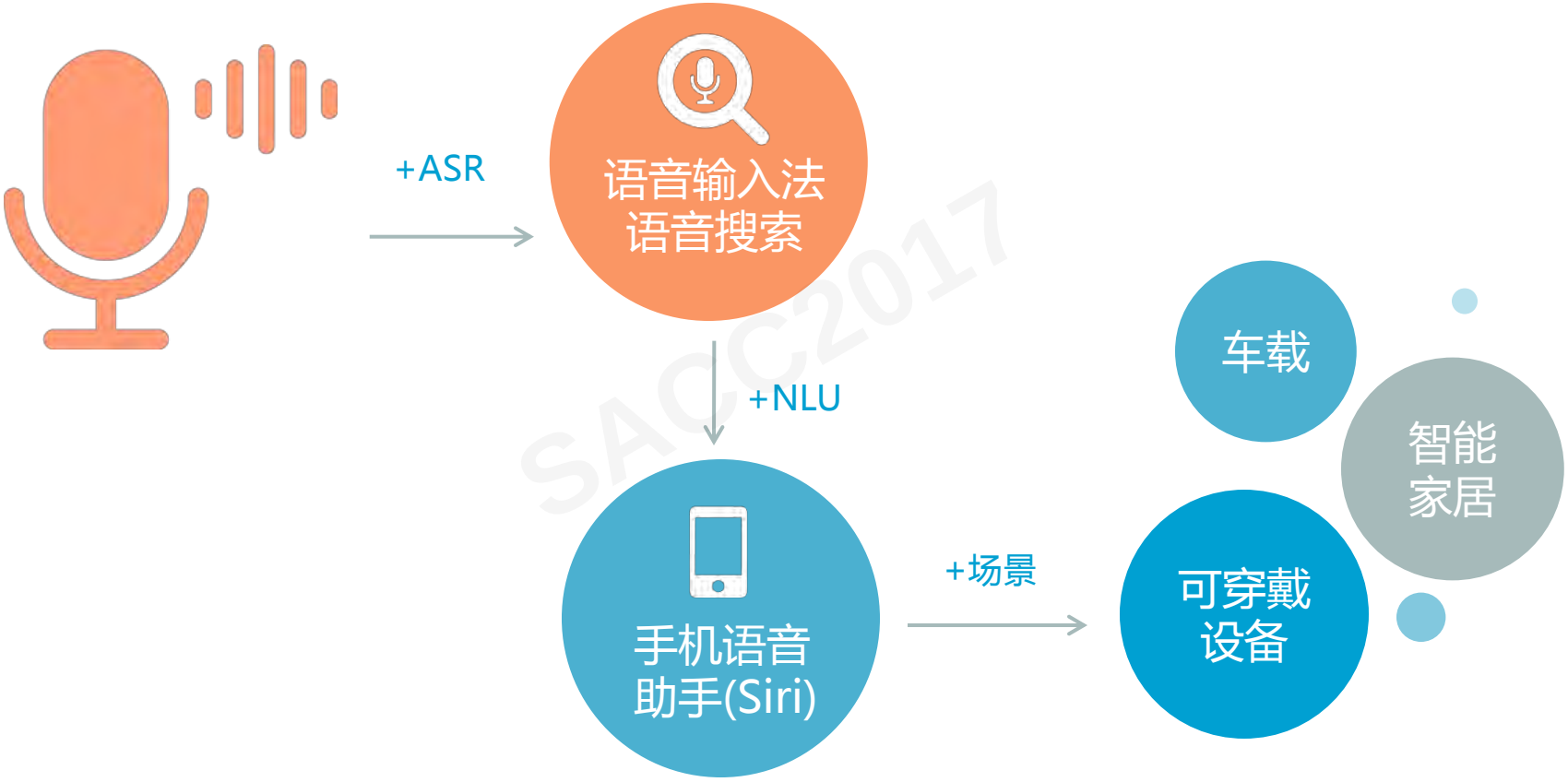
机器翻译

1.Good morning Everyone
2.My name is Lei Li
3.Nice to meet you

搜狗语音同传技术


Sogou 搜狗
搜索的未来
就是人工智能时代的皇冠
In the future, search will be the CROWN
of the AI Era
搜索的未来是什么？
· What is the future of search ?
人工智能的未来是什么？
· What is the future of AI ?
为什么是皇冠？
· Why will search be the crown ?
3rd WORLD INTERNET CONFERENCE
crown of the Artificial Intelligence era. Why do you say that? What is the future of
search? What's the future of Artificial Intelligence?

语音交互产品的不断演进



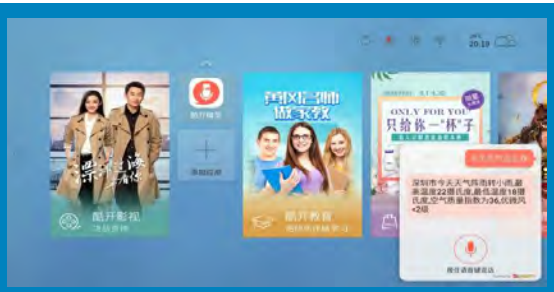
针对刚需场景提供更自然的交互体验

刚需场景下的语音交互产品

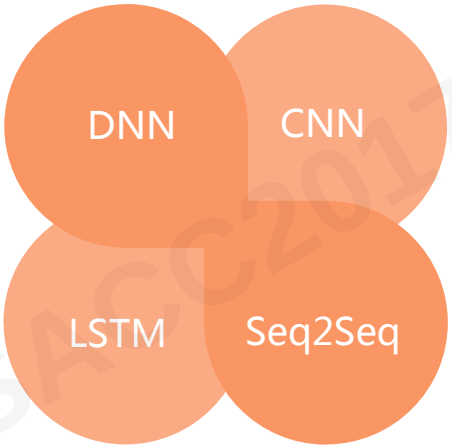
移动
可穿戴

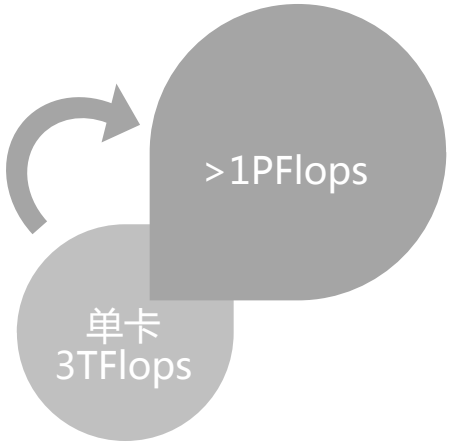
车载

智能
家居



搜狗语音深度学习规模演进



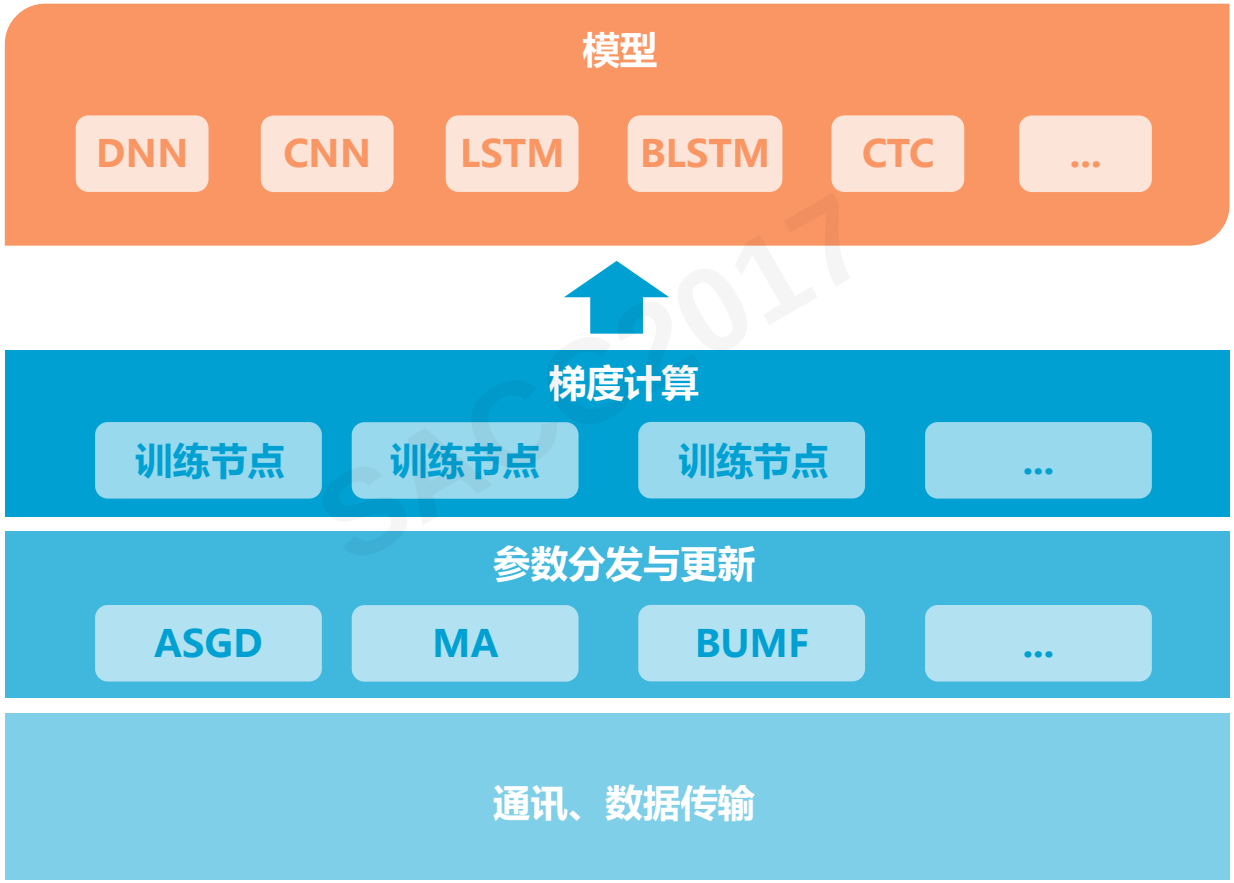


超大规模的语音数据

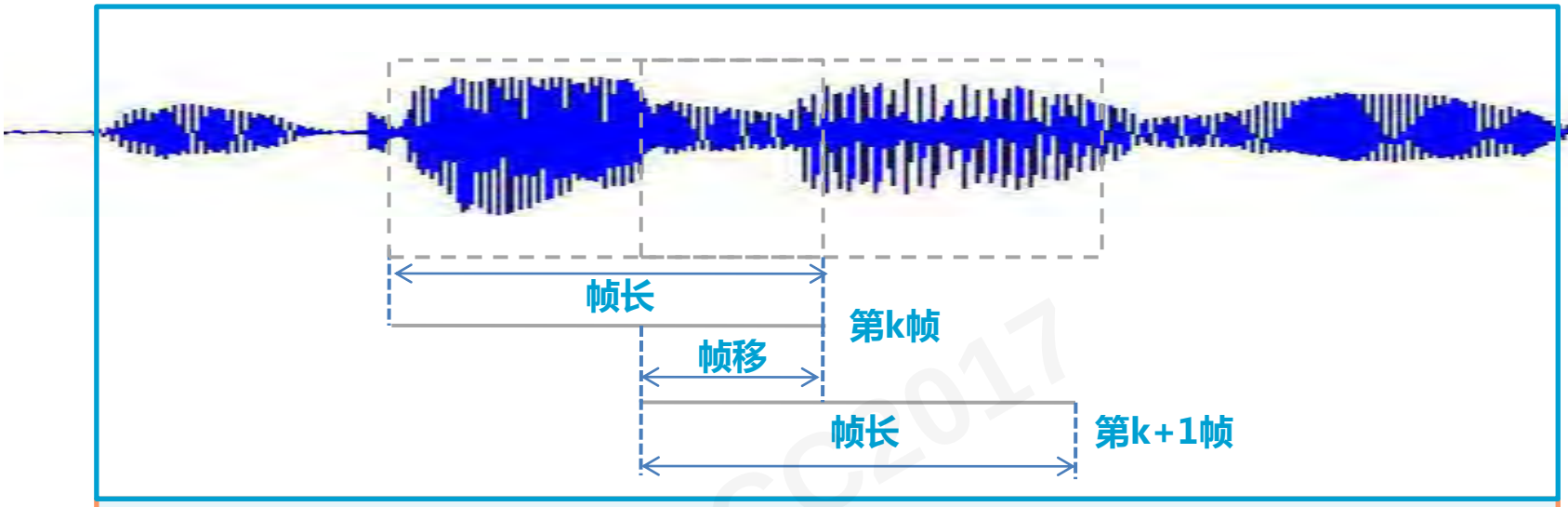
复杂的深度学习算法

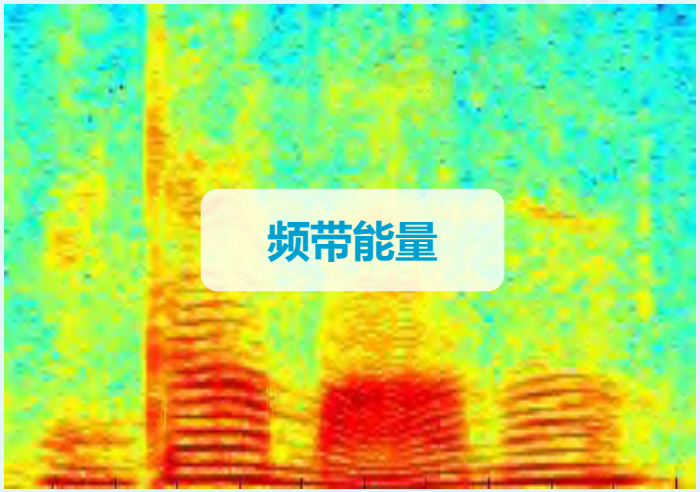
超强的运算平台

搜狗语音深度学习平台



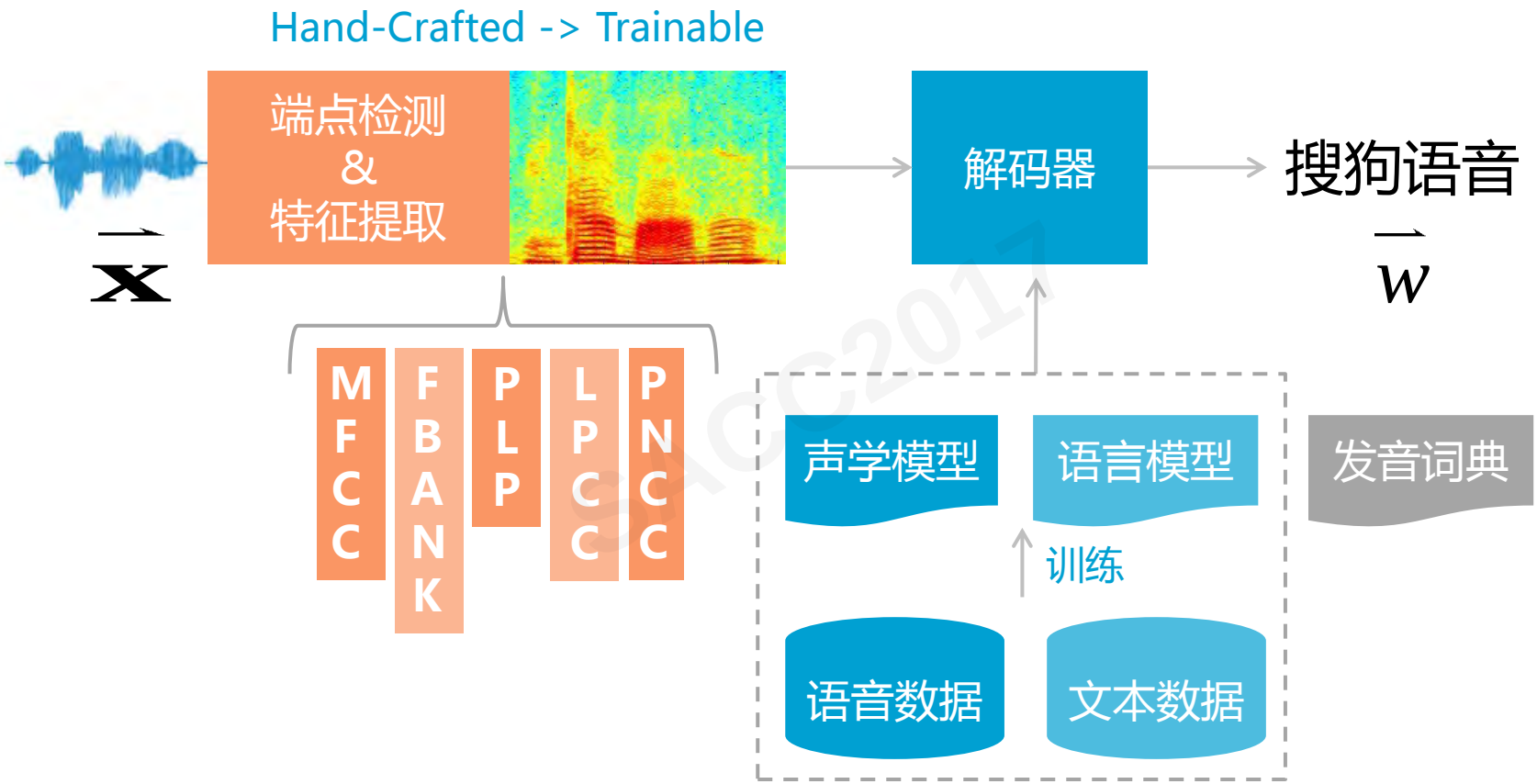
语音是时变+短时平稳的信号





- 语音的维度
  时域信号->语谱图
- 语音的短时平稳
  历史->未来

语音识别整体框架



语音识别-贝叶斯公式

$$\arg\max_{\vec{w}} p(\vec{w}\,|\,\vec{\mathbf{x}}) = \arg\max_{\vec{w}} \sum_{\vec{q}} p(\vec{w},\vec{q}\,|\,\vec{\mathbf{x}})$$

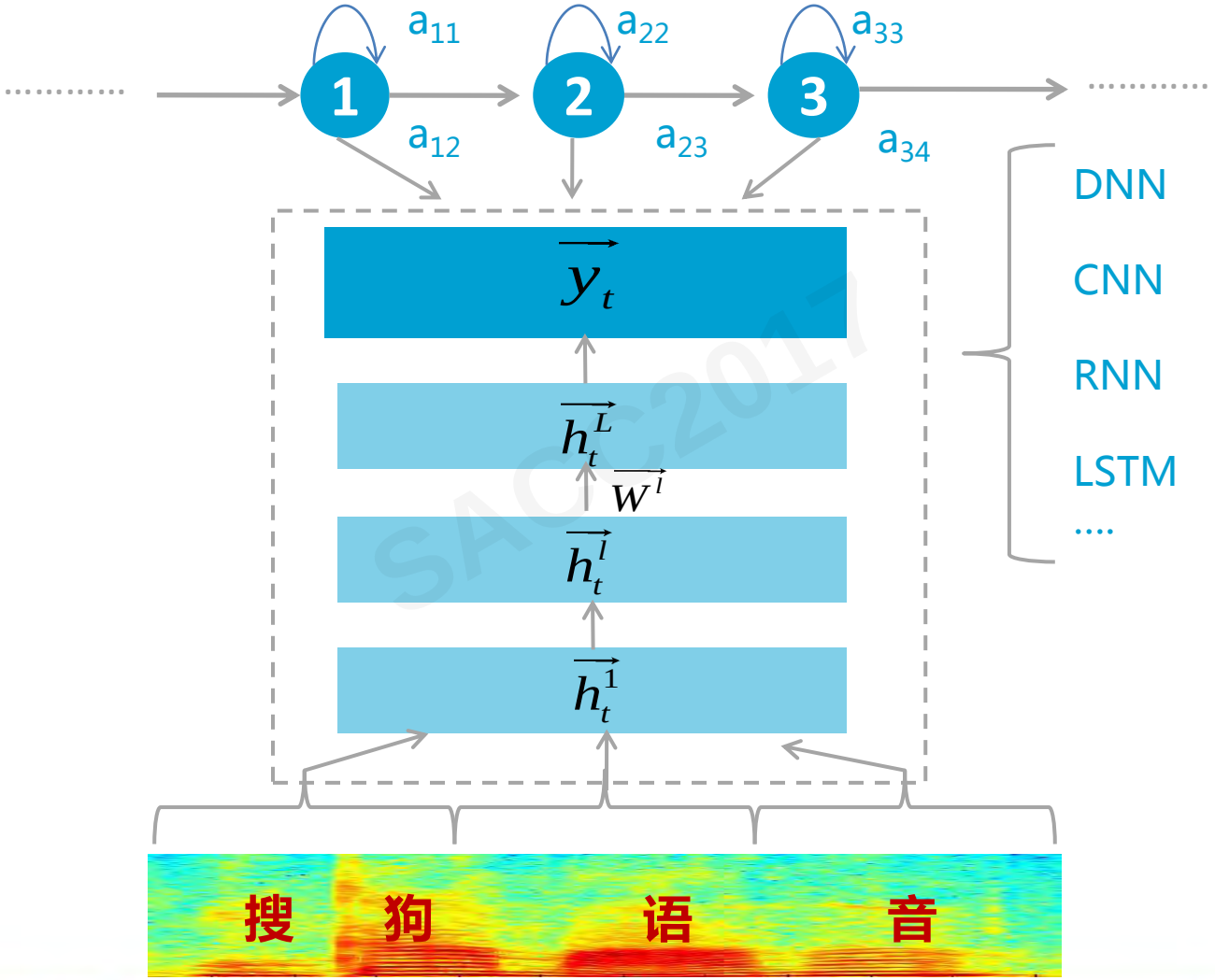
输出词序列　输入特征矢量　音素序列

$$= \arg\max_{\vec{w}} \sum_{\vec{q}} \frac{p(\vec{\mathbf{x}}\,|\,\vec{w},\vec{q})P(\vec{w},\vec{q})}{P(\vec{\mathbf{x}})}$$

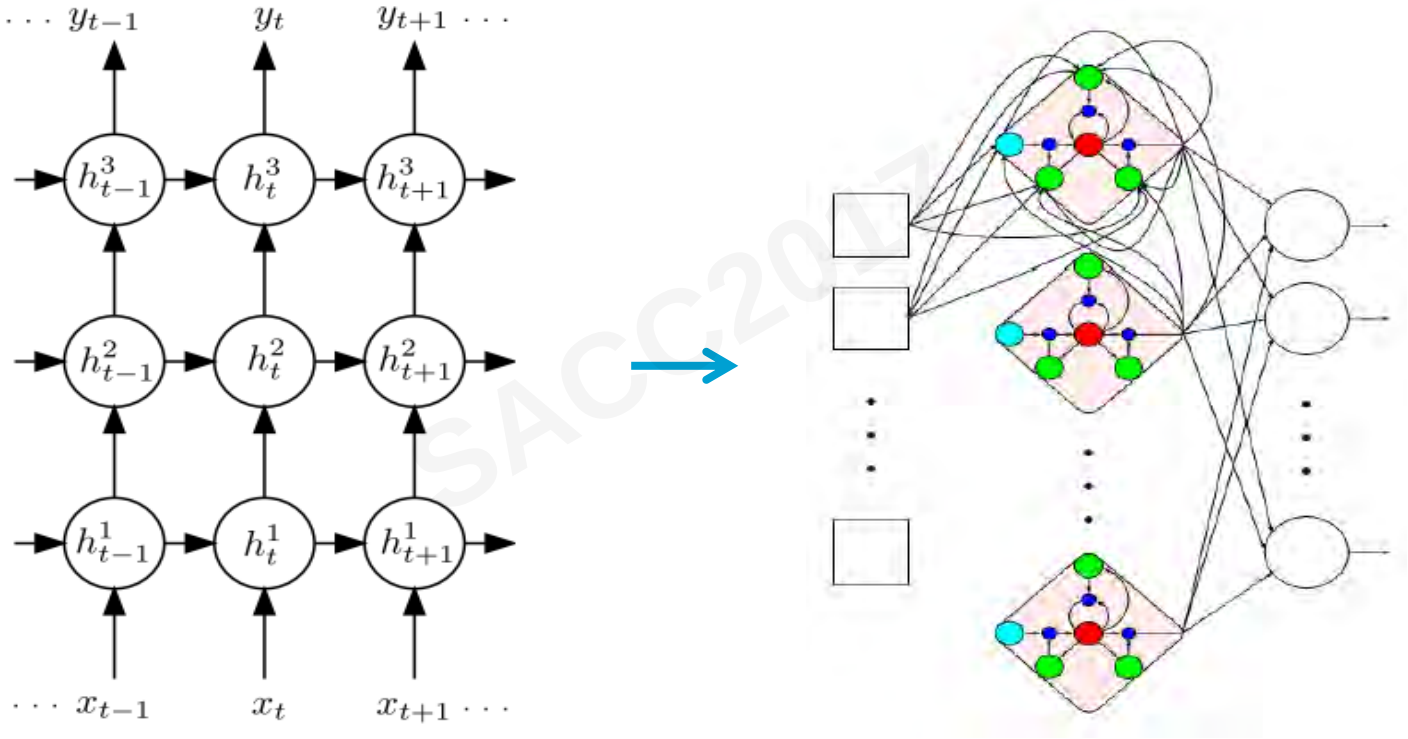
$$= \arg\max_{\vec{w}} \sum_{\vec{q}} p(\vec{\mathbf{x}}\,|\,\vec{q})P(\vec{q}\,|\,\vec{w})P(\vec{w})$$

声学模型　发音词典　语音模型

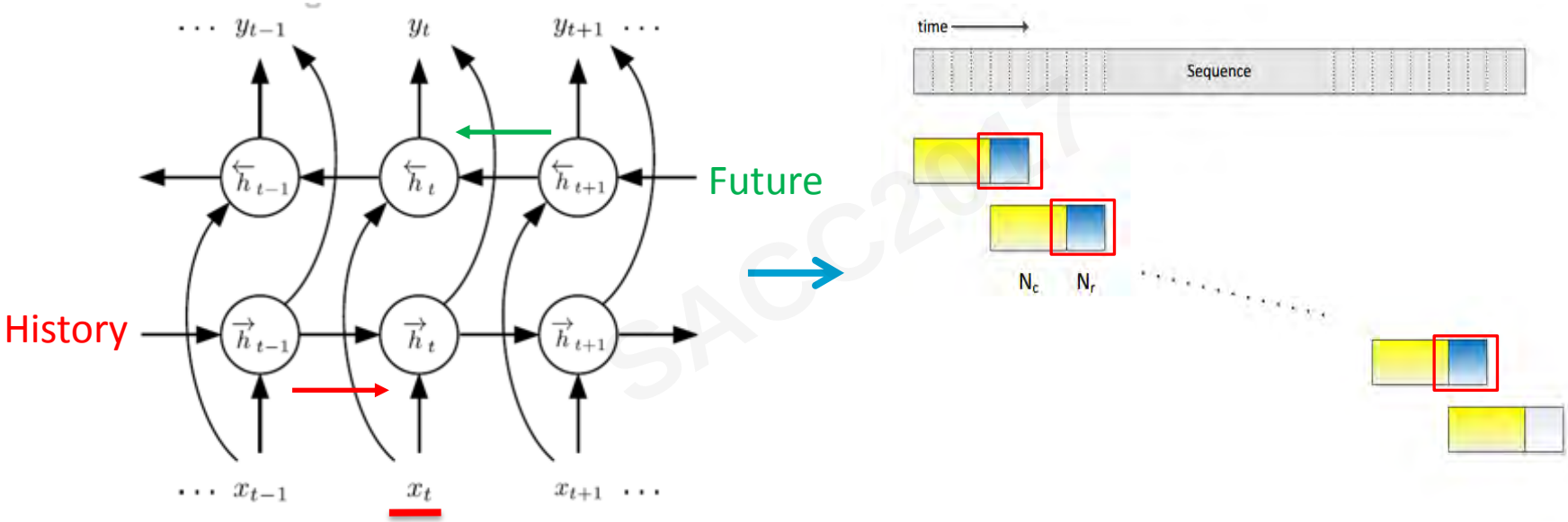
基于深度神经网络的声学建模



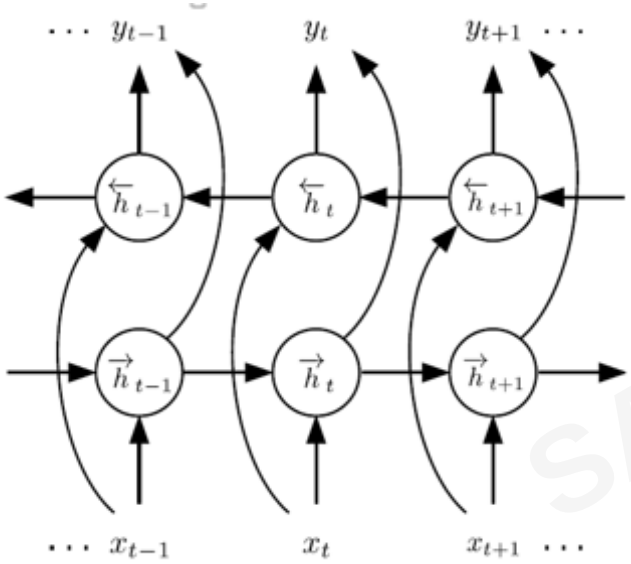
LSTM-RNN是声学模型建模的主要结构

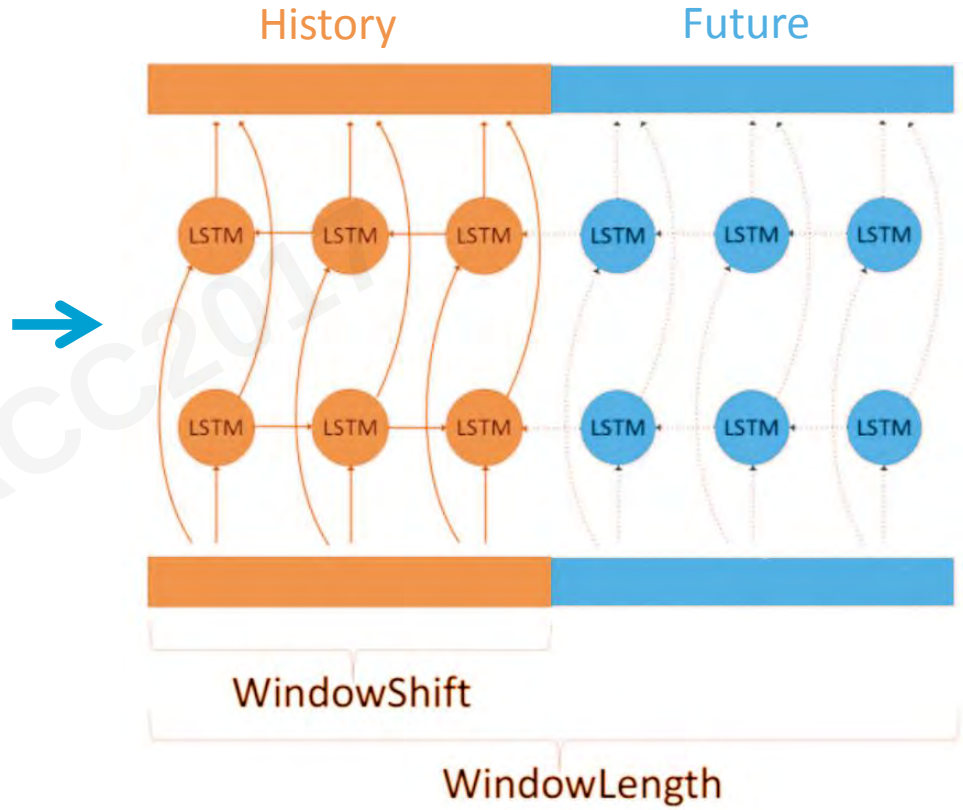


LSTM-RNN+双向特性

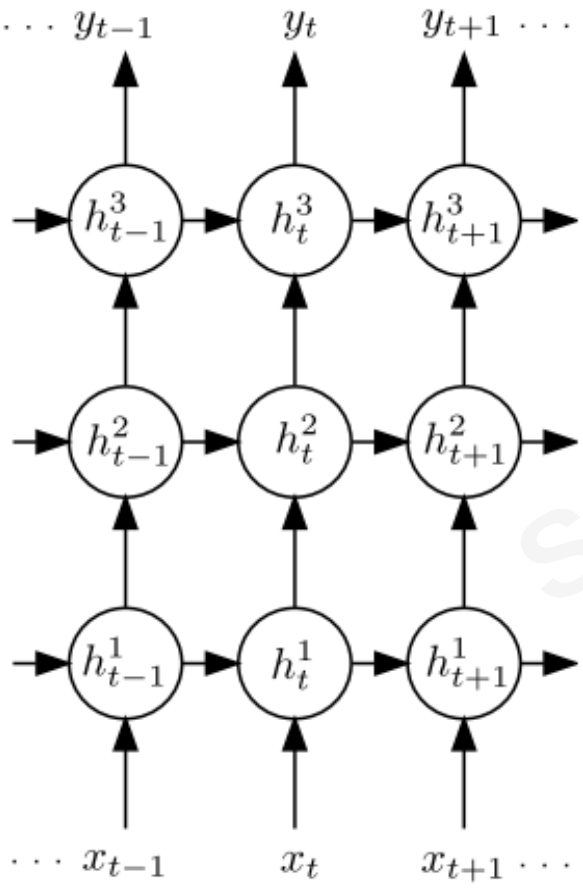


LSTM-RNN+双向特性

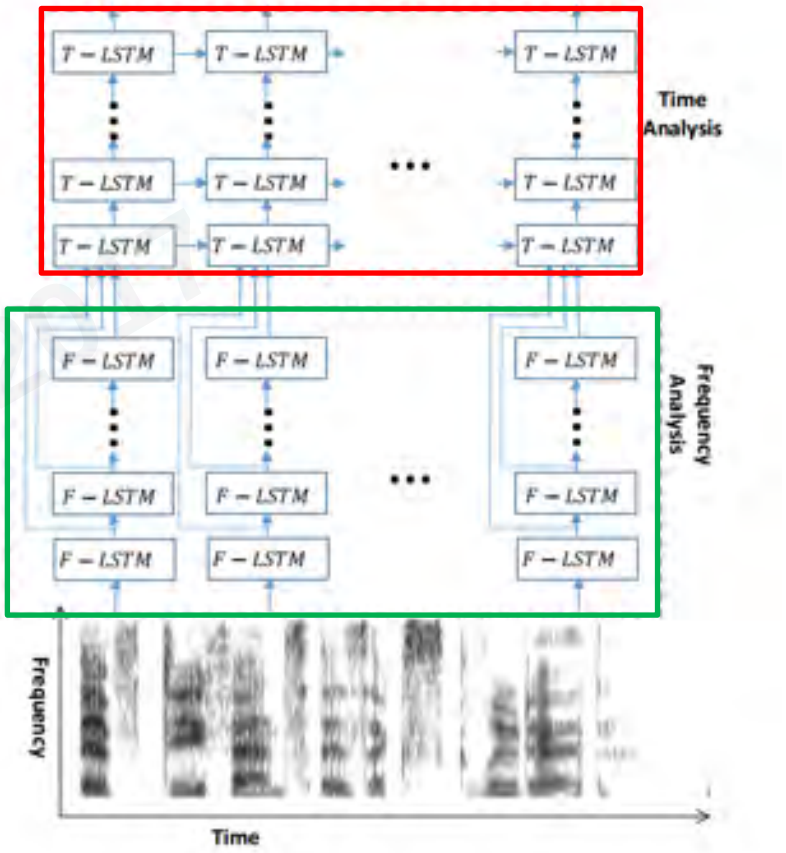




LSTM-RNN+频域扩展

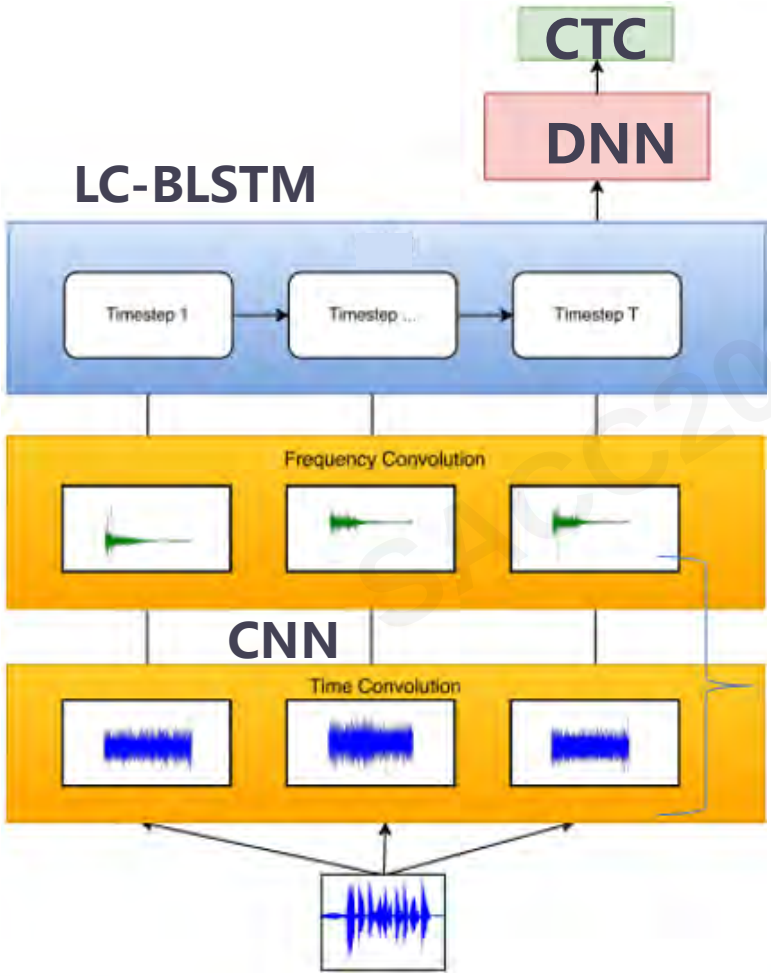


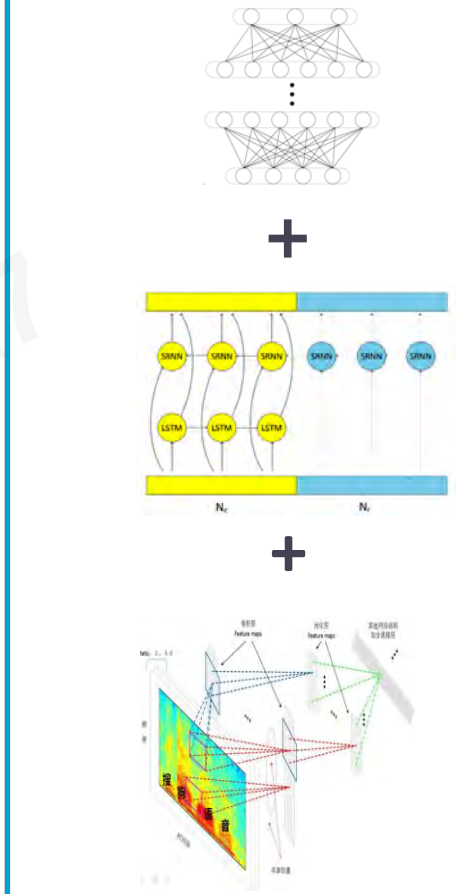
T-LSTM



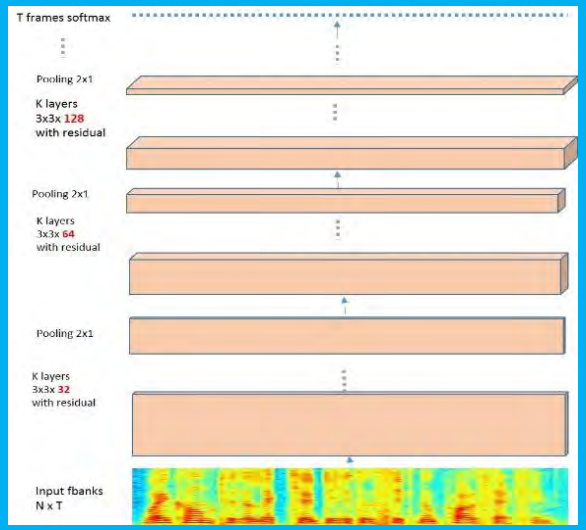
TF-LSTM

LSTM-RNN+多模型融合

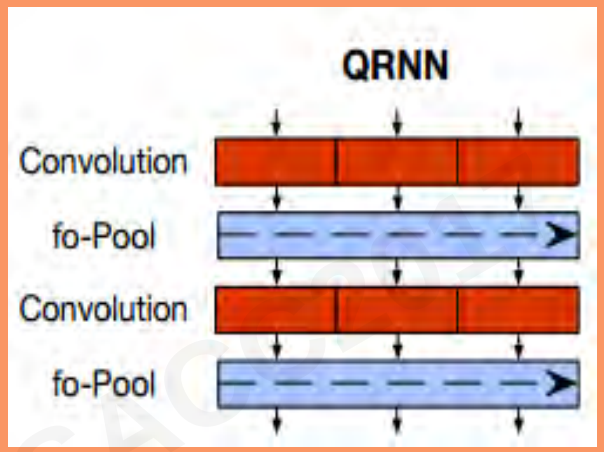




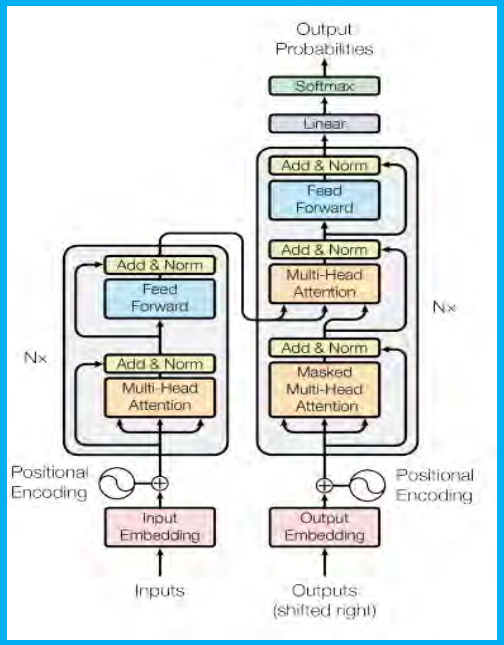
其他网络结构



DeepCNN



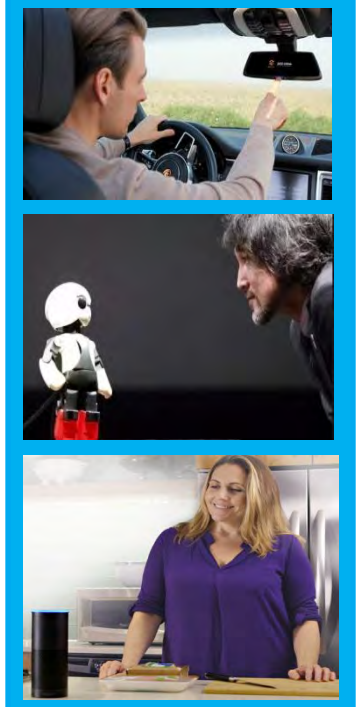
QRNN



Transformer-seq2seq

复杂场景下的语音识别问题仍未解决

近场的语音听写



高噪环境下的
语音交互

混响噪声

空调噪声

风噪

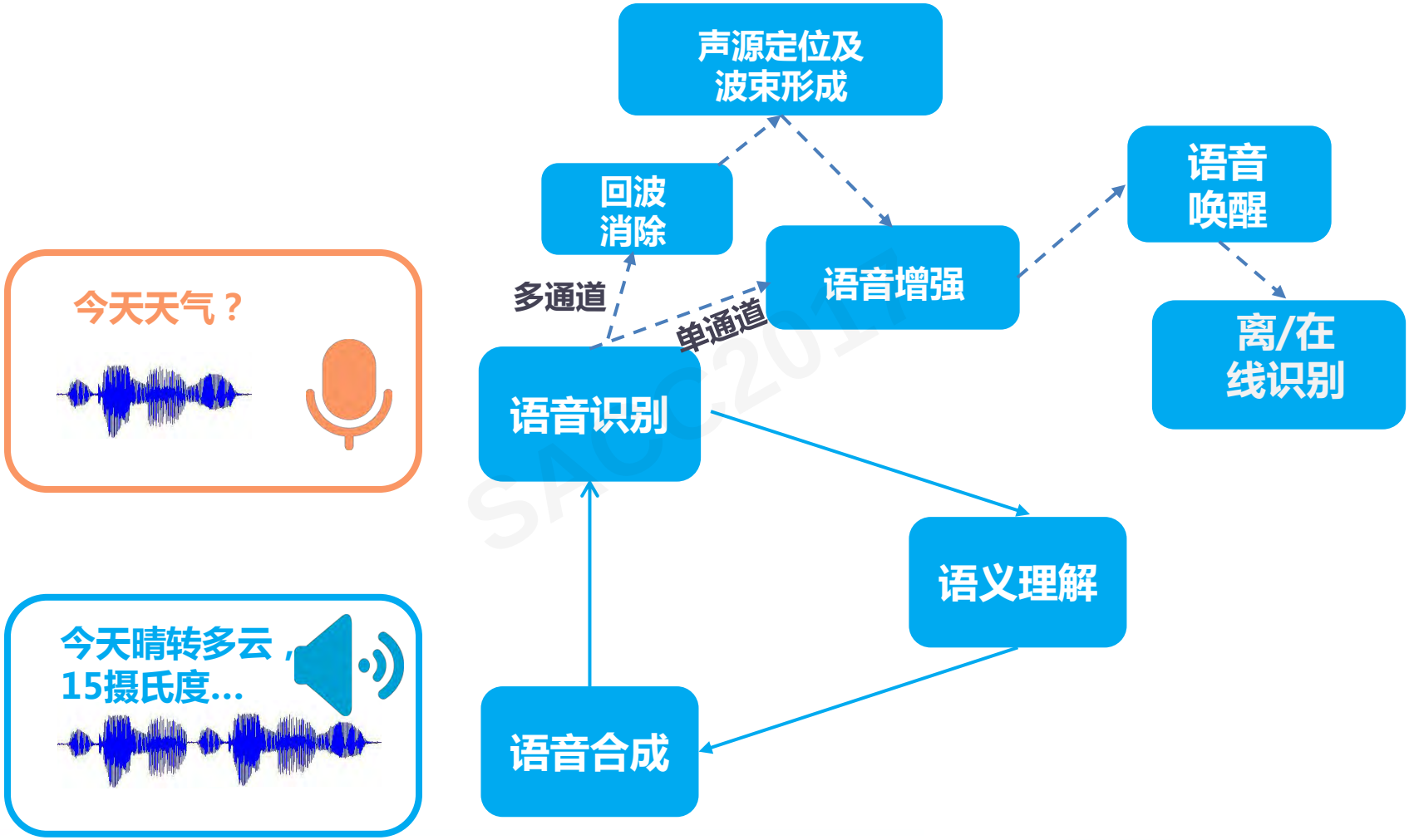

拨号

如此搞笑的效果只是口音产生吗？

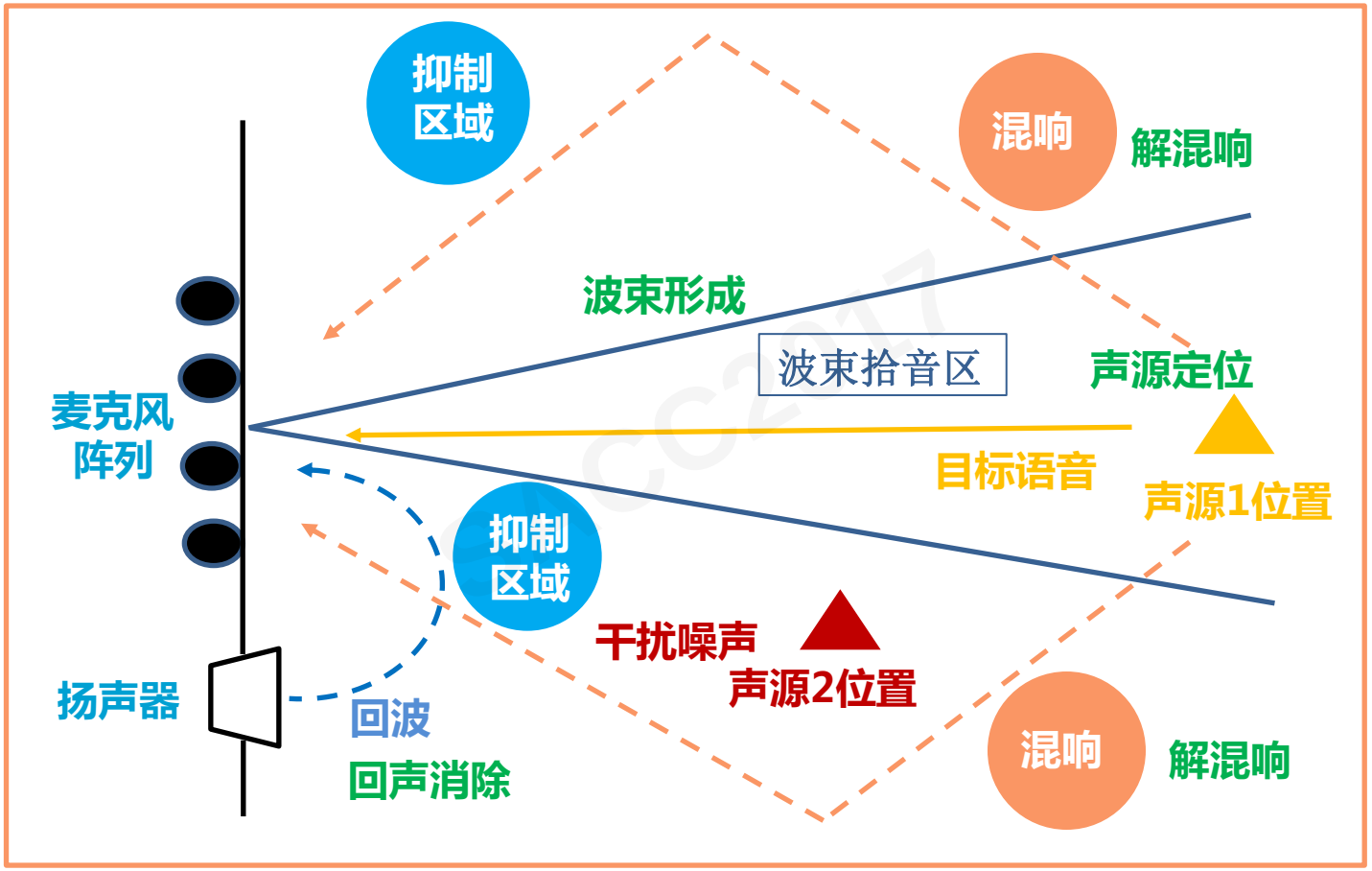
- **语音识别：** 口音、噪声、远场
- **语义理解：** 多轮对话、纠错容错
- **语音交互：** 全双工持续交互

单点能力 -> 系统整体能力

语音交互完整链路



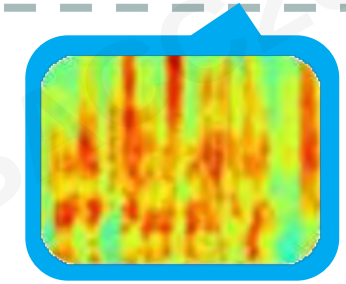
麦克风阵列技术


抑制
区域
混响
解混响
波束形成
波束拾音区
声源定位
麦克风
阵列
目标语音
声源1位置
抑制
区域
干扰噪声
声源2位置
扬声器
回波
回声消除
混响
解混响

基于深度学习的单通道语音增强

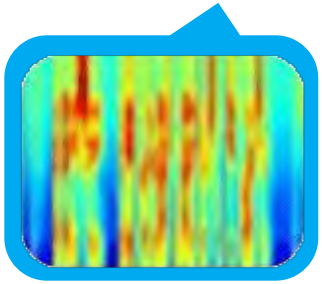
Noisy time-domain signal

STFT



Noisy spectral features

DNN feature mapping



Denoised spectral features


0 km/h
17:50
0km/h

WHAT IS POSSIBLE IN PRINCIPLE IS NOT ALWAYS
WHAT IS SIMPLE IN PRACTICE

THANKS