



第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017

搜狗智能语音之路

搜狗语音交互技术中心 | 陈伟

SACC
2017

北京·新云南皇冠假日酒店

IT168.com

ChinaUnix

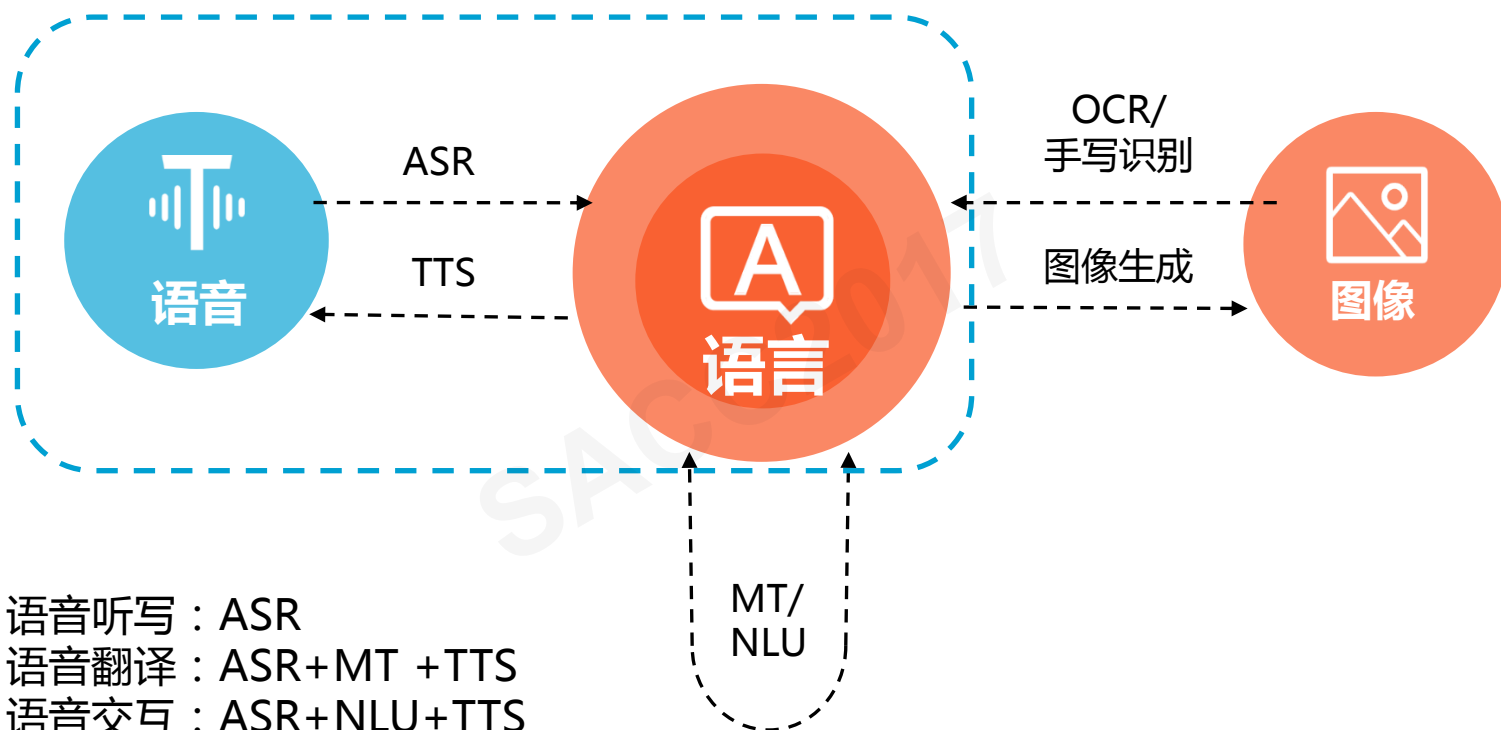
ITPUB

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



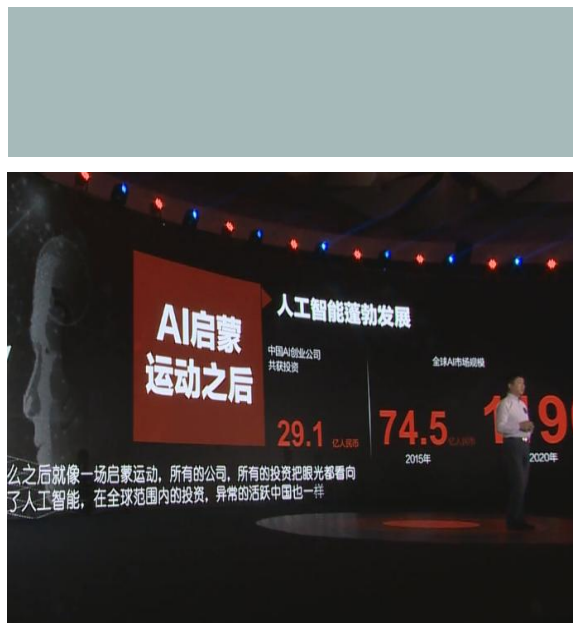
语音听写技术已走向实用

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



搜狗听写

语音转文字的速记工具

Android v1.1.3

iPhone v1.1.1

一边听一边写

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

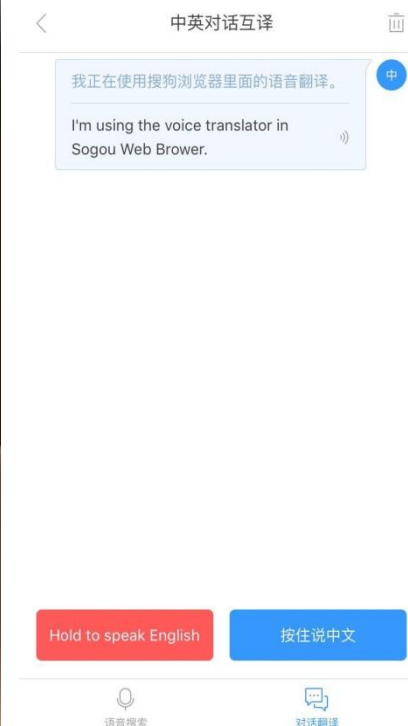
应用场景

出国
旅行

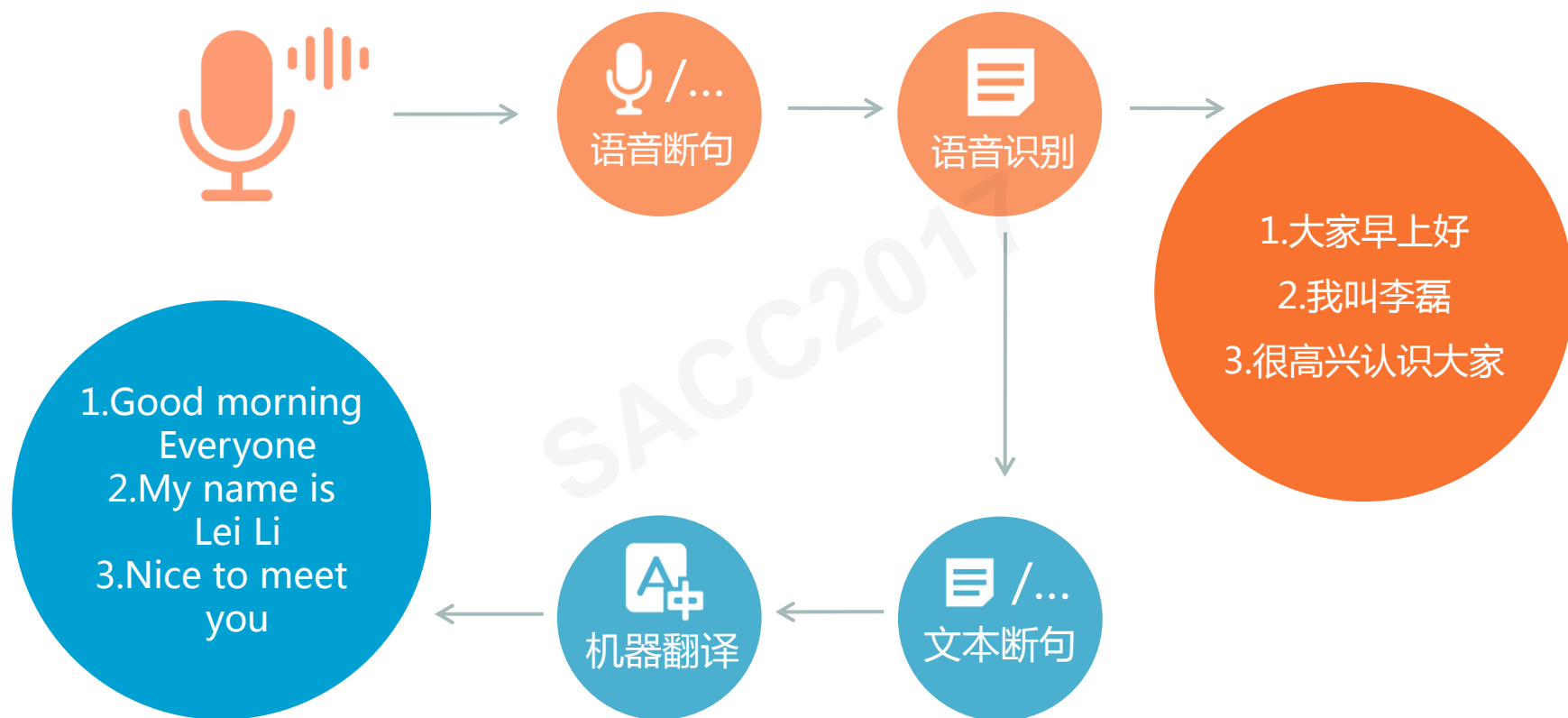
国际
交流

演讲
同传

视频
字幕



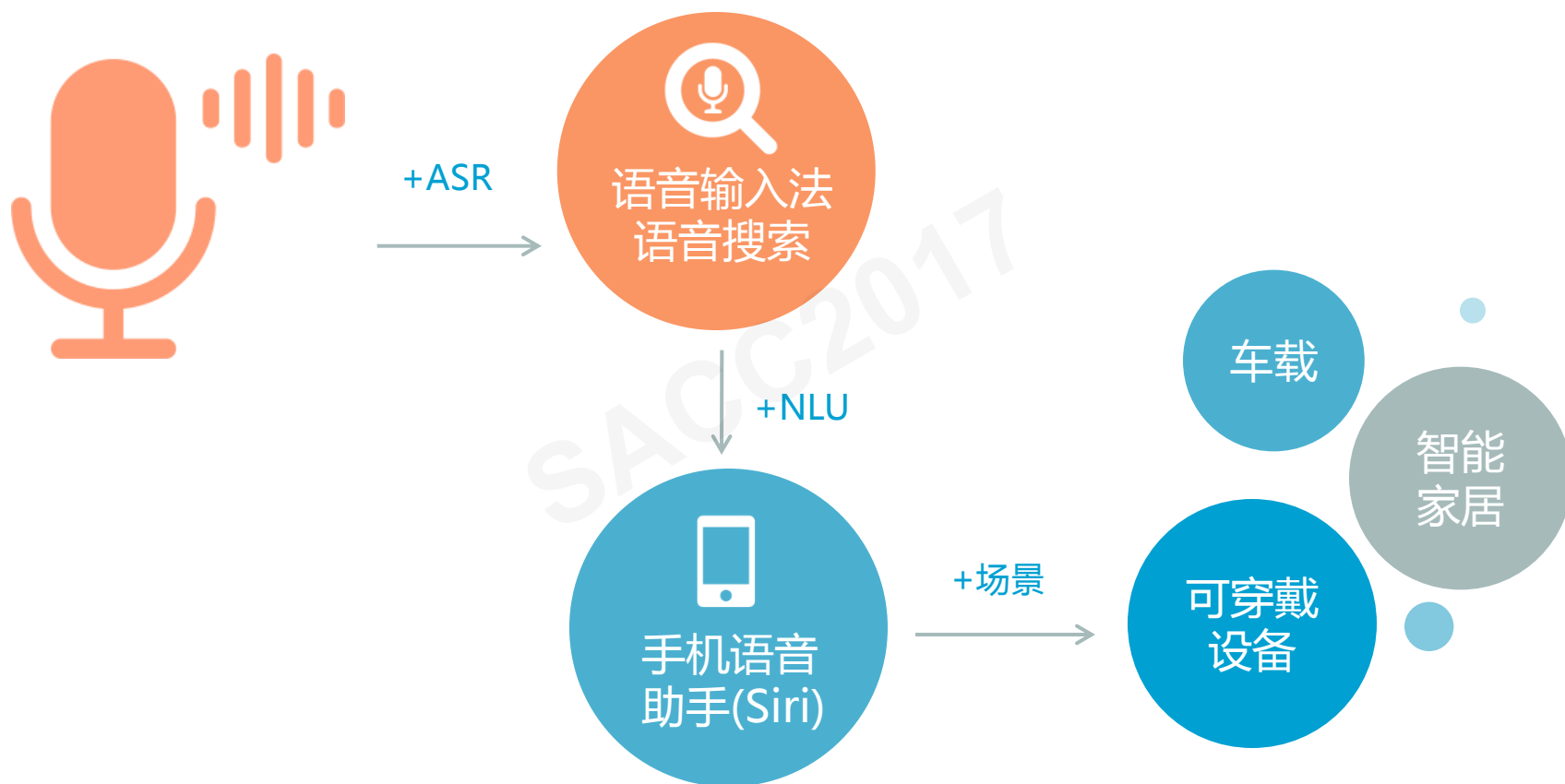
搜狗语音同传技术



搜狗语音同传技术



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



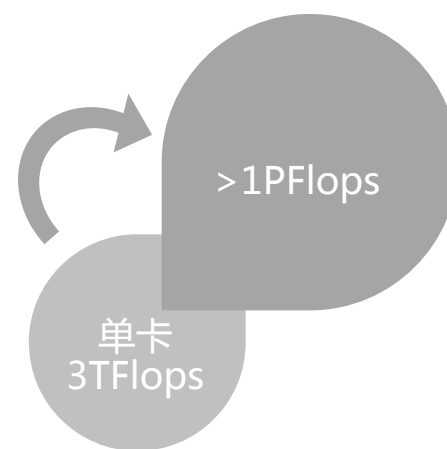
搜狗语音深度学习规模演进



超大规模的语音数据

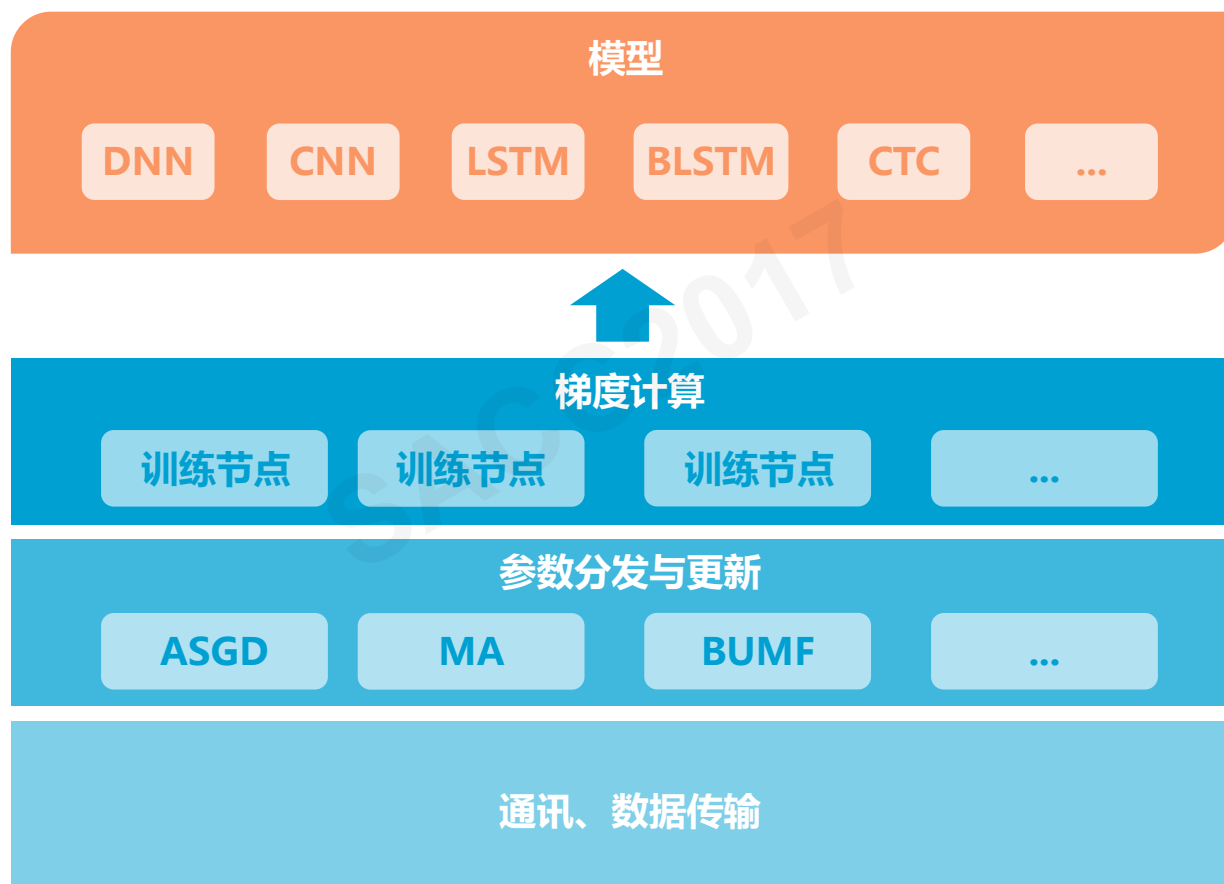


复杂的深度学习算法

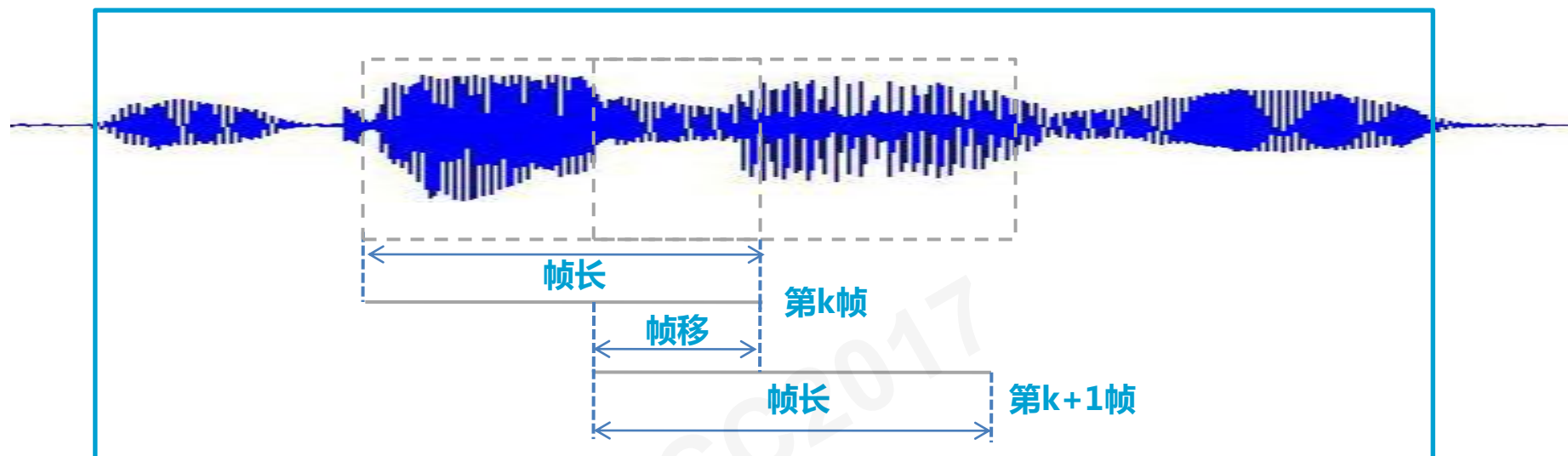


超强的运算平台

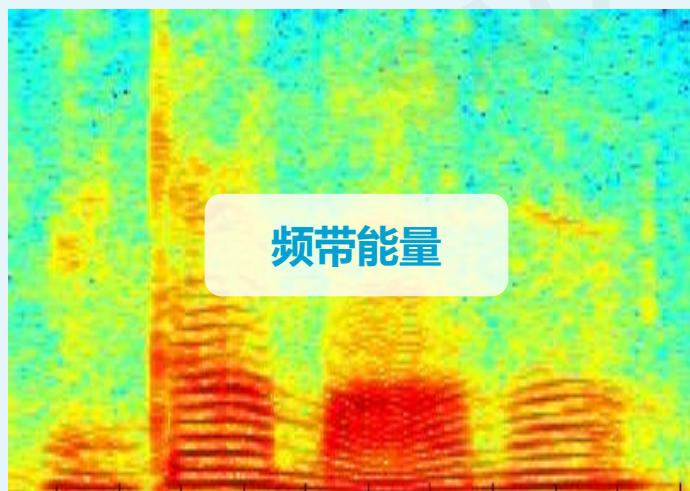
搜狗语音深度学习平台



语音是时变+短时平稳的信号



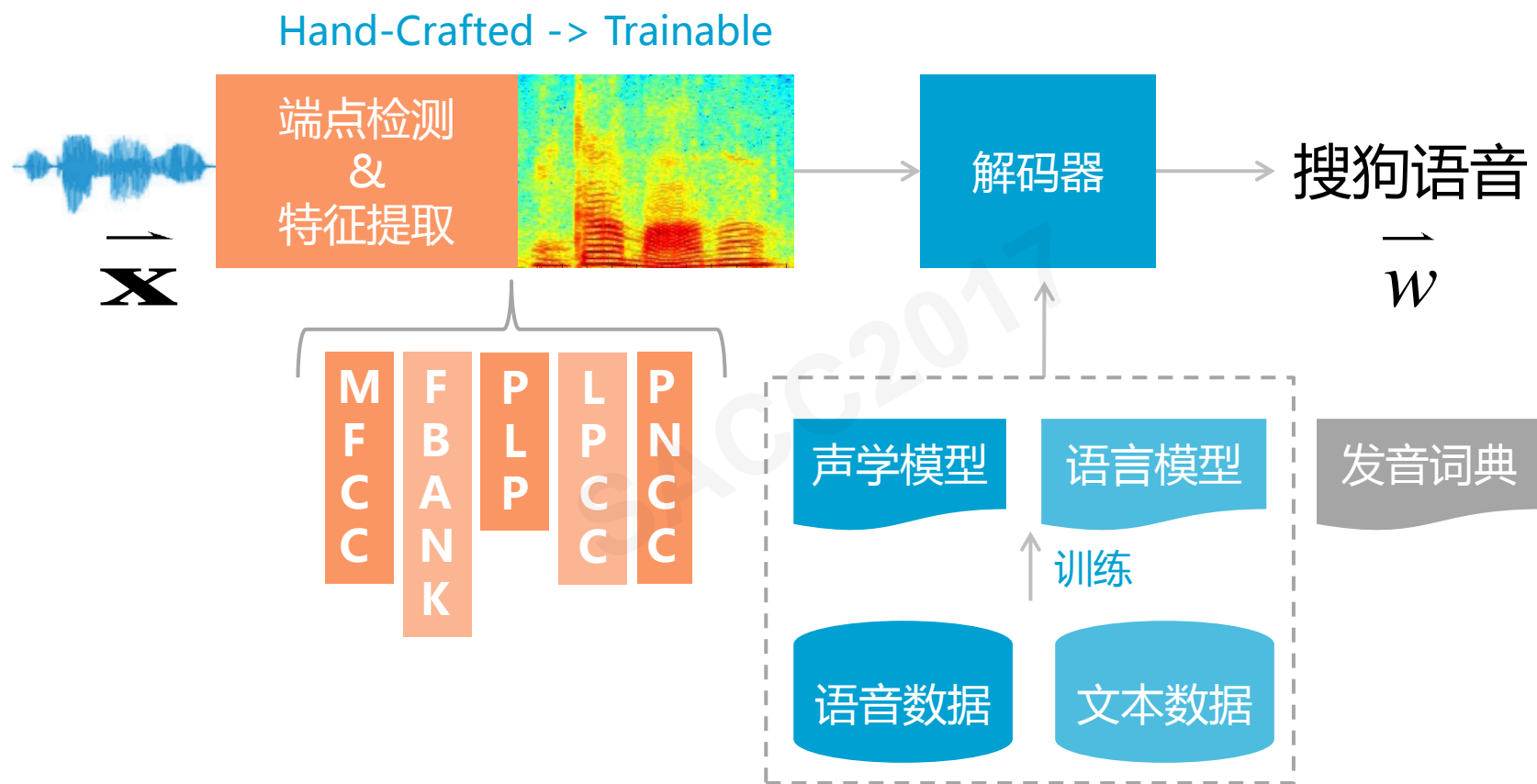
频率



时间

- 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

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

输出词序列

输入特征矢量

音素序列

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

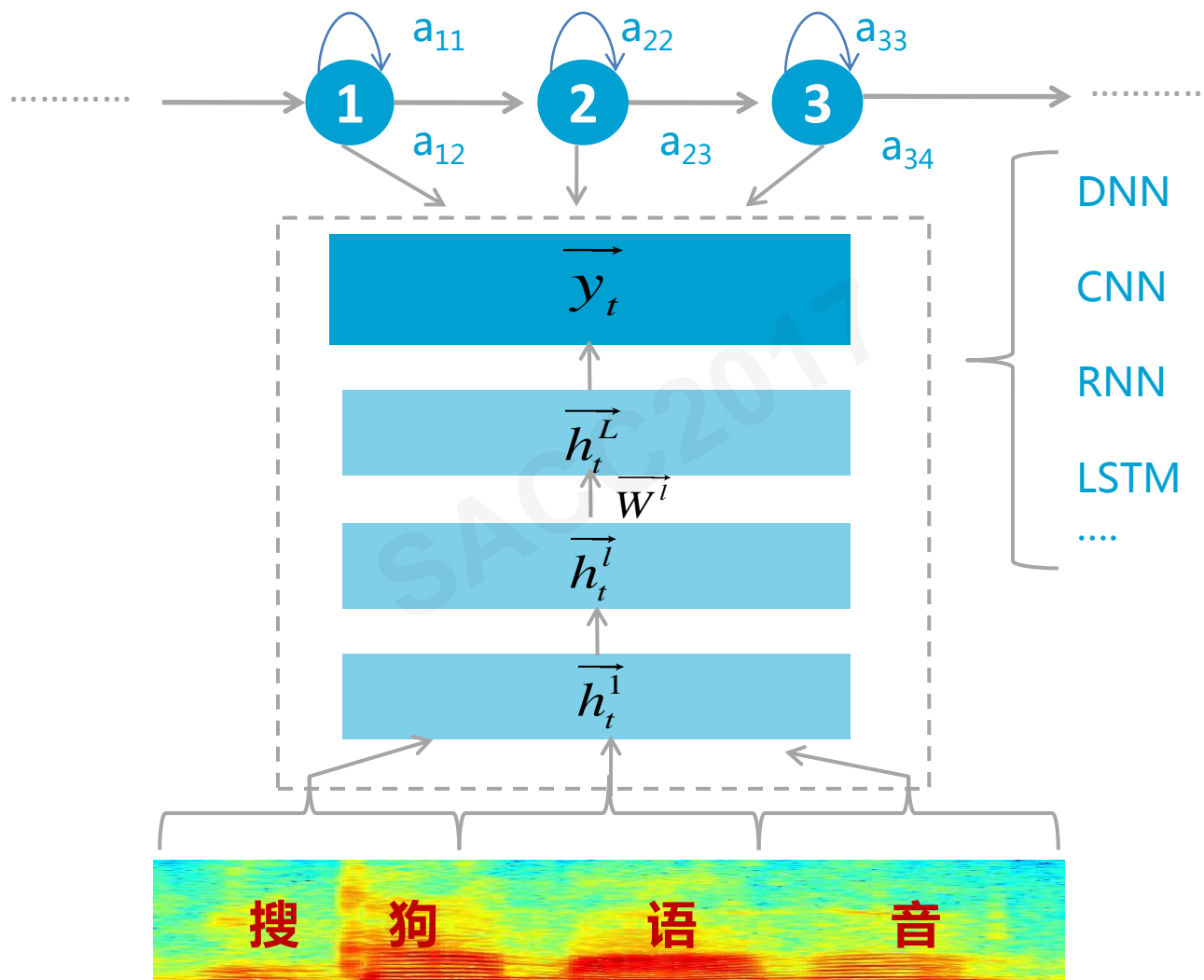
$$= \arg \max_{\vec{w}} \sum_{\vec{q}} p(\vec{x} | \vec{q}) P(\vec{q} | \vec{w}) P(\vec{w})$$

声学模型

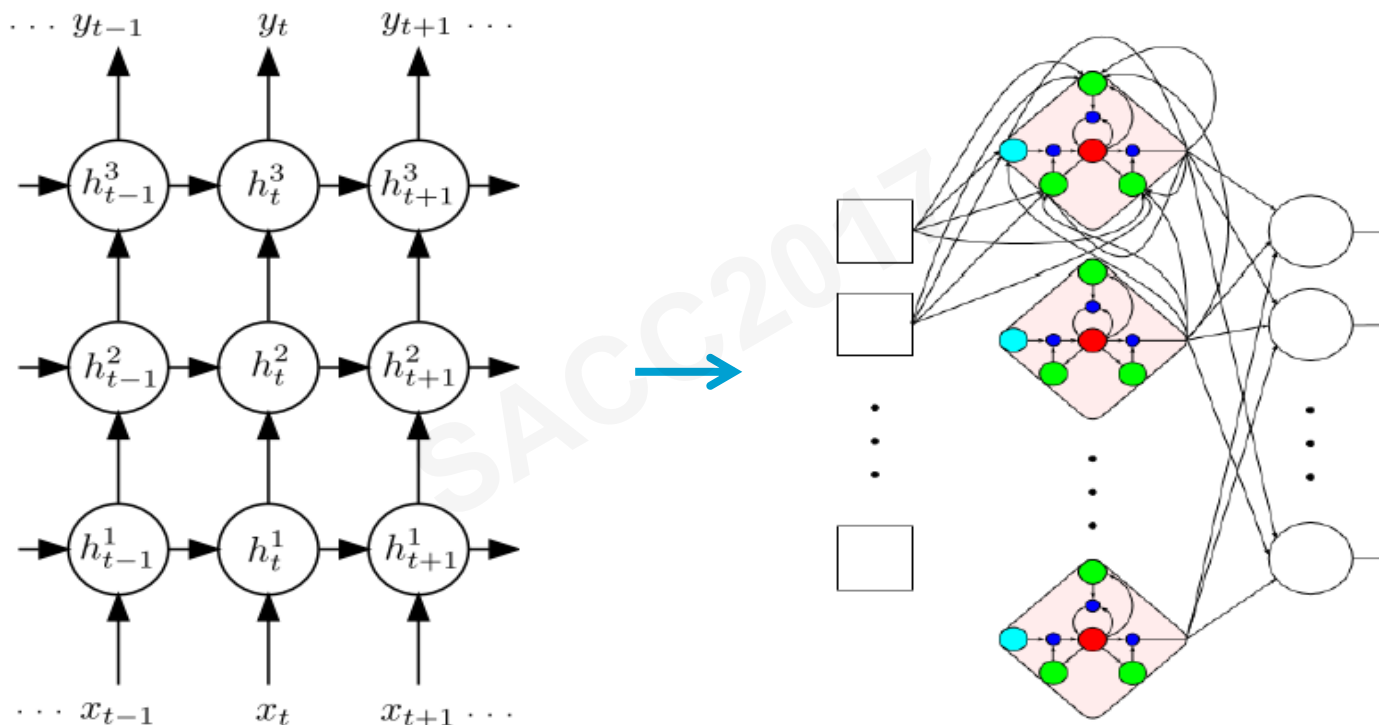
发音词典

语音模型

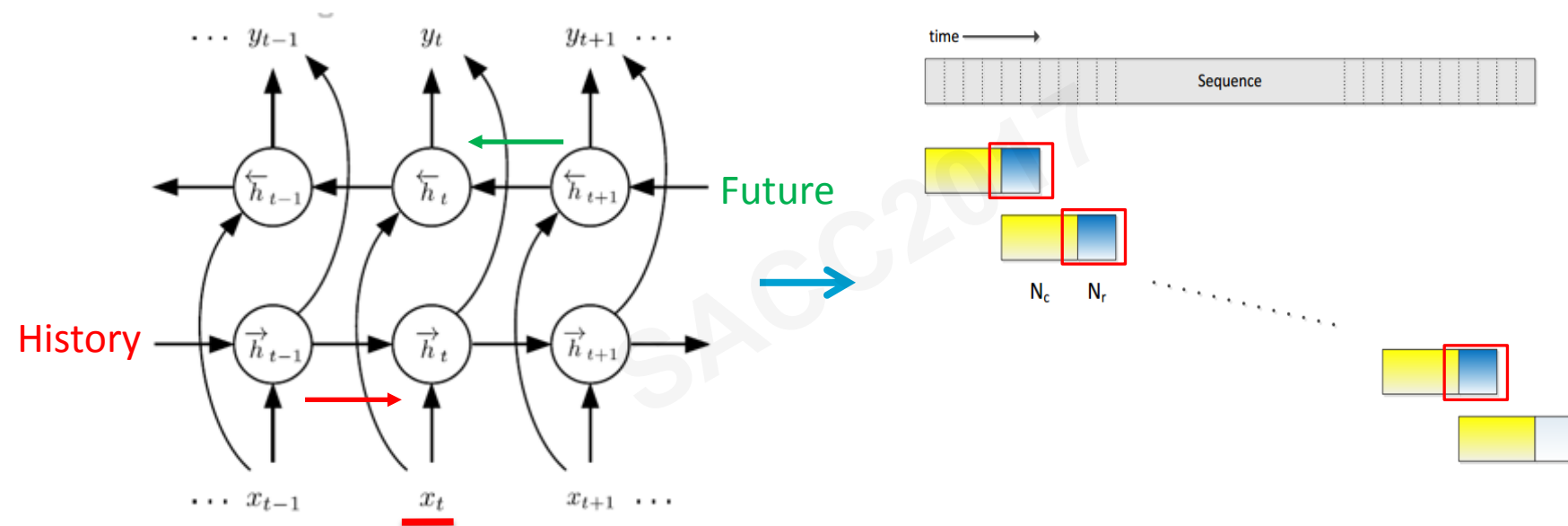
基于深度神经网络的声学建模



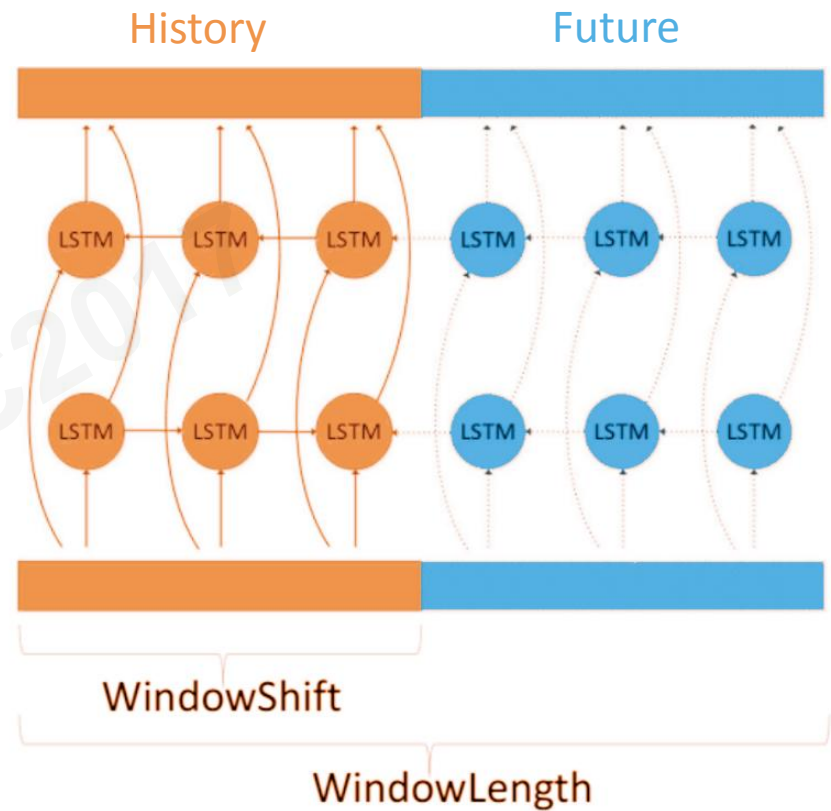
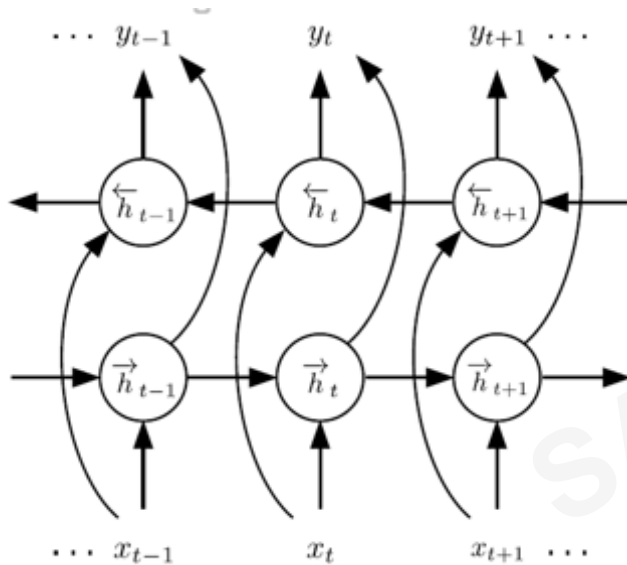
LSTM-RNN是声学模型建模的主要结构



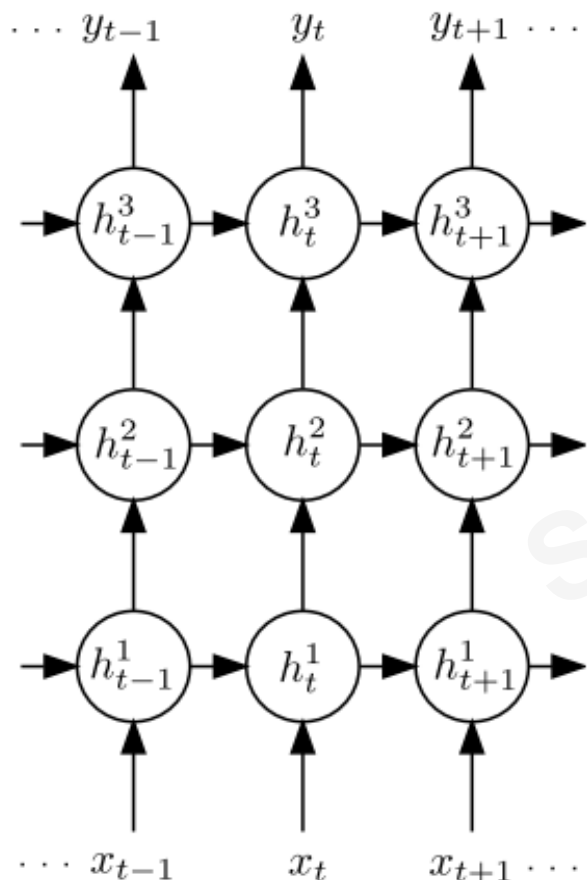
LSTM-RNN+双向特性



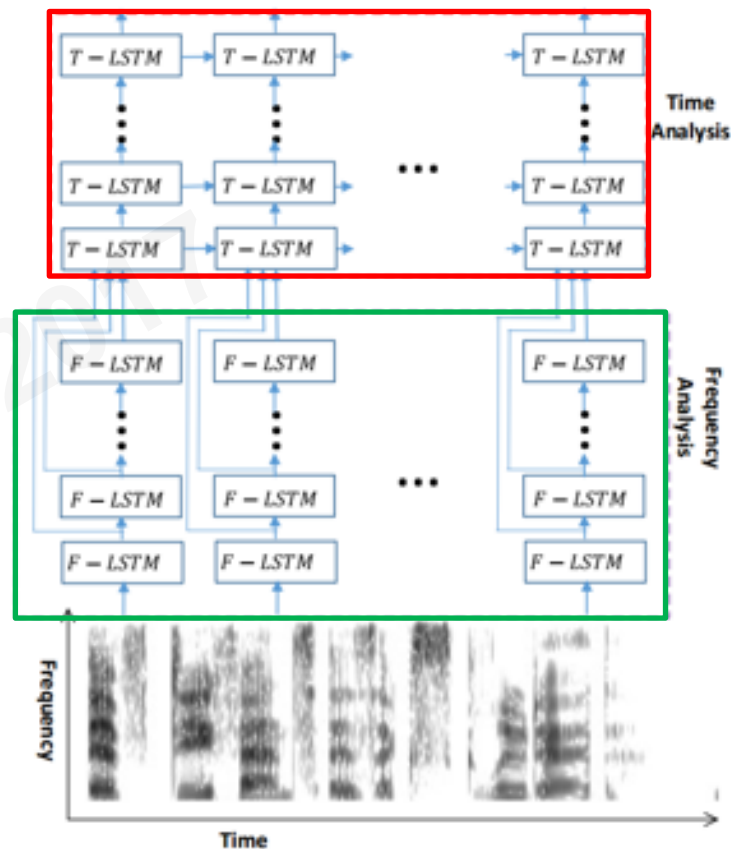
LSTM-RNN+双向特性



LSTM-RNN+频域扩展

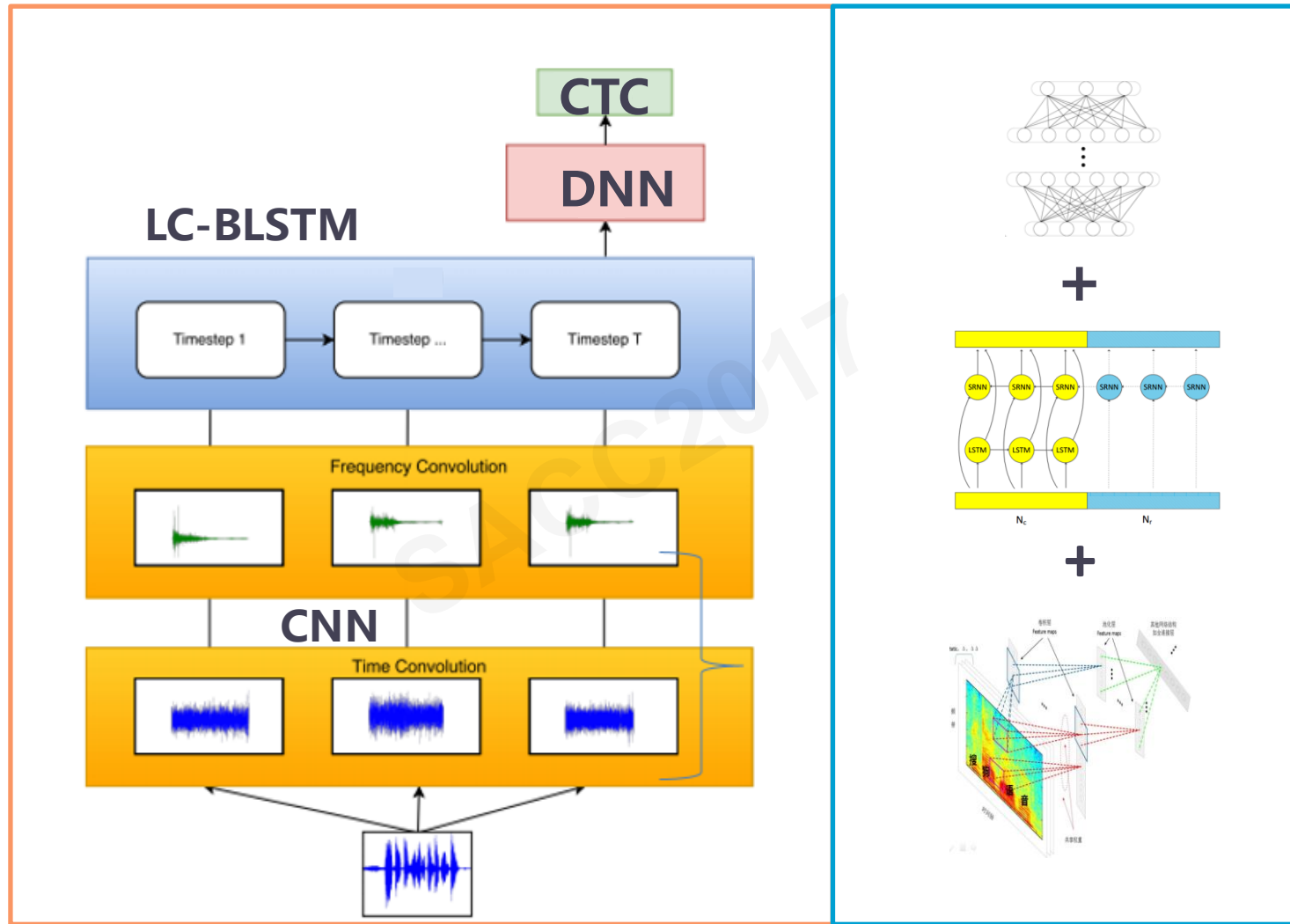


T-LSTM

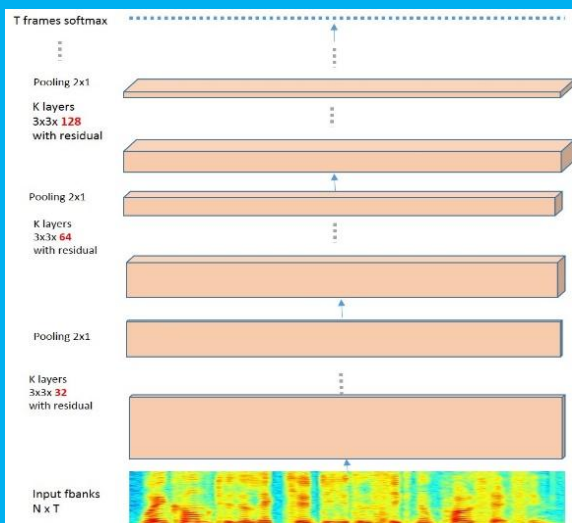


TF-LSTM

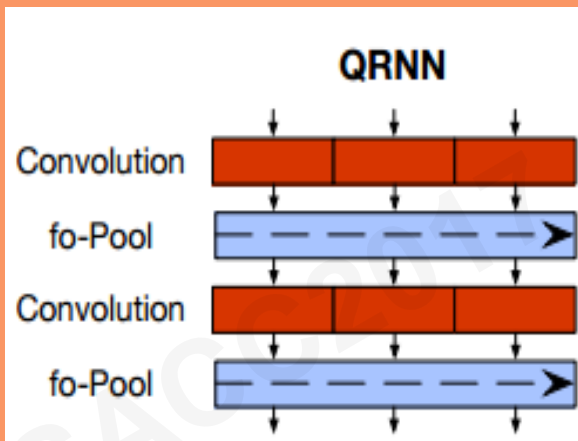
LSTM-RNN + 多模型融合



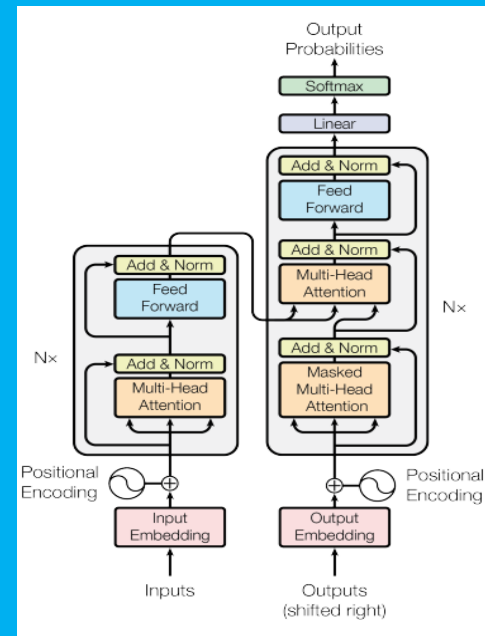
其他网络结构



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

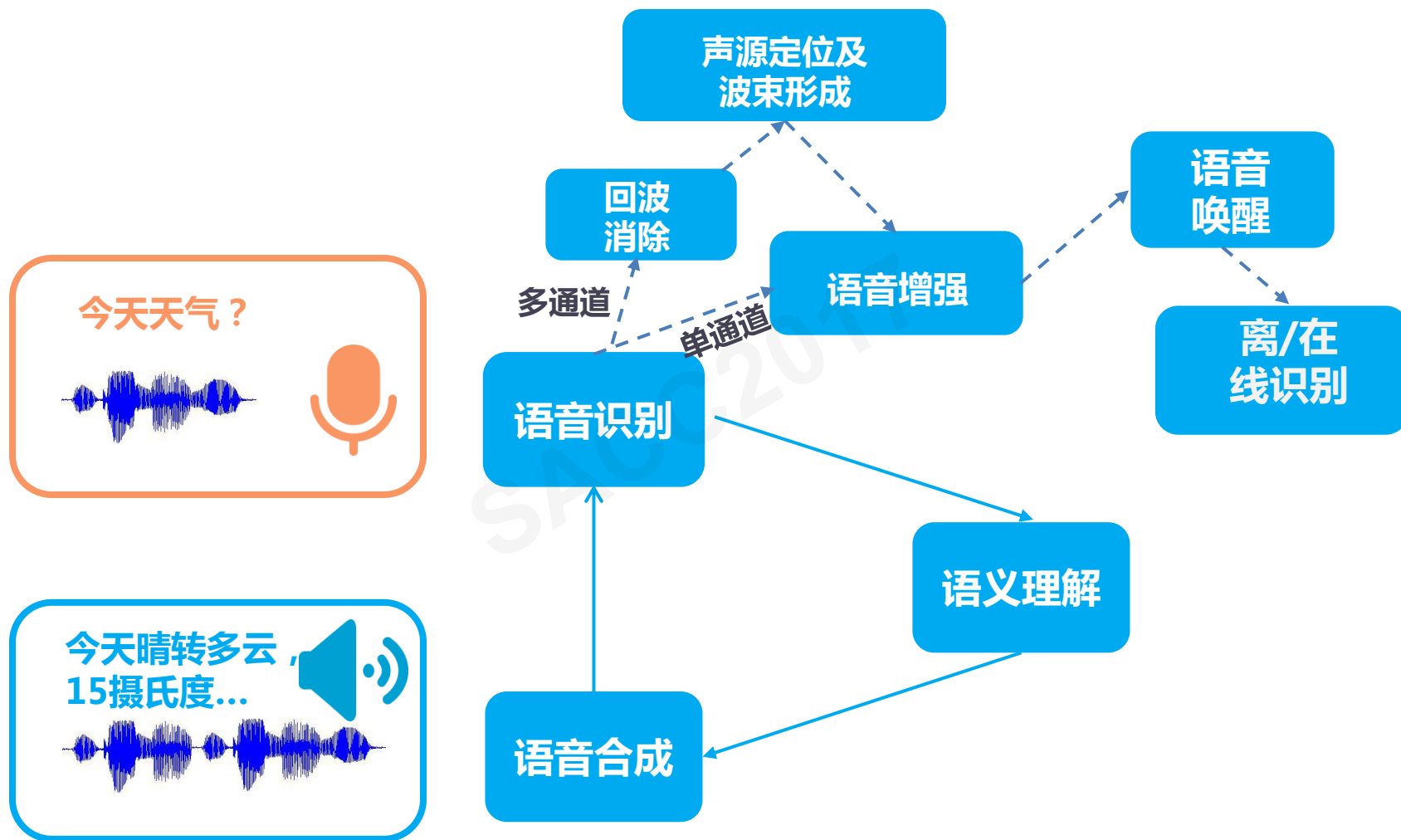
如此搞笑的效果只是口音产生吗？



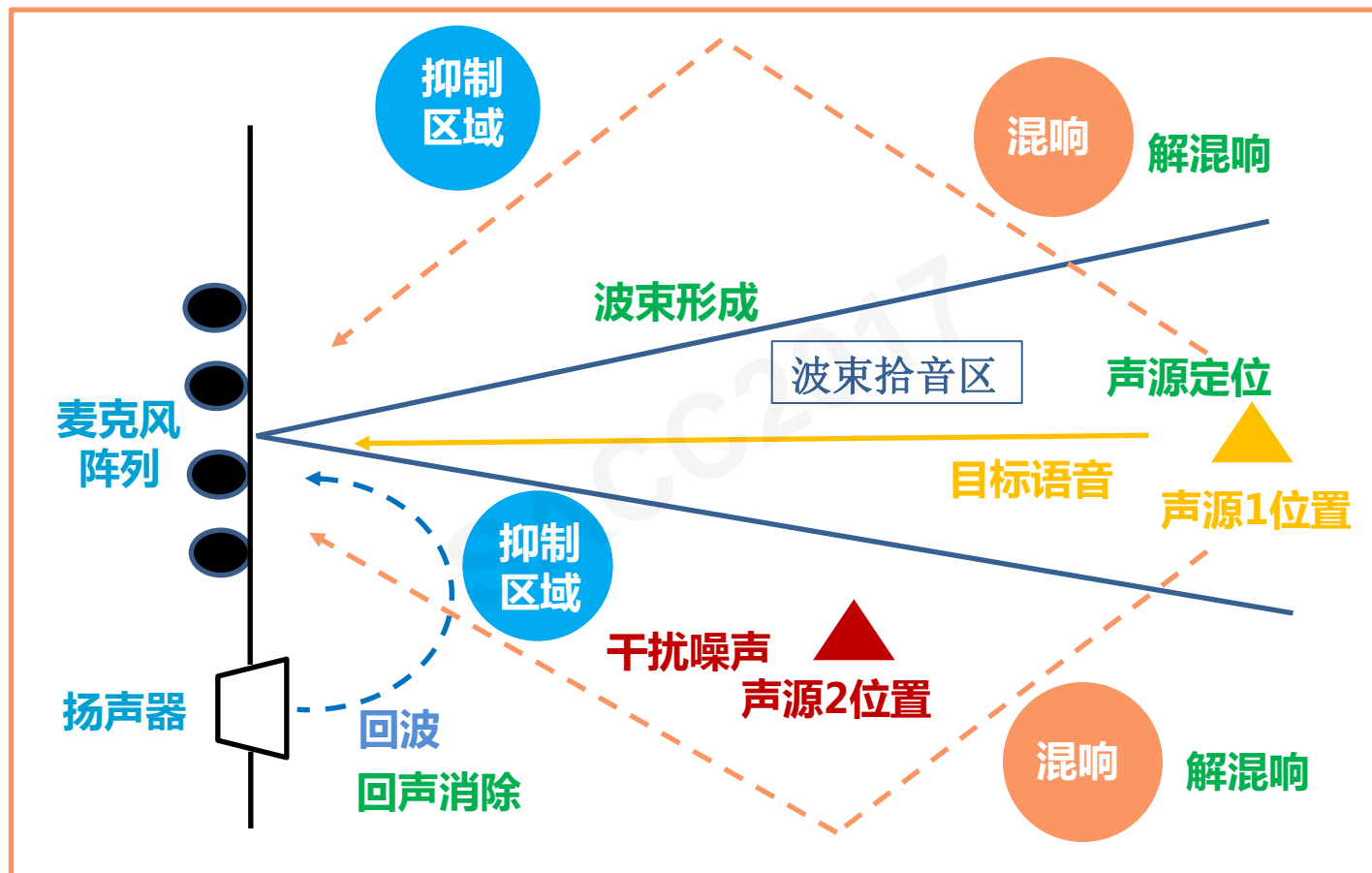
- 语音识别：口音、噪声、远场
- 语义理解：多轮对话、纠错容错
- 语音交互：全双工持续交互

单点能力 -> 系统整体能力

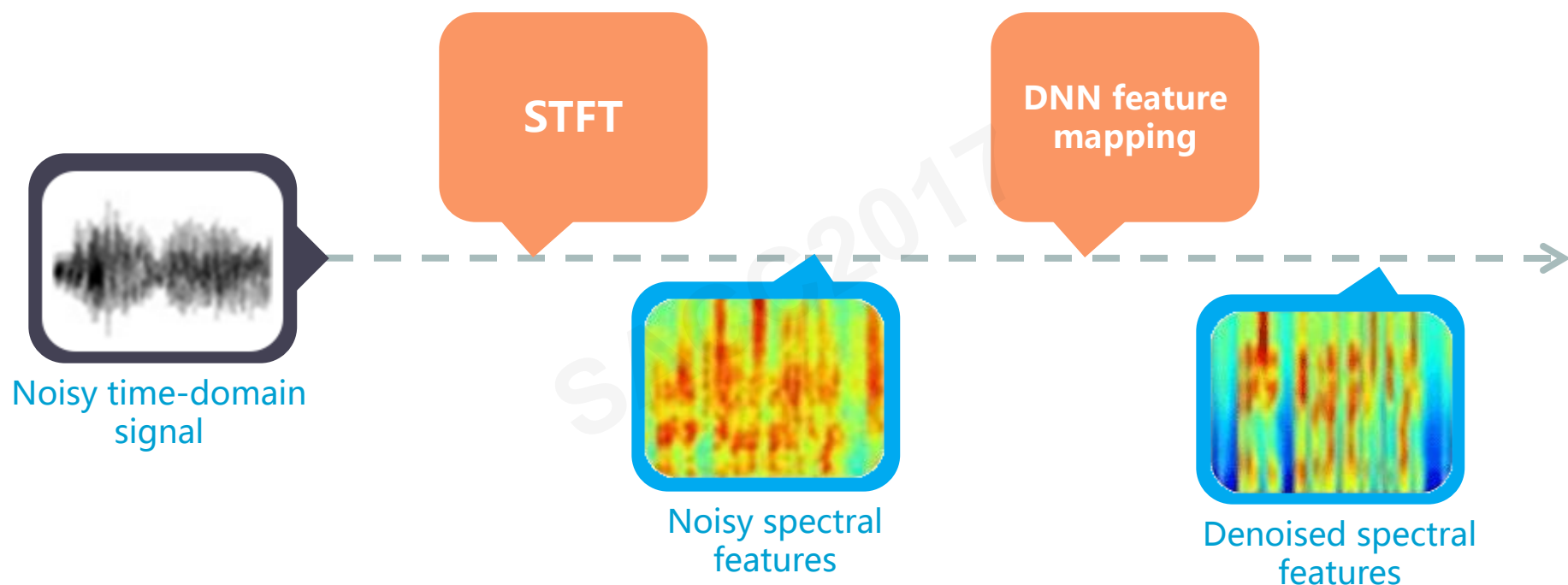
语音交互完整链路



麦克风阵列技术



基于深度学习的单通道语音增强





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

THANKS

The background features a dark, deep blue gradient. A bright, glowing light source is positioned behind the word 'THANKS', creating a lens flare effect. From this light source, several streams of glowing blue particles or dots emerge, trailing off towards the right and bottom corners of the frame. These trails have a wavy, organic quality, resembling smoke or data flow. The overall aesthetic is futuristic and digital.