

# 百度 远场ASR技术

语音技术部 李超

20170902

自我介绍

内容要展开了讲

百度的远场识别方案，方案有什么效果  
多方几遍？

举例子，（1）听出来音频，我们的识别结果  
（2）聪姐的视频

工程实例 （1）联想的  
（2）车载

远场的开放进度，暂时不支持，有计划

互联网时代



移动互联网时代



物联网时代

GMM-HMM

DNN-CTC

JOINT optimization

近场

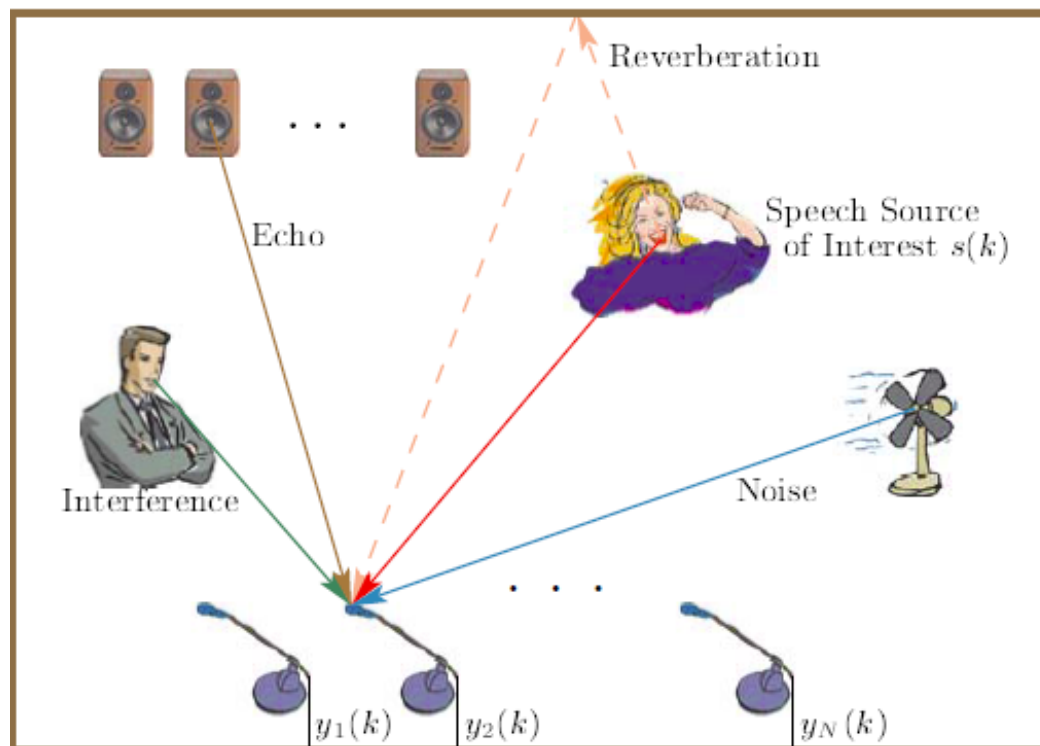
远场



噪声：由目标声源之外声源发出的声信号

回声：智能终端接收目标信号的同时，也发出声音所造成的干扰

混响：目标语音信号经过墙壁等再次或者多次发射信号



干扰语音信号

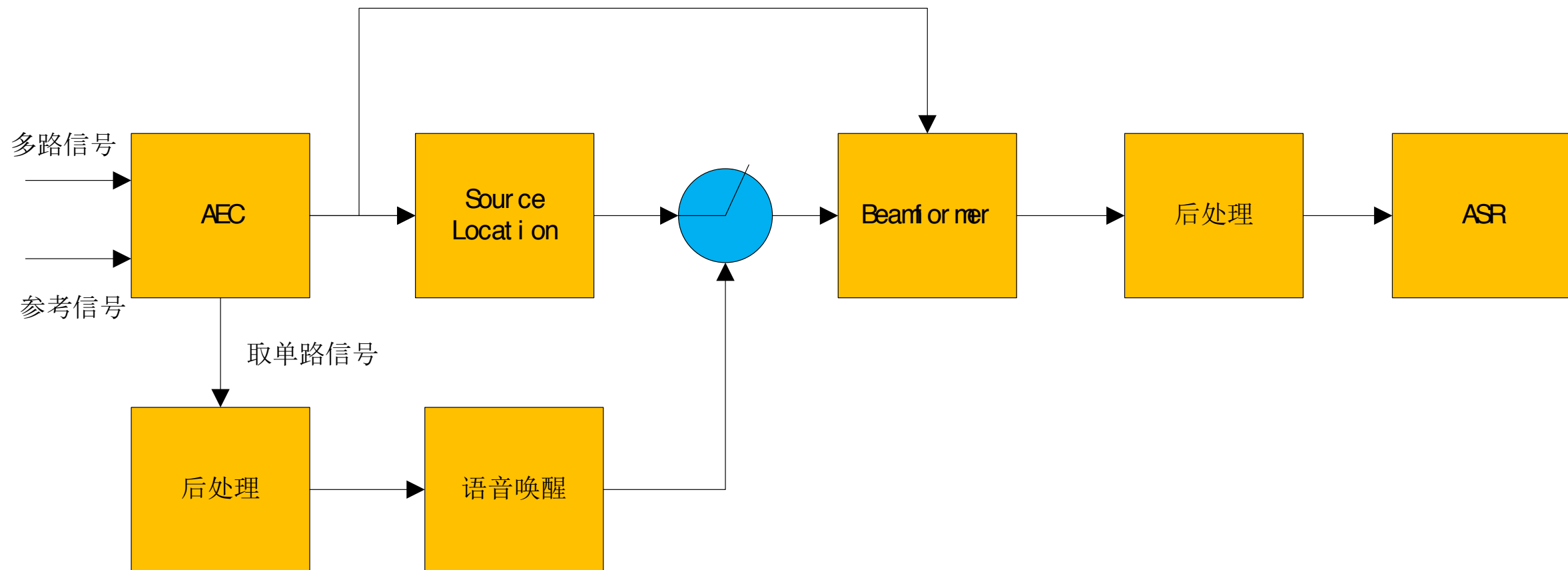
降低可懂度

语音交互

语音感知

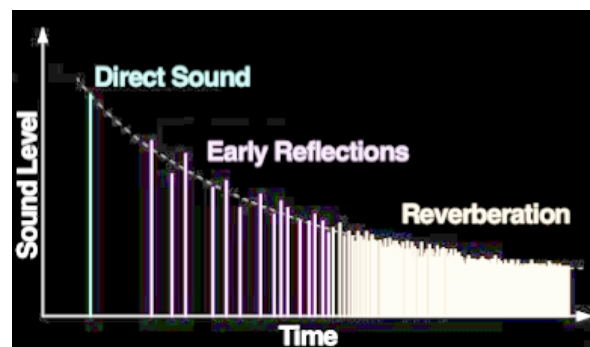
语音增强  
语音分离





Near-field  
speech

\*

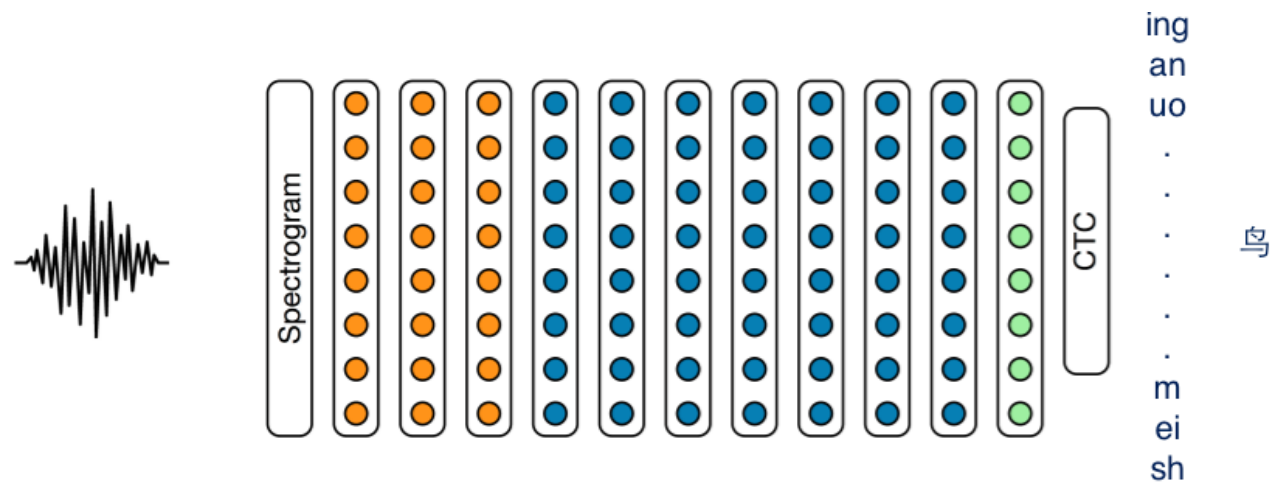


=

Augmented  
far-field speech

real far-field  
speech





- Convolution Layer
- Recurrent Layer
- Fully Connected Layer

## □ Transfer learning

1. 使用已经充分训练的进场模型做初始化
2. 几万小时的远场增强数据做训练
3. 加入一定比例的近场数据进行正则化





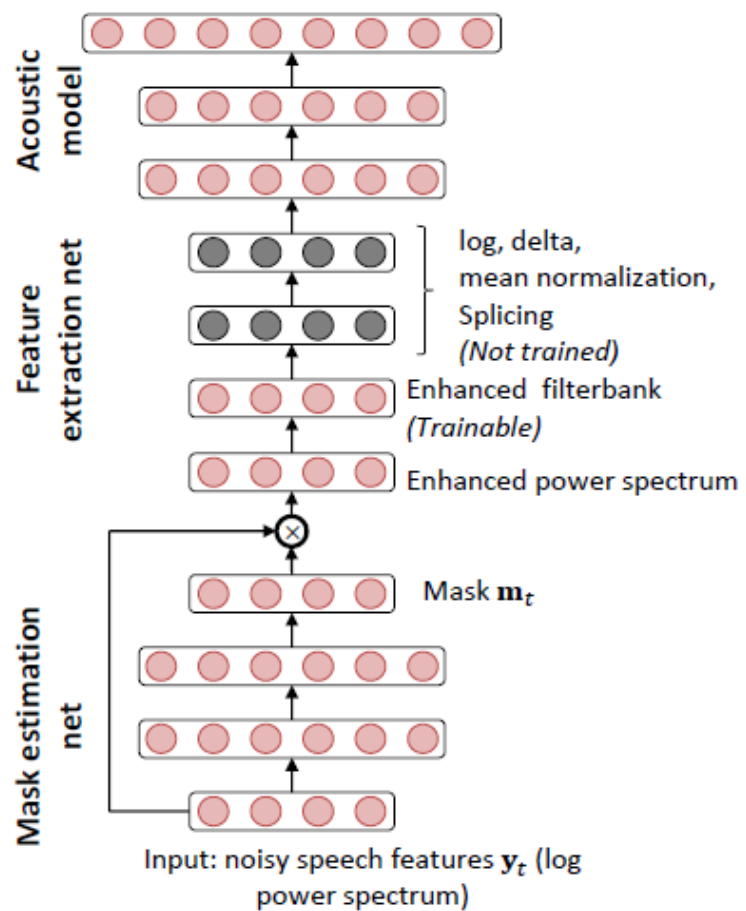
我要听90年代的音乐



帮我放1下有关娱乐的直播电台

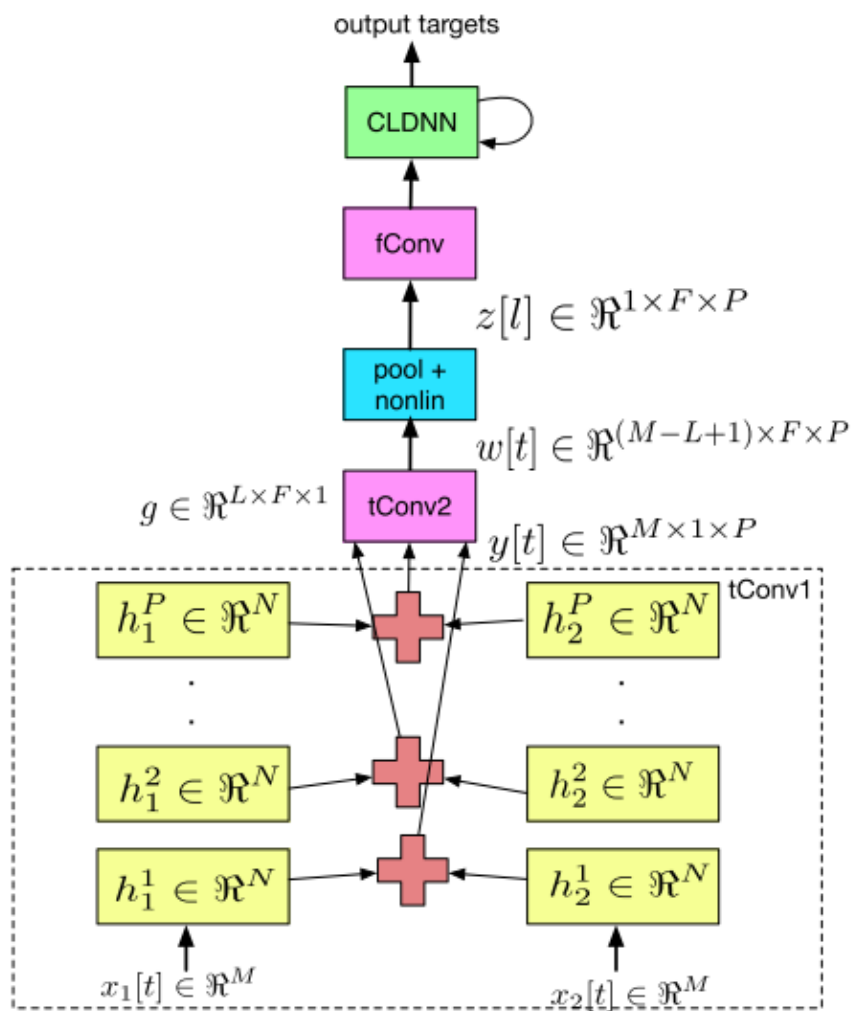


播放硅谷的新闻



## □ Mask & Joint Training

1. 前端信号处理网络和后端声学模型组成一个网络
2. 先训练前端，然后在一起训练



## □ end-to-end Fairfield acoustic model

1. Google home
2. 2Mic 媲美 6Mic
3. 更低的硬件成本，广泛的适用性

