



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

深度学习在AEC中的应用探索

大象声科高级算法工程师 闫永杰

info@elevoc.com

出品: liveVideoStack CSDN
—— 音视频技术社区 ——



深圳
2019

遨游“视”界 做你所想
Explore World, Do What You Want

LiveVideoStackCon 2019 深圳

2019.12.13-14



成为讲师: speaker@livevideostack.com

成为志愿者: volunteer@livevideostack.com

赞助、商务合作: kathy@livevideostack.com

出品:  LiveVideoStack CSDN
—— 音视频技术社区 ——



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 什么是AEC
2. 传统AEC的实现方式
3. 深度学习解决AEC问题
4. 思考&总结
5. QA



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

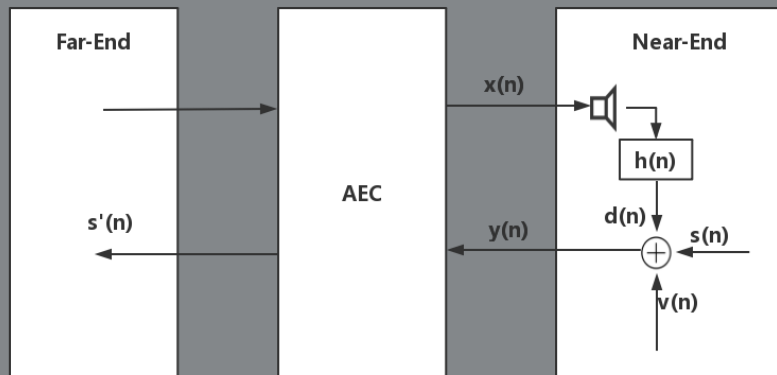
什么是AEC? AEC问题的引入

AEC问题定义



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



$$y(n) = d(n) + s(n) + v(n)$$

$y(n)$ – 麦克风信号

$d(n)$ – 回声

$s(n)$ – 目标语音

$v(n)$ – 背景噪声

传统自适应算法原理



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 在只有远端信号段估计参考信号到回声的传输路径
(回声路径) ;
2. 通过估计的回声路径和参考信号得到估计的回声;
3. 麦克风信号减去估计的回声分量, 得到AEC的结果;

传统自适应算法存在的问题



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 需要Double-Talk检测，漏检会导致滤波器发散；
2. 实际场景中自适应算法会受到背景噪声的干扰，
导致回声路径估计不准确；
3. 现实场景中播放设备的非线性问题。

从另一个视角看AEC



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

目标语音

回声

背景噪声



目标语音



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

我们把回声消除看成一个语音分离问题。

计算目标：理想二值掩膜(Ideal Binary Mask, IBM)

$$IBM(t, f) = \begin{cases} 1 & \text{if } s(t, f) - n(t, f) \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

- $s(t, f)$: Target energy in unit (t, f)
- $n(t, f)$: Noise energy
- θ : A local SNR criterion (LC) in dB, which is typically chosen to be 0 dB

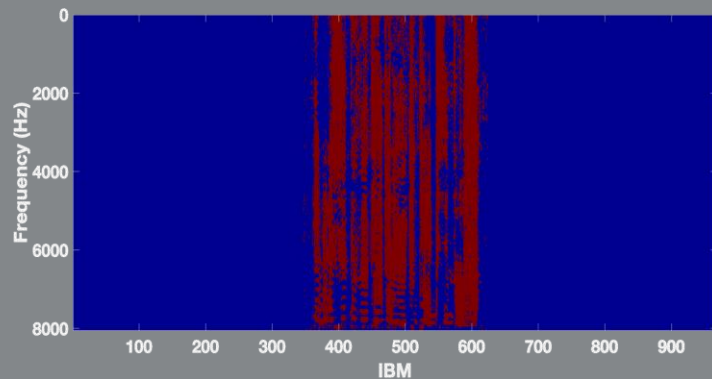
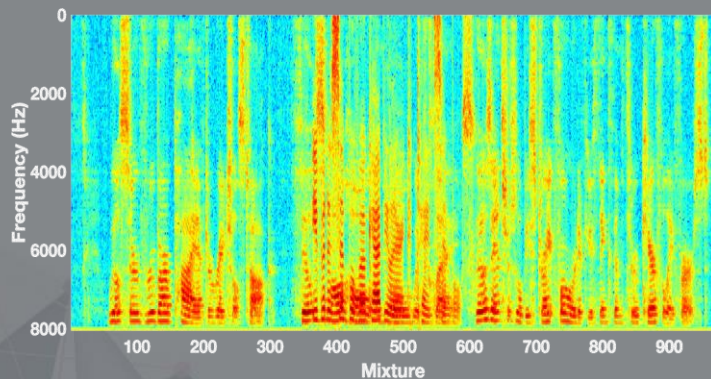
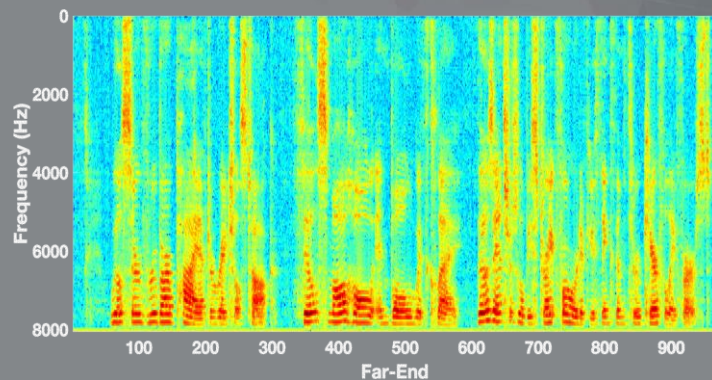
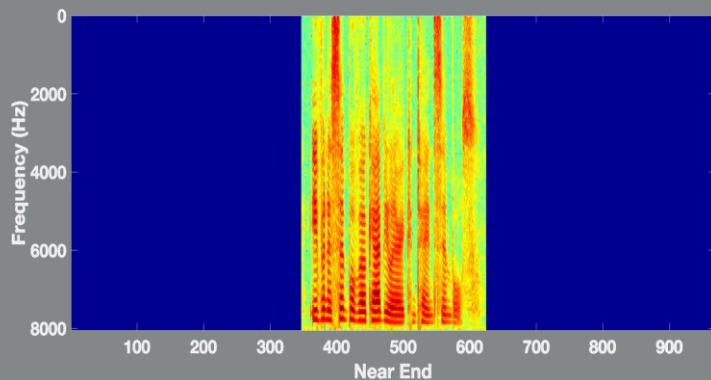
随着语音分离课题的不断发展，后续陆续出现了TBM，IRM等新的目标。

IBM



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



学习的本质

$$y = f(x)$$

如何构建一个模型，可以模拟任意的函数f

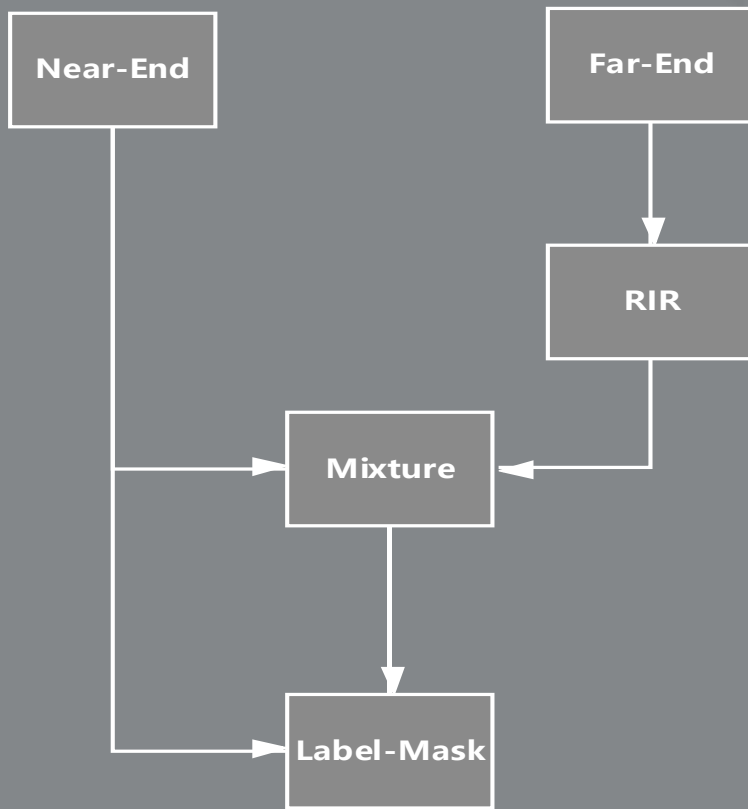
- 高斯混合模型
- 支持向量机
- 多层感知机
- 深度神经网络 (DNN)

训练数据的构建



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



深度学习解决AEC问题



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

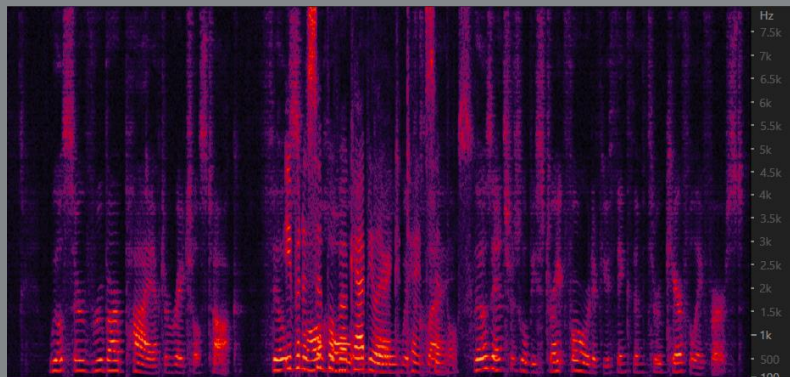
1. 目标-IBM
2. 特征-混合语音及参考信号 STFT后幅度谱
3. 工具-Tensorflow、Pytorch...
4. 数据驱动

实验结果示例

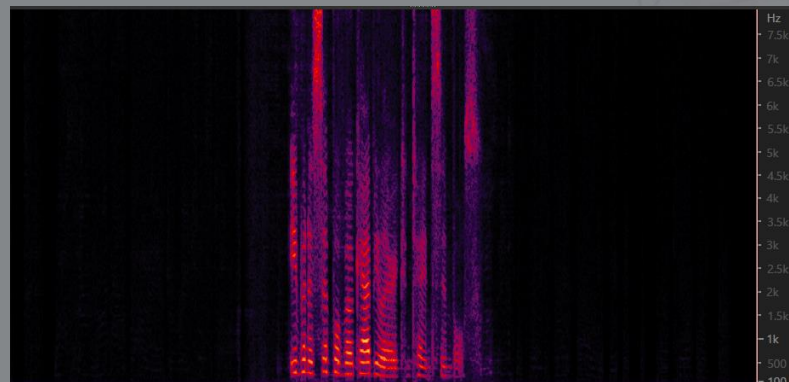


北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



处理前



处理后





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

实验场景和现实场景存在巨大差距

实验与真实场景的距离



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

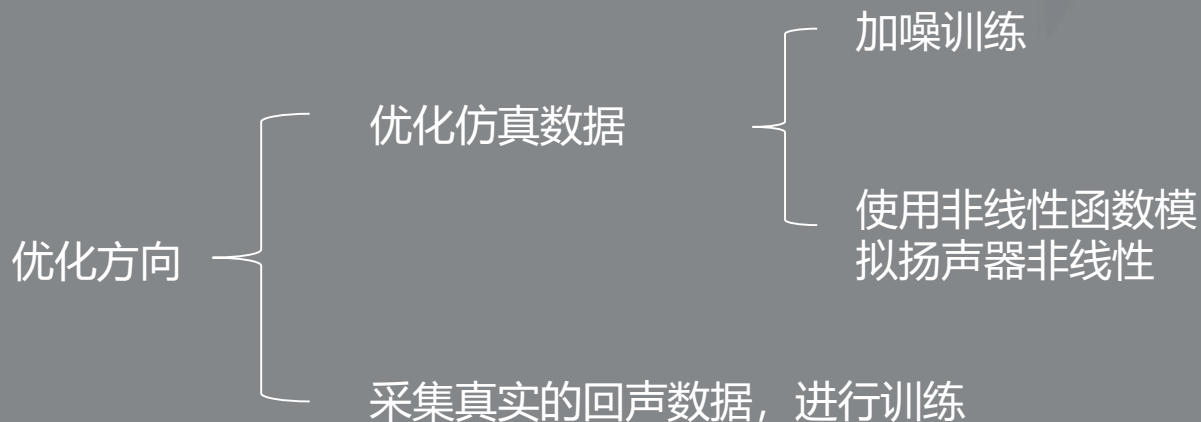
1. 现实场景需要考虑背景噪音的干扰;
2. 非线性带来的不匹配;
3. 现实中的RIR和实验生成RIR的不匹配;

解决实验数据与真实数据间的差距



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want





北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

这种方式解决AEC问题的优势和局限

相比传统算法的优势



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 无需考虑Double-Talk;
2. DL具有强大非线性拟合能力, 能够更好的覆盖AEC系统中的非线性因素;
3. 无需复杂的调参过程;

存在的局限性



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 仅恢复了目标语音的幅度，未恢复目标语音的相位；
2. 在低SER条件下，Masking方式恢复幅度能力有限；
3. 训练数据和真实场景数据的分布差异，会导致模型能力下降；

一些探索的方向



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

1. 同时预测目标语音的幅度谱和相位谱;
2. 直接在时域信号上进行预测;



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

我们坚信深度学习会对语音信号处理带来更好的效果

深度学习的优势



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

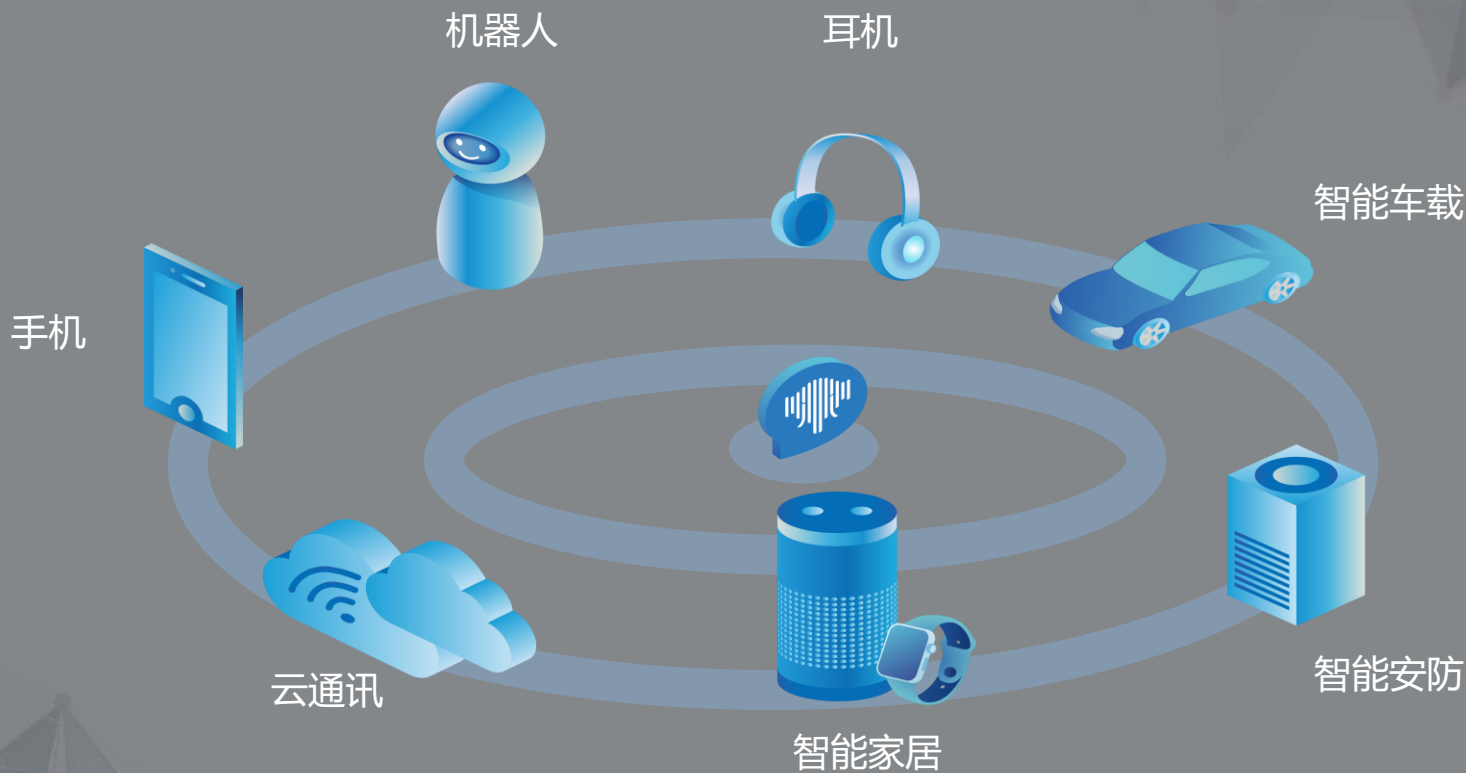
1. 数据驱动，一定条件下，数据越多性能越好；
2. 相比传统算法手工统计的模式，DL可学习到更加鲁棒的模式；
3. 深度学习具有记忆的能力；

应用场景



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



部分应用案例



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



OPPO Reno



vivo Z5x



中兴Axon10Pro



努比亚红魔3



黑鲨游戏手机2



美图 V7



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want

Thank you



出品: LiveVideoStack CSDN
—— 音视频技术社区 ——