Report

Understanding of Overlapped Speech Research

**Zhaohui Yin**

**Tue, Nov 17, 2020**

1. 从CHIME6竞赛任务来看技术问题

参考资料：

1. https://towardsdatascience.com/the-chime-6-challenge-review-15f3cbf0062d （CHIME6网站）

2. https://zhuanlan.zhihu.com/p/138800436（CHIME6介绍）

3. https://chimechallenge.github.io/chime6/，数据音频下载

1.1 竞赛任务（实际应用场景）

Track1：给定说话人边界的情况下重点考察参赛团队的多通道信号处理能力及复杂场景语音识别的能力。

Track2：在自动说话人分离的基础上再进行语音识别。

Ranking A: 基于常规声学模型和官方语言模型。

Ranking B：不限制模型。

1.2 任务难点（实际应用场景）

1. 大量的语言交叠；【重叠语音检测，重叠语音区分，重叠语音识别】

2. 远场混响和噪声干扰对录音的影响；【语音增强，降噪】

3. 对话风格非常自由、近乎随意；

4. 训练数据有限。【数据增强】

1.3 涉及技术及样例学习

语音增强（降噪），语音增强（多类数据），重叠音检测（When），重叠音区分（Who），重叠音识别（What）。

2015年的GSS方法因为效果良好，在本次比赛中被广泛使用。

|  |  |
| --- | --- |
|  |  |

以**USTC-NELSLIP**为例来初步理解从前端到后端的一系列流程。（看起来很吃力）

✅ 建模：

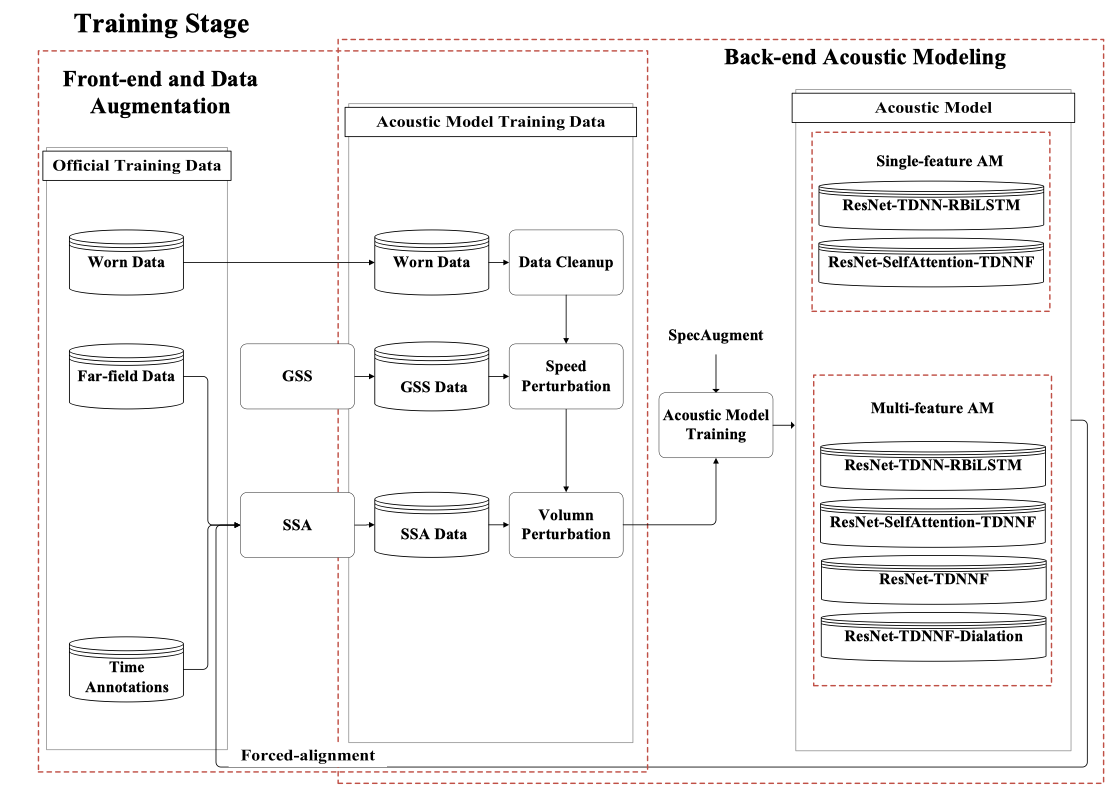


Figure1:技术框架

=> 概述：

前端先基于加权预测误差（WPE）的对数频谱振幅应用常规的多通道降噪，紧接着基于深度学习技术建立SSA模型，获得不同类型的训练数据。后端通过融合五种声学模型来进行训练。

解码语句则被分为四个阶段：1）波束成形初始化；2）基于SSA的信号统计估计；3）波束成形；4）识别

=> 前提：

官方允许利用有关音轨开始和结束时间，扬声器说话者标签和扬声器位置标签。允许使用双耳数据（binaural data）和远场数据（far-field data）。

=> 前端（Front-end）：

提出基于空间-说话人同步感知的迭代掩码估计算法SSA-IME（space-and-speaker-aware iterative mask estimation），结合传统信号处理和深度学习的优点，利用空时多维信息进行建模，迭代的从多个说话人场景中精确捕捉说话人的信息。该算法可以有效的降低环境干扰噪声，而且可以有效消除干扰说话人的语音，从而大幅降低语音识别的处理难度。

=> 后端（Back-end）：

提出了基于空间-说话人同步感知的声学模型(Spatial-and-Speaker-Aware Acoustic Model, SSA-AM)，通过在声学模型输入端拼接多维度空间信息和不同说话人信息，使其能自适应区分目标说话人和干扰说话人。因此，声学模型不仅依赖前端算法的处理结果，也能够自适应完成对目标说话人语音特征的提取，大幅提升多人对话场景下语音识别声学模型的容错率和鲁棒性。

✅ 评价结果：

1.4 竞赛论文学习

|  |  |
| --- | --- |
|  | <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_du.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_chen.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_medennikov.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_zorila.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_tang.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_yang.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_arora.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_ren.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_zmolikova.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_lee.pdf>  <http://chimechallenge.github.io/chime2020-workshop/abstracts/CHiME_2020_abstract_sreeram.pdf>  <http://chimechallenge.github.io/chime2020-workshop/papers/CHiME_2020_paper_watanabe.pdf> |
|  |
|  |
|  |

2. 从INTERSPEECH 2020论文集来了解各方向针对问题和研究进展

* **Scientific Area Topics**  
  1. Speech Perception, Production and Acquisition  
  2. Phonetics, Phonology and Prosody  
  3. Analysis of Paralinguistics in Speech and Language  
  4. Speaker and Language Identification  
  5. Analysis of Speech and Audio Signals  
  6. Speech Coding and Enhancement  
  7. Speech Synthesis and Spoken Language Generation  
  8. Speech Recognition – Signal Processing, Acoustic Modeling Robustness and Adaptation  
  9. Speech Recognition – Architecture Search, and Linguistic Components  
  10. Speech Recognition – Technologies and Systems for New Applications  
  11. Spoken Language Processing – Dialog, Summarization and Understanding  
  12. Spoken Language Processing – Translation, Information Retrieval and Resources