**Acoustic & speech & signal processing顶会顶刊**

顶刊

APPLIED ACOUSTICS

Signal Processing Letters

IEEE Transactions on Audio Speech and Language Processing

Signal Processing

顶会

ICASSP

INTERSPEECH

WASPAA

IWNEAC

**Database**

Free internet databases

* [**Freesound**](https://freesound.org/browse/): crowdsourcing database

([ref-Freesound: a Platform for the Creation of Open Audio Datasets](database%20research/Freesound_ISMIR2017.pdf), ISMIR2017, citation 73)

优：数据量大；是很多dataset的母数据库

缺点：众包导致音频质量参差不齐；

* [**SoundBible**](http://soundbible.com/)
* [**UrbanSound & UrbanSound8K**](https://urbansounddataset.weebly.com/urbansound8k.html)**: urban sound taxonomy, with baseline**

(ref-[A Dataset and Taxonomy for Urban Sound Research](database%20research/Sound8K_ACMMM2014.pdf), ACM MM2014)

1. 4 top level groups: human, nature, **mechanical**, music
2. 10 low level classes: … gun shot …
3. Based on Freesound
4. UrbanSound8K: a subset of short audios (4s is enough for source identification)
5. Baseline对于纯gunshot的正确率在90%左右（MFCC+SVM）

优：有专门gunshot一类

缺点：基于Freesound

By google

* [**AudioSet**](https://research.google.com/audioset/dataset/gunshot_gunfire.html) **with baseline system**

([ref-AudioSet: an Ontology and Human-labeled Dataset for Audio Events, ICASSP2017, citation 599)](database%20research/Audioset_ICASSP2017.pdf)

1. video + audio

优：数据量最大

缺点：基于视频，需要视音频分离；AudioSet给出的只是Youtube ID；

Challenge

* [**DCASE2016**](http://www.cs.tut.fi/sgn/arg/dcase2016/) **with baseline system (MatLab implementation) & ranking**

1. Acoustic scene classification (TUT Acoustic **Scenes**2016)
2. Sound event detection in synthetic audio (Task2)
3. Sound event detection in real life audio (TUT Sound **Events**2016)

([ref-TUT Database for Acoustic Scene Classification and Sound Event Detection](database%20research/TUT_EUSIPCO2016.pdf))

1. Domestic audio tagging: home activity (CHiME-Home)

优：拾取工具较统一

缺点：综合性

* [**TRECVID**](https://www-nlpir.nist.gov/projects/trecvid/)**: multimedia event detection**

1. video + audio

缺点：需要音视频分离

* [**ESC**](https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/YDEPUT)**: environmental sound**

([ref-ESC: Dataset for Environmental Sound Classification](database%20research/esc_ACM%20MM2015.pdf))

1. Constructed from Freesound sound clips

缺点：基于Freesound

**Related works**

**Urban Sound Event Classification-local & global feature, Jo. Applied Acoustics2017**

[**(ref-Urban Sound Event Classification Based on Local and Global Features Aggregation)**](file:///E:\Git-Repository\Acoustic-Detection\跑分\1-s2.0-S0003682X16302274-main.pdf)

**Urban sound event -跑分**

**UrbanSound baseline**

MFCC + SVM\_rbf

70%左右，针对gunshot有90%以上

**Urban Sound Event Classification-local & global feature, Jo. Applied Acoustics2017**

Local & global feature -> Mixture of Expert (feature fusion scheme) -> (f\_local,f\_global)

mean Average Precision(mAP) = 77.36%

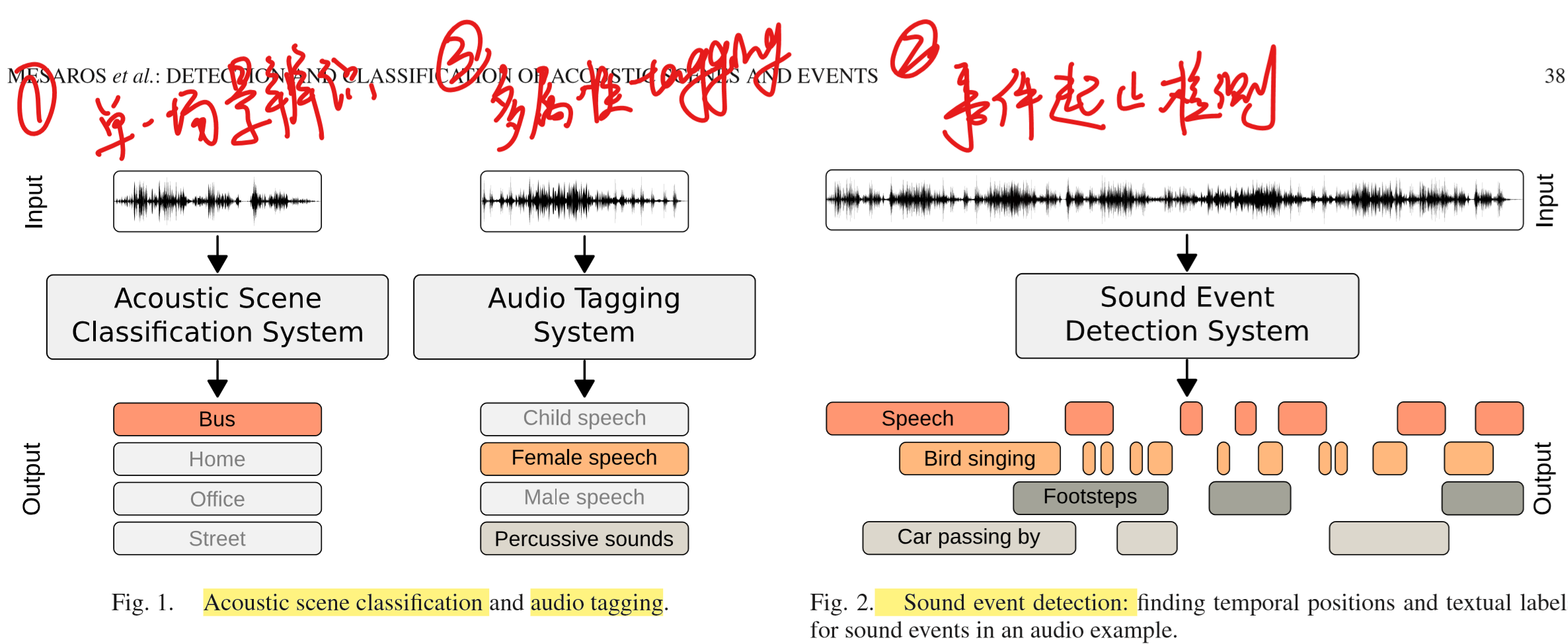
**Event detection: state-of-the-art & pipelines**

**Motivation**

1. **与现有的声学事件检测pipeline相比（需要调研），我们针对枪声，长处在哪里，因为general acoustic event detection也可以检测枪声非枪声？**
2. **特别对对单纯gunshot的分辨能力很强，针对gunshot已经正确率达到90%以上;**

**可以针对emergency事件做文章，系统针对emergency事件进行检测，单纯的枪击检测过于侠义**[**(ref-Urban Sound Event Classification Based on Local and Global Features Aggregation)**](跑分/1-s2.0-S0003682X16302274-main.pdf) **-》 做acoustic scene classification(emergency)**

**VS**

**sound events detection(gunshot)** 

**方向1：gunshot detection – sound events detection**

**方向2：alert event detection(emergency) – acoustic scene classification**

**方向4：枪种类分析（数据集较难获取）**

**方向5：并发计数（overlapping）**

1. **做好data augmentation**
2. **选题关键：1.做的人少；2.可提升空间大**
3. **鸟声识别（DCASE2018，知网有部分研究，已有专利）**
4. **稀有声学事件识别（枪击）**
5. **抗欺骗系统**
6. **数据增强+枪声分类+枪声合成+回响消除+考虑实际传声器可能削峰**

Robert C Maher

**论文略读**

**数据库 & baseline**

**UrbanSound原始论文**

**A Dataset and Taxonomy for Urban Sound Research, ACM MM2014**

Adobe，NYU-MARL

**Justin Salamon(4k+ citation，SONYC成员，UrbanSound成员)**

Christopher Jacoby

**Juan P Bello(7k+ citation，SONYC成员，UrbanSound成员)**

* **关键点：**

1. **4个top level类别，10个low level类别**
2. **在数据集上跑了baseline**
3. **最大的郊外环境声urban sound数据库**
4. **基于Freesound**

* Intro

1. **4个top level类别：**
2. **human**
3. **nature**
4. **mechanical**
5. **music**

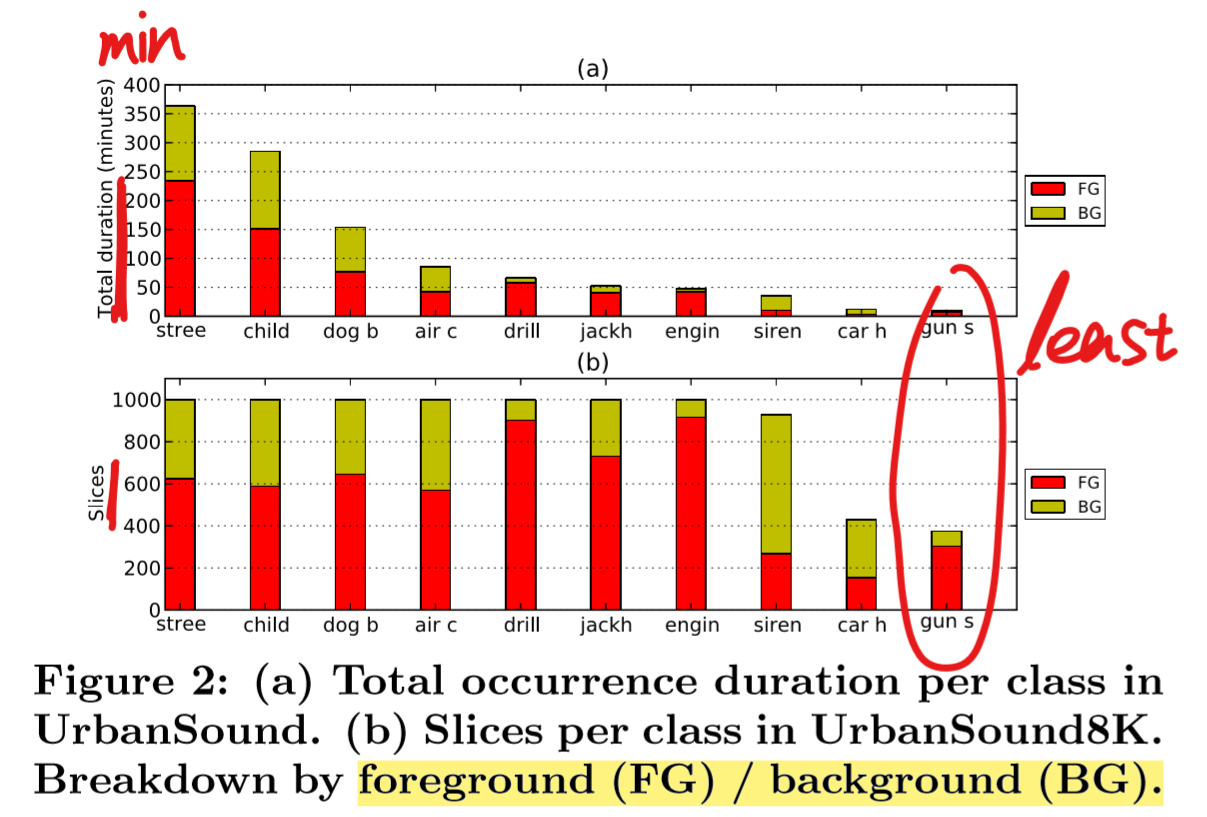
* dataset

1. 10个low level类别：

air conditioner, children playing, drilling, engine idling, street music

gunshot, dog bark, car horn, jackhammer, siren

1. 使用Freesound构建数据库，使用Freesound API可以搜索并下载仓库的子集，并发掘使用上传者所提供的metadata（名字、描述、标签）
2. 对于每个类，在Freesound中下载所有按类别名称搜索返回的音频
3. **对所有音频进行了人工核查，只收录真实现场录制的**
4. **UrbanSound：使用Audacity对事件起止点进行标注**
5. **UrbanSound8K：由于4s音频能达到82%正确率的声源识别（ref），将音频限制最长为4s**

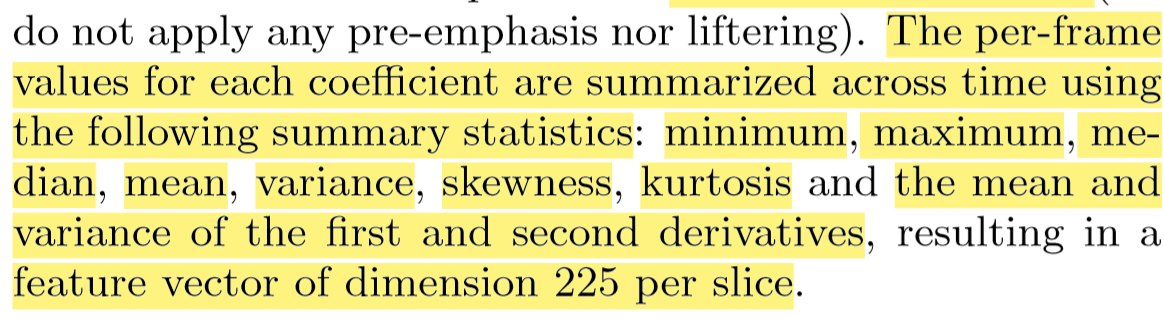


* Sound classification

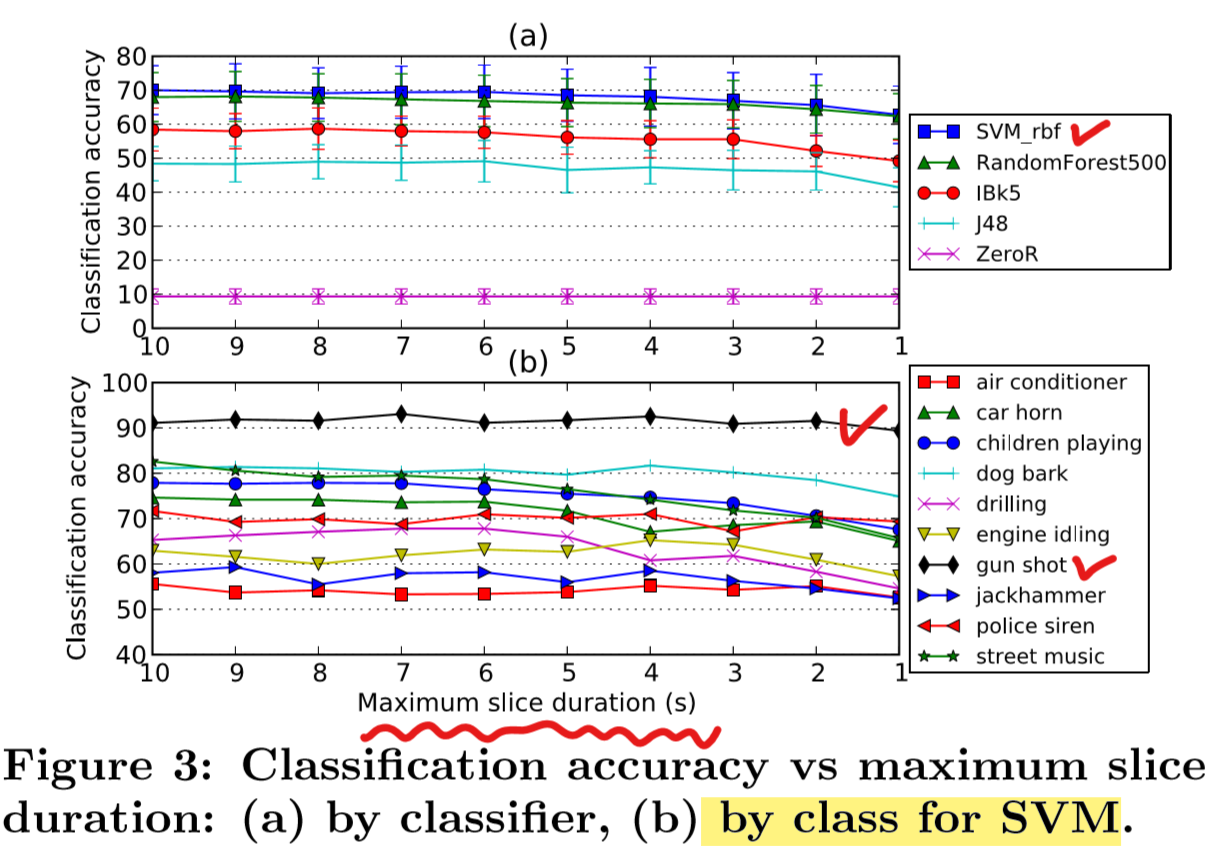
1. 用baseline在数据集上测试
2. **Baseline：MFCC+classifier**

MFCC：**用Essentia audio analysis library(C++库)实现**，23.2ms加窗，50%的frame overlap

用0-22050Hz的40个Mel bands，使用前25个MFCC系数



1. **使用Weka data mining工具测试不同的分类器**
2. 音频时长对准确率的影响：



1. 3对最容易混淆的数据：air conditioner & idling engines, jackhammer & drills, children & street music
2. **MFCC对噪声敏感**

**DCASE2017**

TUT-Lab of Signal Processing, CMU-ECE & Language Technologies Institute, DCASE Workshop2017

Annamaria Mesaro, Toni Heittola, Aleksandr Diment

* **关键点**

1. **Task2: rare sound events中有3类目标，其中一类是gunshot**
2. **基于Freesound，人工清洗过数据**
3. **随后有overview文章对各系统进行总结**
4. **基于MLP的baseline，Python+Keras**

* Challenge setup

1. **Datasets**
2. **Development dataset(training set + test set): 用于训练模型**
3. **Evaluation dataset: 用于测试模型**
4. **Baseline**
5. **基于mel band energies+MLP**
6. 前端：40ms frame，50% overlap，40 mel bands(0-22050Hz)
7. 特征向量：长200
8. MLP：2个dense layer，每个有50 hidden units，20% dropout
9. 训练：Adam for gradient-based optimization(ref25)

Maximum 200 epochs, learning rate 0.001

Early stopping  **detection was done by applying binary classification frame-wise**

1. 输出层：task specific; **task 2:每个class都有一个binary classifier，每个classifier输出层为sigmoid**
2. **实现：Python+Keras integrate the classification decisions into event activity by median filtering**

* **Task 2: detection of rare sound events**

1. 训练数据都是**音频混合出来的**，混合时有专门调整不同信噪比(EBR)
2. **每段音频中目标出现1次/0次**
3. **对3类的rare events，每类搞一个分类器**
4. **source file：用于混合制造mixture**

**recipe：混合好的音频**

提供了信号混合的软件工具

**background：TUT Acoustic Scene 2016 development dataset**

1. **3类rare sound**
2. Baby cry
3. Glass break
4. **Gunshot（134 training，53 test，平均长度1.32s）**

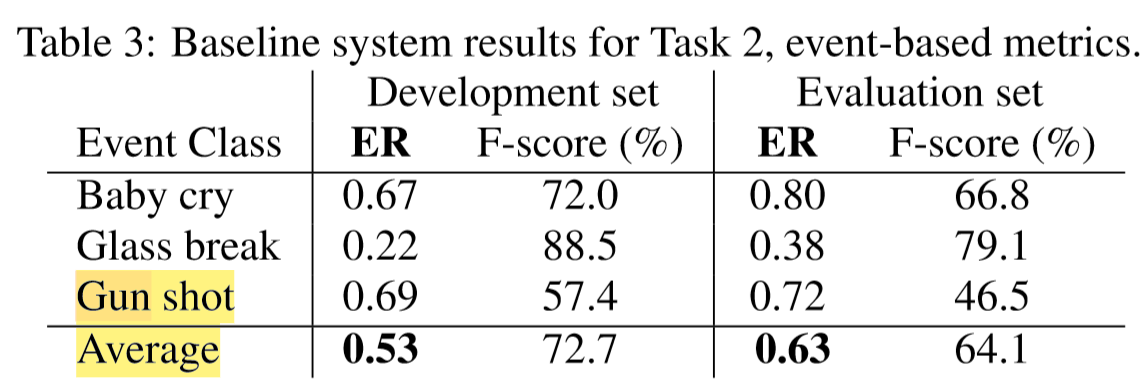
**用Freesound API下载的**

对数据进行了人工清洗和剪裁

1. 每类声音，制造500个mixture，目标出现率为0.5（250个样本有目标，250个没有目标），混合信噪比(EBR)为-6dB, 0dB, 6dB
2. 按照0.71:0.29划分training set和test set，因此Gunshot（134 training，53 test，平均长度1.32s）
3. **Baseline：for each of the target classes, there is a separate binary classifier with one output neuron with sigmoid activation, indicating the activity of the target class**

**评估：(ref27) 排行榜中的结果为3类结果的均值**

1. **event-based error rate**
2. **event-based F-score**



**DCASE2017 Overview**

**data augmentation**

**Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification, Signal Processing Letters 2016, 580+ citation**

Adobe，NYU-MARL

**Justin Salamon(4k+ citation，SONYC成员，UrbanSound成员)**

**Juan P Bello(7k+ citation，SONYC成员，UrbanSound成员)**

* **关键点：**

1. **使用DCNN做环境声检测（基于Urbansound8K）：SB-CNN**
2. **使用data augmentation提高classification accuracy**
3. **分析了5种data augmentation对Urbansound8K中10个类别数据的classification accuracy影响**

* **分析：**

1. **MUDA由NYU的音乐科技团队推出，现在是否用专门用于audio的工具？用MUDA是否合适？**
2. **整体和MUDA无二致，不过MUDA用于音乐/乐器，这个用于环境声检测**

* Intro

1. 提出基于DCNN的环境声检测：基于spectrogram，很好捕捉能量模式，分辨噪声类声音的特性；CNN的receptive field-即使有sound masking也能从spectro-temporal patterns中识别出，而MFCC可能会失效
2. Data augmentation：给出了之前在环境声检测中用DA的ref（ref11，ref23）

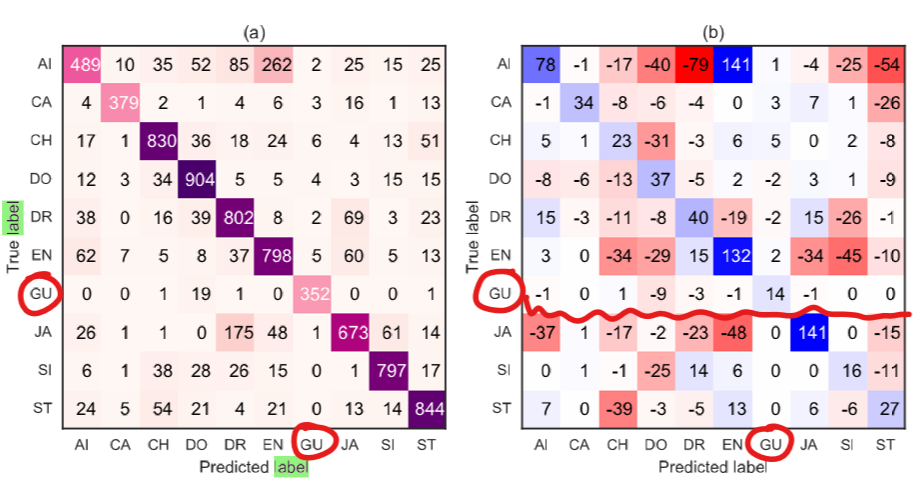
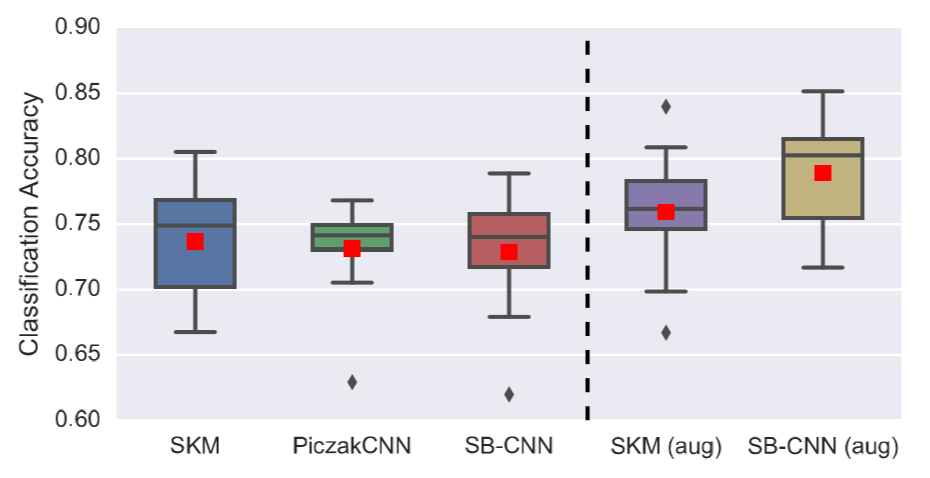
* Method

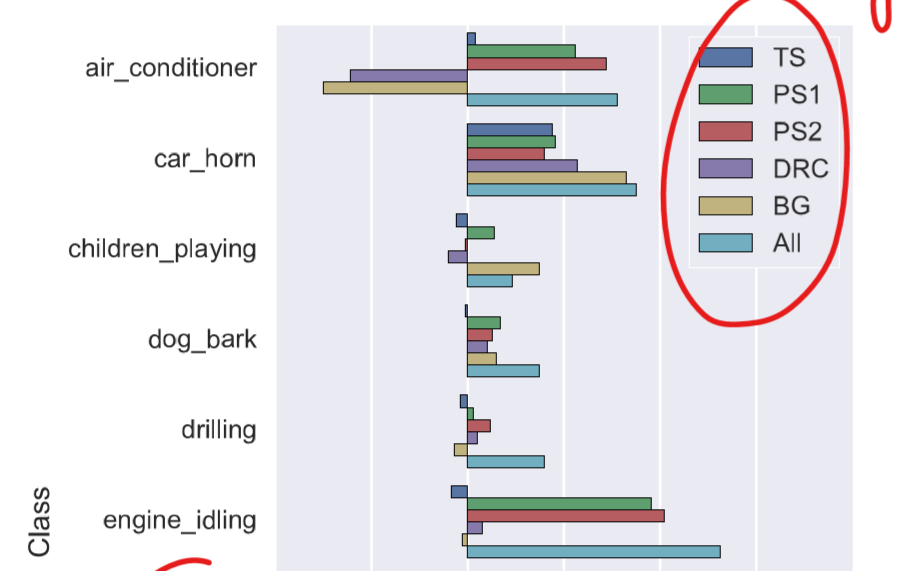
1. DCNN架构
2. **Data augmentation：5种操作（从MUDA抄来的）**
3. **time stretching**
4. **pitch shifting 1**
5. **pitch shifting 2**
6. **dynamic range compression**
7. **background noise**

**使用MUDA（ref22）来具体实施**

* evaluation

1. 数据使用UrbanSound8K
2. 几个测试模型：基于k-means的SKM，基于CNN的PiczakCNN，**本文的DCNN叫SB-CNN**
3. 测试结果：左上不同模型对比，右上对比DA对SB-CNN的影响，左右下分析5种DA对各类数据影响





**MUDA原始论文**

**A Software Framework for Musical Data Augmentation, ISMIR2015**

NYU-MARL

Brian McFee, **Juan P Bello(7k+ citation，SONYC成员，UrbanSound成员)**

* **关键点：**

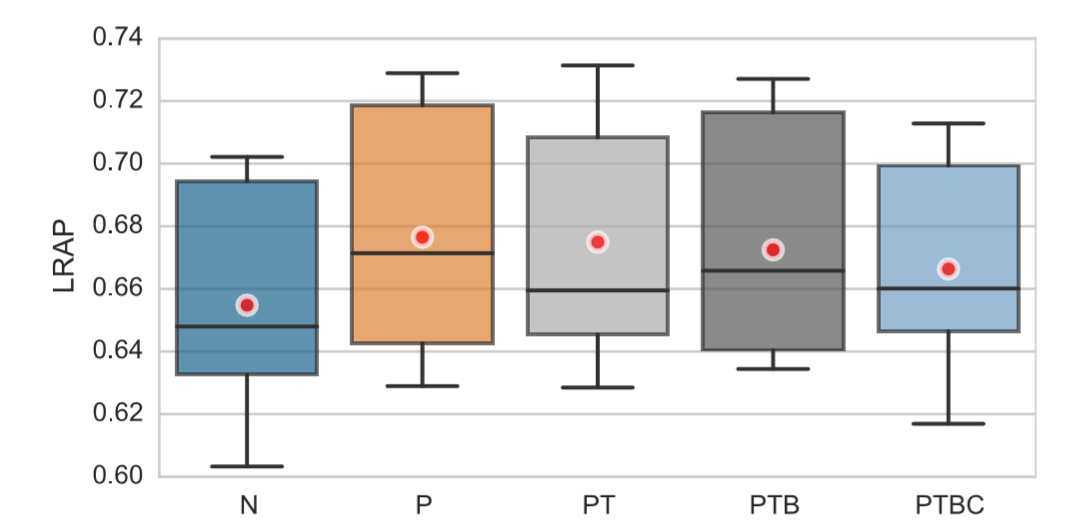
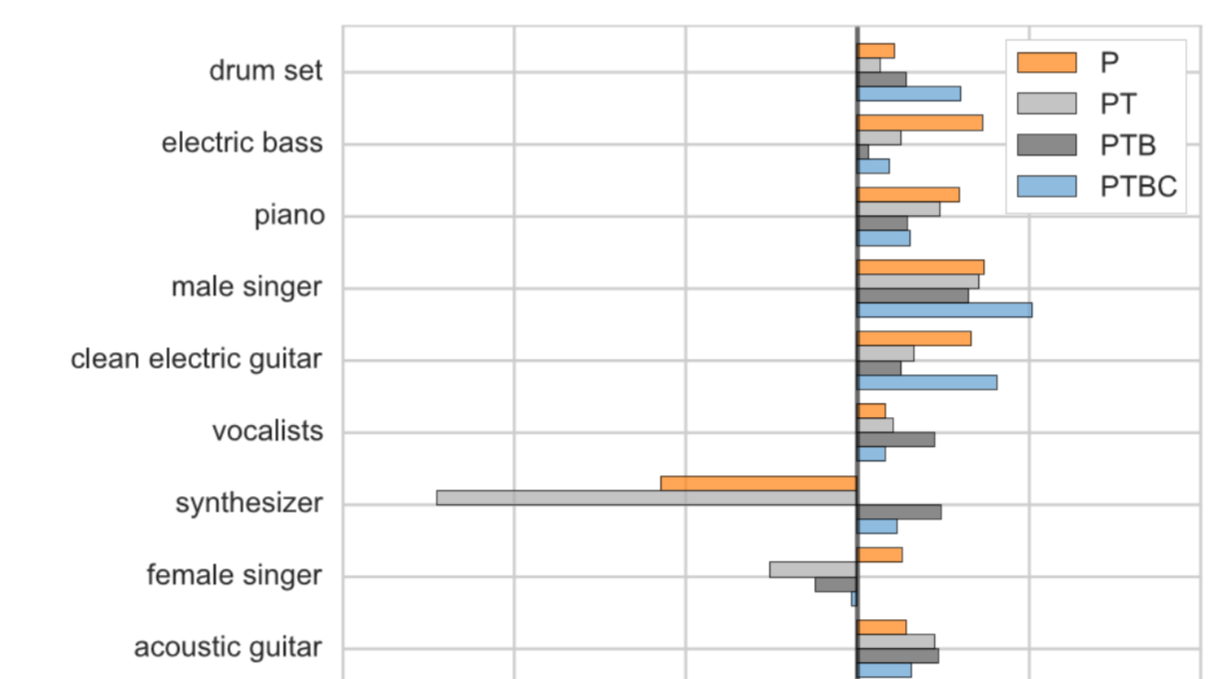
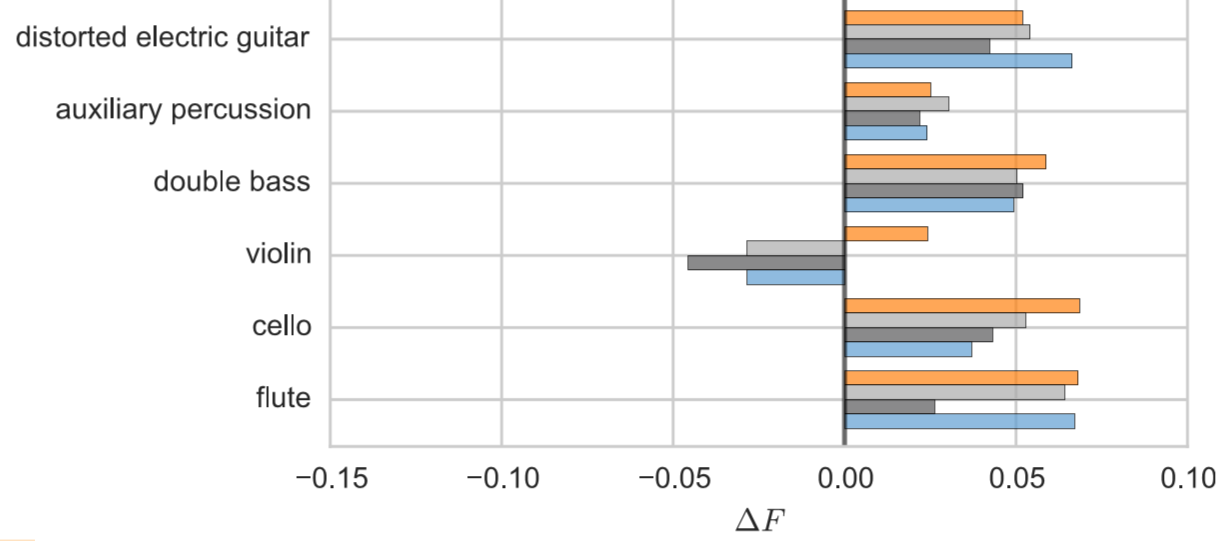
1. **提出了基于Python的data augmentation工具MUDA**
2. **用MUDA对乐器识别进行了测试**

* Data augmentation architecture:

1. 基于Python
2. 基于JAMS
3. 输入(S,x)，输出x‘，S是代表变换的state object，x是信号，x‘是变换后的信号

* Example: instrument recognition

1. 4种data augmentation:
2. time stretching
3. pitch shifting
4. dynamic range compression
5. background noise
6. 声学模型：DCNN
7. 只对比了4种操作累加的效果：N,P,PT,PTB,PTBC依次代表没有DA，只是用P，使用P+T,…,所有DA都用上
8. 左：不同DA累加的影响，中右：对不用乐器的影响

1. 数据集：MedleyDB

**选题：**

1. **环境事件检测：在已有模型上做修补+数据增强**
2. **枪声（稀有事件检测）：数据增强-》枪声分类-》枪声合成，回响消除+考虑现场/野生数据的削峰问题**
3. **用理论枪声波形增强野生数据库数据来提高枪声baseline**
4. **用Maher数据库中，信号随角度的变化规律来分析UrbanSound/Freesound中的录音角度**

**用数据训练一个声音生成器，使得模型输入为角度，输出为虚拟声音**

**枪声检测目前存在的问题：**

* **正确率随信噪比下降**
* **传声器失真**

**近期：**

1. **跟老师讨论题目：基于先验声学模型+data的机器学习模型**

**Pipeline：**

**自适应滤波-》深度学习模型**

1. **论文/专利查重**
2. **建立baseline！！！跑UrbanSound baseline、DCASE baseline**

**环境配置+理解代码使用有难度，需要一定时间**

**先用Matlab+小规模数据进行算法验证？**

**颜永红：在Maher上训练，迁移到UrbanSound/DCASE，如果自适应数据量更少而正确率更高就赢了**

**探究UrbanSound和DCASE baseline迁移到对方数据库的效果**

**现场采样背景声混合后对模型做adaptation**

**LPC和MFCC的混用**

1. **数据库：TUT rare sound events 2017**

**2020-06-13**

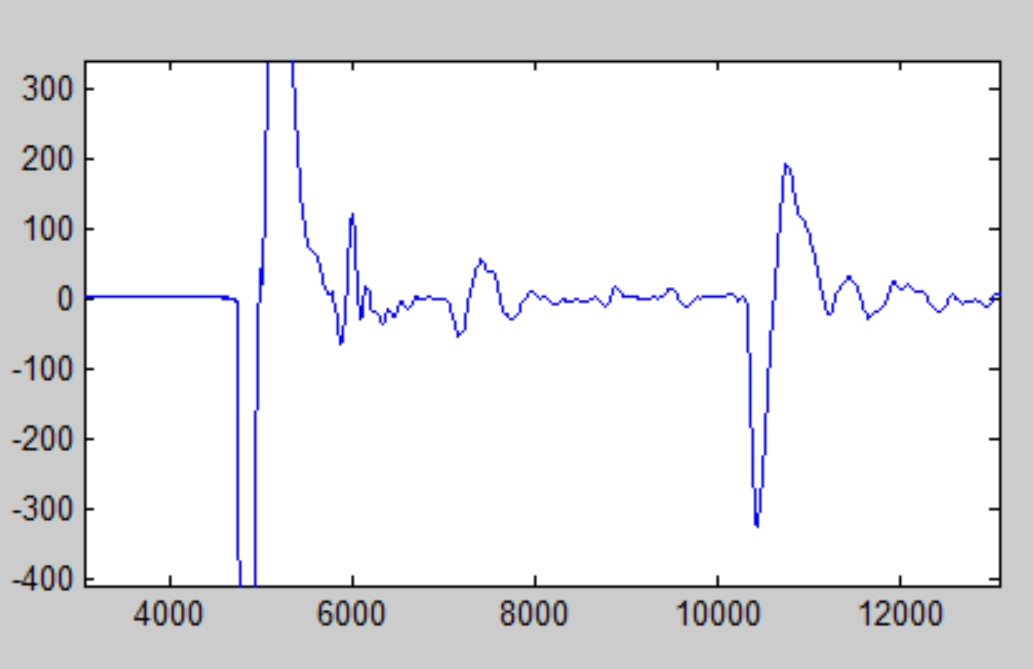
[Mahar枪声数据库：](http://www.montana.edu/rmaher/gunshots/gunshot_data.html)

* 音频播放有很大杂音
* Matlab分析看信噪比很高，非常接近理想波形（使用MaherDatasetSignalAnalysis.m分析），结果fig所在目录同数据所在目录
* 6个数据包，10种枪，共104个mat，每个mat有12声道

Sig-10个mat，sp101 38-9个mat，sp101 357-10个mat，glock19-10个mat

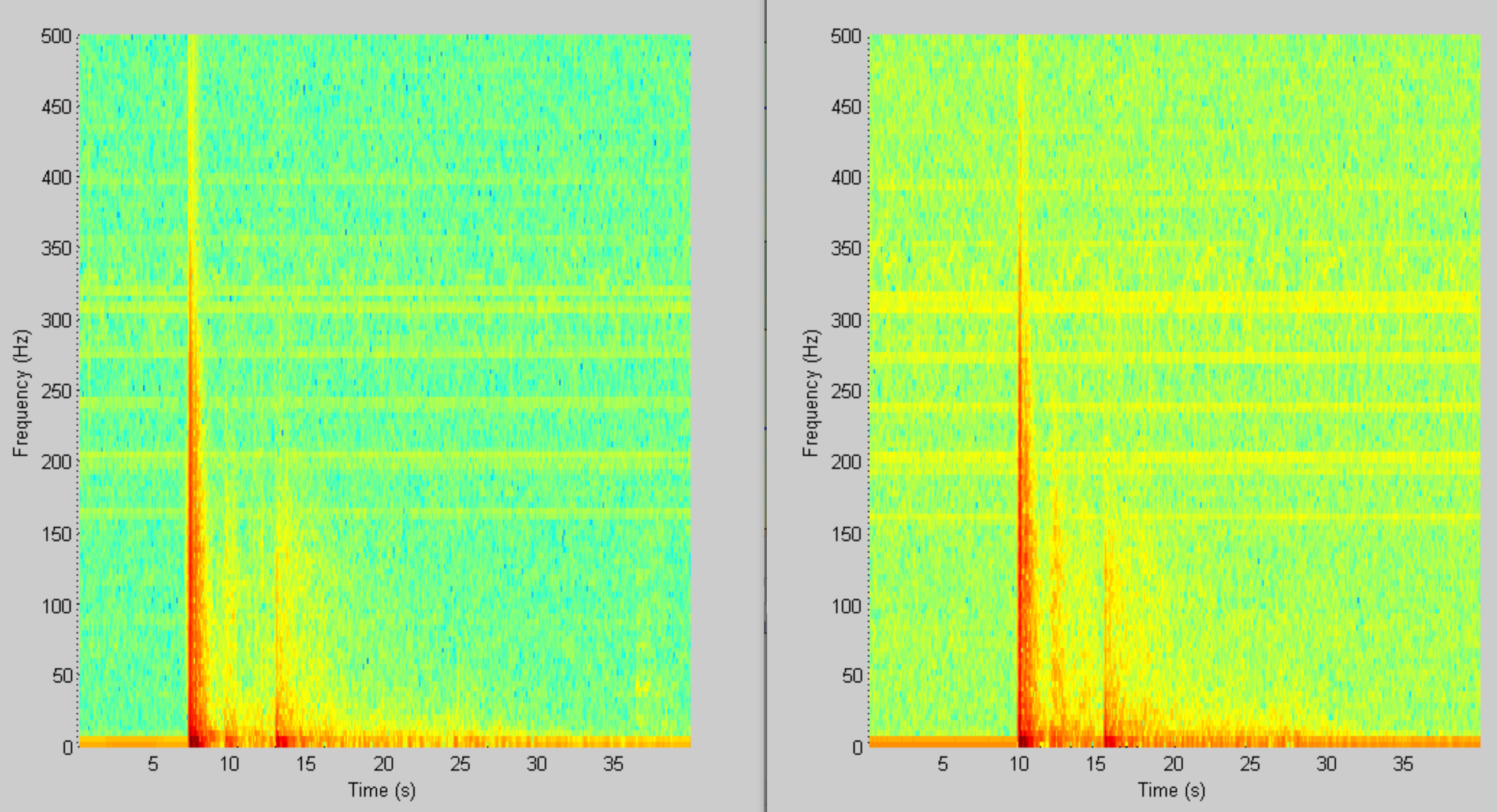
Glock23-10个mat，colt-10个mat，22rifle-10个mat，20dB增益22rifle-12个mat

308rigle-10个mat，12ga-3个mat

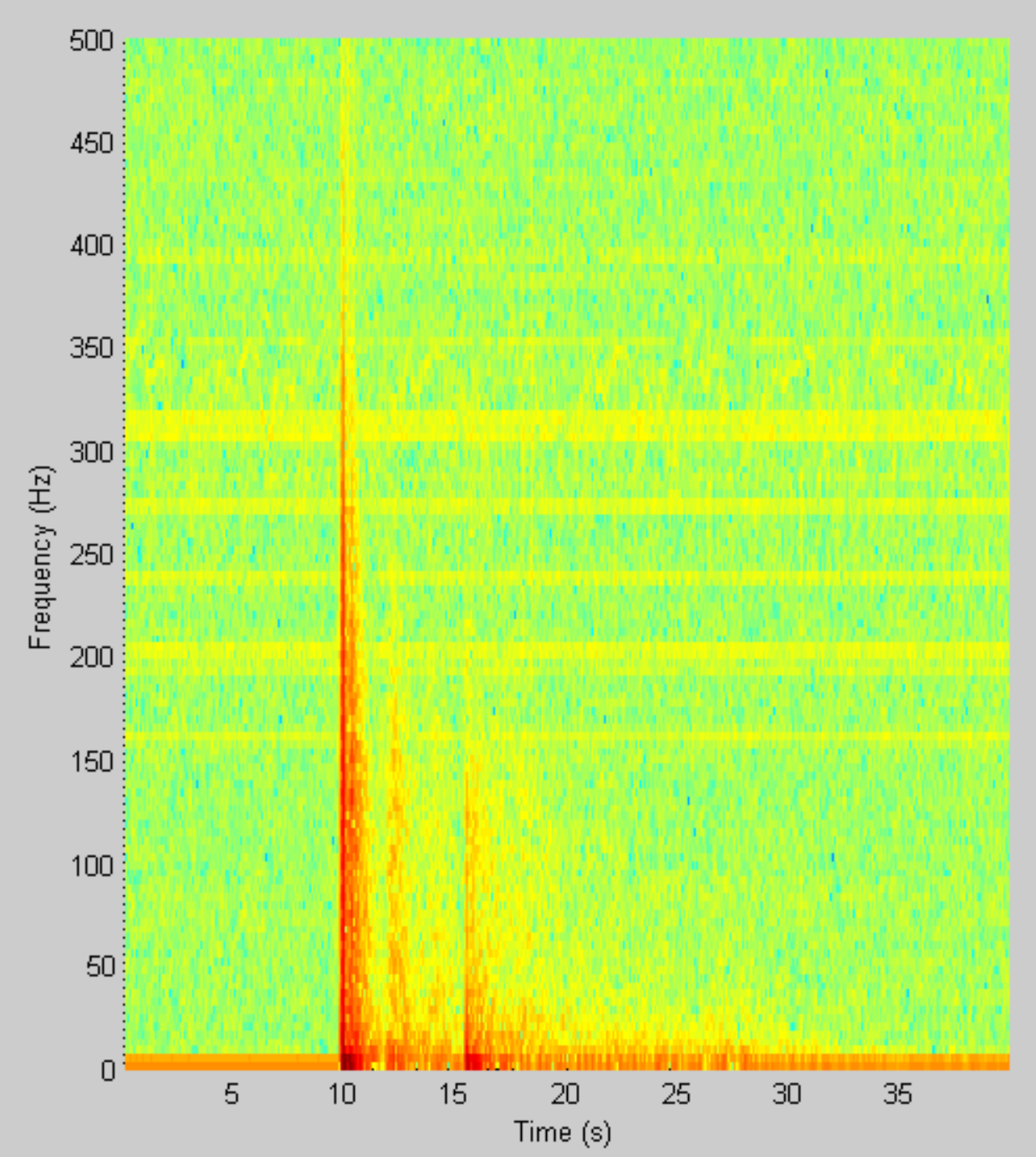
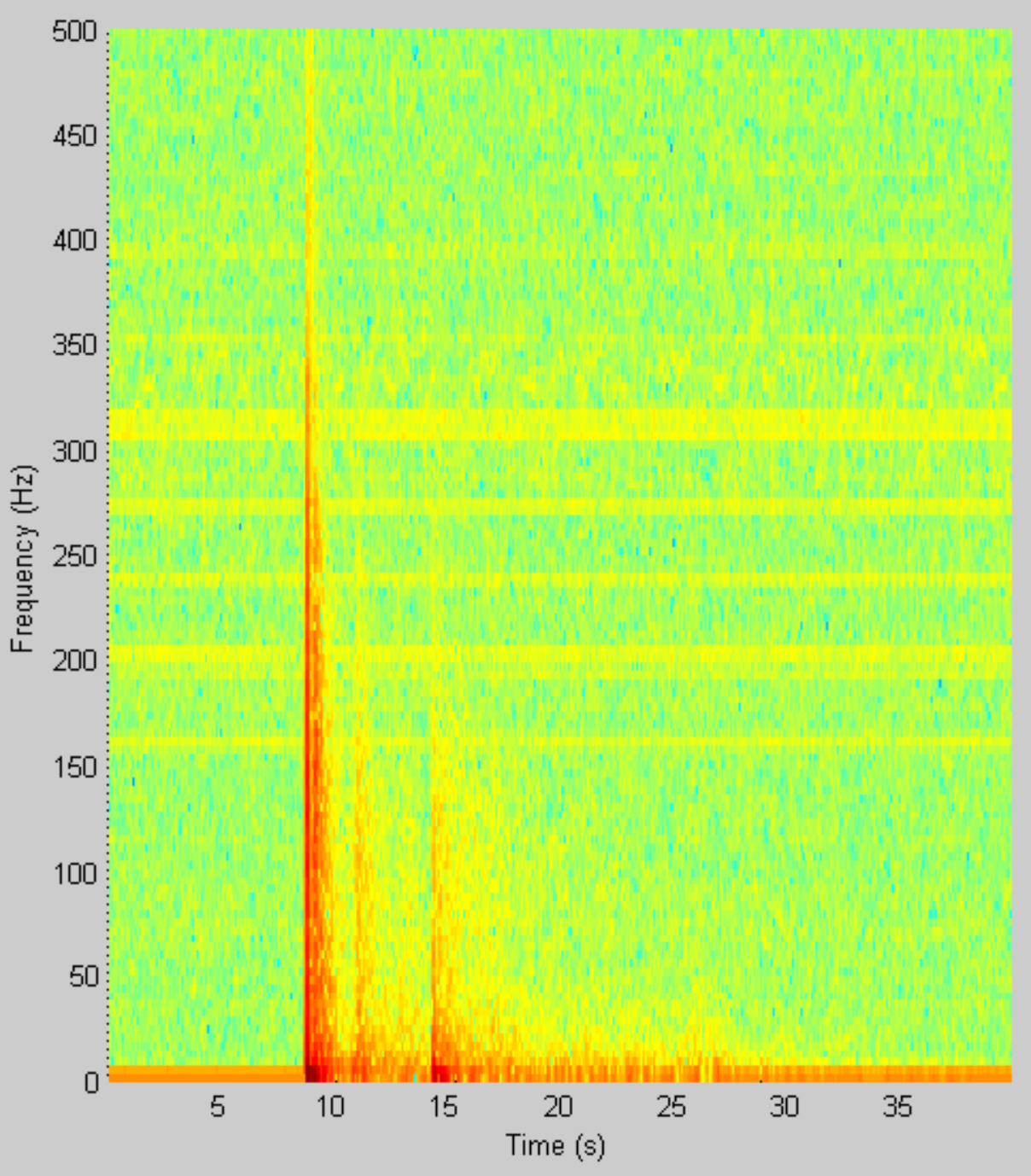


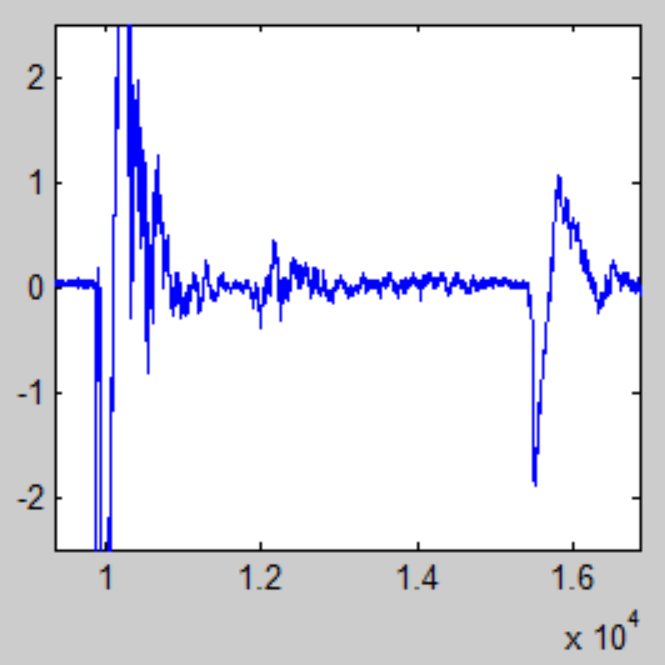
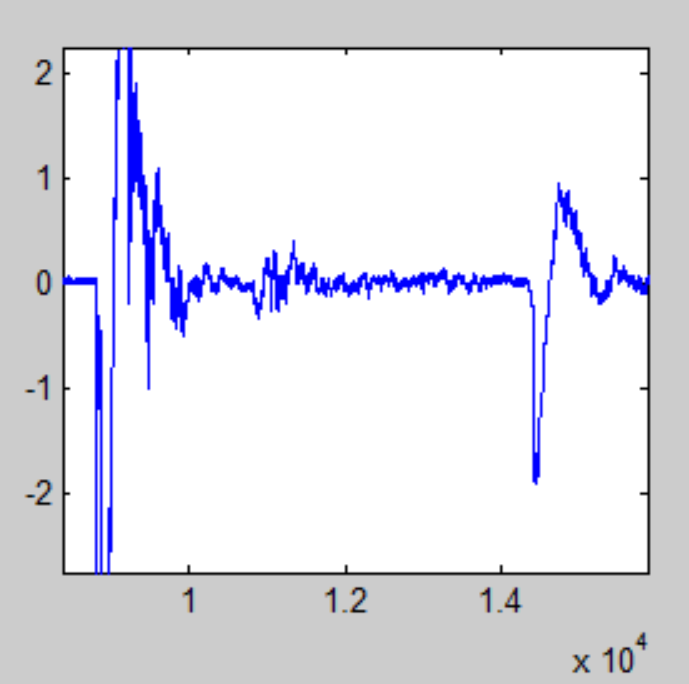
**2020-06-14**

声谱图纵向对比：左sig，右ruger（使用MaherDatasetSignalAnalysis.m分析）



声谱图横向对比：左ruger01，右ruger05（使用MaherDatasetSignalAnalysis.m分析）



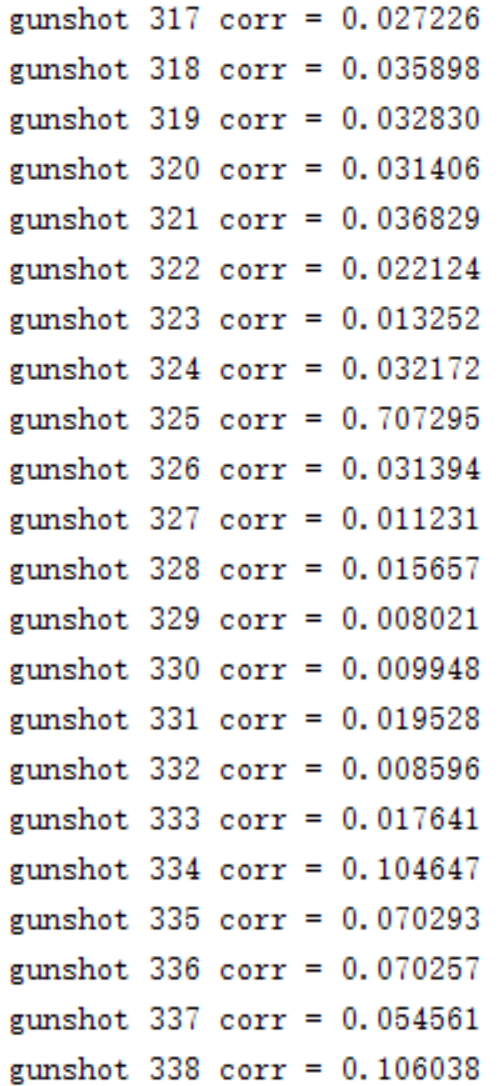


**2020-06-18**

* 略读了论文Context-dependent sound event detection(EURASIP Journal2013)：场景作为先验限定SED中event的选区范围，感觉价值不是太大，我觉得文中的先验应该被概率化
* Lin解压了UrbanSound整个数据包并做了枪声文件提取的程序（Python），结果在Gunshot Research Group云盘中
* 装了weka，在看教程(Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 4th edition, 2016“)，接下来准备用weka复现UrbanSound的baseline

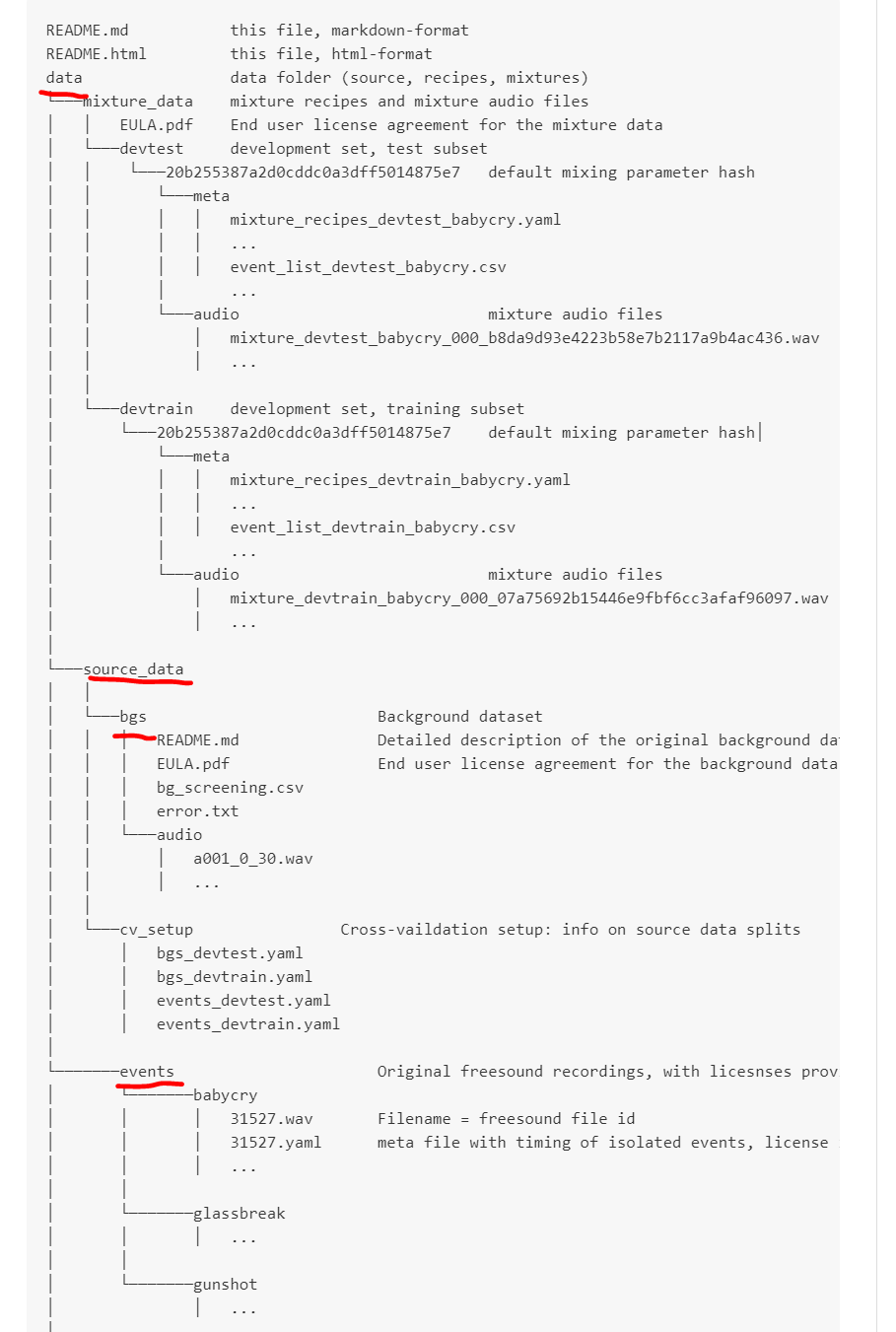
**2020-06-19**

* 数据之间的互相关很小：



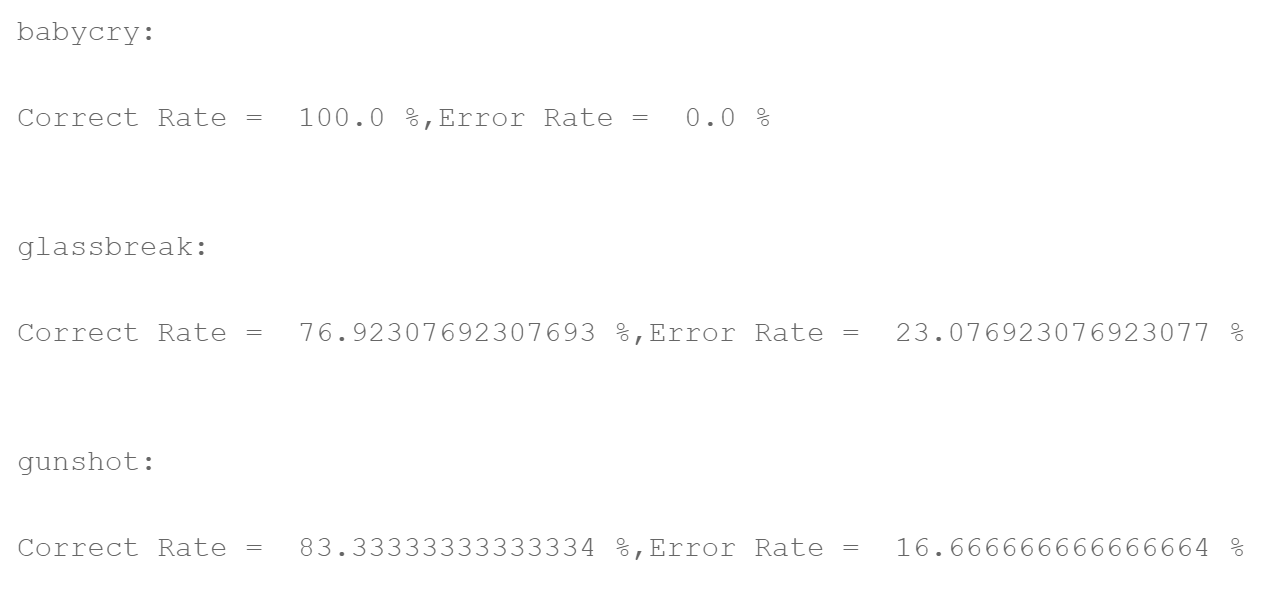
**2020-06-20**

* DCASE2017文件结构



**2020-06-21**

* **MFCC+GMM迁移到了AIStudio的Python版，使用sklearn做机器学习框架，在DCASE2017 task2的纯event data上测试：**



1. MFCC+GMM（备胎，用于做初步测试）
2. CRNN复现（主力）
3. 在UrbanSound/DCASE2017(Freesound+DCASE2016)/Maher上测试
4. 探究点：

* 模板粗匹配加速识别（聚类，每个类出一个模板）
* 数据自适应
* 特征的选用：MFCC/mel-band energy/LPC/混合特征
* 容易导致误检的数据

数据-模型：实验平台