

## 進捗報告

### 空間情報に着目した 音響イベント検出

筑波大学 マルチメディア研究室音声班 4年  
溝口和輝  
指導教員 山田武志准教授

2017/09/19

1

## 研究背景

- ◆ 環境音認識の研究の活発化
  - 高齢者見守りシステム
  - 特定の音響イベントの感知

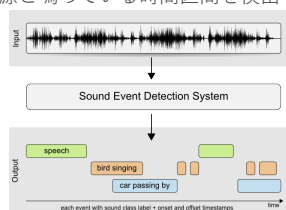
2017/09/19

2

## 実生活を想定した音響イベント検出

### ◆ 概要

日常生活と同様の音源が重なり合っている状況下で、  
個々の音源と鳴っている時間区間を検出する



→DCASE2017を用いた比較評価

2017/09/19

3

## DCASE2017

1. Acoustic scene classification
2. Detection of rare sound events
3. Sound event detection in real life audio
4. Large-scale weakly supervised sound event detection for smart cars

2017/09/19

4

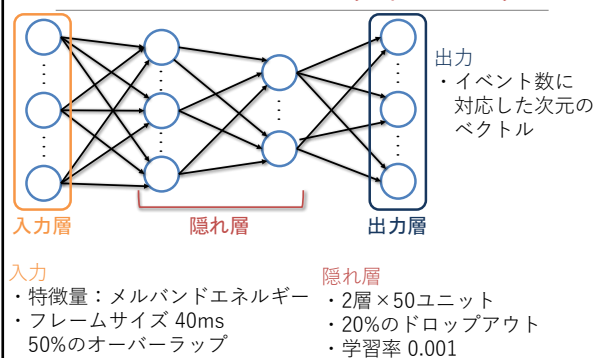
## 先行研究

- ① mel energy+RNN(ステレオ入力)  
advanne\_IID/DCASE2016
- ② MFCC + GMM(モノラル入力)  
DCASE2016\_baseline/DCASE2016/

2017/09/19

5

## DCASE2017のベースラインシステム



2017/09/19

6

### ◆ Sound event detection in real life audio

#### ◆ データセット

- 録音データが24個
- 各3分～5分の長さ
- いずれも屋外(道路沿い)
- 44.1kHzサンプリングレート  
24ビット

#### ◆ 音響イベントの数

サウンドイベントラベル	イベントカウント
ブレーキ音	59
車	304
子供	58
大型車両	61
会話音	117
歩行音	130

2017/09/19

7

### ◆ Sound event detection in real life audio

#### ◆ 出力

イベント開始時刻[s]	イベント終了時刻[s]	イベントラベル
0.186989	2.558792	car
2.421011	4.536935	car
4.271214	10.530413	car
11.475197	13.51239	car
13.837159	22.064659	car
15.500374	18.187105	brakes squeaking
18.531558	22.00561	car
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮

2017/09/19

8

### DCASE2017の評価方法

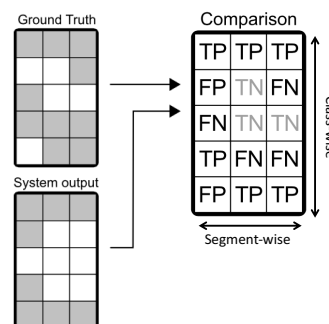
DCASE2017では2種類の評価方法が存在

- Segment-Base による評価方法
  - task3&task4
- Event-Baseによる評価方法
  - task2

2017/09/19

9

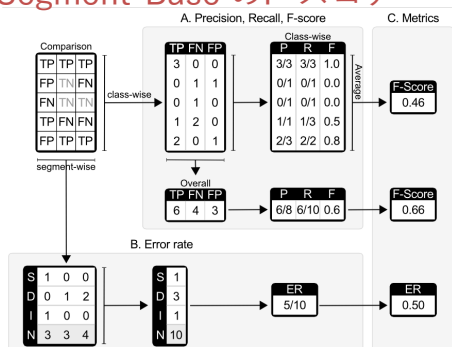
### Segment-Base のF-スコア



2017/09/19

10

### Segment-Base のF-スコア



2017/09/19

11

### Segment-Base のF-スコア

$$F = \frac{2(P * R)}{P + R}$$

$$P = \frac{\sum TP(k)}{\sum TP(k) + \sum FP(k)}$$

$$R = \frac{\sum TP(k)}{\sum TP(k) + \sum FN(k)}$$

実数	概要
TP	正しく検出されたイベント
FP	システムの出力はアクティブであるが本来はアクティブでないもの
FN	システムの出力はアクティブではないが、本来はアクティブなもの

2017/09/19

12

## Segment-Baseのエラー率

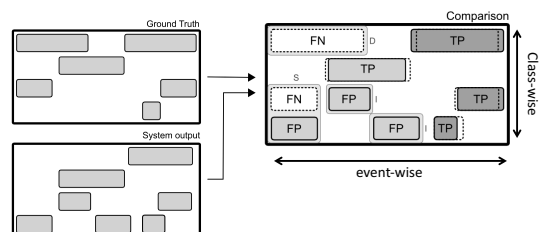
$$ER = \frac{\Sigma S(k) + \Sigma D(k) + \Sigma I(k)}{\Sigma N(k)}$$

変数	概要
TP	正しく検出されたイベント
FP	システムの出力はアクティブであるが本来はアクティブでないもの
FN	システムの出力はアクティブではないが、本来はアクティブなもの
S(k)	正しいイベントが出力されなかった参照イベントの回数 $S(k) = \min(FN(k), FP(k))$
I(k)	本来は存在しないのに誤検出してしまったシステムのイベントの回数 $I(k) = \max(0, FP(k) - FN(k))$
D(k)	本来は存在するが正しく識別されなかったイベントの回数 $D(k) = \max(0, FN(k) - FP(k))$
N(k)	セグメントkにおいてアクティブだと判断されたkの総数
k	セグメント

2017/09/19

13

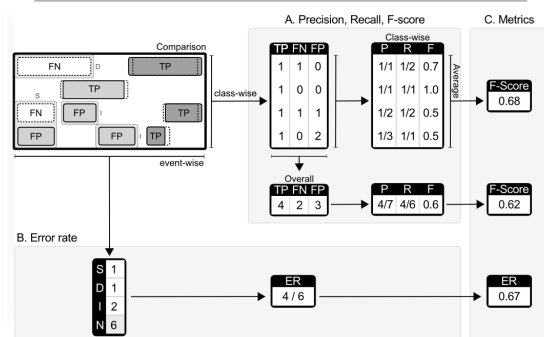
## Event-Baseのエラー率とF-スコア



2017/09/19

14

## Event-Baseのエラー率とF-スコア



2017/09/19

15

## Event-Baseのエラー率とF-スコア

$$ER = \frac{S + D + I}{N} \quad F = \frac{2(P * R)}{P + R}$$

変数	概要
TP	正しく検出されたイベント
FP	システムでのみ間違っって出力されているイベント
FN	システムからの出力に同じラベルのイベントが存在しないイベント
S	間違っったラベルをアクティブとして出力 (正しいイベントを検出出来なかったが、何らかのイベントを検出した場合)
I	Sを除いたFP(偽陽性)
D	Sを除いたFN(偽陰性)
N	総数

2017/09/19

16

## 実験結果

### ◆DCASE2017

Developments dataset	SegmentBase		EventBase	
	Fスコア	Error率	Fスコア	Error率
Baseline	56.70	0.69	(6.61)	(3.75)

### ◆(参考)DCASE2016

Evaluation dataset	SegmentBase		EventBase	
	Fスコア	Error率	Fスコア	Error率
task3 advanne_IID	47.8	0.8	4.8	5.1
task2 komatu	80.2	0.3	73.8	0.5

2017/09/19

17

## DCASE2017の結果

Segment base/Evaluation dataset

入力	音響特徴量	識別器	Fスコア	Error率
モノ	scattering transform clustering	Neuroevolution	44.9	0.8979
モノ	log-mel energy	CRNN	44.2	1.0318
バイノーラル	log-mel energy	CNN	43.6	0.8985
バイノーラル	log-mel energy	CNN	42.9	0.8061
モノ	log-mel energy	MLP	42.8	0.9358

2017/09/19

18

## 今後の方針

- ◆入力のステレオ化

2017/09/19

19

## 年間計画

4月		10月	卒研中間報告
5月	就職活動	11月	音響学会申込
6月		12月	
7月		1月	音響学会提出 卒論完成
8月		2月	研究発表
9月		3月	音響学会発表

2017/09/19

20