

# 中間報告

## 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

# 研究背景

---

## ◆ イベント音検出の研究が活発化

### ➤ 高齢者見守りシステム



例)生活音の中から異常音を検知

### ➤ 動画の自動タグ付け

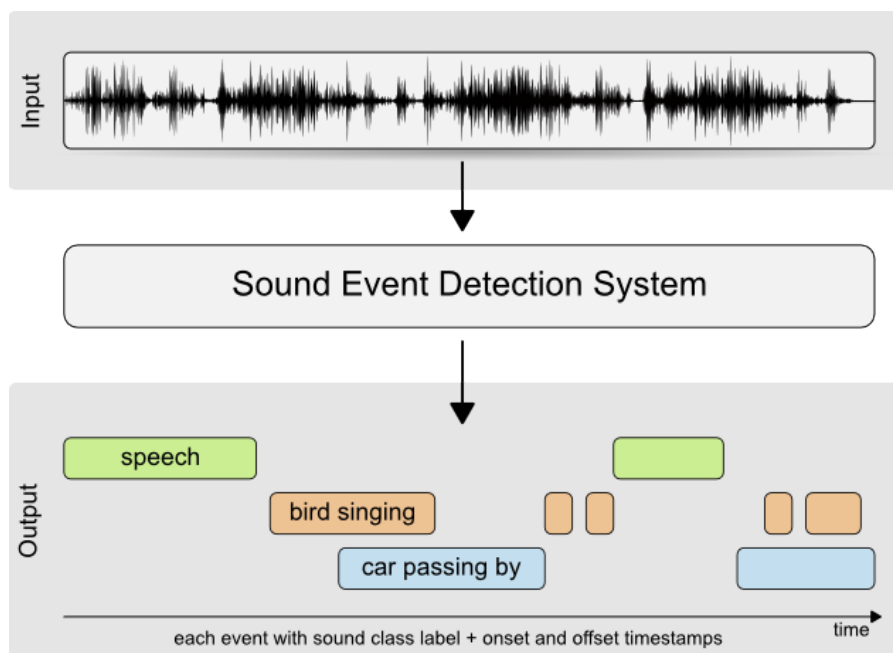


例)動画のタグ付けのために音響イベントを検出

# 実生活を想定した音響イベント検出

## ◆概要

複数のイベント音が重なり合って鳴っている状況下で、個々の音の種類と鳴っている時間区間を検出する



- DCASE2017[1]task3のデータセットを用いた比較評価  
[1]<http://www.cs.tut.fi/sgn/arg/dcase2017/>

# Sound event detection in real life audio

本研究で扱うDCASE2017task3のデータセットは以下のようになっている

## ◆ データセット

	開発用データセット	評価用データセット
録音環境	binaural録音 屋外(道路沿い)	
録音データ	24個	8個
	各3～5分程度	各2分～4分程度
サンプリングレート	44.1kHz 24bit	
教師データ	あり	なし
評価法	4-fold cross-validation	

# Sound event detection in real life audio

## ◆音響イベント ◆教師データ

サウンドイベント	発生回数
ブレーキ音	59
車	304
子供	58
大型車両	61
会話音	117
歩行音	130

ファイル名	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
b099.wav	0.186989	2.558792	car
b099.wav	2.421011	4.536935	car
b099.wav	4.271214	10.530413	car
b099.wav	11.475197	13.51239	car
b099.wav	13.837159	22.064659	car
⋮	⋮	⋮	⋮

- ラベルは 1 s～60s超と様々(平均9s)
- 1ファイル辺り平均27個のラベルがあり内半数がオーバーラップ

# 先行研究

DCASEでは例年ベースラインシステムが公開されており、参加者はこのベースラインを改良するかオリジナルなシステムを組むことでスコアの向上を目指す

## ◆ mono + mel energy

+ Multilayer perceptron architecture (MLP)

[DCASE2017\_baseline/DCASE2017]

今年のベースラインとして公開されている手法

## ◆ binaural + mel energy + RNN

[Sharath Adavanne 他/DCASE2016]

入力に左右の音のアクセントやピッチ情報を取り込む  
左右の音の到着時間差を利用

## ◆ binaural + mel energy + CRNN

[Sharath Adavanne 他/DCASE2017]

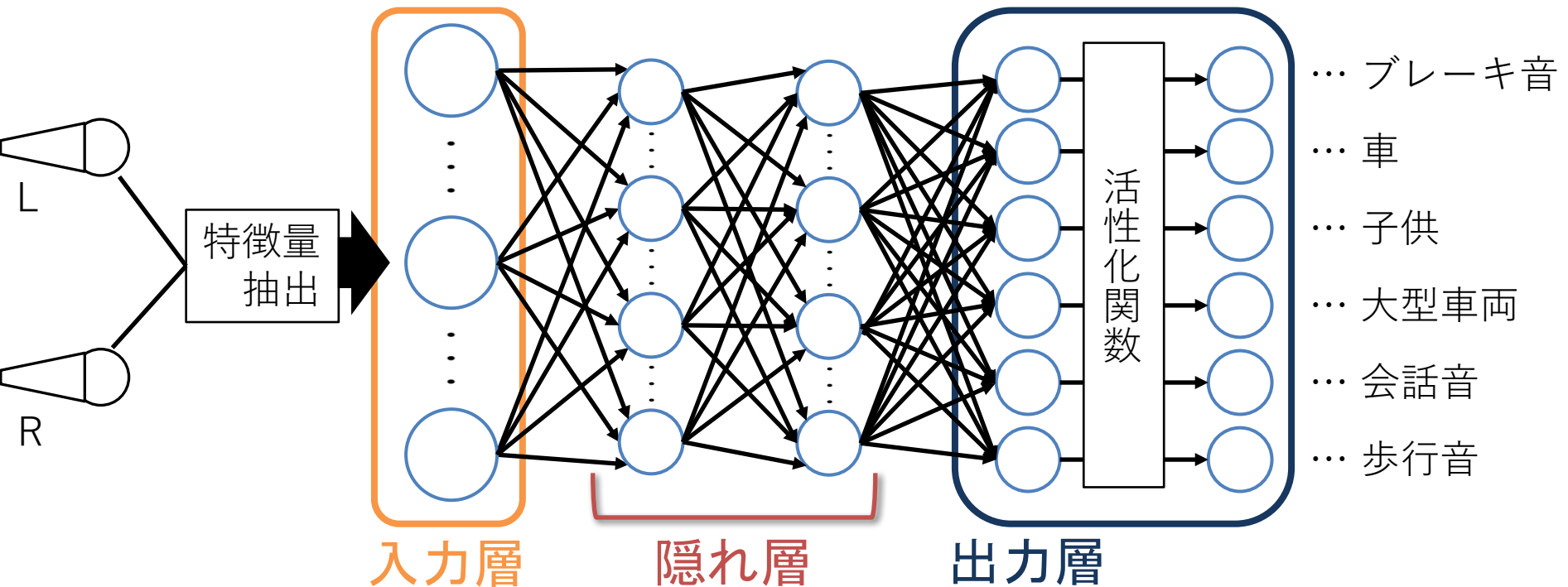
抽出した左右の空間情報に短時間フーリエ変換をかける

# 研究目的

---

- ◆ 本研究では空間情報を効果的に活用することにより、音響イベント検出の識別器の性能を改善することを目的とする
  - 昨年のDCASE2016において入力をステレオで取り扱ったものがbaselineを上回るスコアを出している
  - 先行研究では空間情報を生かした前処理を活かしてスコアを上げることが出来なかった
  - この点に着目することで性能の改善が見込める

# DCASE2017 baselineシステム



## 入力

- ・チャンネル：LとRの特徴量を連結
- ・特徴量：メルバンドエネルギー
- ・フレームサイズ：40ms

## 隠れ層

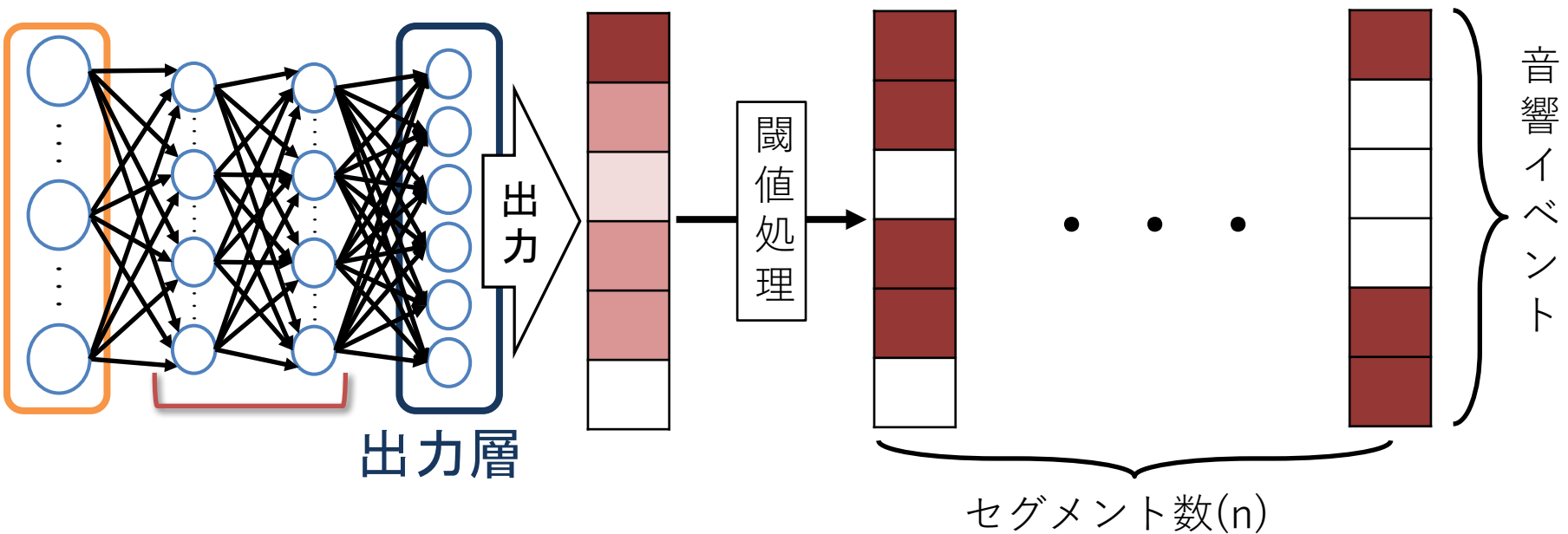
- ・2層×50ユニット

## 出力

- ・音響イベント毎に存在確率を出力
- ・活性化関数：シグモイド



# DCASE2017 baselineシステム



## 出力

- MLPからの出力された確率 $p$ を閾値処理

$$p(t_i) = 1 (0.5 \leq p \leq 1)$$

$$0 (0 \leq p < 0.5)$$

- 音響イベント数 $\times$ セグメント(音響フレーム)の行列

# task3の評価法

---

task3では2種類の評価法によってシステムを評価

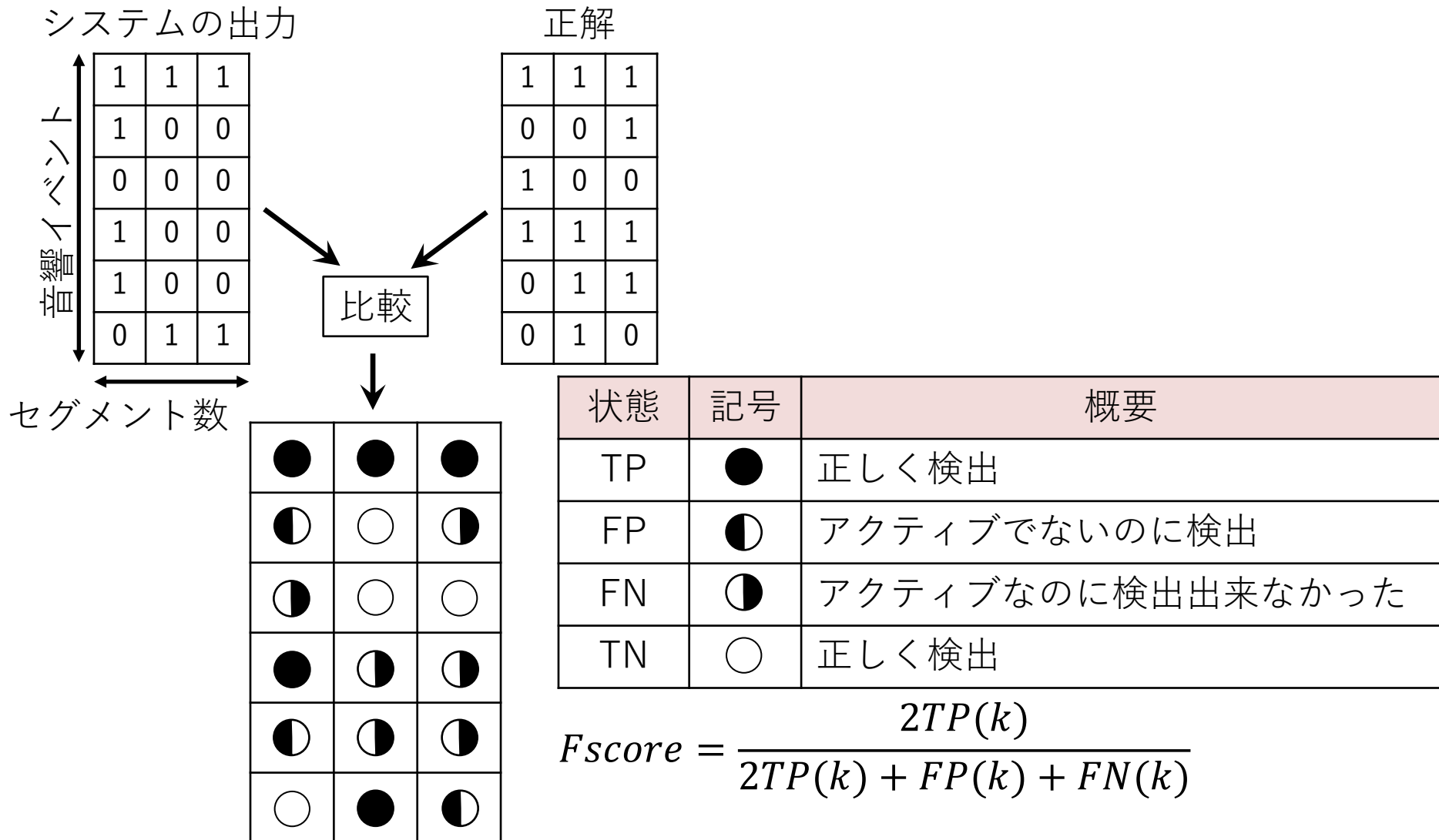
## ◆F-スコア

- 出力結果を正解と比較することによって導出
- 正確性と網羅性を総合的に評価可能

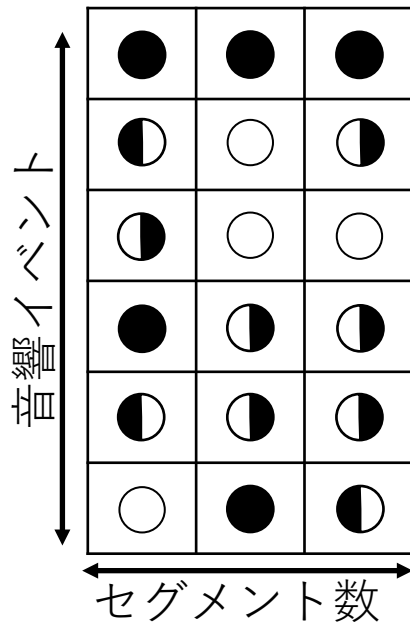
## ◆エラー率

- 割合ではなくスコア値であるため1を超える
- 誤った結果を出力した場合だけでなく、正しい結果を出力出来なかった場合にも加算される

# task3の評価法(F-スコア)



# task3の評価法(エラー率)



S	1	0	1
I	1	0	0
D	0	2	2
N	3	4	4

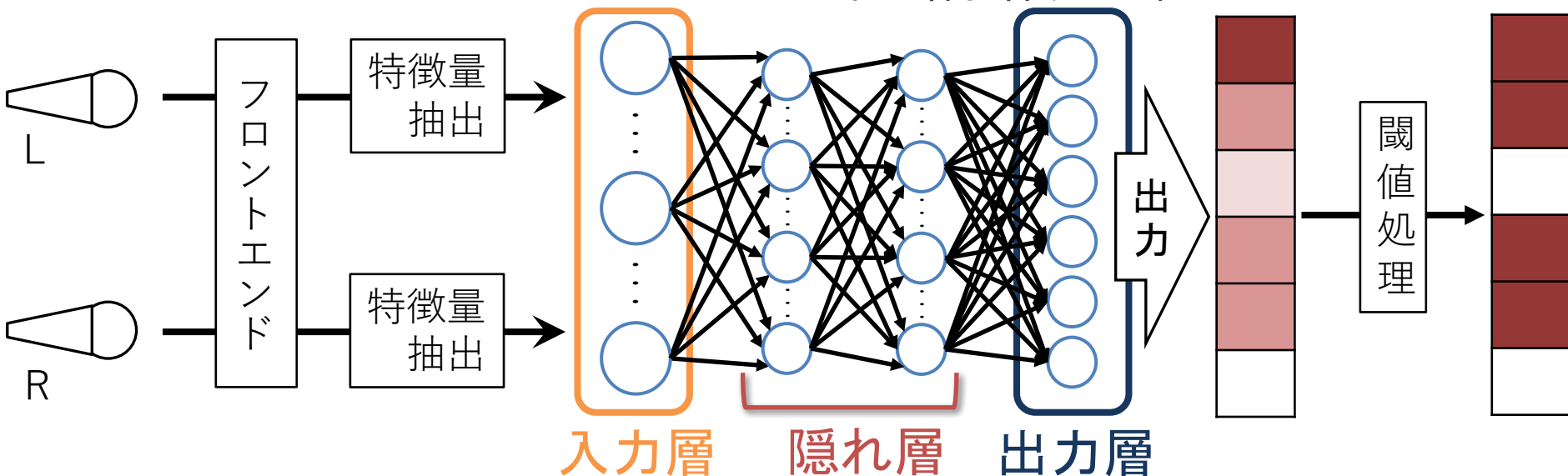
状態	記号	概要
TP	●	正しく検出
FP	◐	アクティブでないのに検出
FN	◑	アクティブなのに検出出来なかった
TN	○	正しく検出

変数	概要
S	正しいイベントを検出出来なかった回数 $S = \min(FN(k), FP(k))$
I	存在しないものを誤検出した回数 $I = \max(0, FP(k) - FN(k))$
D	存在するが正しく認識出来なかった回数 $D = \max(0, FN(k) - FP(k))$
N	本来の存在するイベントの数

$$ER = \frac{\sum S + \sum I + \sum D}{\sum N}$$

# 提案手法

## ◆ baselineの入力に空間情報を取り込む



- 特徴量は位相情報を持たないので、NNでは位相情報を活用した処理が出来ない
- フロントエンドで空間情報を抽出することでNNで利用出来るようにする

# 今後の予定

---

## ◆実装及び実験

- 10月末までにフロントエンドを組み込めるようにベースラインを改良
- 11月以降はフロントエンドの処理を試行

## ◆フロントエンドの手法

- DCASE2017task3のCompetitionを更に読み込みつつ、取り入れられそうな部分があれば敵船取り入れる

# 年間計画

4月			10月	卒研中間報告
5月	就職活動		11月	
6月			12月	
7月			1月	卒論完成
8月			2月	研究発表
9月			3月	