メディアプロジェクト演習２

RITSUMEIKAN

山下研究室（音声言語研究室）

http://www.slprits.com

１．研究紹介

（音声情報処理研究の現状）

２．作品制作のヒント

2021年4月13日

山下 洋一 ・ 福森 隆寛

音声情報処理技術

- 目的
  - 音声や音楽などの音情報をコンピュータで処理して，もっと便利に／もっと楽しく
- 応用
  - 車内でのカーナビ操作，検索キーワードの音声入力(voice search)，自動受け付け，音声翻訳，個人認証，ゲーム，音声データ検索，語学学習支援，障害者向けインタフェース，見守り，，，
- 研究分野
  - 音声認識，音声合成，音声対話，環境音認識，音声符号化，音響信号処理，情報検索，音楽情報処理，，，
- 支える技術・理論
  - 確率・統計，情報理論，ディジタル信号処理，パターン認識，プログラミング，，，

RITSUMEIKAN

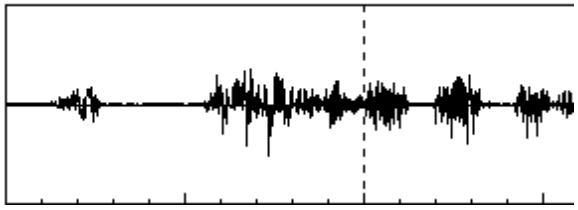
研究紹介

山下研究室での研究

- 研究内容
  - 音声・音を対象とした研究
    - 音声は，表現力豊かで人に優しいメディア
- 研究テーマ
  1. 音声の認識・合成
  2. 環境音の認識・合成
  3. 音情報を使った異常検出
  4. 音楽情報処理

1. 音声認識の現状と課題

- 音声認識
  - 計算機による人間の声の認識



音声波形

「大量のデータに基づいた統計的手法を用いることによって，，，，

  - 音声認識研究の現状
    - 大量データを利用する統計的手法による性能向上
    - 音声認識サービスの実用化
  - 今後の研究課題
    - 認識率の向上
    - 意図・態度・感情などパラ言語情報の認識
    - 頑健性(robustness)の向上
      - 雑音下での認識

1. 音声認識

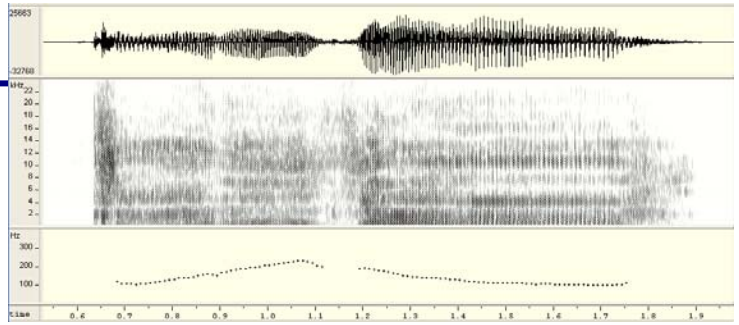
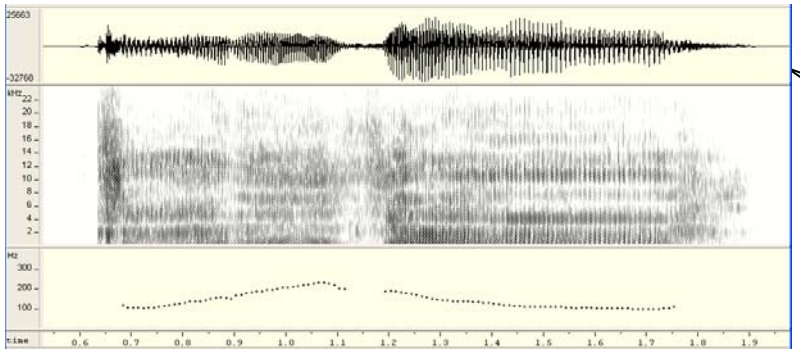


図　発話「カレーかー」の音声波形と分析結果

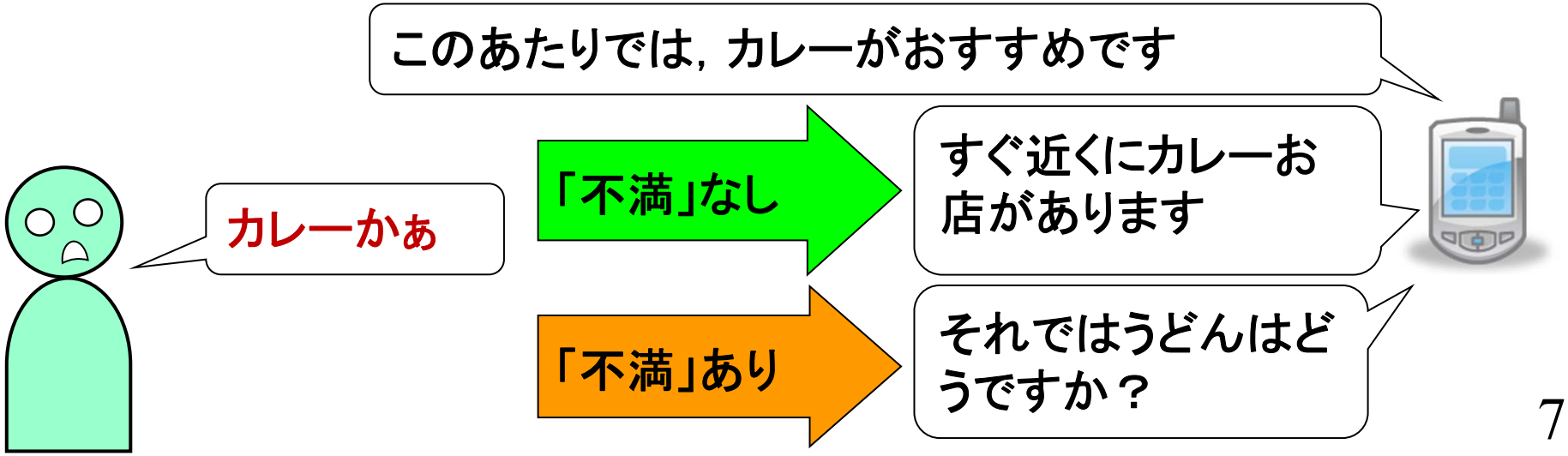
- 音声が伝える情報
  - 言語情報：発話の内容
    - 「カレーかー」
  - パラ言語情報：意図，態度，感情，など
    - 同意，喜び，…
  - 非言語情報：性別，年齢，など
    - 女性，20歳代，…
- パラ言語情報による多様な表現
  - 「そうですか」
    - 納得した？　　質問している？
  - 「わかりました」
    - 喜んでくれた？　　しぶしぶ了解した？

1. 音声認識

- ユーザの“気持ちを察する”音声対話の実現
  - 音声におけるパラ言語情報の認識



時間波形，スペクトログラム，基本周波数



1. 音声における感情認識

- 音響情報と言語情報を合わせて用いる音声感情認識
  - 喜び，悲しみ，怒り，平静の4クラスの認識

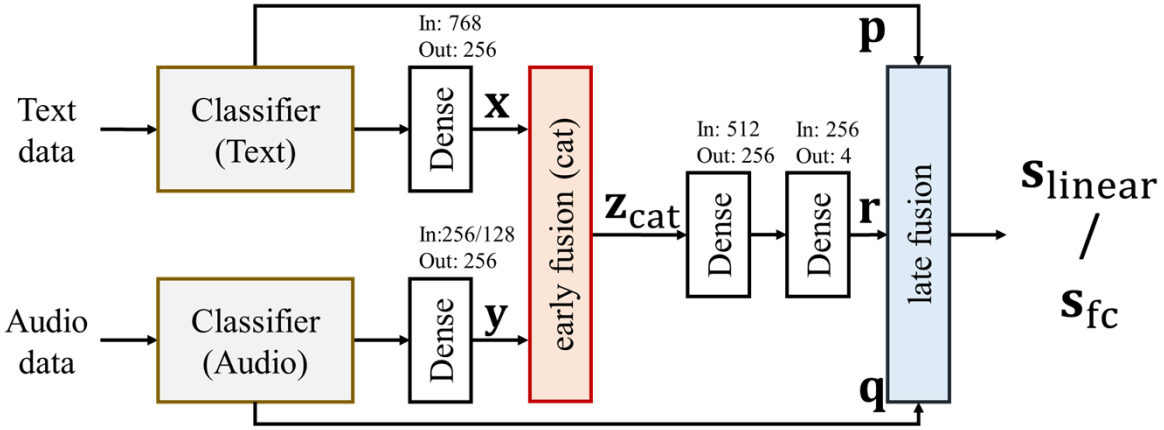


図　音声・テキスト情報を用いたネットワーク

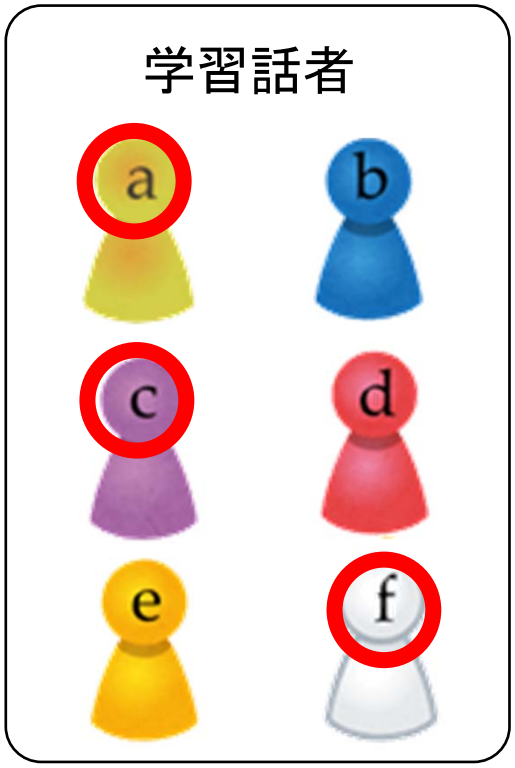
表　音声・テキスト情報を用いたネットワーク

用いる情報	認識率 (%)
音響情報のみ	71.31
言語情報のみ	66.45
音響情報 + 言語情報	87.39

1. 音声における感情認識

- 学習話者の選択

認識対象の話者X



話者Xに似た
感情表現を行う
話者を選択して
感情認識モデ
ルを学習する。

1. 音声における感情認識

- 2種類の話者選択手法
  - 部分空間を用いる手法
  - 平均ベクトルを用いる手法

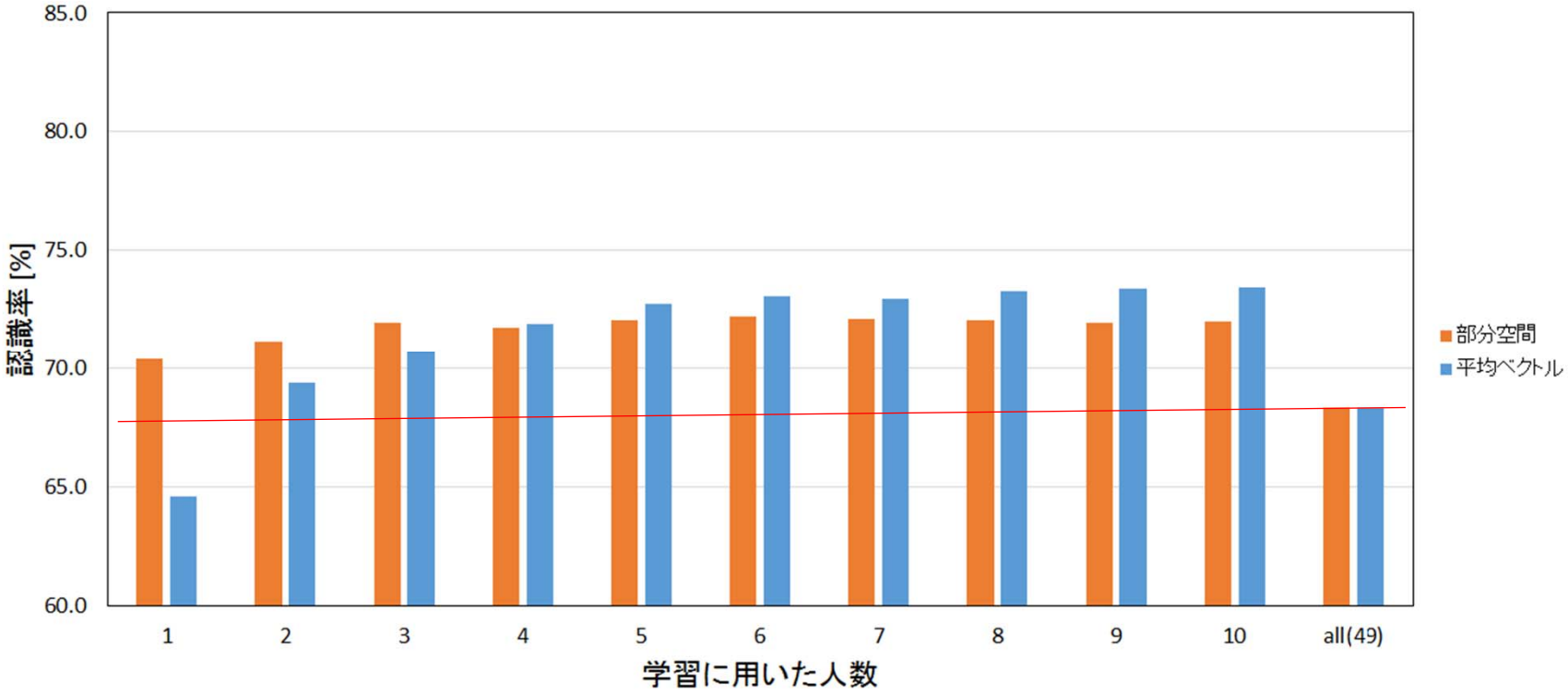


図　学習話者を選択することによる感情認識率の向上

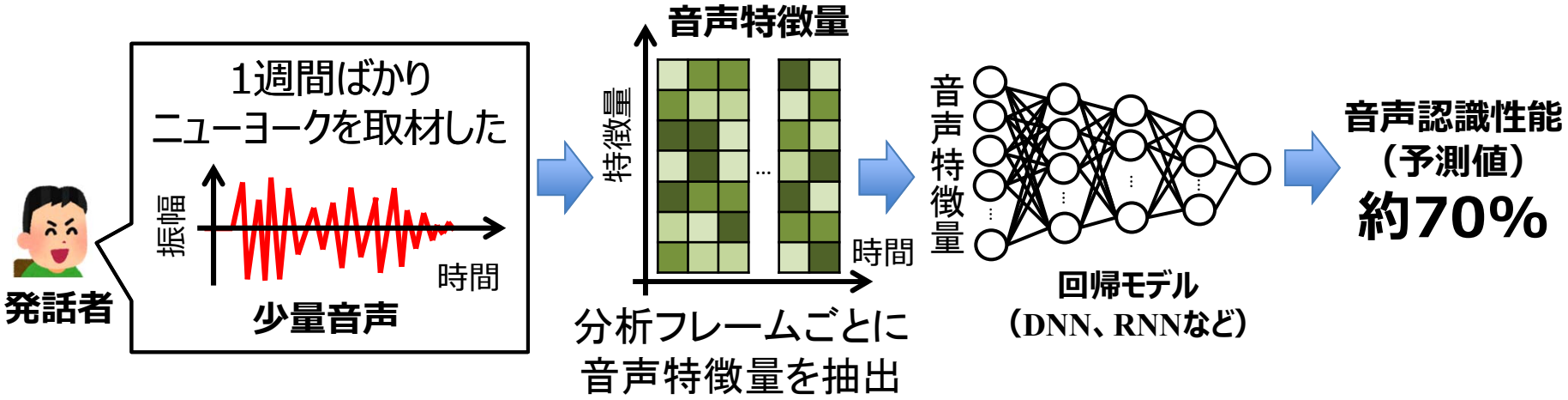
1. 音声認識性能の予測（簡易測定）

- 音声認識性能
  - 正しく認識された音声を定量的に評価する指標
  - 評価手順
    1. 利用環境で発話した**膨大な音声データを収集**
    2. 収集した音声データを音声認識システムに入力
    3. AccuracyやWERなどから音声認識性能を算出
  - 課題：**音声収集や認識実験に長い時間を要する**

研究目的：**少量の音声**で、発話者の音声認識性能を評価したい

1. 音声認識性能の予測（簡易測定）

- DNNやRNNを用いた音声認識性能の予測
  1. 利用環境で少量の音声を計測
  2. 計測音声から音声特徴量を抽出
  3. 音声特徴量をDNNに入力・伝播させて音声認識性能の予測値を算出

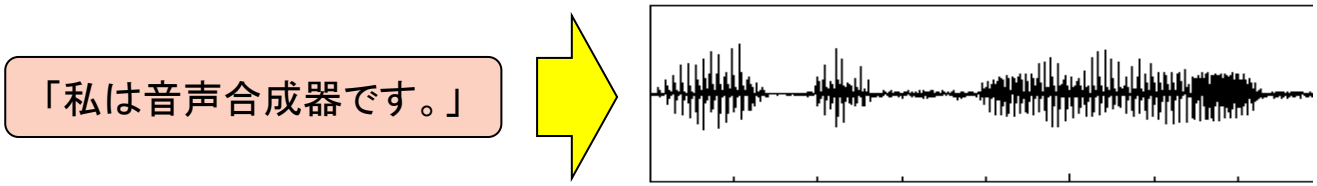


評価実験結果

残響環境において、約5%以内の誤差で音声認識性能を予測可能

1. 音声合成の現状と課題

- 音声合成に関して
  - 計算機による人間の声の生成



合成音声波形

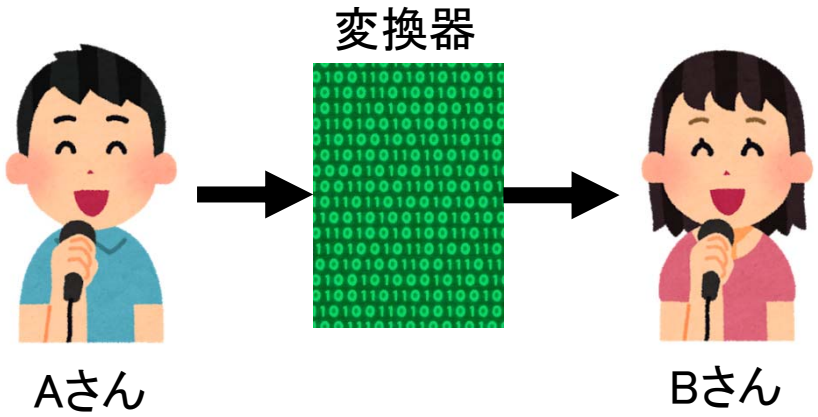
  - 音声合成研究の現状
    - 現在では，かなり高い品質

  - 今後の研究課題
    - 品質向上
      - 音質と自然性
    - 多様な声質での合成
      - 声質（話者性），感情，…

表現力豊かな合成音声を

1. 声質変換

- 話者の声質を異なる話者の声質に変換する技術

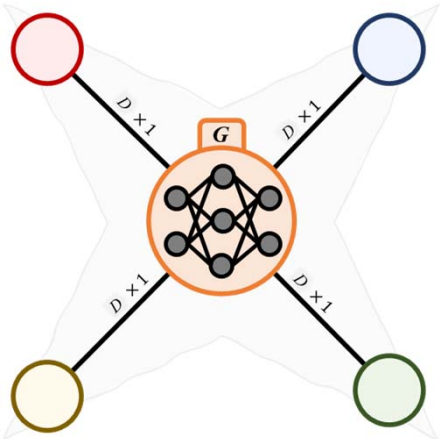


- 変換例

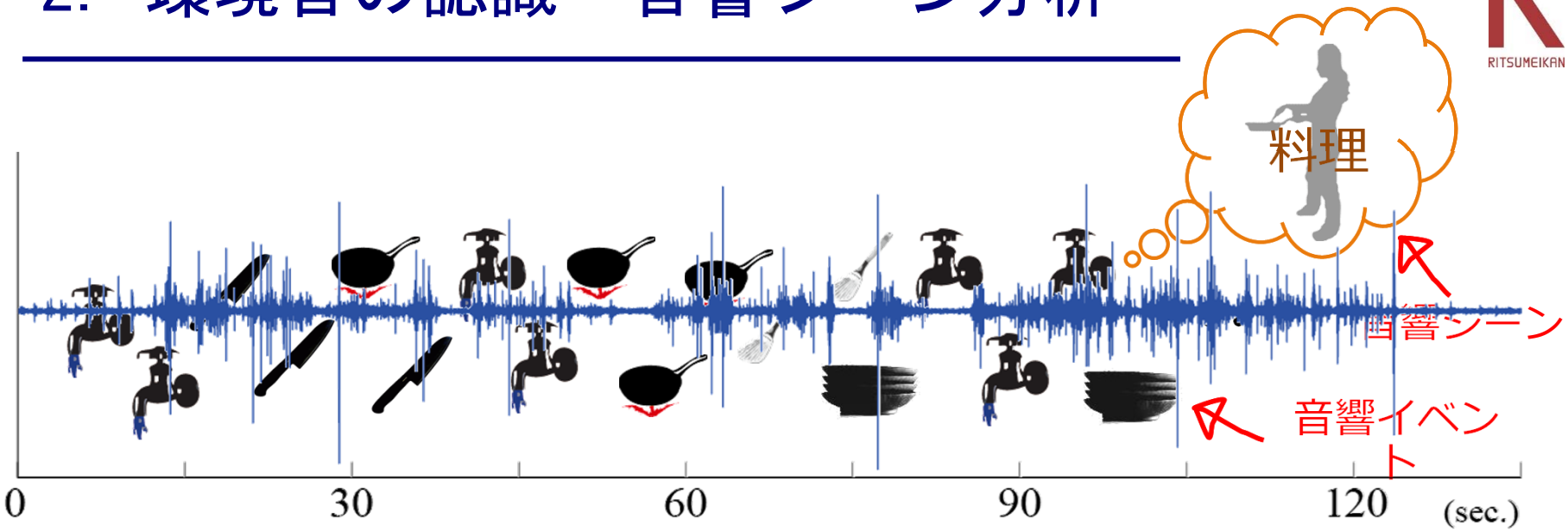
元話者	目標話者	変換音声

StarGAN

複数ドメイン間の変換を生成器一つで可能にするGAN



2. 環境音の認識 -音響シーン分析-



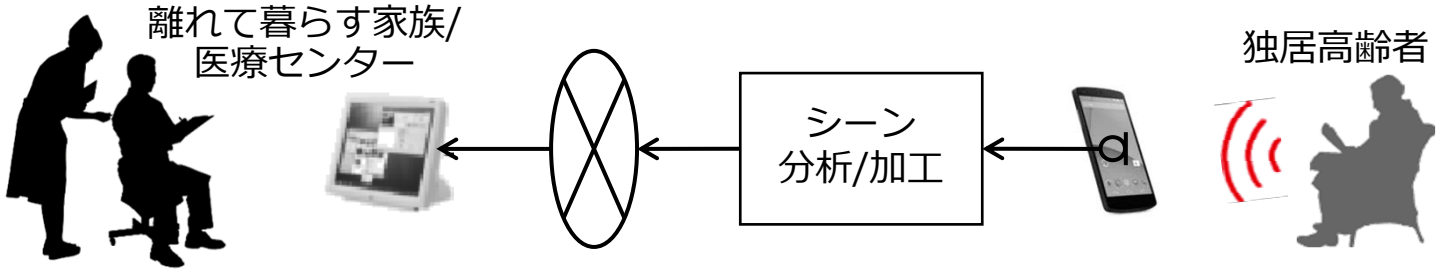
- どんな「シーン」でしょうか → 「料理」
  - 音響シーン：音響信号が収録された場所，状況，周囲にいるユーザの行動など
  - 音響イベント：音源の種類名(水の音, 包丁の音, 足音, 音声など)

音響シーン分析

音響信号から自動的に(計算機によって)シーンを推定する

2. 環境音認識結果の利用例

- 音響信号を利用した見守りやライフログ
  - 起床/睡眠などの自動推定を行い，異常時に遠隔地に通知

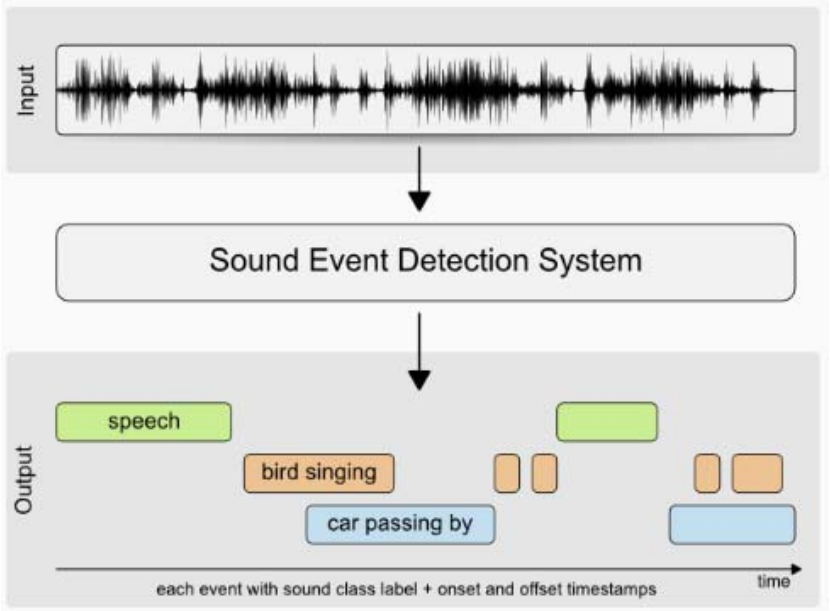


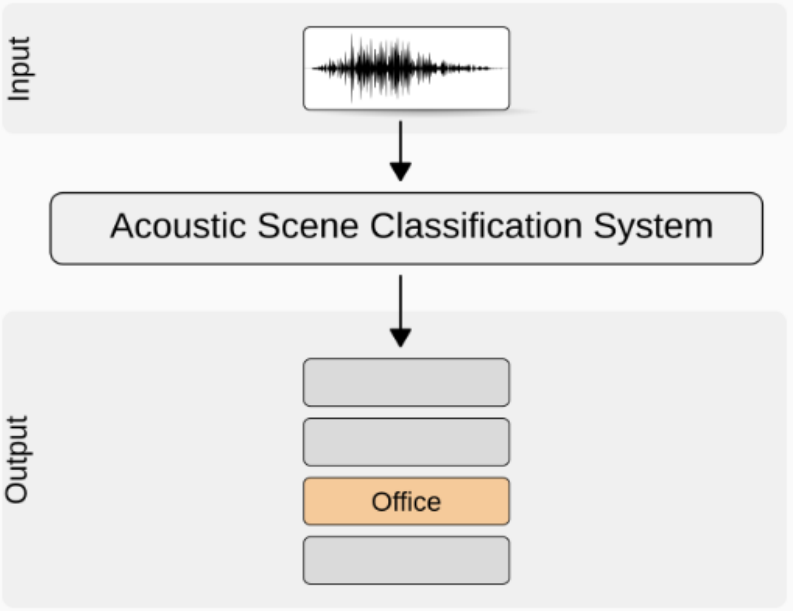
- 投稿型動画サイトやクラウドストレージのコンテンツ分析
  - YouTubeのコンテンツを自動解析/分類し，タグを自動で付与

- 聴覚障がい者支援システム
  - 屋内生活音（アラーム音，インターホン）を検出して振動などで通知

2. 環境音の認識

■ 音響イベント検出と音響シーン分類





フレーム単位で推定

① 音響イベントラベル
② タイムスタンプ(onset + offset)

音クリップ単位で推定

① 音響シーンラベル

2. 音響イベント検出

- マルチタスク学習を用いた環境音認識

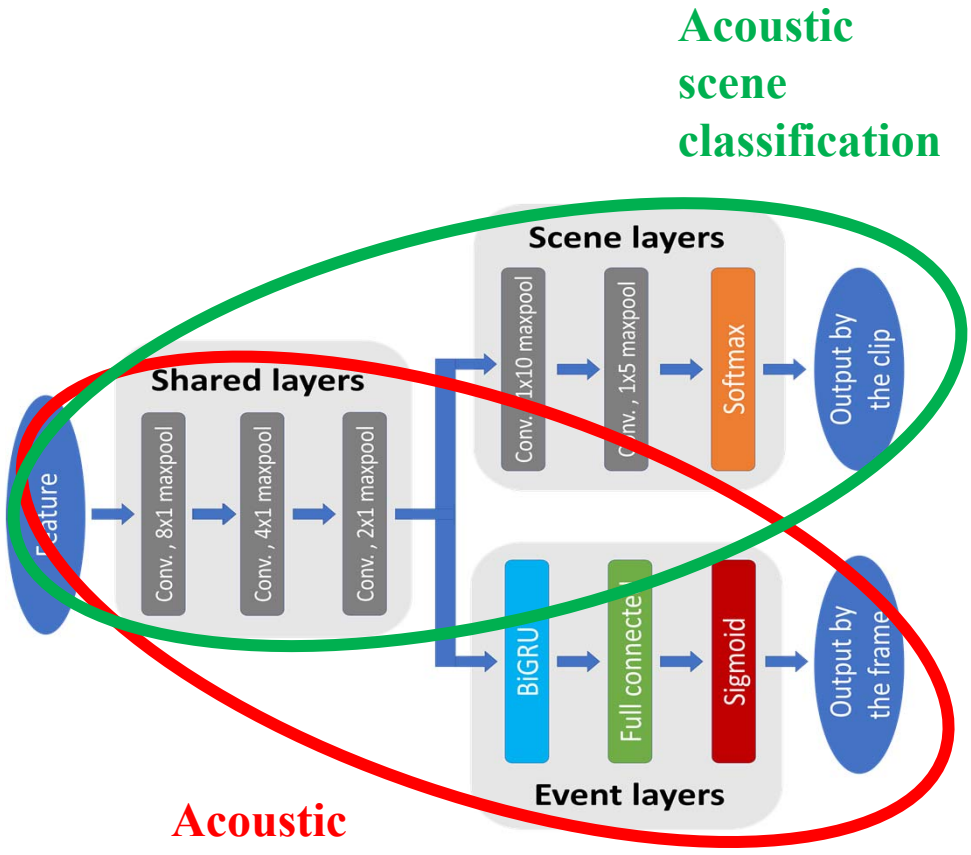
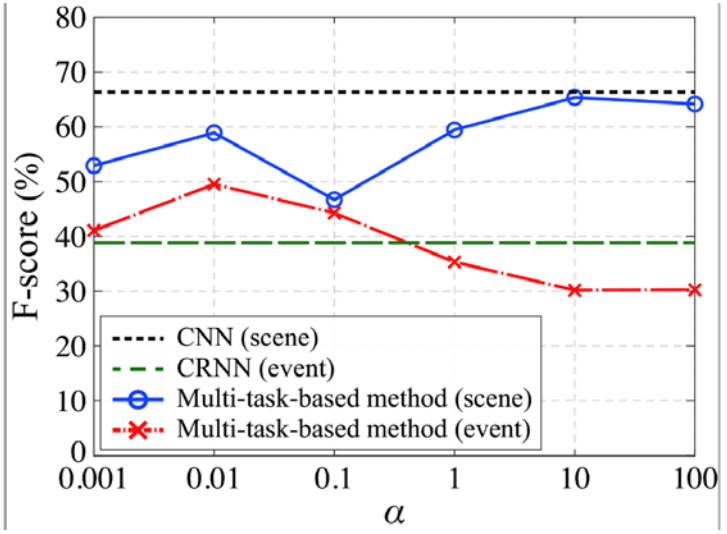


Table 2　音響イベント検出性能とシーン分類性能

Method	Event		Scene
	F-score	ER	F-score
CRNN (event)	38.90%	0.776	-
CNN (scene)	-	-	**66.36%**
Multi-task (α=0.1)	44.31%	0.721	46.72%
Multi-task (α=0.01)	**49.56%**	**0.695**	58.94%
Multi-task (α=0.001)	41.11%	0.760	52.93%



3. 音情報を使った異常検出

■ 危機的状況を示す要素である叫び声を検出

マイクロホンに入力された音声の叫び声らしさを計算

読売新聞 朝刊 2018.03.30

人間の発声メカニズム

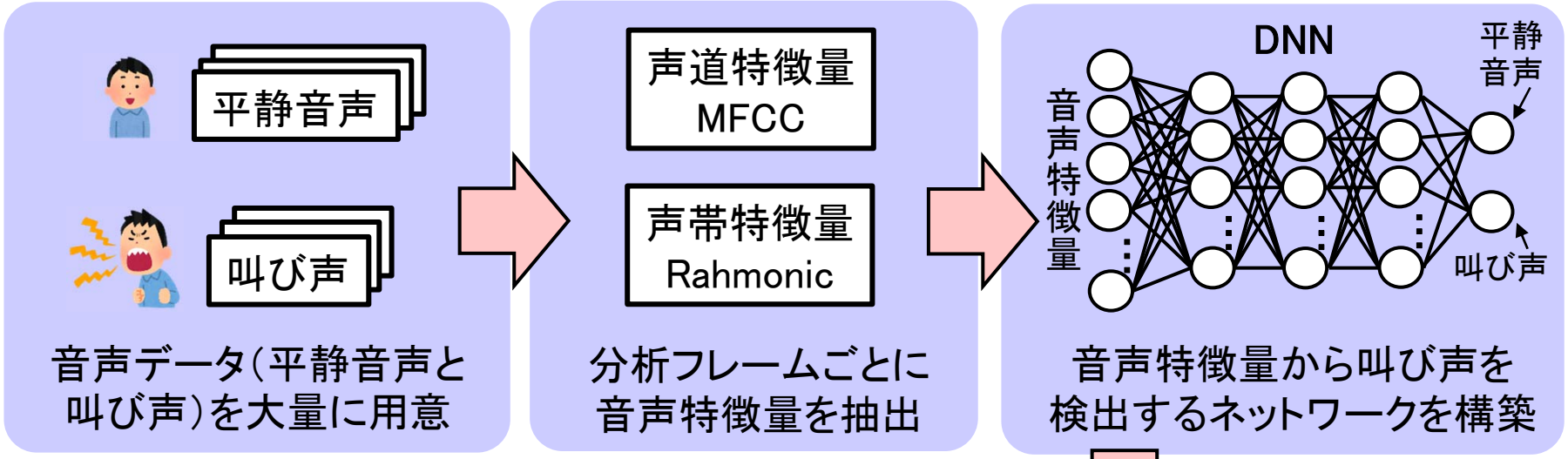
喉元の **声帯** で発生させた音を **声道**(喉や口など) に通過させて様々な言葉を発声

「**通常音声と叫び声で、声帯と声道の動きが大きく異なること**」を解明し、これらの運動量と叫び声らしさの関係をDeep Learningを用いて学習

3. 叫び声検出

① 叫び声検出モデルの構築

※ MFCC：聴覚特性を考慮したケプストラム係数
※ Rahmonic ：基本周波数（F0）の低調波成分



② 叫び声の検出

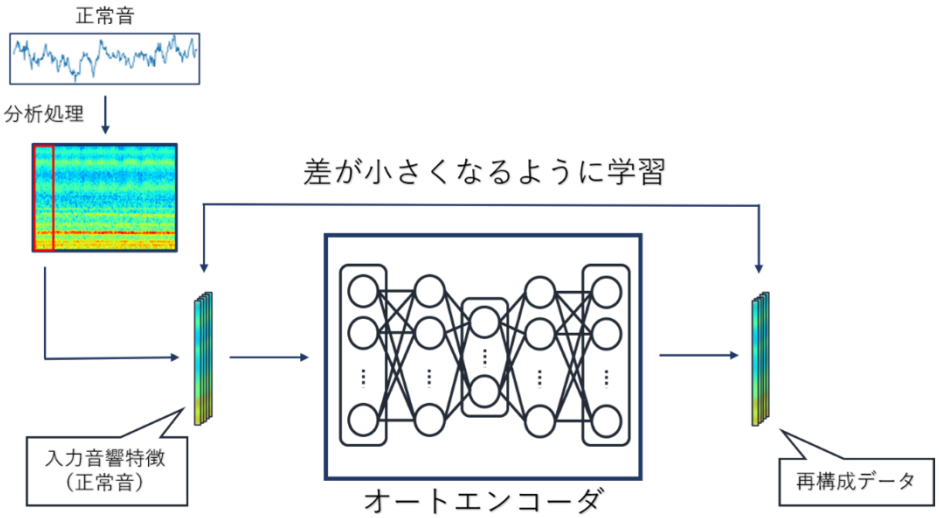
評価実験結果

多くの人が集まるような騒々しい環境で95%以上の精度で検知

3. 音情報を使った異常検出

オートエンコーダを使った異常検出

- 正常なデータだけで学習



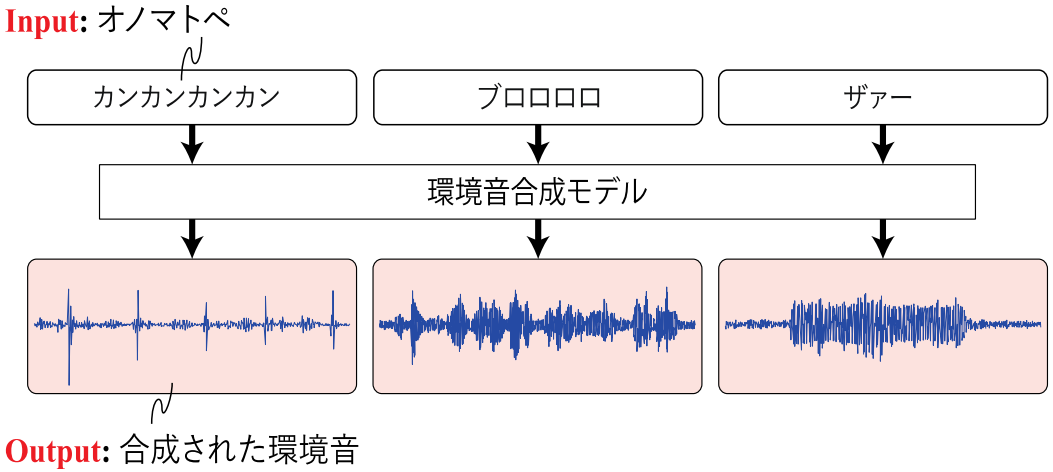
- 6種類の機械音，既存のデータセット

表： 機械の種類ごとの異常検出率

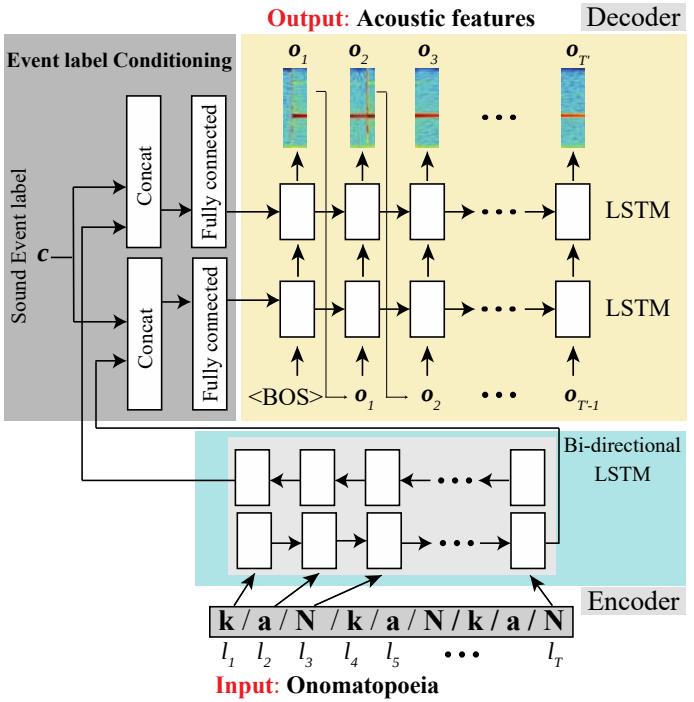
	DXF#(‘					
	Wr|Fdu	Wr|Frqyh|ru	Idq	Sxps	Vdg#dlb	Ydyh
従来手法（データ拡張なし） DH	:<B:	:41:;	981;3	:5B4	;71:4	99B3
提案手法（データ拡張あり） DH#1#GD	:;153	:5B4	989<	:5B:	;7B6	977;
DH#2#GD	::1<:	:41;5	9814:	:59<	;8B8	98B5

3. 環境音の合成

- オノマトペを入力とする音合成
  - オノマトペ：音の特徴を自然言語を使用して表現したもの
    - 例：カンカンカン



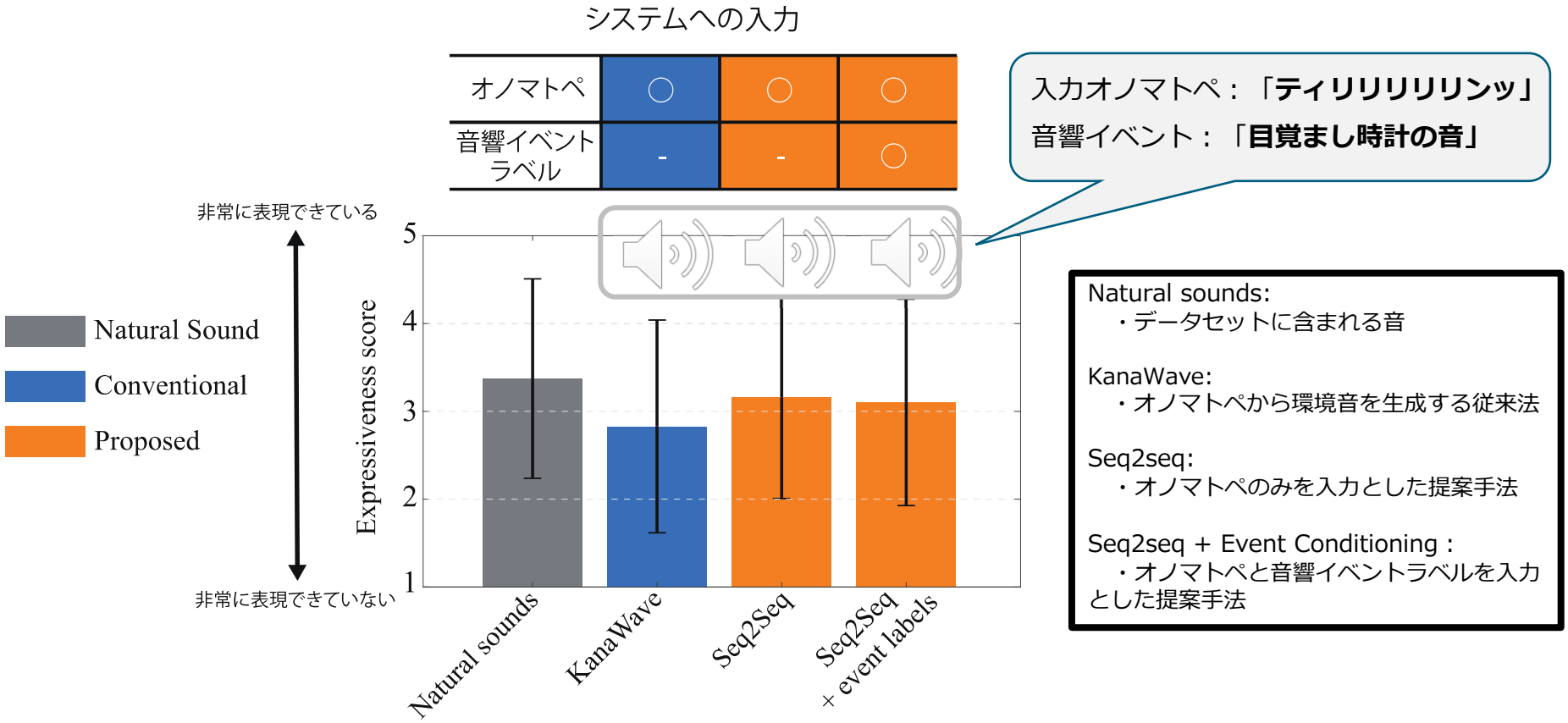
図：オノマトペを入力とする環境音合成



図：環境音合成モデル

3. 環境音の合成

■ 表現性の評価結果



図：オノマトペに対する環境音の表現性の平均スコアと標準偏差

4. 音楽情報処理

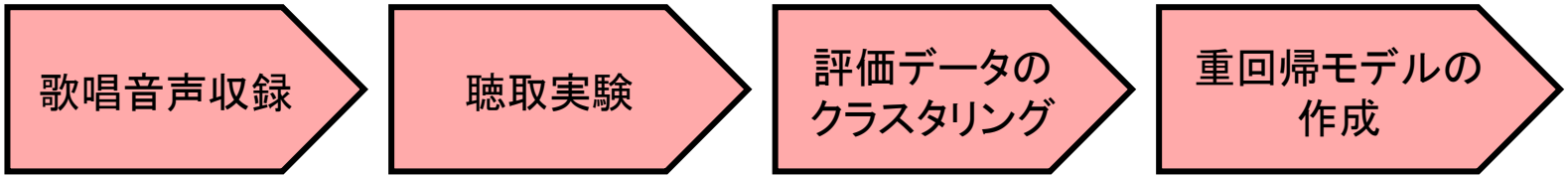
- 効果音・BGMの自動付与
  - 効果音の合成
  - 自動作曲
- 音源分離
- 楽曲のジャンル推定
- 楽曲の印象評定
  - 国による違い
- 楽曲評価の分析

4. 音楽情報処理

- 歌唱データの収録と分析
  - アカペラサークルの協力
    - 男10人，女7人
  - 歌詞での歌唱，スキャット（Lala･･･）での歌唱
  - 音韻バランス文の読み上げ
  - 楽曲選定基準
    - 男女が同じキーで歌える
    - 楽曲の印象に左右されない
    - 歌唱者が歌いやすい
  - 声の印象評価，歌声合成，，，

4. 歌唱評価の分析

- 主観における歌唱力評価を行う
  - 歌唱力評価データの収集（クラウドソーシング）
    - 評価者： 281名
  - 評価データのクラスタリング
    - 評価基準の個人性の分析
- 音響特徴量との関係の分析
  - 各クラスタの評価データをスコア化
  - 各クラスタのスコアを音響特徴量から予測する重回帰モデルの作成



4. 歌唱評価の分析

- クラスタの分析例
  - クラスタに特徴的なパラメータ

特徴量名	RMSenergy	Brightness	Inharmonicity	Pulseclarity	MFCC(1次元)	MFCC(2次元)
偏回帰係数	0.76	1.23	−0.10	0.01	−0.06	0.73

  - クラスタの特徴
    - 発声や発音が明瞭な歌唱者が上位
    - 音量と音色に関する特徴量の寄与率が高い

研究のすすめかた

- 研究スタイル
  - プログラミング（Python, MATLAB, C, Perl, …）
  - ツールの利用
    - 音声認識(julius), MATLAB, WEKA, , ,
  - パターン認識, 信号処理, , ,
  - 統計的手法
    - データの収録・収集・利用
    - DNN, HMM, 主成分分析, クラスタリング, MDS(多次元尺度構成法), SVM, 決定木, , ,
  - 聴取実験
    - クラウドソーシングの利用
  - 手法の提案とシステムの開発

RITSUMEIKAN

作品制作のヒント

作品

(1) ソフトウェア

- 教員の講義の中で特に興味を持った分野における「ソフトウェア」

(2) 調査研究

- 教員の講義の中で特に興味を持った分野をさらに掘り下げ，同級生たちにその面白さを伝えるための「調査研究」

(3) 未来創造

- 教員の講義の中で示された今後の課題や，自分が重要であると考える情報社会の諸問題の解決策，さらには情報技術を使った望ましい社会のあり方などについて提言を行う「未来創造」

作品制作のヒント

(1) ソフトウェア

- 音声認識・音声合成を利用したアプリケーションの作成
  - Web Speech APIの利用 (HTML5で)
- WORLD (フリーの分析合成ツール) の利用
  - http://www.kki.yamanashi.ac.jp/~mmorise/world/

分析合成:
音声を特徴パラメータに変換し，それを音声に合成する処理

(2) 調査研究

- 音声情報処理技術の実用化について
  - カーナビ，ゲーム，個人認証，スマートホンでのサービス，，，
  - 音を使った通信技術
    - 「おもてなしガイド」など
- 音データの圧縮方式
  - 方式の比較，なぜ圧縮できるのか

(3) 未来創造

- 音声認識の新しい使い道は？
  - より便利に or より楽しく