

```
emacs-25.2
File Edit Options Buffers Tools Python Help
import KaldiReceiver
import json
import subprocess
import time

import os
import os.path

class KaldiDecoderDetector:
    def __init__(self, kaldiport=10500, options=["--c\
onfig=kaldi_conf/online.conf"], path=""):
        self.options = list(options)
        self.runKaldi(path, jconfs=self.options)
        time.sleep(10)
        self.proxy = KaldiReceiver.KaldiReceiver(port\
=kaldiport)

    def runKaldi(self, path, jconfs):
        cmd = "{}kaldidecoder {}".format(path, ".jo\
in(jconfs))

--UU-:----F1 Kaldi.py Top (1,0) (Python) -----
```

```
itonoselab@eiwww:~ (emacs-25.2)
File Edit Options Buffers Tools Text He
<NOISE> <NOISE> nsn
<SPOKEN_NOISE> <SPOKEN_NOISE> spn
<UNK> <UNK> spn
<eps> <eps> sil
ア ア a
アー アー a:
アーサー アーサー a: s a:
アーチ アーチ a: ch i
アーチェリー アーチェリー a: ch e r \
i:
アーティスト アーティスト a: ch i s u t \
o
アート アート a: t o
アマド アマド a: m a d o
アール アール a: r u
アールエヌエイ アールエヌエイ a: r \
u e n u e i
アールビーエフ アールビーエフ a: r \
u p i: e f u
アールブイ アールブイ a: r u b u i
--UUU:----F1 align_lexicon.txt Top (1
```



# 伊藤・能勢研究室 研究室紹介



# 研究室の構成

- 教授 伊藤彰則

- 音声認識、音声対話システム、  
音楽情報処理、外国語教育システム

- 准教授 能勢隆

- 音声合成、声質変換、歌声合成、顔動画像生成

- 学生・その他

- 秘書1名
- 学術研究員1名
- 修士課程12名、学部5名

# 居室の様子

## ● 居室

- 542号室：小ゼミ室、防音室、お茶室、准教授室
- 520号室：学生部屋
- 515号室：教授室



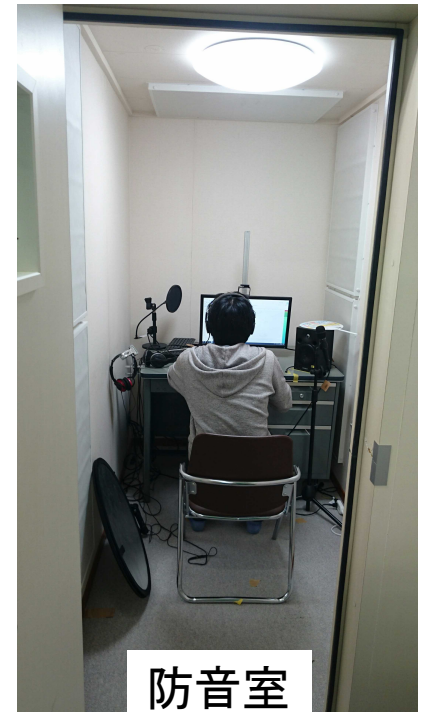
お茶室



小ゼミ室



学生部屋



防音室

# 研究分野

## 人と機械のコミュニケーション

信号処理、統計モデル、機械学習(人工知能、深層学習)

### 音声・言語

音声認識  
音声合成  
音声対話システム  
声質変換

### 音楽・歌声

歌声合成  
自動作曲  
作詞補助  
再生速度変換

### 動画・画像

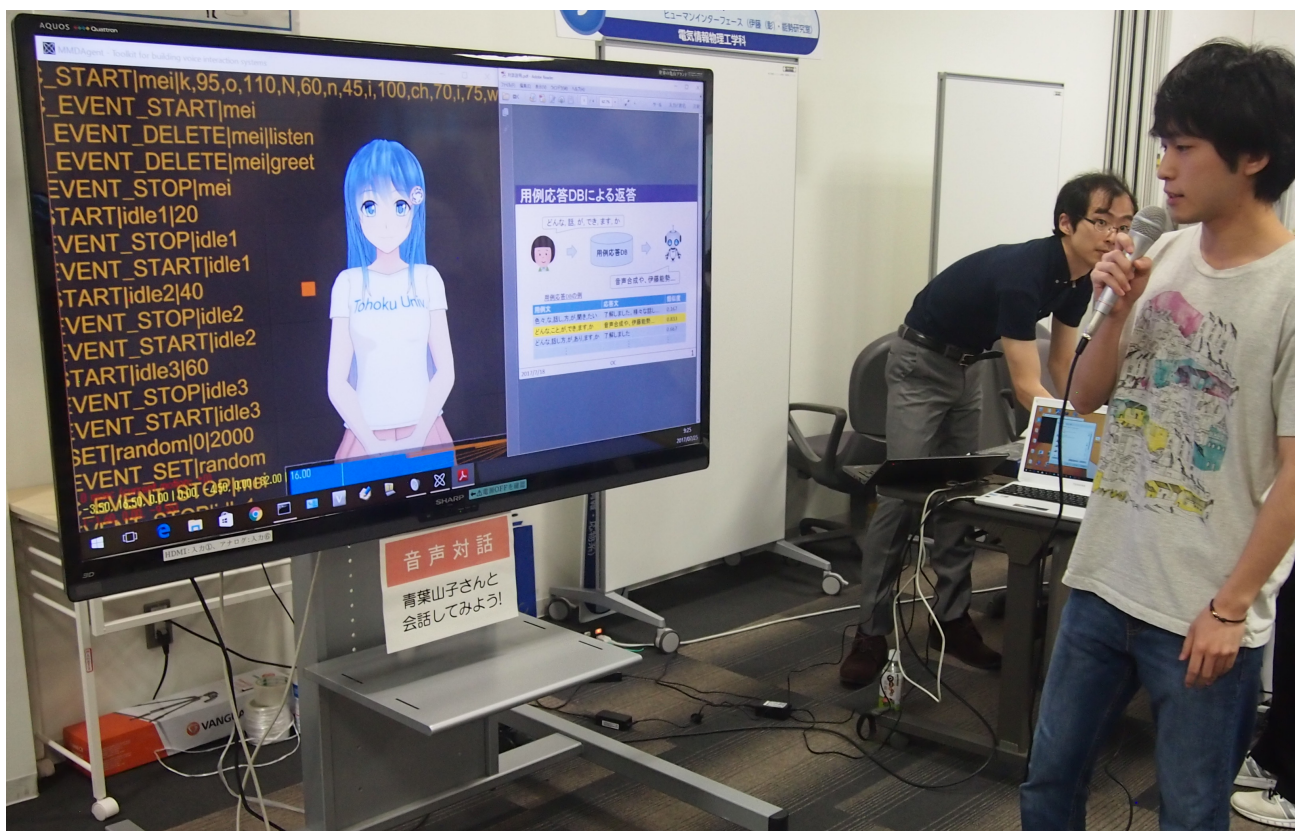
マルチモーダル対話  
顔動画像合成・変換

### 学習支援

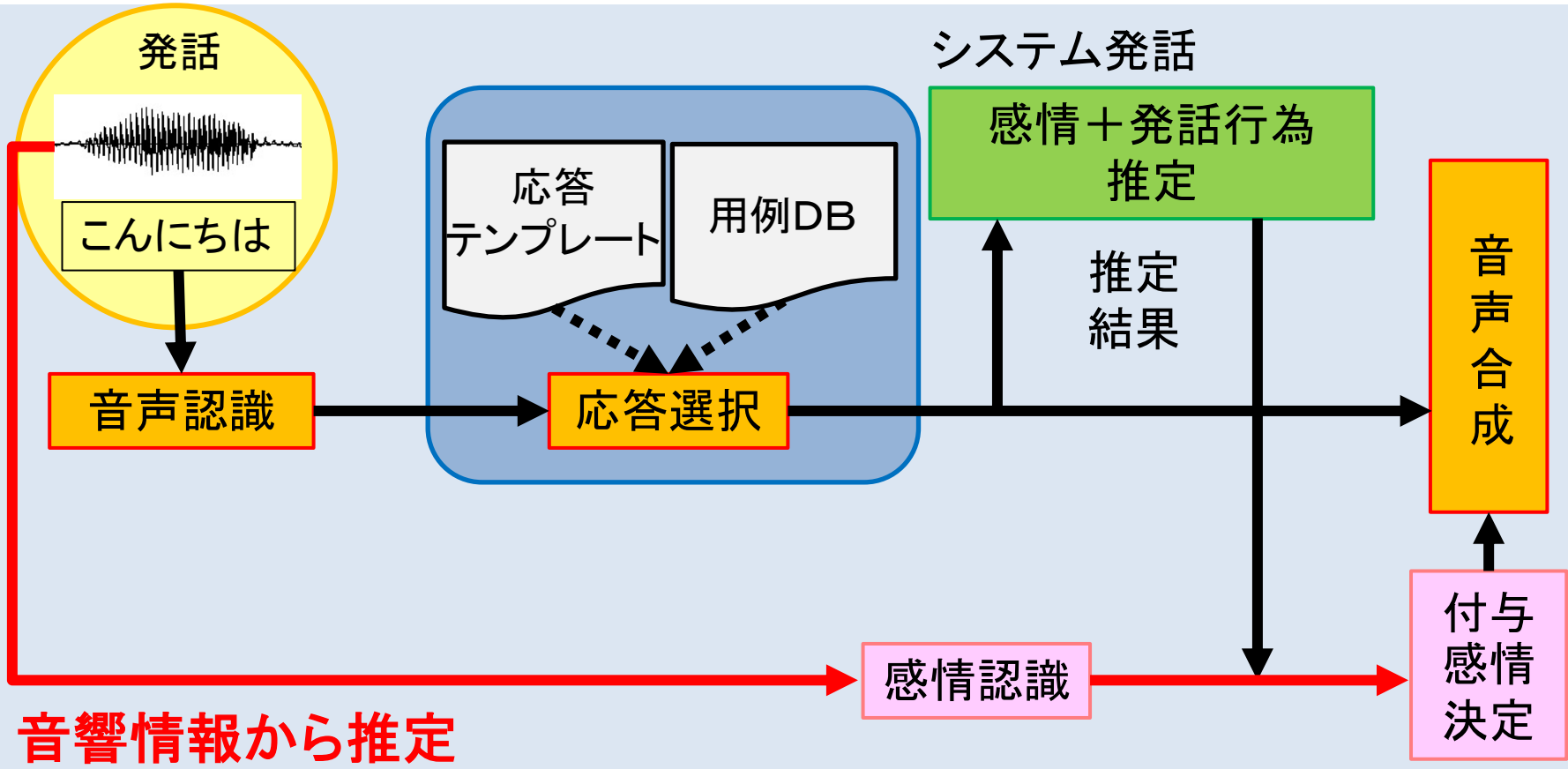
英語発音評価  
文法誤り検出  
やさしい日本語

# より人間らしい音声対話システム

キーボードやマウスなどに頼らずに、対話を用いた、より人間らしい  
**コンピュータとの自然な対話システム**の研究・開発を目指す。  
研究・開発されたシステムなどは、**オフィスロボット**や**介護ロボット**に  
応用される。

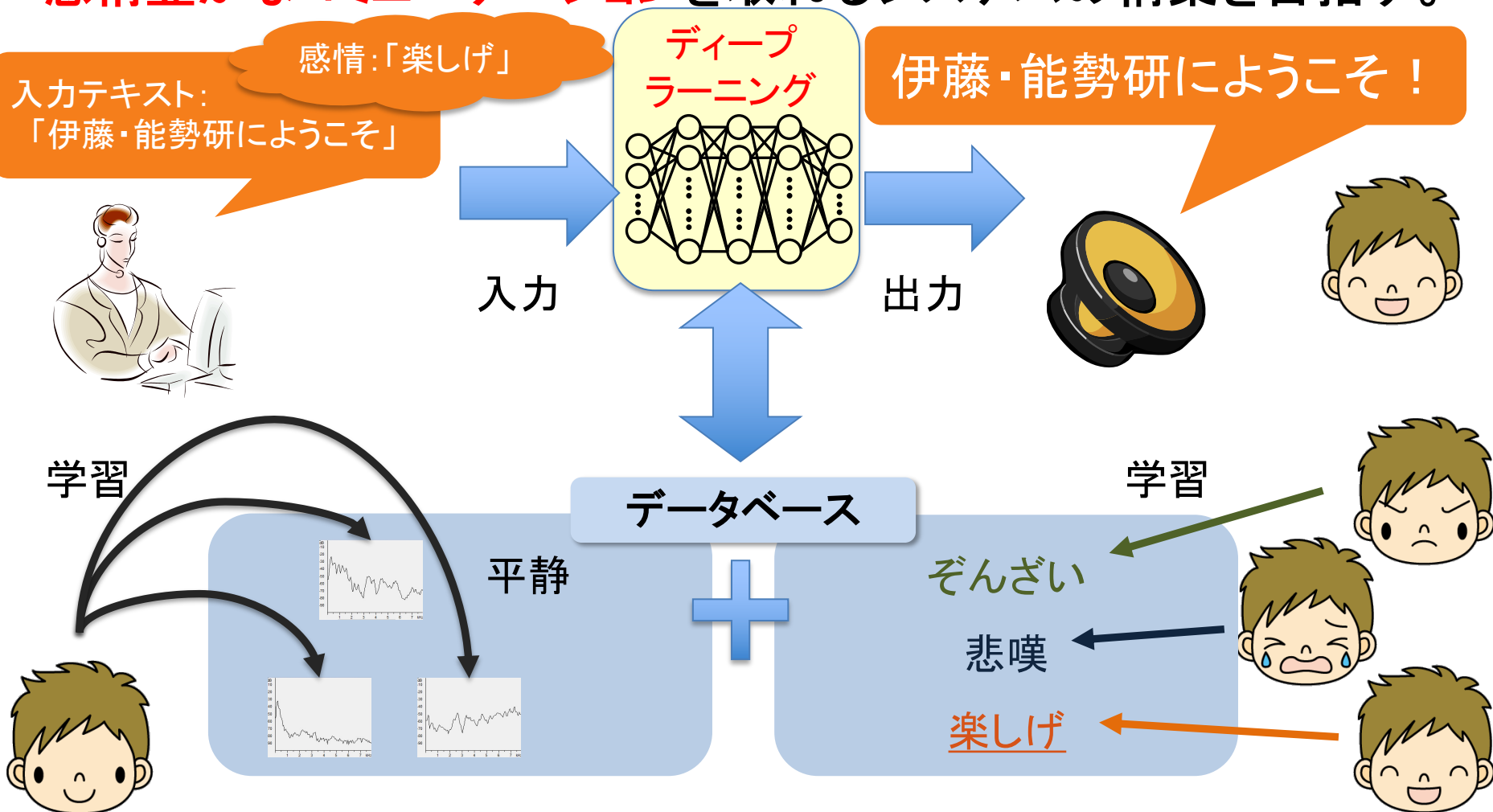


# 感情を考慮した音声対話システム



# 音声合成における感情の付与

喜び、驚き、悲しみなどの声に現れる感情を表現し、人間と感情豊かなコミュニケーションを取れるシステムの構築を目指す。



# デモ:スタイルの連続的な制御

ぞんざい



悲嘆

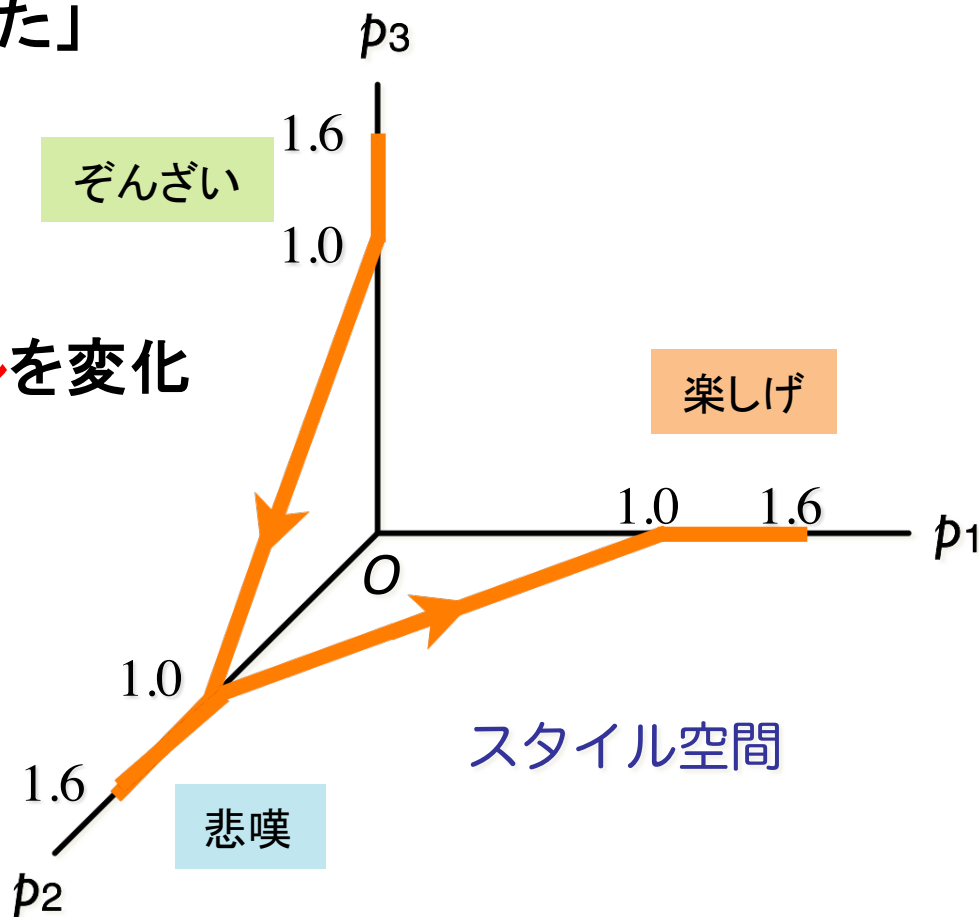


楽しげ



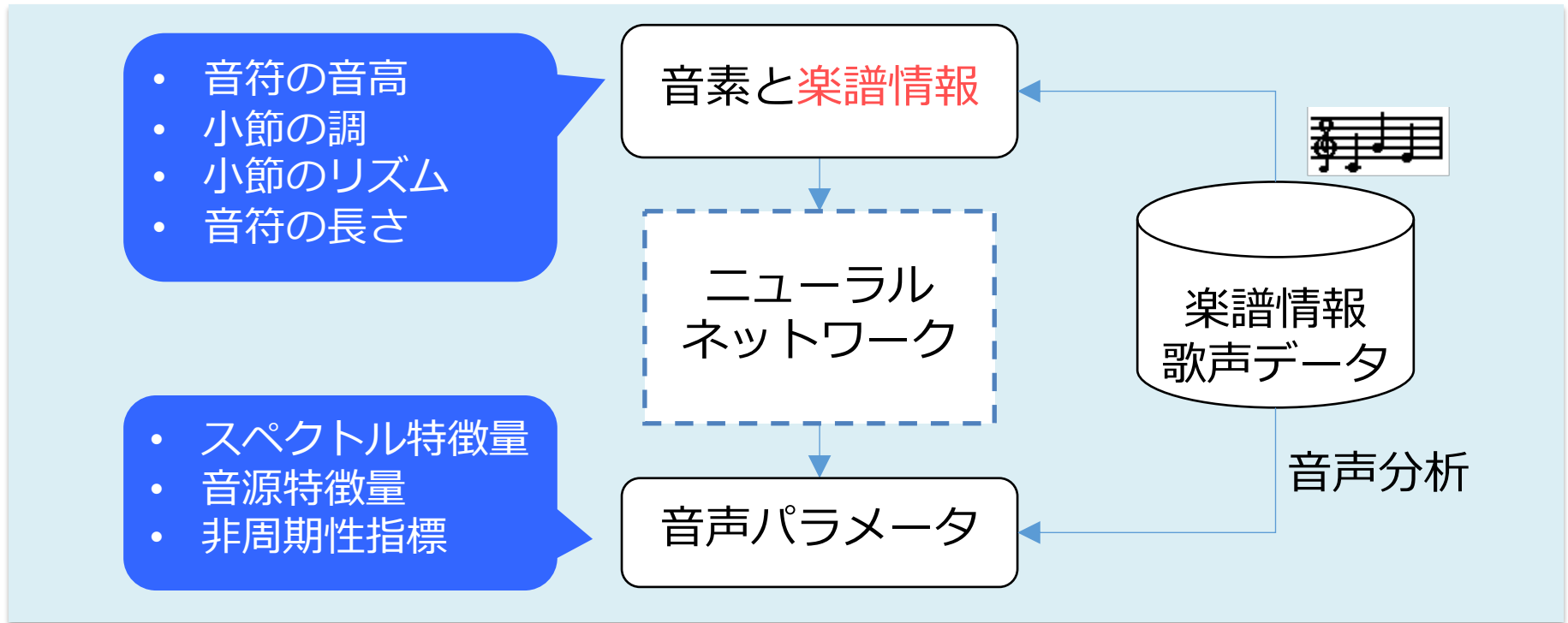
「実習中は雪が降っていたけれど、ほっぺの痛さももう味わえなくなるかと思うと、全然平気だった」

音素単位でスタイルベクトルを変化





# 深層学習に基づく歌声合成

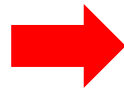


初音ミク (Vocaloid)



声が機械音っぽく聴こえる

深層学習ベース



とても自然な歌声に！

# 深層学習に基づく声質変換

音声合成

「こんにちは」

ニューラル  
ネットワーク

こんにちは



声質変換



こんにちは

ニューラル  
ネットワーク



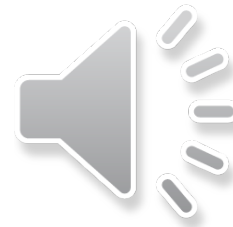
こんにちは



変換前



目標話者音声



変換後

# 顔動画生成

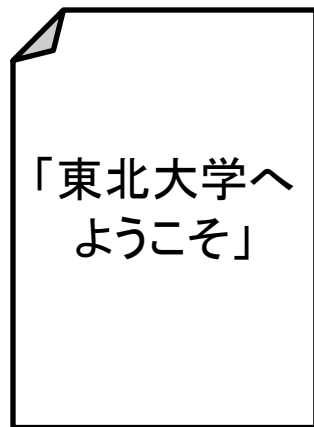


目標話者

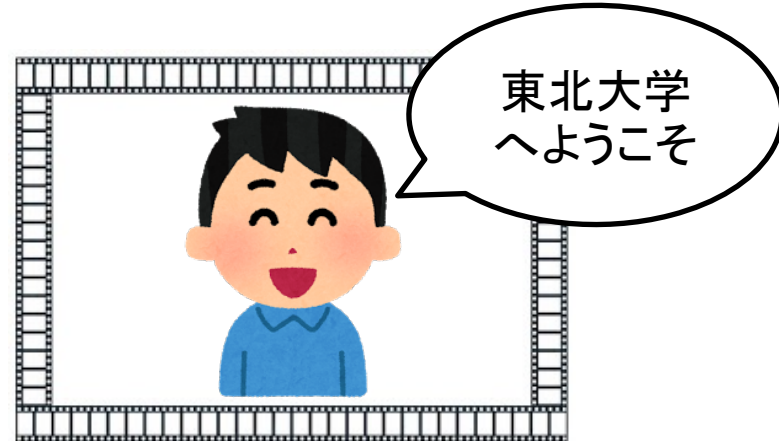


コンピュータ

目標話者の喋り方(顔)をコンピュータが自動で学習!



テキストに合った顔を生成



# デモ：顔動画生成



オリジナル

テキストから生成

# デモ：顔動画生成（まばたきの制御）



閉じたまま



ずっとまばたき

# 外国語教育システム

外国語の学習では、自分一人で発音を練習するのは難しい。  
そこで、音声認識技術を利用して**日本人の外国語学習を**  
**サポートするシステム**を作成。



We can change!

（  
・イントネーション  
・リズム  
・アクセント etc...  
）  
入力音声のパラメータ化

見本のデータと  
マッチングを  
とって比較する。

教師データと比較

[change]ノ [é]ノ ハツオンガ  
[a]ニ ナツテマス。  
キミノ ハツオンヲ チェンジ！

フィードバック



# 音響情報を用いた韻律評価

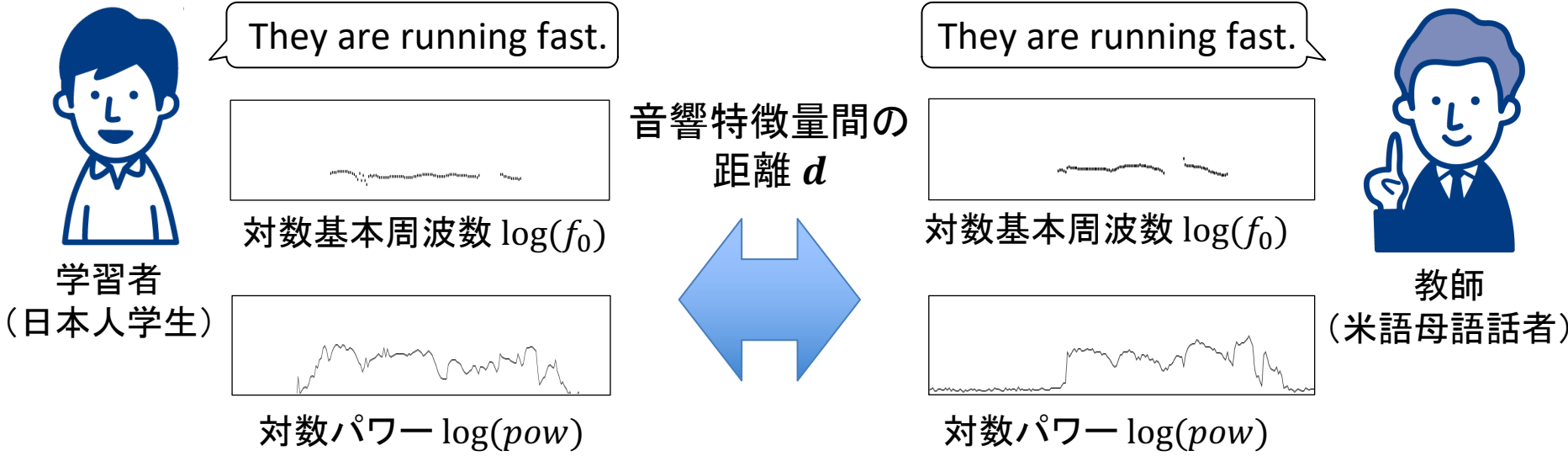
## ◆ イントネーション(Intonation)

- 発話における声の高さ(ピッチ)の変化の様相
  - **ピッチ**: 人間が聴覚的に感じる音の高さ(基本周波数 $f_0$ に相当)



Cauliflower, broccoli, cabbage, sprouts, and onions.

## ◆ 学習音声と教師(米語母語話者)音声の音響情報を比較



音響特徴量間の距離が短い(教師音声に似ている)ほどスコアが良いと判断

# 研究室の研究環境

- 計算サーバ

- Ubuntuマシン：22台

- GPUボード

- RTX 3090を中心に42枚  
→現在拡充中（昨年比+16）

- 学生端末

- WinまたはMacのノートPCを配布
- 研究室外からもVPNによりアクセス可

- 防音室x 2

- 評価実験、対話実験などに利用

host(gpu_id)	gpu name	load	process	memory(used)	memory(max)
mitsuhide(0)	NVIDIA_GeForce_RTX_3090	99	1	18243	24268
grimm(0)	NVIDIA_GeForce_GTX_1080_Ti	98	1	8109	11178
grimm(1)	NVIDIA_GeForce_GTX_1080_Ti	98	1	8151	11178
grimm(2)	NVIDIA_GeForce_GTX_1080_Ti	97	1	7937	11178
yukimura(0)	NVIDIA_GeForce_RTX_3090	78	1	8987	24268
elgar(0)	NVIDIA_GeForce_RTX_3090	37	1	8141	24268
ryoma(0)	NVIDIA_GeForce_RTX_3090	34	1	17197	24268
alexander(0)	NVIDIA_Quadro_RTX_6000	4	0	0	24219
gandhi(0)	NVIDIA_GeForce_GTX_1050_Ti	2	0	0	4040
masamune(0)	NVIDIA_GeForce_RTX_3090	0	1	12329	24268
shingen(0)	NVIDIA_GeForce_RTX_3090	0	1	23907	24268
toshiie(0)	NVIDIA_GeForce_RTX_3090	0	1	12369	24268
newton(1)	NVIDIA_GeForce_RTX_2080_Ti	0	1	8	11019
newton(0)	NVIDIA_GeForce_GTX_1080	0	1	18	8118



# 研究室生活の1日



水落智 修士2年当時

研究内容: 絶滅危惧言語の音声検索

趣味 : アニメ、ゲーム、読書

ある1日の行動例

研究室の学生と一緒に食堂に行ったり。  
時間がないときはお茶室で。

9:00 研究室着  
ゼミ発表準備

10:00~12:00  
ゼミ

昼食

13:00~  
研究

17:00 帰宅  
(いつのまにか  
20:00とかも?)

資料とプレゼンテーションを準備  
ゼミでは進捗を全員の前で発表

現在は、DNNなどの機械学習を用いて音声から特徴量を抽出するシステムを作成中。同時並行で、10月に行われる学会の発表準備も。

現在、研究は家で、ゼミやミーティングはオンラインで実施

# 年間スケジュール

4月 お花見

5月 新入生勉強会

6月 スポーツ大会

8月 院試

9月 研究スタート

10月 芋煮

11月 駅伝大会

年末年始 忘年会・新年会

3月 卒論発表 & 温泉旅行



伊藤・能勢研究室では…

**機械学習**や**信号処理**を利用した

**音声・音楽・画像情報処理**に興味のある人  
(プログラミング経験がなくてもOK!)

をお待ちしています!



アニメ・ゲーム・音楽・  
アイドル・スポーツ好きも大歓迎!!



興味を持たれた方は、ぜひ一度見学にお越しください。  
担当者：修士2年 内城沙月 [satsuki.naijo.t2@dc.tohoku.ac.jp](mailto:satsuki.naijo.t2@dc.tohoku.ac.jp)

# よくある質問

Q. コアタイムはありますか？

A. オンラインのためコアタイムはありません。  
しかし**1日8時間**が目安です。

Q. プログラム経験は必要ですか？

A. 新入生勉強会で**基礎から勉強**します。  
また、教員や先輩方も**適宜フォロー**します。

Q. 院試勉強休みはありますか？

A. **院試1ヶ月前から**、勉強に集中してもらいます。

Q. 海外で発表の機会はありますか？

A. **もちろん、あります！**