

```
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



研究室ではこんなことをやっています！

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

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

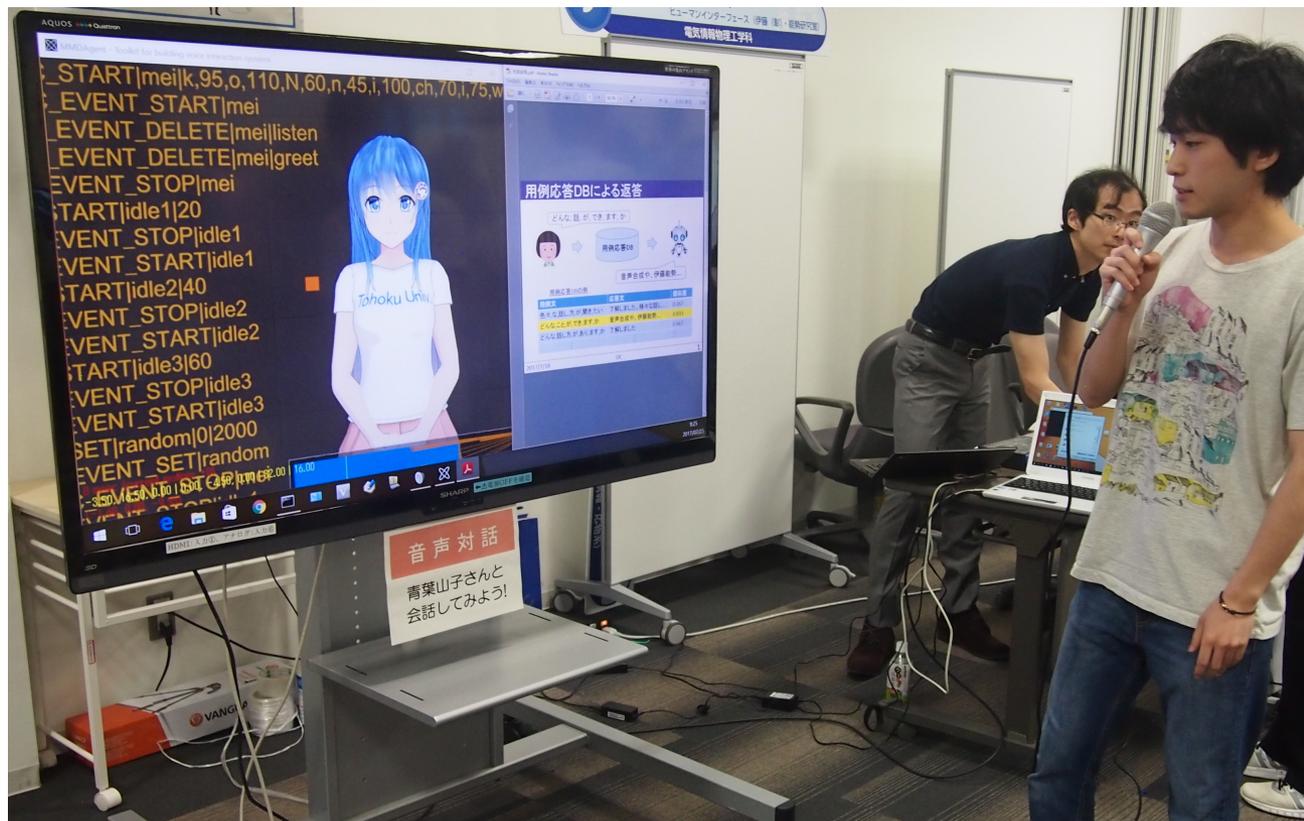
● 学生・その他

- 秘書1名
- 学術研究員1名
- 博士課程3名、修士課程14名、学部5名

対話・認識チームの研究紹介

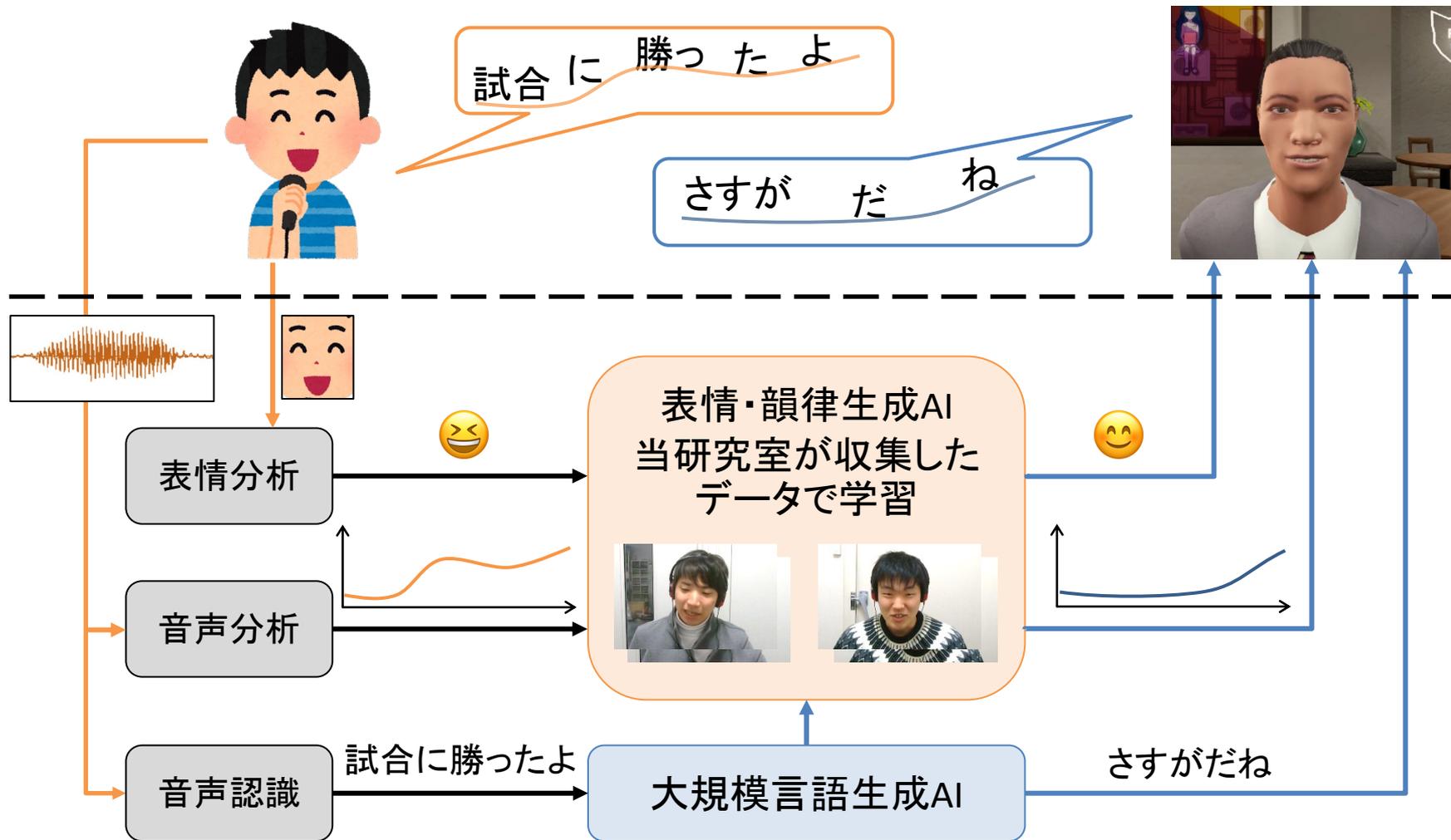
より人間らしい音声対話システム

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。



マルチモーダル対話システム

人間らしい表情・韻律(抑揚)のやり取りを伴う対話システム ➡ カウンセラーAI等への応用を期待



感情認識



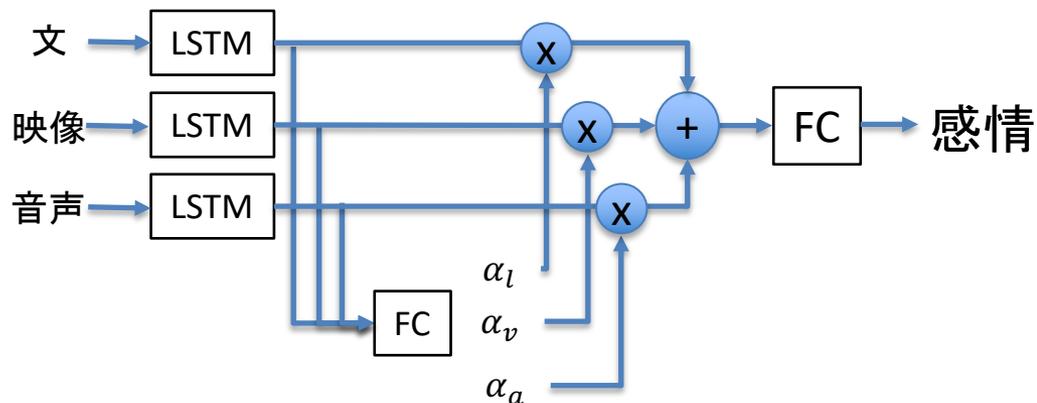
従来

経験則的な特徴量で予測

- えくぼができたなら
- 声が大きくなったら
- 「嬉しい」と言っていたら

現在

ニューラルネットワークを活用して高精度に！
6感情(怒り, 悲しみ, 幸福, 驚き, 嫌悪, 恐怖)で
78%予測 [Ando+, '22]

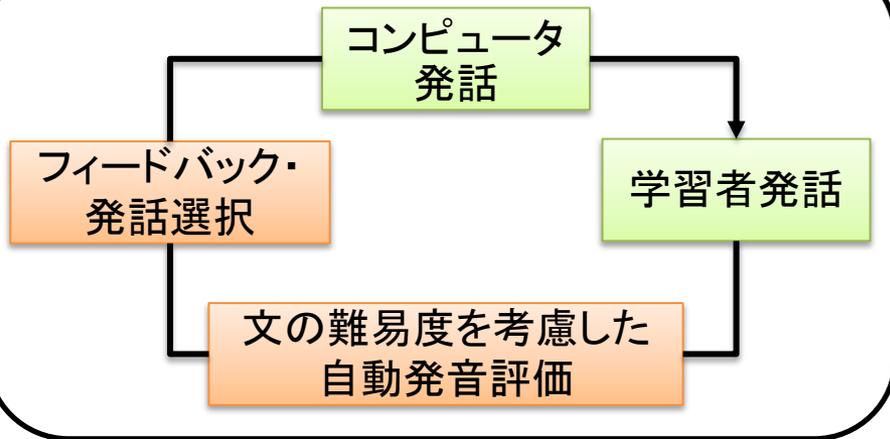


顧客満足度推定や対話システムなど**応用先が豊富**

学習支援システム

● 語学

発音学習



シャドーイング

見本音声



各地に劇団を支える応援団が...

各地に根北を支える応援団を...

劇団

が...

学習者



● オンデマンド講義



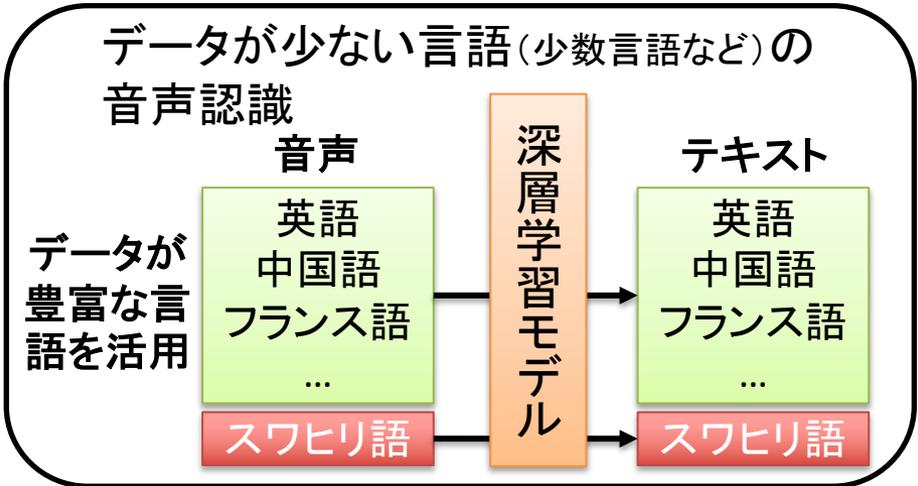
今のところ大事やったな！
ちゃんと聞いてたか？

バーチャル・クラスメイト

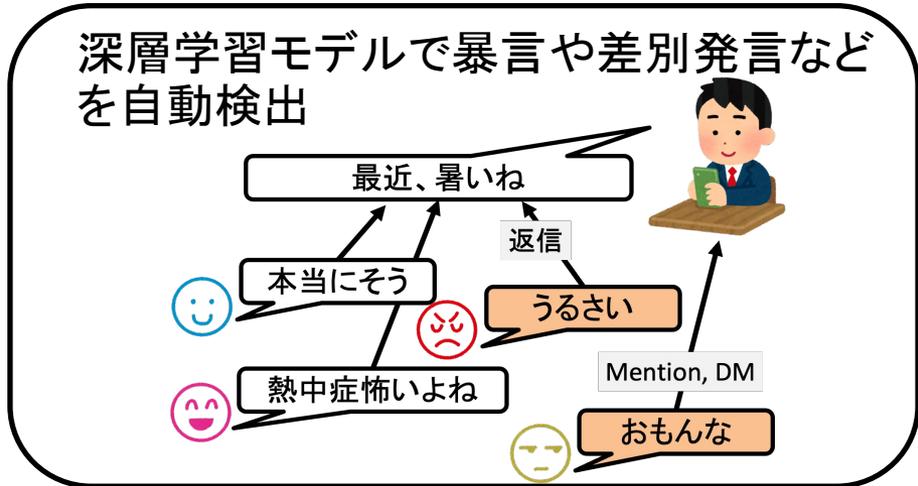
- 学習者の集中度自動推定
 - 講義の重要箇所自動推定
- システムから能動的に注意・アドバイス

その他、このような研究も扱っています

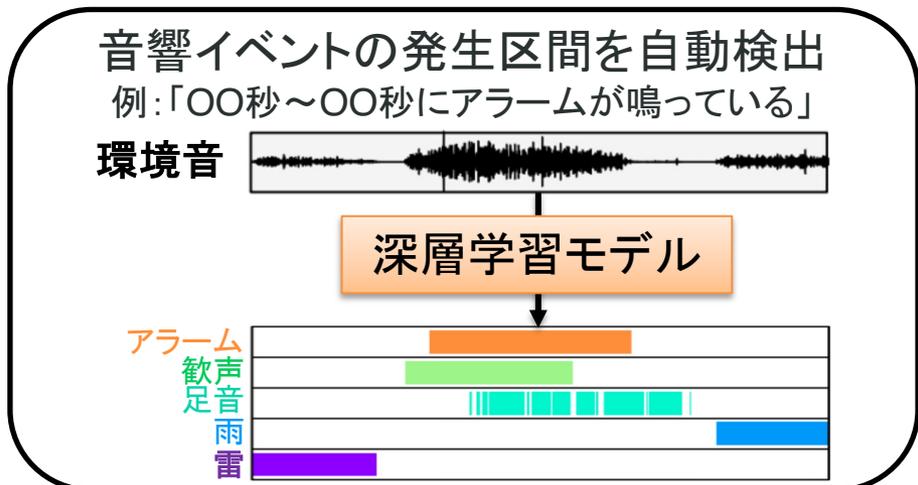
● ゼロ資源言語の音声認識



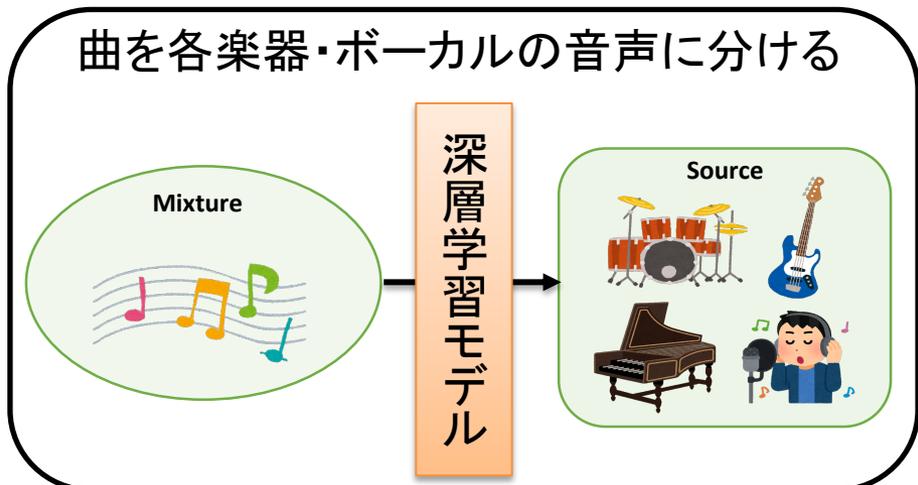
● 攻撃的な言葉の自動検出



● 音響イベント検出



● 音源分離



合成チームの研究紹介

深層学習に基づくテキスト音声合成

テキスト音声合成(Text-to-Speech): テキストを入力として音声を生成

深層学習に基づくテキスト音声合成の例

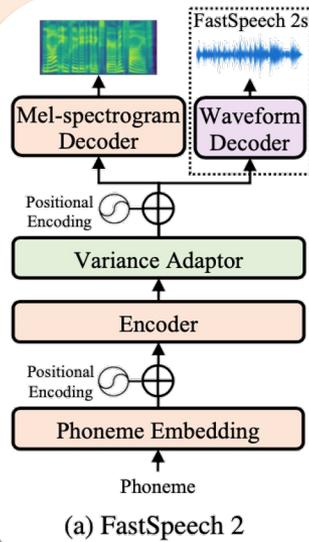
主な利用例



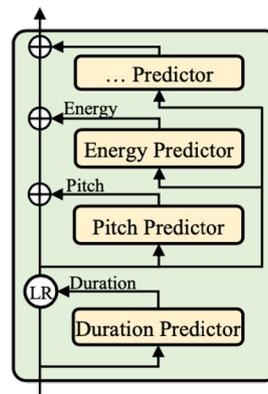
スマートスピーカー



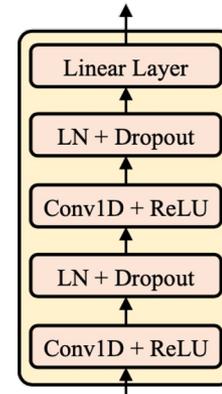
対話エージェント



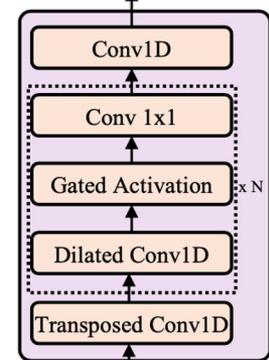
(a) FastSpeech 2



(b) Variance adaptor



(c) Duration/pitch/energy predictor



(d) Waveform decoder

- **Encoder-Decoderモデル, Transformer** を使用
- Variance Adaptorによる話速・エネルギー(声の大きさ)・F0(声の高さ)といった**音声の細かい制御が可能**

最先端の深層学習の研究で得られた知見を活かすことで、違和感の少ない**肉声感のある人間らしい声**を生成できる！

音声合成の応用と未来

◆ バーチャル対話エージェントへの応用



能勢 隆 (Takashi Nose)

@takashi_nose

...

私が所属している東北大の電気情報物理工学科通信工学コース

ecei.tohoku.ac.jp/eipe/の紹介動画ができました。バーチャル音声対話エージェントSakiが学生さんにコースの内容についてインタビューしています 😊

◆ 今後の研究における目標・眺望

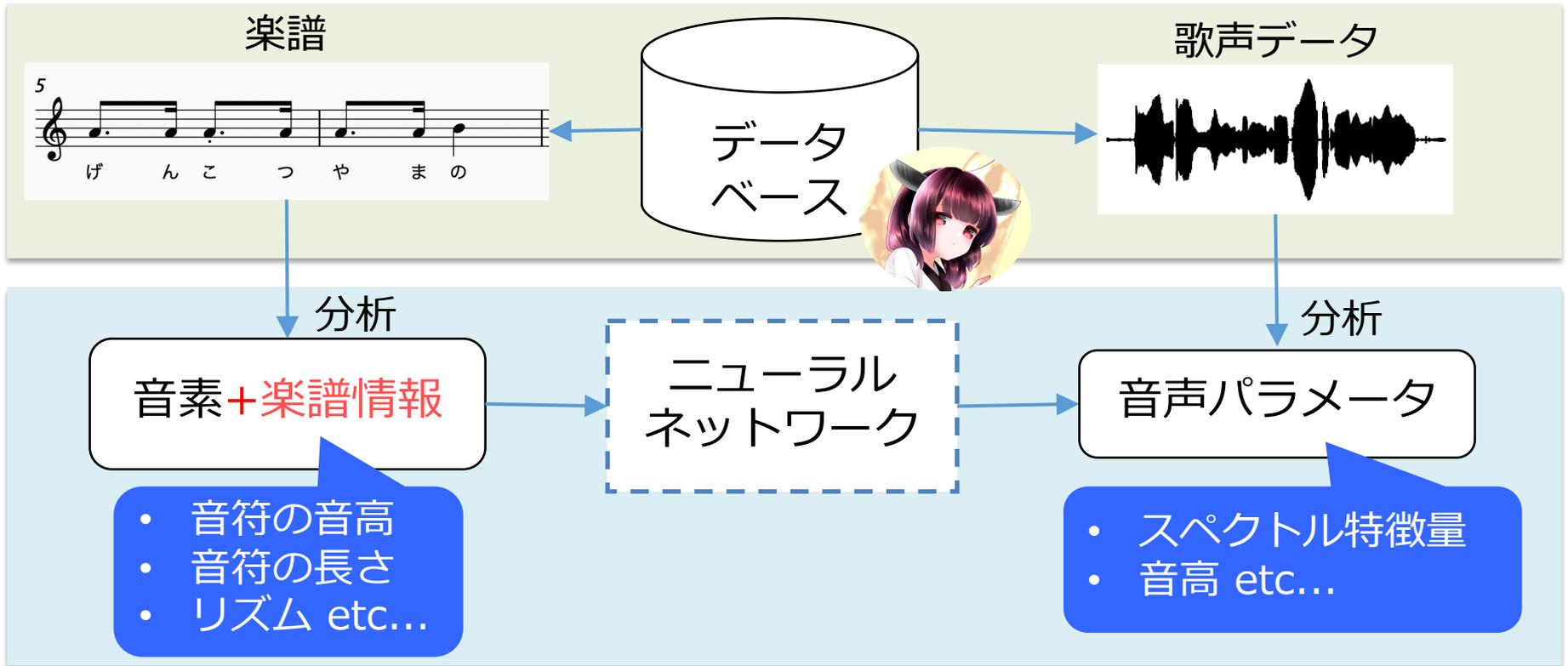
- より滑らかで自然な音声
- 現実には存在しない声質を持つ音声
- 発話が難しい人たちのための音声合成システム

比較的自由度が高い研究室なので、これら以外の研究でも大丈夫です！
(明確にやりたい研究があれば、是非相談してみてください！)



深層学習に基づく歌声合成

歌声合成(Singing Voice Synthesis): 楽譜を入力として歌声を生成



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

声質変換(Voice Conversion): 音声の発話内容を変えずに声質を別人に変換



アバターに合わせて声を変換



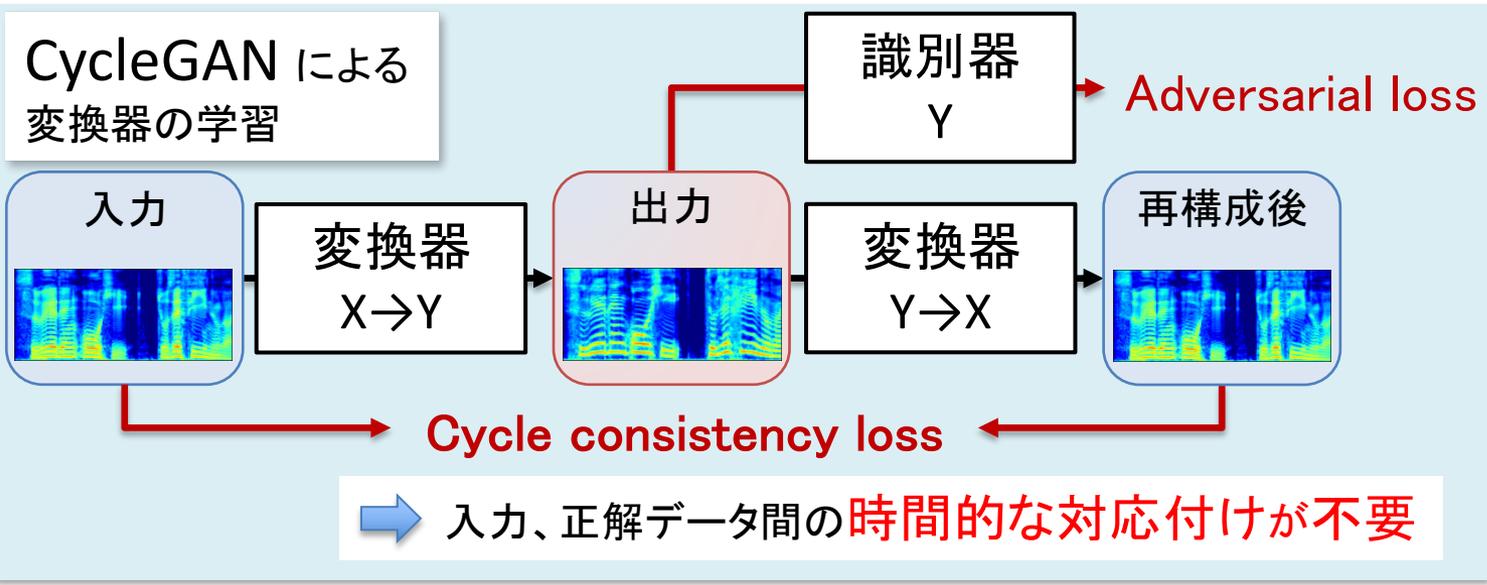
好きな声で配信



キャラクターボイスの吹き込み



歌手の声でカラオケ



音声デモ	
目標話者 音声	
変換前	
変換後	

深層学習に基づく発話動画像生成

特徴点画像を用いることでAIが口や表情以外の顔全体を高性能に再現

線画から写真のような画像も作れる！



入力テキスト
こんにちは！

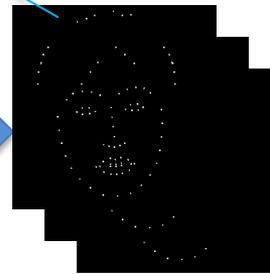
or

入力音声

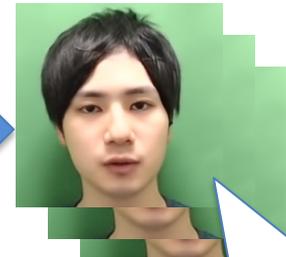


顔特徴点
予測AI

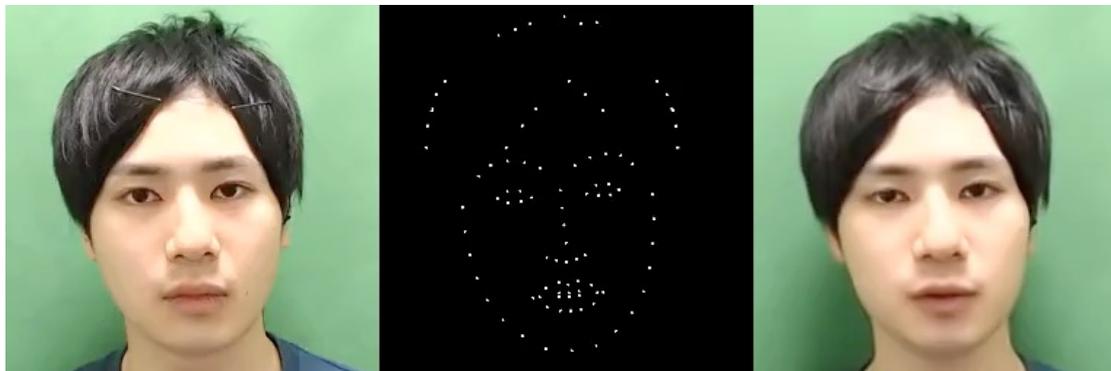
画像化



画像変換
AI



こんにちは！



オリジナル

入力(点動画)

出力(生成動画)

研究室での生活について

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

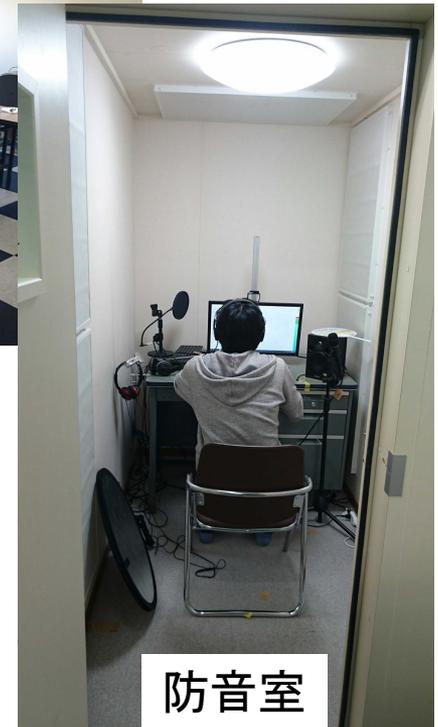
小ゼミ室



お茶室



学生部屋



防音室

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



堀井大輔 修士2年

研究内容: マルチモーダル感情認識, コーパス作成

趣味 : ゲーム, サウナ, バイク

ある1日の行動例

10:30
研究室着

10:30~12:00
研究

昼食

13:00~17:30
研究, ミーティング, 授業

帰宅

必要なら
家で作業

着いたらまずは一日の
作業の整理

現在はクラウドワーカーを活用して
コーパス作成作業
10月の発表の構成も検討中

基本的には自由時間
学会やゼミ前は作業することも

研究室生活の1日:リモートワーク



藤原知樹 修士1年

研究内容：攻撃的な言葉の自動検出

趣味：料理、ゲーム

- 家事の時間が取りやすい！
- 煮詰まってる時は研究室に行くようにしている

リモートワークの1日

～10:00
起床・家事

10:00～12:00
研究

昼食

13:30～1700
研究

～20:00
家事

20:00～21:30
研究

21:30～
ゲーム

実装作業
結果の収集

結果の分析とまとめ
追実験

文献調査
今後・明日の研究内容の整理

研究室生活の1日：ゼミがある日



川西翔貴 学部4年

研究内容：モノログ発話と対話のリップシンク差異の調査
趣味：最近はスプラトゥーンにどハマリ

ゼミがある日

先生からの提案など、ゼミで出た意見をまとめて
今後の研究に活かそう！

～10:00
起床・家事等

10:00～12:00
ゼミの準備

昼
食

13:00～16:00頃
ゼミ

内容
整理

17:00 帰宅
自由時間

発表用の資料を準備
家でやることも研究室でやることも

進捗を先生や他のメンバーに発表！
メンバーの発表も聞いて勉強
家からリモートでも研究室で現地参加でもOK

配属後のスケジュール



11月
12月

顔合わせ

先生・先輩との自己紹介後，研究室のルールや基本的な事柄についての説明も行います

1月
2月

新入生課題

Latexの環境構築，サーバー上の環境構築など研究のための下準備を行います。

3月

4月

勉強ゼミ

音声信号処理・機械学習の基本的な知識について学習します

5月

6月

花見

コロナでなくなってしまいましたが，そろそろ再開できるかも...？

7月

8月

院試

大体1カ月前くらいから院試勉強を始めてもらいます

9月

10月

芋煮・駅伝

コロナでなくなってしまいましたが，そろそろ再開できるかも...？

11月

12月

卒業研究

先輩に指導されながら，2週間～1カ月間くらいで執筆します

1月

2月

3月



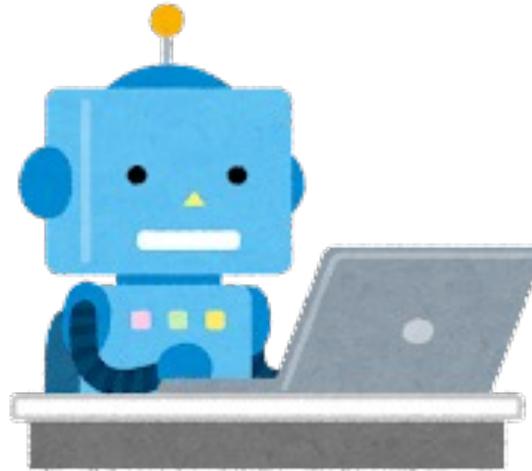
伊藤・能勢研究室のアピールポイント

伊藤・能勢研究室は音声だけの研究室ではありません!!!

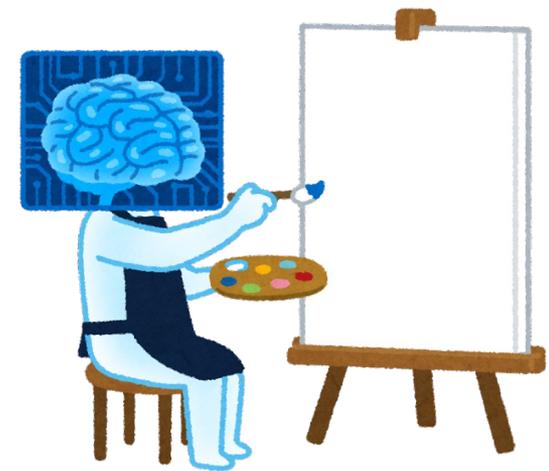
以下の3分野について研究しています



音声



言語(対話システム)



画像(画像生成)

◆ 伊藤・能勢研究室におすすめの人

- 音声・言語・画像(マルチモーダル)に興味がある人
- 人工知能・機械学習・深層学習に興味がある人
- 先生・先輩とコミュニケーションをとりながら研究を通して成長したい人

伊藤・能勢研究室では…

人工知能・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

をお待ちしています！



教授
伊藤 彰則 先生



准教授
能勢 隆 先生

アニメ・ゲーム・音楽・声優・
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 遠藤真裕 masahiro.endo.t8@dc.tohoku.ac.jp

Q&A

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**程度の活動が目安です。

Q. プログラミング経験は必要ですか？

A. プログラミングは必須となりますが、経験が少ない人も勉強会で**基礎から勉強**するので大丈夫です！
また、教員や先輩方も**適宜フォロー**します。

Q. 長期休暇・院試勉強期間はありますか？

A. 夏季2ヶ月あります。**院試1ヶ月前から**は院試勉強に集中できます。

Q. 海外で発表の機会はありますか？

A. もちろん、**あります**！