

AIがあなたの声になる？

GPT soVITSの仕組み

情報工学科4年
マルチェンコ ダニール

1. ロジックベース vs AIの仕組み

ロジックベース
(昔のやり方)

VS

AI

(今どきのやり方)

入力 (1, 2, 3...)

入力 (1, 2, 3...)



ルール: $y = 2x + 1$
"自分で決める"



出力 (3, 5, 7...)



"ルールを見つける"

ルール: $y = 2x + 1$

2. 音声AIってなに？

人の声を出力する数式はない！

入力

文字

こんにちは！ 音声ファイル

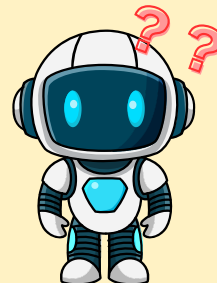
出力



ルール



AIモデル



でも、文字の「強調」や「感情」は再現むずかしい

3. GPTsoVITSを使って、より人間らしい声へ



GPTモデル

(脳: 文脈を判別)



soVITSモデル

(口: 音質を調整)



「自分らしい声」
の完成！



Reference音声
(その人の声)

学習の秘密: Few-shot

事前に大規模データで汎用的な音声生成能力を獲得しているため、少量の参照音声から「声質ベクトル」を抽出するだけで、未知話者でも自然な音声を再構成できる。

4. さらなる提案: 「心」も込めるには？

① 参照音声を感情でわけ



HAPPY

ANGRY

SAD

色々な感情の声を聞かせる

② LLMで生成した文に「感情ラベル」をつける

「happy」これ、すごく楽しい！

LLMで文章生成するAIアバターでも人間のような感情だせる！



Project DANYA
現在開発中！

AIを単なる道具にとどめず、より意味のある存在として活用したいと考えています！