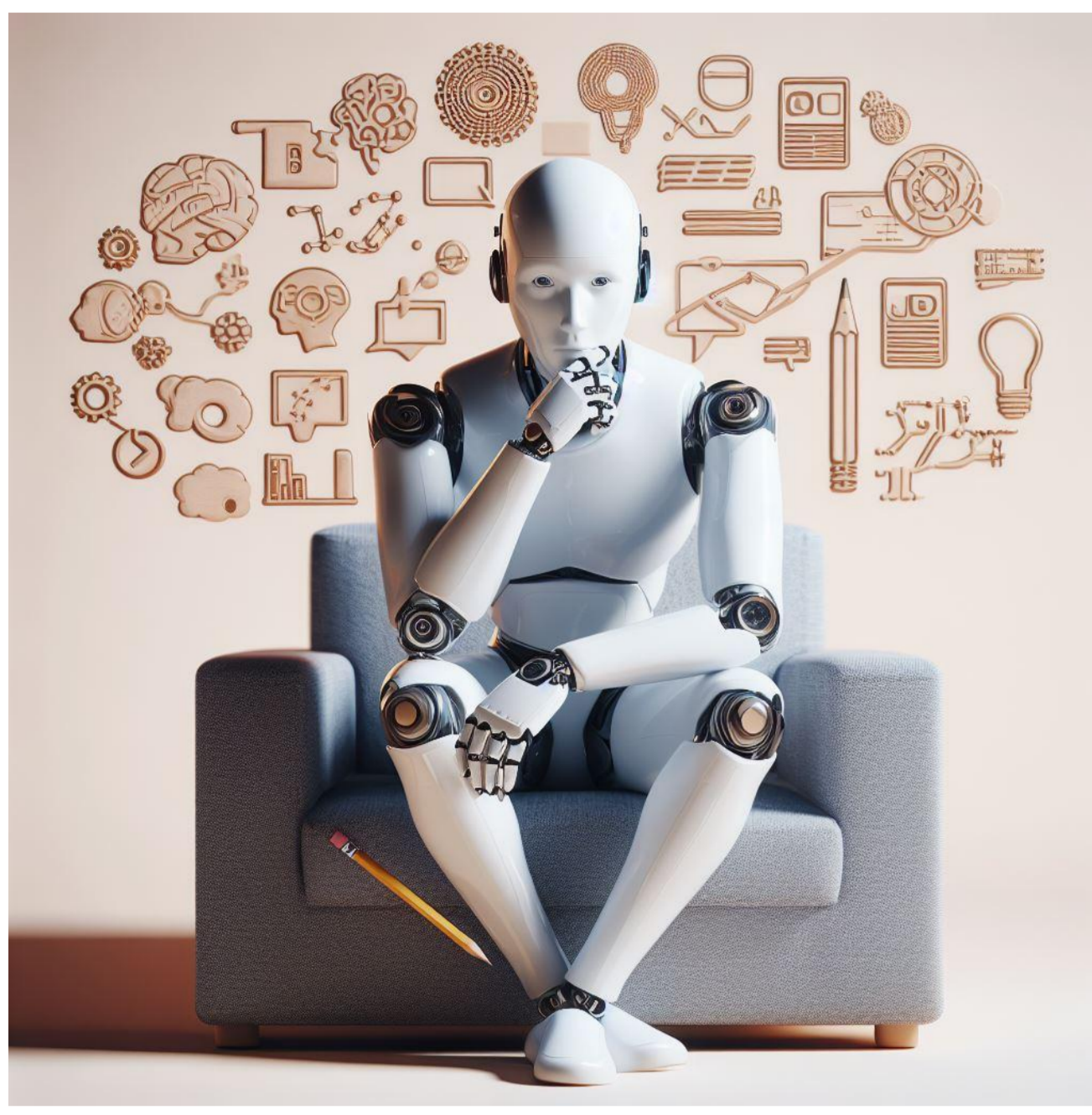




東京大学 猿渡・高道研究室では、NII 音声資源コンソーシアム (NII-SRC) への委託分を含め 30 以上のコーパスを提供しています。本ポスターではその一部を紹介いたします。



## 言葉と絵を理解して生じる声とは？ 平絵付き紙芝居朗読コーパス

NII-SRCで提供中

絵本や紙芝居を朗読する時、人間は文章や挿絵(平絵)を理解することで、場面や人物の情景を声に載せる。その音声表現を分析、あるいは人工的に再現することは可能だろうか？

→ 日本語単一話者オーディオブック・紙芝居朗読音声コーパス (J-KAC: Japanese Kamishibai and audiobook corpus) [2021]

- 作品の選定
  - 文と絵の著作権が消滅した小説… 青空文庫など
  - 著作権の存続している紙芝居… 著者の許諾を得て利用

- 音声収録と後処理
  - 文章を章・段落などに構造化
- 挿絵・平絵
  - キャラクターや情景の画像
  - 許諾のもとでデジタル化

```
[蜘蛛の糸.yaml (芥川龍之介)]
chapt000: # 章
parag000: # 段落
style000: # 文種
sent: ある日の事でございます。
time: [0.96, 3.32] # 開始終了時刻
```



宮沢 賢治(作) 茂田井 武(絵) 福音館書店(刊) “ゼロひきのゴージュ” 終盤の演奏場面

著者(文)	タイトル
新美南吉	ごん狐
橋山正雄	花咲かじい
太宰治	桜桃

著者(文・絵)	タイトル
田中六太	むしにごようじん
大島妙子	おぼけいぬのおハナちゃん
ようふゆか	ありくんとかえるくん

要素	数・条件
話者	プロ話者1名
収録環境	スタジオ
作品数	小説25作品, 紙芝居17作品
絵付き作品数	小説1作品, 紙芝居17作品
音声時間長	9時間



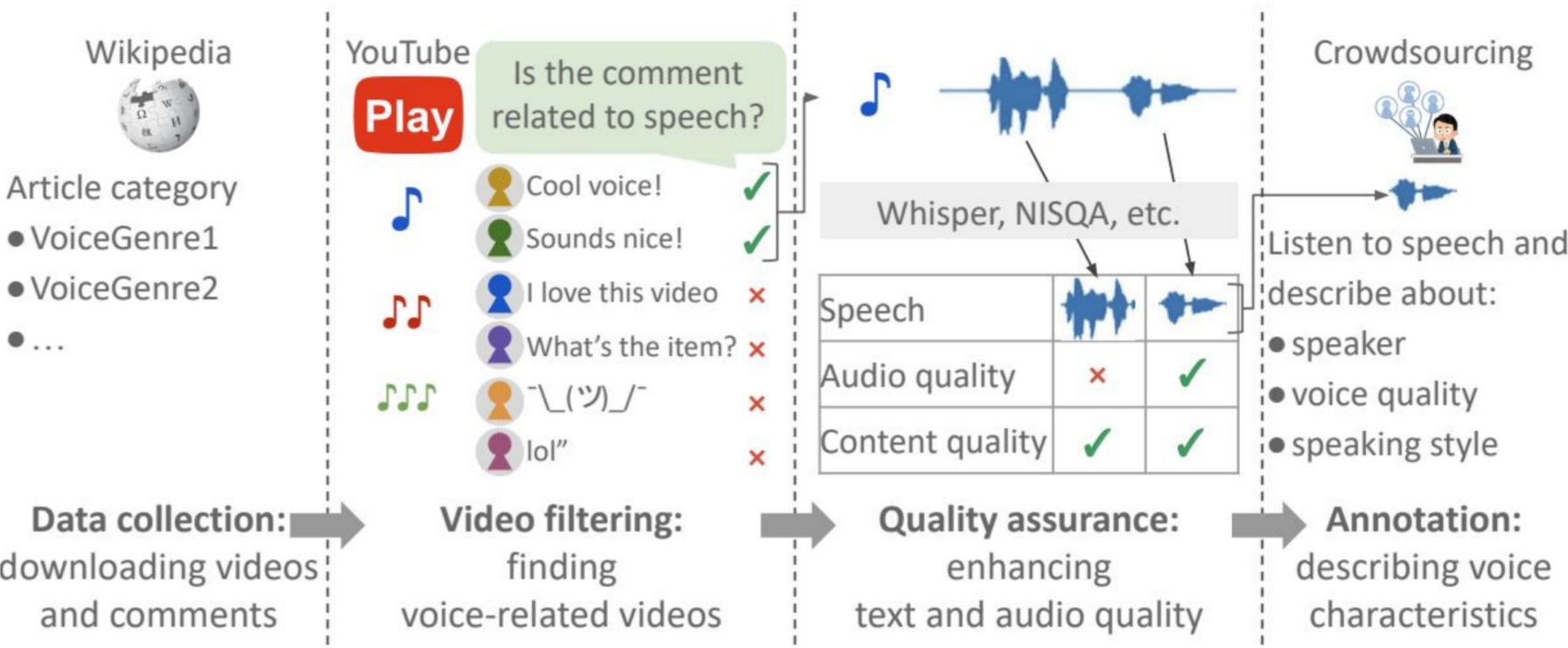
## 声の個性を言葉で表せる？ 声質表現文付き多話者音声コーパス

NII-SRCで提供中

人間は声質をどのように言葉で表すのだろうか？ また、昨今の text-to-image のように、人工音声の声質を言葉でデザインすることは可能だろうか？

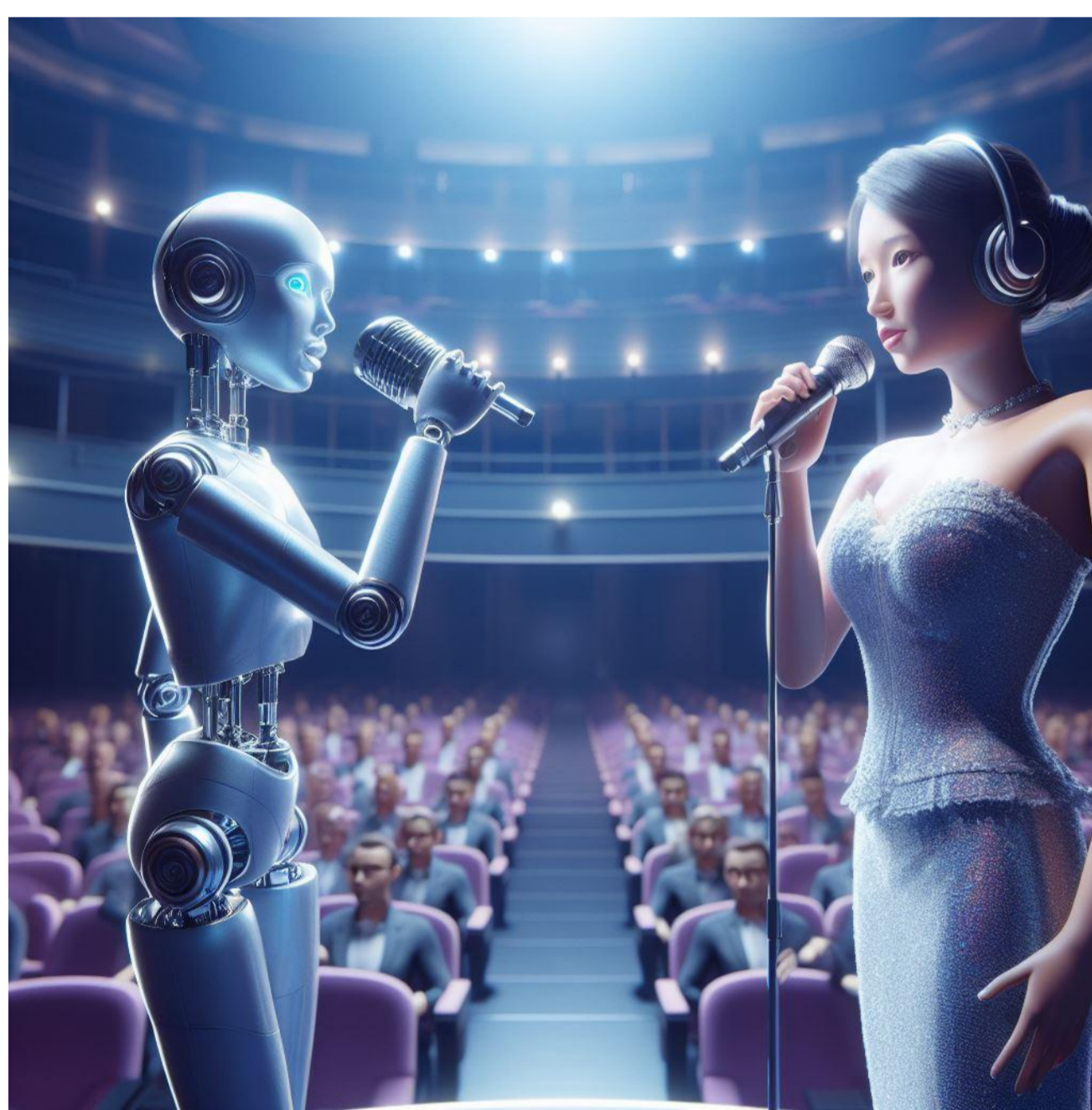
→ 日本語声質表現文音声ペアコーパス (Coco-Nut: Corpus of connecting NIHONGO utterance and text) [2023]

- 多話者音声と声質表現文の作成
  - YouTubeから「魅力的な声」を自動収集
  - クラウドソーシングで声質表現文を付与



声質表現文サンプル  
30代くらいの男性の声。ゆっくりと穏やかな話し方でした。苦悩に満ちた、けだるそうな声でした。元気な男性が明るい声で、テンション高く発表をするように喋っている。若い女性が抑揚のある声でゆっくりと喋っている。

要素	数・条件
内容物	・日本語母語話者音声データ ・書き起こし文 ・声質表現文
発話数	約7,600発話 8時間

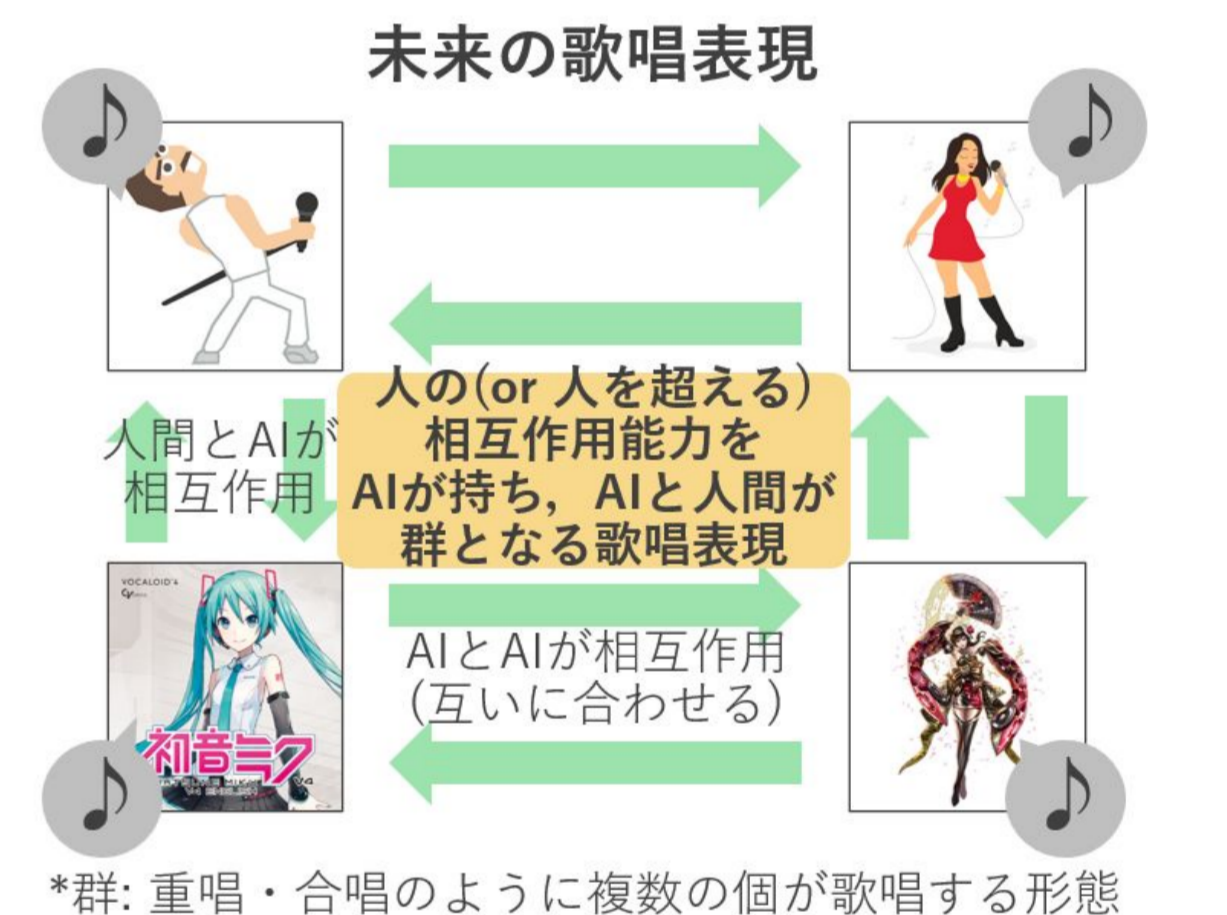


## 一緒に歌うと何が起こる？ 再配布可能なアカペラ重唱コーパス

人間は「他者に合わせる」能力を持つ。他者に合わせて歌うことも出来る(例えばアカペラ)。その歌唱相互作用を分析し、また、歌唱相互作用を持つAIシンガーは作れるだろうか？

→ 日本語アカペラ重唱コーパス (jaCappella: Japanese a Cappella Vocal Ensemble Corpus) [2022]

- 再配布可能なアカペラ曲を新たに作曲、歌唱したデータベース
  - 童謡のアカペラアレンジ
  - 10ジャンル 50曲 6パート
- 相互作用するAIシンガー (研究中)
  - 人間・AIと一緒に歌えるシンガーへ

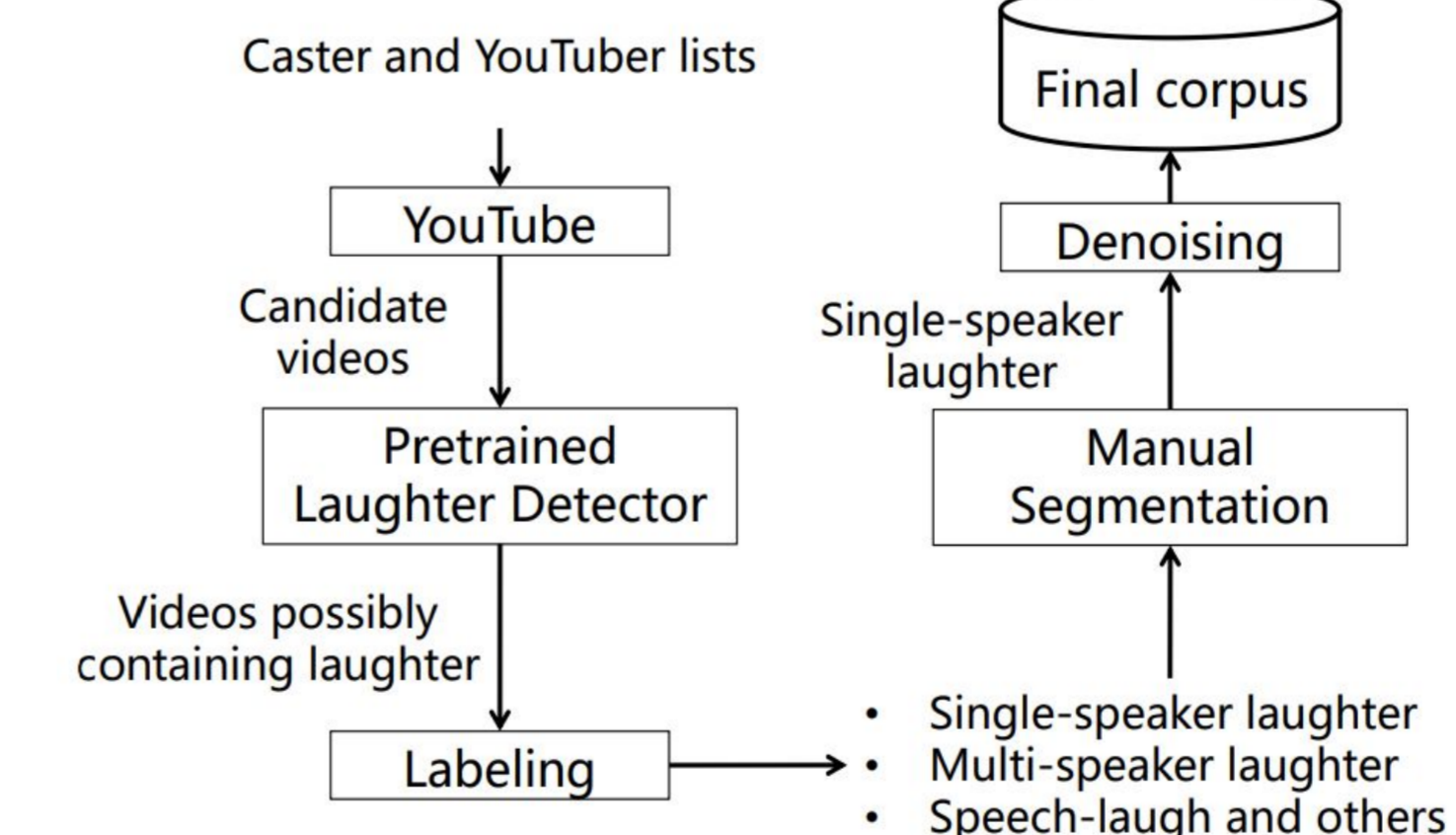


## 人はどう笑う？ AIは笑う？ 世界最大規模笑い声コーパス

笑いは感情表出の手段でありコミュニケーションの手段でもある。人間はいつ、どう、なぜ笑うのか？ また、AIも人間と同じように笑うのだろうか？

→ 日本語母語話者による大規模笑い声コーパス (Laughterscape: large-scale in-the-wild Japanese laughter corpus) [2023]

- 笑い声の収集
  - YouTube動画を元にして、単一話者による笑い声を大量に収集、セグメンテーション。
- 笑い声の人工合成 (笑うAI)
  - 笑い声の「音の塊」を自動学習
  - GPTライクな構造で「音の塊」を人工的に合成可能



要素	数・条件
内容物	・笑い声音声データ、複数人で同時に笑う声は含まない。 ・各笑い声の話者ID
話者	500名以上の日本語母語話者
発話数	11,000発話以上。1名あたり平均22発話以上。ただし偏りあり。
音声時間長	6時間超。オープンコーパスとして世界最大規模
アノテーション	無し

## その他にも東京大学 猿渡・高道研究室では音声コーパスを提供しています。

### #テキスト読み上げ

- JSUT: 単一話者による日本語読み上げ 10 時間
- JVS: 100名話者による日本語読み上げ 30 時間。裏声、ささやき声も含む

### #環境音

- DCASE2023-task7: 環境音合成用のオープンコーパス
- ML-AudioCaps: 同じ環境音を日本語と英語で説明した環境音説明文

### #音声認識 #話者認識 #自己教師あり学習

- JTubeSpeech: YouTubeから収集した日本語コーパス
- JTubeSpeech-ASV: YouTubeの話者IDあり日本語 500 時間
- YODAS: 書き起こし付き多言語音声 500,000 時間

### #歌声

- JVS-MuSiC: 100名歌唱者によるパラレル歌声 2 時間。1曲は全歌唱者共通
- PJS: 単一話者による音素バランス歌声 1 時間。歌声音源としても配布あり

### #非言語音声 #自発音声

- JrecSponSpeech: 非流暢性アノテーション付き講義音声
- JNVN: 非言語音声と言語音声から成る音素バランス文

### #方言音声

- JMD: 単一方言話者による大阪方言・熊本方言 3 時間
- CPJD: クラウドソーシング収集の方言話者による20方言パラレル 9 時間
- Tohoku-folktale: 東北地方の歴史的昔話 14 時間