音声クローン クロコさん用プロンプト

タイプA(CPU)で試し、必要ならGoogle ColabでタイプBへ進む2段階構成(2026年10月19日作成)

使い方のポイント

- 下の枠内の文章を、そのままクロコさん(Claude Code)に貼り付けて使う。
- 最初に**候補モデルの比較表を見せてもらい、自分で選んでから**進む形にしてある。この分野は入れ替わりが早いため。
- **声の録音は、クロコさんに録音のコツを聞いてから**行う。段階1は数十秒〜数分、段階2に進む場合は30分〜1時間ほど必要。

プロンプト本文

```
自分の声でナレーションを読み上げる「音声クローン」の環境を、
このPCに作ってほしい。YouTube動画やプレゼンのナレーションに使う予定。

【このPCの環境】
- MINISFORUM AI X1(ミニPC)、Windows
- メモリ32GB、SSD 512GB
- GPUはAMDの内蔵GPUのみ(NVIDIAのGPUなし=CUDAは使えない)
- 処理はCPUで行う前提

【絶対に守ってほしい条件】
- すべてこのPCの中で完結させる。私の声のデータを外部のサービスに
  アップロードしない(段階2のGoogle Colabだけは例外)
- 追加の費用がかかるもの(有料APIなど)は使わない
- クローンに使う声は、私自身の声だけ
- インストールの前に、必要なディスク容量を見積もって報告し、
  SSDの空き容量が足りるか確認する

═══════════════════════════════════════

段階1:タイプA(ゼロショット型)をCPUで試す

═══════════════════════════════════════

1. 調査
   - 日本語に対応し、数十秒〜数分の声のサンプルだけで声を真似できる
     オープンソースの音声合成モデルを調べる
     (候補の例:GPT-SoVITS、Style-Bert-VITS2、その他の新しいもの)
   - CPUだけで動くこと、Windowsで動くこと、日本語の品質を重視して、
     候補を2〜3個に絞り、それぞれの特徴・必要な容量・
     生成速度の目安を表にまとめて私に見せる
   - 私が選んでから、次に進む

2. 環境構築
   - 選んだモデルを、専用のフォルダ(例:C:\voice-clone)に
     Pythonの仮想環境を作ってインストールする
```

- 手順は、あとで私が見返せるようREADME.mdに日本語で残す

3. 声のサンプルの準備
   - 私が録音した声のファイル(samples/my_voice.wav)を使う
   - 録音のコツ(静かな部屋・同じマイク・自然なトーン、
     必要な長さ)を先に私に教えてほしい
   - 必要なら、ノイズ除去や音量の調整、形式の変換も行う

4. 試しに読み上げ
   - 次の文章を私の声で読み上げ、output/test_01.wav に保存する:
     「こんにちは。今日は、奥日光の冬景色をお届けします。」
   - 生成にかかった時間も報告する

5. 本番用の仕組み
   - テキストファイル(script.txt)に書いた原稿を、
     1文ずつ読み上げて、1つの音声ファイル(mp3)にまとめる
     スクリプトを作る
   - 使い方をREADME.mdに追記する

6. 報告
   - 仕上がりの品質について、私が聞いて判断するので、
     比較用に設定を変えた音声を2〜3パターン作っておく

段階2:タイプB(ファインチューニング型)をGoogle Colabで
(段階1の品質に私が満足できなかった場合だけ進める)

- 私の声で追加学習するための、Google Colab(無料版)用の
  ノートブック(.ipynb)を作る
- 学習には30分〜1時間程度の私の声の録音を使う前提
- 無料版のGPU時間の制限内で終わるよう、学習の設定を調整する
- 途中で切断されても続きから再開できるよう、
  学習結果をこまめにGoogleドライブに保存する
- 学習が終わったモデルをこのPCにダウンロードし、
  段階1の読み上げスクリプトで使えるようにする
- Colabの操作手順を、初めての人でも分かるようREADME.mdに書く

まずは段階1の「1. 調査」から始めて、候補の比較表を見せてください。