

## Slide Voice Maker 完全仕様書

バージョン: 1.0.0 日付: 2026-01-05 リポジトリ: <https://github.com/J1921604/Slide-Voice-Maker>

### 目次

- 概要
- システムアーキテクチャ
- 機能仕様
- データモデル
- CLI仕様
- Web UI仕様
- ファイル構造
- テスト

### 概要

Slide Voice MakerはPDFスライドと原稿CSV（`index.html`, `script.js` 形式）からAI音声レーション付き動画（WebM/MP4）を自動生成し、PPTXや原稿CSVとともに出力できるツールです。フロントエンドはGitHub Pagesで静的配信し、バックエンドはローカルFastAPIで動作します。

### 主要機能

機能	説明	優先度
解像度選択	720p/1080p/1440pから選択可能	P1
temp上書き更新	毎回tempフォルダを自動クリア	P1
CSV読み込み	UTF-8/Shift_JIS/EUC-JP対応	P1
音声合成	Edge TTSで実行	P1
WebM/MP4動画出力	FFmpegでVP8/VP9/H.264を選択（WebM/MP4）	P1
PPTX出力	スライド画像をPPTXとしてダウンロード	P2

### 機能一覧 (src/server.py起動時)

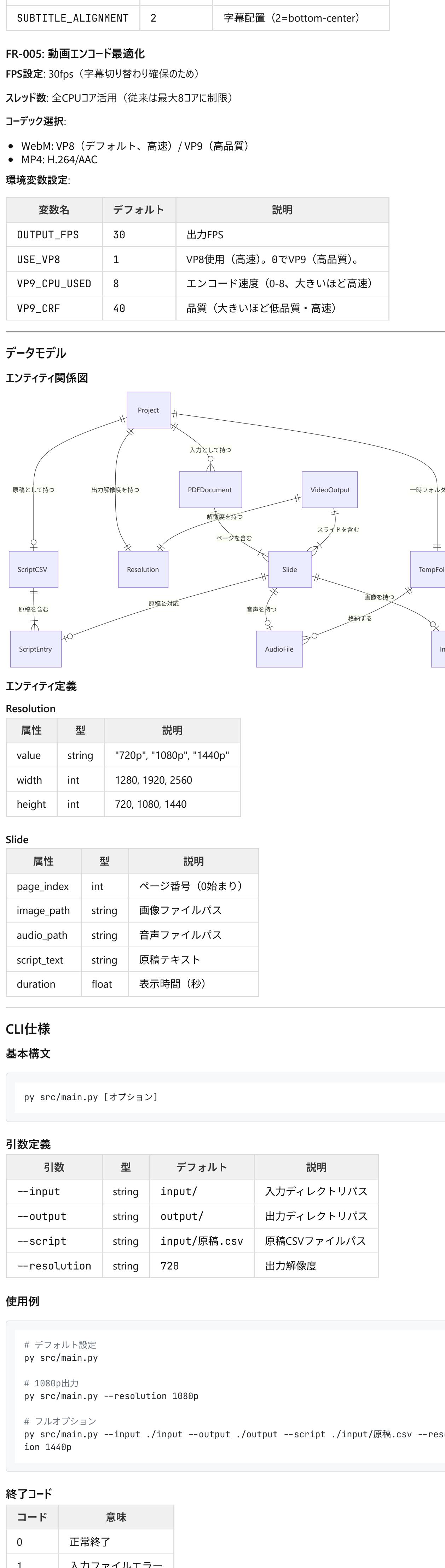
機能	説明
PDF入力	PDFファイルをアップロードしスライドとして展開
原稿CSV入力	inputフォルダにCSVファイルを上書き保存
発音辞書	word,alias形式のCSVで特定単語の読み方をテキスト置換（オプション）
解像度選択	720p/1080p/1440pフルダウン（画像解像度）
音声選択	画像・音声生成時の音声が4種類のプリセットから選択（実声はNanami/Keitaの2種を母体に、rate/pitch/volumeで聽感差を作る）。x1.5倍速デフォルト（rate +50%）、女声1/男声。
再生速度選択	0.5x～2.0x（全スライドに適用）
字幕ON/OFF	動画に字幕を埋め込むかどうかを選択
画像・音声	Edge TTSで音声を生成、output/tempに画像・音声を保存
動画生成	output/tempから動画WebM/MP4を生成（PDFと同名で保存）
原稿CSV出力	編集した原稿をCSVでダウンロード
動画出力	outputフォルダから選択したWebM/MP4をダウンロード
PPTX出力	スライド画像をPPTXとしてダウンロード

### 動作環境

項目	要件
OS	Windows 10/11
Python	3.10.11（推奨）
ブラウザ	Chrome / Edge（最新版）

### システムアーキテクチャ

#### 全体構成図



#### データフロー



#### 状態遷移



### 機能仕様

#### FR-001: 解像度選択

選択肢	幅 (px)	高さ (px)	環境変数
720p	1280	720	1280
1080p	1920	1080	1920
1440p	2560	1440	2560

#### 実装:

```
RESOLUTION_MAP = {  
    "720p": 1280,  
    "720p": 1280,  
    "1080p": 1920,  
    "1080p": 1920,  
    "1440p": 2560,  
    "1440p": 2560,  
}
```

#### FR-002: temp上書き機能

```
def clear_temp_folder(temp_dir: str) -> None:  
    """tempフォルダを上書きクリア"""  
    if os.path.exists(temp_dir):  
        try:  
            shutil.rmtree(temp_dir)  
            print(f" Cleared temp folder: {temp_dir}")  
        except PermissionError as e:  
            print(f"Warning: Could not clear temp folder: {e}")  
        os.makedirs(temp_dir, exist_ok=True)
```

#### FR-003: CSV読み込み (文字化け対処)

対応エンコーディング:  
1. UTF-8  
2. Shift\_JIS  
3. EUC-JP  
4. ISO-2022-JP

#### CSV形式:

```
word,alias  
JERA,ジラ  
成果物,せいかぶつ  
AI,エーアイ
```

#### データモデル

#### エンティティ関係図



#### エンティティ定義

属性	型	説明
value	string	"720p", "1080p", "1440p"
width	int	1280, 1920, 2560
height	int	720, 1080, 1440

#### FR-004: 字幕生成 (ASS形式)

字幕フォーマット: ASS (Advanced SubStation Alpha)

#### 字幕分割方式:

- 句読点（。！？！？）で分割
- 文字数比率でタイミングを計算

#### 字幕ファイル:

- YouTube風の半透明グレー背景
- 太字白文字
- 太字白文字

最小セグメント時間: 0.15秒 (30fpsで約4.5フレーム、字幕切り替わりを確実にする)

#### 環境変数設定:

変数名	デフォルト	説明
SUBTITLE_MARGIN_V	10	字幕の縦マージン（下寄せ調整）
SUBTITLE_ALIGNMENT	2	字幕配置（2=bottom-center）

#### FR-005: 動画エンコード最適化

FPS設定: 30fps (字幕切り替わり確保のため)

フレームレート: 全CPUコア活用（従来は最大8コアに制限）

コーデック選択:

- WebM: VP8 (デフォルト、高速) / VP9 (高品質)
- MP4: H.264/AAC

#### 環境変数設定:

変数名	デフォルト	説明
OUTPUT_FPS	30	出力FPS
USE_VP8	1	VP8使用（高速）。0でVP9（高品質）。
VP9_CPU_USED	8	エンコード速度（0-8、大きいほど高速）
VP9_CRF	40	品質（大きいほど低品質、高速）

### データモデル

#### エンティティ定義

属性	型	説明
value	string	"720p", "1080p", "1440p"
width	int	1280, 1920, 2560
height	int	720, 1080, 1440

#### Slide

属性	型	説明
page_index	int	ページ番号（0が始まり）
image_path	string	画像ファイルパス
audio_path	string	音声ファイルパス
script_text	string	原稿テキスト
duration	float	表示時間（秒）

#### CLI仕様

#### 基本構文

```
py src/main.py [オプション]
```

#### 引数定義

引数	型	デフォルト	説明
--input	string	input/	入力ディレクトリパス
--output</td			