

A I 音声認識システム仕様書

1 システム仕様

1-1 基本機能

- (1) Web 会議や商談、会見等、対面・非対面の幅広いシーンで利用できるスタンドアローン型文字起こし支援アプリケーションを提供すること。
- (2) リアルタイムでの音声認識に加え、録音した音声データを取り込んで文字化するバッチ認識にも対応していること。
- (3) 音声認識で文字化されたテキストから順次、修正作業に着手する事ができること。
- (4) 音声の録音からテキストの編集、文字起こし内容の出力まで1つのアプリケーション内で行えること。
- (5) 視聴用に音声認識で文字化されたテキストや文字起こし内容を別の画面に表示できること。
- (6) 音声の再生やテキストの編集は、キーボード操作だけでも行えること。
- (7) 音声認識エンジンを簡易カスタマイズできる単語登録機能が搭載されていること。
- (8) オフライン（インターネット未接続）の状態で音声認識が可能なスタンドアローン型であること。
- (9) アプリケーションは、端末（Windows パソコン）にインストールして利用できること。
- (10) ライセンスの提供形態は、以下の2種類から選択できること。
 - ・決められた台数の端末にアプリケーションをインストールして利用できるスタンドアローンライセンス（ノードロック）。
 - ・インストール台数に制限がなく、端末にUSB キーを接続した時にのみアプリケーションが利用できるスタンドアローンライセンス（USB キー）。

2 AI 音声認識・編集機能

2-1 AI 音声認識機能

<共通機能>

- (1) AI 音声認識機能は、不特定話者対応で事前に話者の音声の登録・学習が不要であること。
- (2) AI 音声認識機能は、単語認識ではなく連続音声認識であり、発話内容を一字一句文字化できること。
- (3) 話し言葉（特に会議特有の話し言葉）が音声認識できること。
- (4) 音声認識の速度を調整できること。
- (5) ユーザー辞書に登録された単語を用いて音声認識ができること。
- (6) 音声認識を停止せずにユーザー辞書に単語を追加登録でき、かつ、即時反映され文字化できること。

- (7) AI 音声認識エンジンやユーザー辞書に登録されていない単語が認識、編集されたとき、未知語として表示され、単語登録ができること。
- (8) 自動的に句読点の出力ができること。
- (9) 話者振り分けができること。また、振り分けられた話者情報から自動で話者識別できること。

2-2 リアルタイム AI 音声認識機能

- (1) 複数のサウンドデバイスを選択し、音声を入力できること。
- (2) 入力した音声をリアルタイムで音声認識できること。
- (3) 選択したサウンドデバイスごとに発言者の名前を割り付けることができ、その発言者ごとに認識結果を表示することができること。

2-3 音声ファイル認識機能

- (1) 音声ファイル (MP3、WMA、WAV、M4A)、動画ファイル (WMV、MP4) に対し、音声認識できること。

3 表示・編集機能

3-1 表示機能

- (1) 認識結果を発話時間、発話者、発言内容の項目に分けて表示できること。
- (2) 発話時間の昇順で表示できること。
- (3) 認識結果、編集結果を別の画面に表示できること。

3-2 編集機能

- (1) 発言内容単位で音声を聞きながら認識文字列を修正・編集できること。(バッチでの音声認識時)
- (2) 発言内容の削除ができること。
- (3) キーボードショートカットのカスタマイズが自由に設定できること。
- (4) キーボードとマウスを使用して編集作業ができること。
- (5) キーボードのみでも (マウスを使わなくても) 編集作業ができること。
- (6) 編集結果を Microsoft Word、Microsoft Excel、テキストファイルで出力できること。
- (7) 音声データのみを音声ファイル (WAV 又は M4A) へ出力ができること。
- (8) 重要な単語 (キーワード) を設定することで、テキスト化された発言内のキーワードをハイライト表示できること。
- (9) 発言内容ごとに、「重要チェック」「決定事項」等のタグをつけることができ、編集中やテキストファイル出力後に該当箇所を確認しやすくなること。
- (10) 発言内容に対して絞り込み検索ができること。

- (11) 絞り込み検索において、タグ、出席者、キーワード等で検索を絞り込むことができること。
- (12) 絞り込み検索で絞り込んだ内容のみをテキストファイルに出力できること。

3-3 音声再生機能

- (1) 一つの発話の編集が終了したら、自動的に次の発話の編集に入り、音声も自動再生できること。
(バッチでの音声認識時)
- (2) ピッチを変えずに音声の再生速度の変更ができること。
- (3) 選択した発言内容の音声をループ（繰り返し）再生できること。

3-4 保存機能

- (1) 音声認識した音声や認識結果、編集結果は自動的に保存できること。
- (2) 保存したデータは一覧で表示できること。
- (3) 保存したデータを削除できること。
- (4) データ保存フォルダを任意の場所に変更できること。

4 動作環境・ハードウェア仕様

4-1 動作環境

- (1) CPU：インテル® Core™ i5 プロセッサー 2.4GHz 以上
- (2) メモリ：8GB 以上
- (3) ハードディスク：1GB 以上の空き容量（インストール時）
- (4) CD-ROM ドライブ：インストール時に必要
- (5) OS：Windows® 11 日本語版 64bit 版
 - ※ 32bit 版(×86)には対応しておりません。
 - ※ Mac、Linux には対応しておりません。
 - ※ .NET Framework4.6.0 以降がインストールされていること。
 - ※ Office（Word、Excel）がインストールされていること。
 - ※ 仮想化環境（仮想デスクトップ、仮想アプリケーション）では、インストールや動作を保証できません。
- (6) その他ハードウェア：マイク入力端子・ヘッドフォン端子（又はUSB ポート）

5 インストール条件

- (1) Web サイトから最新版のインストーラーをダウンロードできること。
- (2) Microsoft Edge（EdgeHTML 版）、Microsoft Edge（Chromium 版）、Google Chrome（最新版）、Mozilla Firefox（最新版）に対応していること。

- (3) 管理者権限でのインストールが可能なこと。

6 保守サポート

- (1) メールによるサポートサービスを行うこと。
- (2) 不具合発生時に復旧対応を行うこと。
- (3) アプリケーションの定期バージョンアップを行うこと。
- (4) 製品バグフィックス及び最新マイナーバージョンアップモジュールの提供サービスを行うこと。
- (5) Windows の OS や Web ブラウザのバージョンアップに対応すること。
- (6) 円滑な保守体制を必要とするため、過去に県内の自治体に導入実績を有していること。
- (7) 受託者が必要に応じ、システムの操作説明会などを実施できること。