ChatGPT

構成パターン1：仮想オーディオデバイス＋クラウド型音声認識API

ZoomやMeetの音声をPC内でキャプチャし、クラウド型APIに送信して文字起こしする方法です。WindowsではVB-Audio Virtual CableやVoiceMeeter、LinuxではPulseAudio（`module-loopback`）やPipeWireを使い、会議ソフトの出力を仮想マイク入力にループバックします。これで全参加者の音声が仮想マイクに入力されるので、例えばPython等でAPI（音声認識サービス）にストリーミング送信できます。

有料の高精度APIを使う例としてSpeechmaticsがあります。Speechmaticsはエスペラントを含む多言語対応で、エスペラント音声を90%以上の精度でリアルタイム文字起こし可能と謳っています [1] [2]。リアルタイム処理や複数話者の識別（スピーカーダイアリゼーション）にも対応し、ZoomやMeet音声の文字起こしに適しています [1] [3]。Google Cloud SpeechやAmazon Transcribe、AssemblyAIなど他の主要サービスは現時点でエスペラントに対応していないため、エスペラント精度重視ならSpeechmaticsなどが選択肢になります [1] [4]。

設定例: (1) VB-Audioをインストールして仮想ケーブルを作成。Zoom/Meetの音声出力をこのケーブルにルーティングし、同時に音声をスピーカーにも流す。 (2) Python等でSpeechmatics APIに接続し、仮想ケーブルの入力音声をストリーミング送信して文字起こし。 (3) 受信した文字列を字幕ソフトや独自アプリで画面表示する。

メリット／デメリット: 精度は高い一方、API利用にコストがかかります。また、多少のレイテンシ（数百ms〜1s程度）は生じます。システム構築には音声ルーティングとAPI連携の知識が必要ですが、多言語・エスペラント対応という点で最も実用的な方法です [1] [2]。

構成パターン2：仮想オーディオデバイス＋ローカルASR（Voskなど）

パターン1同様、会議の音声を仮想オーディオデバイスでキャプチャしますが、クラウドではなくローカルの音声認識エンジンで文字起こしします。例えば [5] のようにオープンソースのVosk Speech Recognition Toolkitはエスペラント音声モデルを提供しています。Voskは小型モデル（約40MB）で動作し、CPUでもリアルタイム処理が可能です [5] [6]。仮想ケーブルから受け取った音声をPythonでVoskに入力すれば、エスペラント含む多言語の文字起こしができます。Voskはネット不要・クラウド依存なしで動作するためプライバシー面で優れます。

設定例: (1) PulseAudio等で「モニターソース」（仮想ループバック）を有効にし、会議音声をキャプチャ。 (2) `pip install vosk` で環境構築し、[36]のページからエスペラントモデル（`vosk-model-small-eo-0.42` など) をダウンロード。 (3) Pythonスクリプトでキャプチャ音声をVoskに送信し、文字列を取得。 [5] (4) 得られた文字列を画面上に重ねるアプリ（例：TkinterウィンドウやOBSテキストソース）で表示する。

メリット／デメリット: オフラインで動作し費用ゼロですが、精度はSpeechmatics等のクラウドAPIには劣る可能性があります（例えばVoskのエスペラント小型モデルではテストWERが約7% [6] ）。セットアップはやや技術的ですが、自分のPC内だけで完結します。Windows/Linux両対応であり、GPUがあればWhisper系モデルも使用可能ですが、オフィシャルにエスペラント非対応のため精度が不安定です。

構成パターン3：Zoom閉じキャプションAPI連携

Zoomミーティングではホストが外部キャプションサービスと連携できるAPIが提供されています [7]。Zoomの「Closed Captioning REST API」を使えば、外部プログラムが取得したテキストをZoomの字幕として流し込むことが可能です。例えばZoomミーティングにホストとして参加し、前述の音声キャプチャ＋文字起こし（SpeechmaticsやVoskなど）でリアルタイムテキストを生成し、そのテキストをZoom API経由で送信すれば、参加者全員に字幕表示できます [7]。

設定例: (1) Zoom管理画面でミーティングの「手動字幕（Closed Caption）」機能を有効にし、API用のトークンを取得。 (2) 音声キャプチャ＋STT処理を行うプログラムを実装し（パターン1または2参照）、得られた文字列をZoomの `/closedcaption` エンドポイントにPOST送信。 (3) 参加者はZoom内で字幕がリアルタイム表示される。

メリット／デメリット: Zoom画面内に直接字幕が出るため表示が自然です。しかしセットアップが複雑で、API利用の知識が必要です。また、現状Zoom標準の自動文字起こしは英語など限られた言語のみ対応なため、エスペラントを含む場合は外部処理が必須です [7]。Zoom API連携を利用すれば他ツールに依存せずに字幕を流せますが、開発・設定コストは高めです。

構成パターン4：ブラウザ拡張・デスクトップキャプション機能

Google MeetではChromeのライブキャプション機能で字幕が出ますが、対応言語は英語・スペイン語など主要言語に限られ、エスペラントは含まれていません [8]。また、Google Meet独自の「翻訳付きキャプション」も一部言語間変換のみでエスペラントは非対応です。Chrome拡張の「Live Captions」や各種キャプションツールもWeb Speech APIを使いますが、Web Speech API自体がエスペラントをサポートしていないため、同様にエスペラント字幕は得られません。

Windows 11のLive CaptionやOBSのプラグイン（例：LocalVocal）など、デスクトップ上で音声認識する機能もあります。例えばOBSプラグインLocalVocalはローカルでWhisperモデルを動かし100言語に対応すると説明されています [9]。ただしWhisperのエスペラント精度は未知数です。これらは会議音声をシステムサウンドとして取得し、リアルタイムに字幕表示できる点で便利ですが、やはりエスペラント対応には限界があります。

メリット／デメリット: 既存機能なら追加設定が少なく手軽ですが、残念ながらエスペラントへの対応はどの組み込み機能も不十分です [8]。専門の拡張・アプリでもエスペラント非対応や精度低下のリスクがあります。あくまで英語など主要言語のみのサポートが中心であり、エスペラント会話に必ずしも対応しません。

比較まとめ

上記を比較すると、 **音声認識精度** はクラウドAPI（Speechmatics等）が最も高く、次いでVoskなど学習モデル、次にWhisper系、最後にブラウザ組込機能の順となります [1] [5]。**導入の容易さ** はGoogle MeetやChromeライブキャプションのような既存機能が最も簡単ですが、エスペラント対応が必要な場合は外部連携が必須です。Speechmatics等を用いた仮想オーディオ＋API型は構築コストはあるものの、精度・多言語対応の点で優れます。Voskによるローカル処理は費用ゼロでプライバシー面に優れますが精度と安定性はやや劣ります。ZoomのAPI連携は表示面で優れますが設定が複雑です。以上より、**エスペラント精度を重視**するならSpeechmaticsなど高精度API連携が最適ですが、コストと技術要件を考慮するとVosk等の組み合わせも有力な選択肢となります [1] [5] [7]。

参考資料: Speechmatics公式（エスペラント対応・精度） [1] [2]、Vosk公式（エスペラントモデル対応） [5]、Zoomヘルプ（Closed Caption API） [7]、Google Meetヘルプ（対応言語一覧） [8] など。

1 2 3 Free Esperanto Speech to Text | Transcribe Esperanto Voice and Audio to Text | Speechmatics
https://www.speechmatics.com/speech-to-text/esperanto

4 What languages do you support? | AssemblyAI | Documentation
https://www.assemblyai.com/docs/faq/what-languages-do-you-support-

5 GitHub - alphacep/vosk-api: Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node
https://github.com/alphacep/vosk-api

6 VOSK Models
https://alphacephei.com/vosk/models

7 Accessibility - FAQ | Zoom
https://www.zoom.com/en/accessibility/faq/

8 Use live captions in Google Meet - Computer - Google Meet Help
https://support.google.com/meet/answer/15077804?hl=en&co=GENIE.Platform%3DDesktop

9 LocalVocal: Local Live Captions & Translation On-the-Go | OBS Forums
https://obsproject.com/forum/resources/localvocal-local-live-captions-translation-on-the-go.1769/