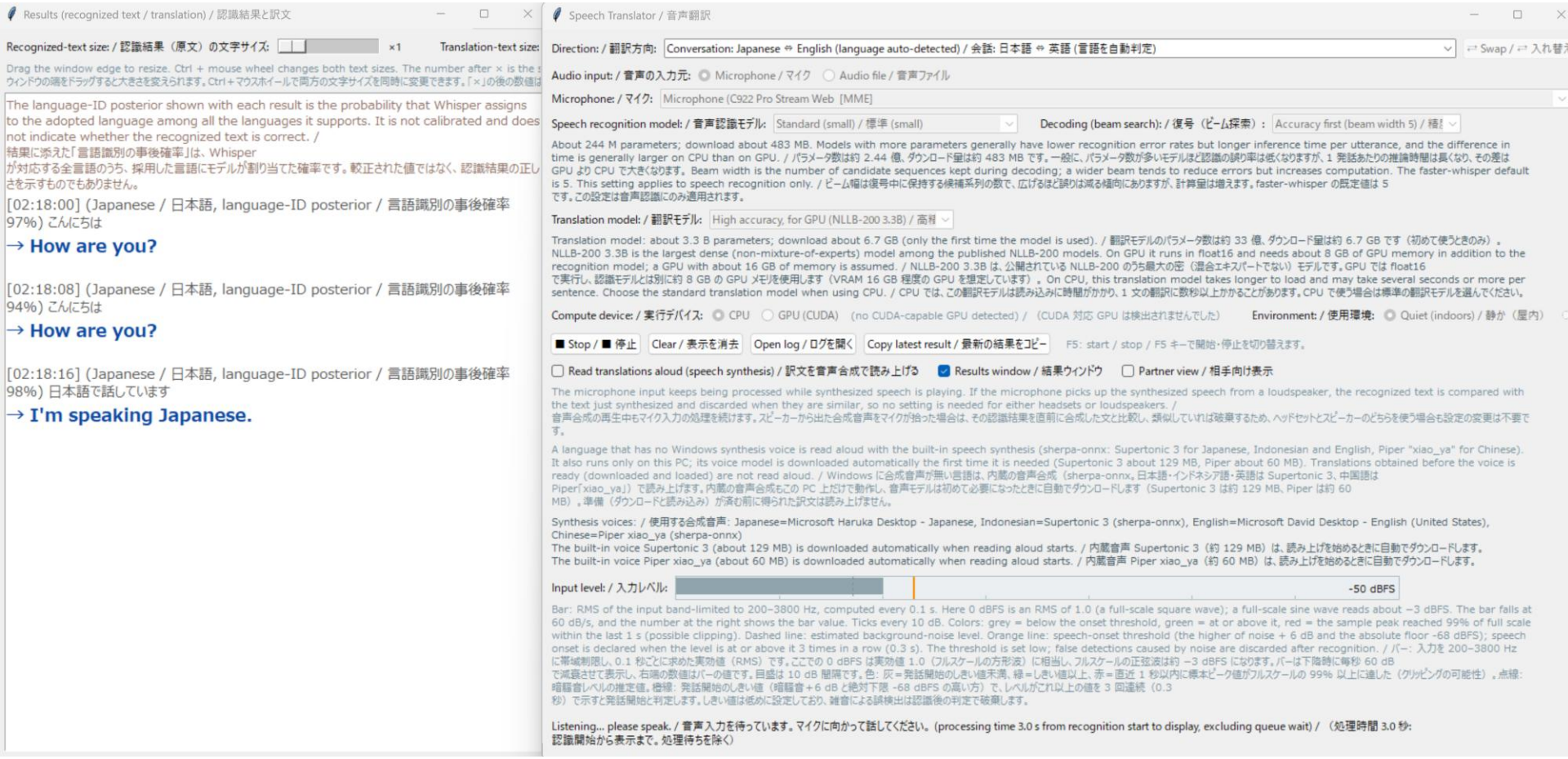
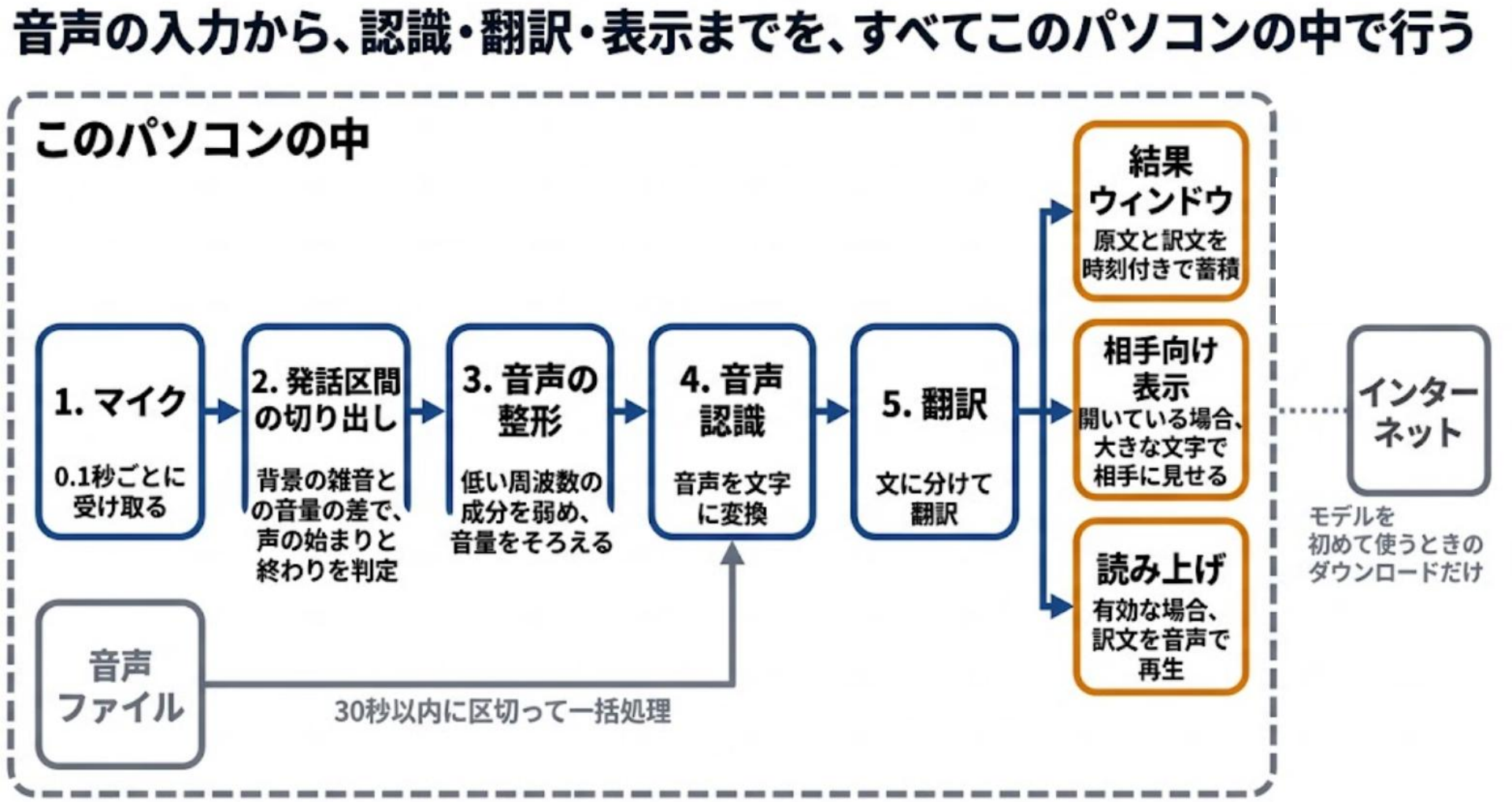
日本語・インドネシア語・英語・中国語のリアルタイム音声翻訳システム



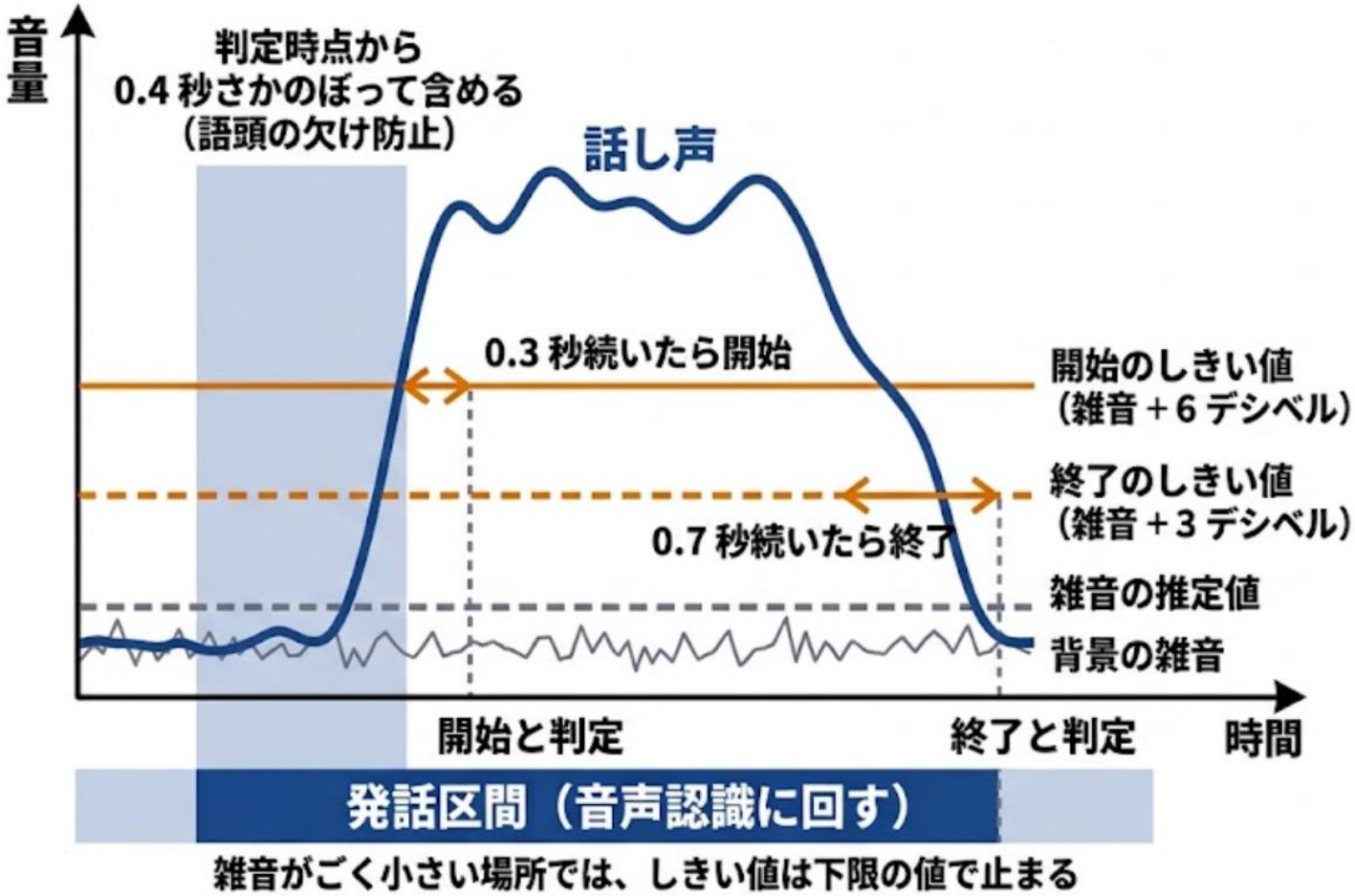
資料中の図などは Google Nano Banana 2 を用いて作成

処理の全体の流れ


音声の入力から、認識・翻訳・表示までを、すべてこのパソコンの中で行う
このパソコンの中
1. マイク
0.1秒ごとに受け取る
2. 発話区間の切り出し
背景の雑音との音量の差で、声の始まりと終わりを判定
3. 音声の整形
低い周波数の成分を弱め、音量をそろえる
4. 音声認識
音声を文字に変換
5. 翻訳
文に分けて翻訳
結果ウィンドウ
原文と訳文を時刻付きで蓄積
相手向け表示
開いている場合、大きな文字で相手に見せる
読み上げ
有効な場合、訳文を音声で再生
音声ファイル
30秒以内に区切って一括処理
インターネット
モデルを初めて使うときのダウンロードだけ

発話区間の切り出し方

背景の雑音を基準に、声が始まった所と終わった所を決める（使用環境「静か」の場合）



雑音がごく小さい場所では、しきい値は下限の値で止まる

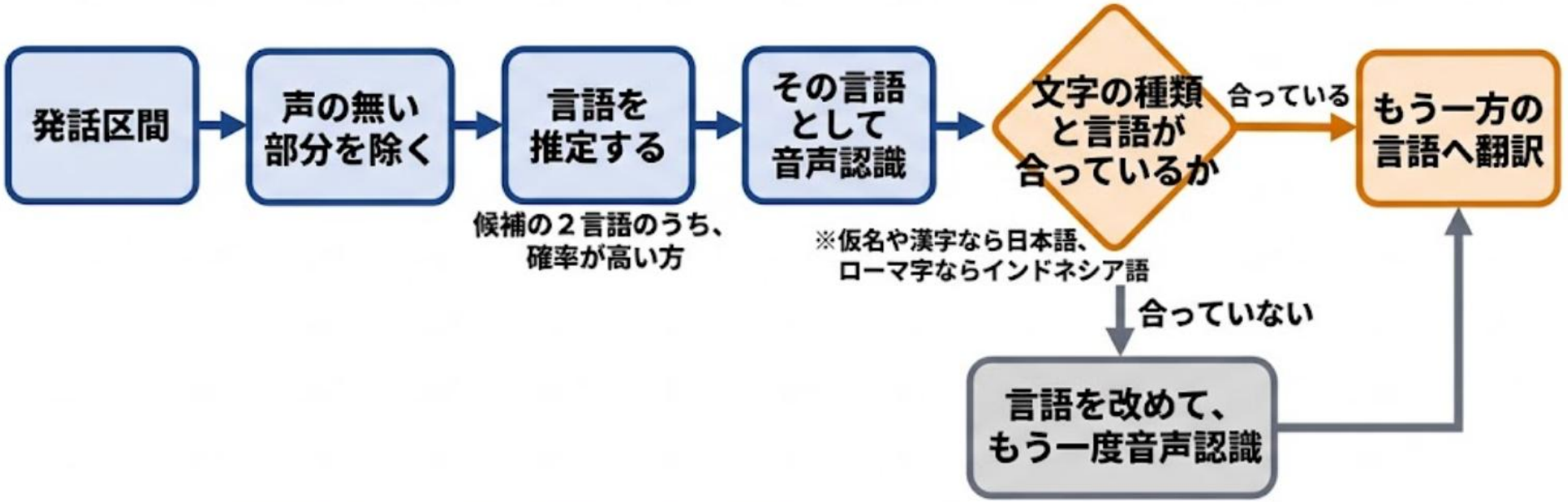
7 秒を超えたら、0.3 秒の息継ぎで区切る

12 秒に達したら、直近 3 秒で最も音量の小さい位置で区切る

表示までの待ち時間を短く保つため

会話モードで話された言語を決める手順

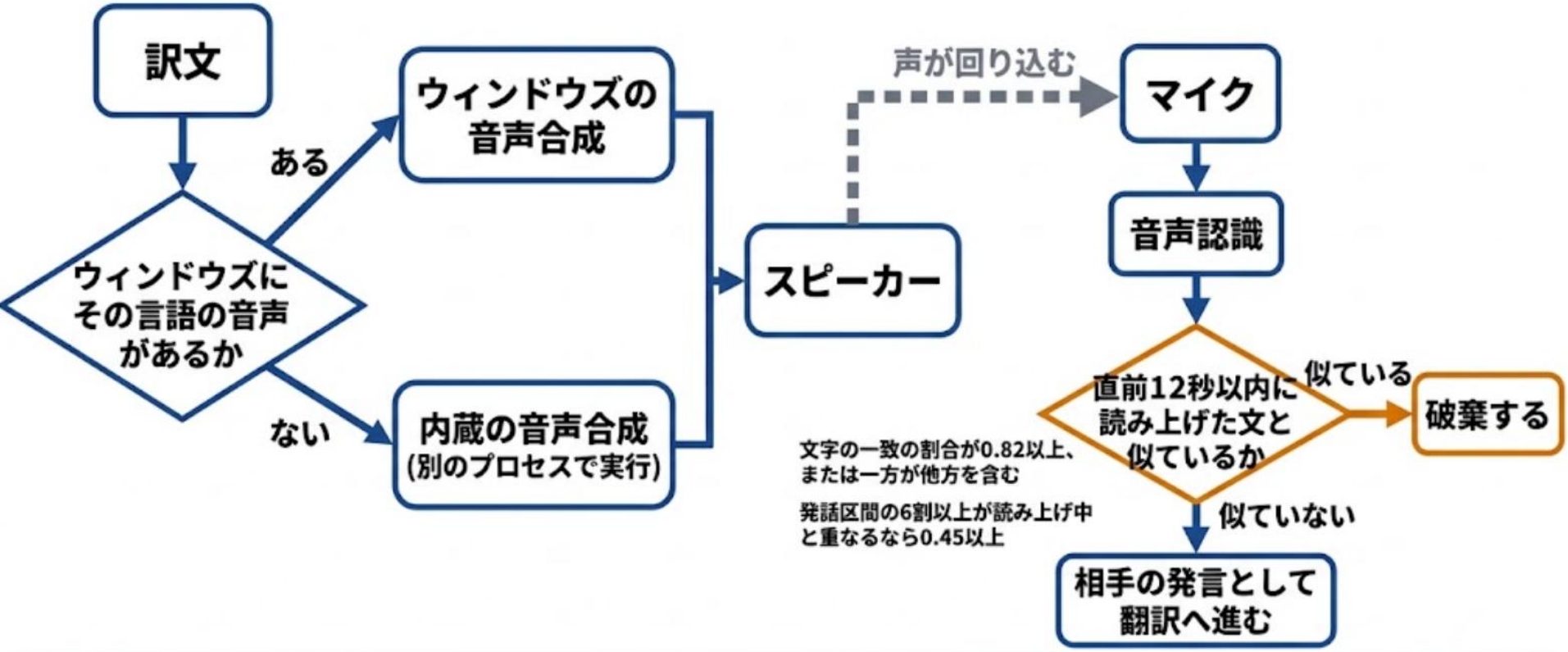
会話モード（例：日本語とインドネシア語）では、
話された言語を２つの候補から決め、もう一方の言語へ翻訳する



結果に添える『言語識別の事後確率』は、全言語に対する
モデルの出力値であり、文字の正しさを保証しない

読み上げと、本アプリの合成音声の除去

訳文を音声で読み上げ、その声をマイクが拾っても、本アプリの声として取り除く



ヘッドセットでもスピーカーでも、設定の切り替えは不要