# Improving Pause Detection in Voice AI Interviews

## Goal :

The goal of this prototype is to **demonstrate a clear improvement in pause detection** using a simple, browser-native approach.

Instead of building a full speech recognition system, the focus is on:

- Detecting when a candidate is *thinking* vs *done speaking*
- Preventing premature interruptions
- Making the interview flow feel more natural

The evaluation is impact-first, so clarity and effectiveness were prioritized over complexity.

## How the Prototype Works:

**High-level flow:**

Browser Microphone → Audio Energy Analysis → Pause Detection → Turn-taking State Machine → UI Debug Panel

**Key components:**

- **Browser microphone pipeline** using Web Audio API
- **Energy-based Voice Activity Detection (VAD)** using RMS values
- **Noise floor calibration** from initial silence
- **Adaptive silence threshold** instead of a fixed timeout
- **Explicit turn-taking states** (AI asking, user speaking, user pausing)
- **Real-time visual debugging panel** showing energy, silence duration, and thresholds
- **Simulated ASR / AI responses** to demonstrate interview flow

This entire system runs fully in the browser and does not require any backend.

## Key Design Decisions & Tradeoffs

**Why energy-based VAD instead of ML?**
The goal was to build something explainable, fast, and browser-friendly. Energy-based detection is easy to debug and sufficient to demonstrate the pause-handling improvement.

**Why adaptive silence thresholds?**
Human speech contains hesitation. A fixed silence timeout treats thinking pauses as end-of-answer. This prototype adjusts the allowed silence duration based on recent speaking activity.

**Why a state machine?**
Turn-taking logic becomes clearer and more reliable when modeled explicitly. This helps avoid incorrect AI interruptions.

**Limitations:**

- Sensitive to heavy background noise
- No real ASR feedback loop

**Future Improvements:**

- ML-based VAD (WebRTC / neural VAD)
- ASR-informed pause handling
- Speaker-specific calibration

## Evidence of Impact

**Before (Interview link):**

- Fixed silence timeout (~1s)
- Frequent interruptions during thinking pauses

**After (this prototype):**

- Longer tolerated pauses during active answers
- Fewer premature cut-offs
- More natural interview turn-taking

Impact is demonstrated using:

- Before/after audio samples
- Real-time debug visualization
- Manual test scenarios (hesitation, long answers, answer completion)