“Speaker ID” Audio ML Project

- Imagine you’re a <20 employee business owner having problems finding new leads & customers.
- You aren’t a SaaS expert and can’t afford to hire OR layoff.
- **You can’t afford +€1000/mo for a business coach who can help you grow with best practices and implementation**
- Voice AI can be the new modern small business coach for an affordable price, **but there’s an AI tech problem: Which voice is talking on the group calls?**

“Speaker ID” Audio ML Project

Teaching AI voice: “Who said what?”

- Current AI audio & meeting tools only keep time-stamp diaries of words, pauses, and anonymous voices *(“Speaker 1, Speaker 2…”)*
- **Need to know who’s talking for person-specific context on multiple calls with the same people** (AI consulting or internal compliance)

“Speaker ID” Audio ML Project

Process — (demo) Speaker ID Engine

- **Demo**: A consulting call between an AI business coach and 3 employees:
  - A fake medical transport company. Describing their problems finding new leads & customers.
  - **1. Charlie**: Business Owner (ElevenLabs: Archer - Conversational)
  - **2. Claire**: Lead Customer Service Rep (ElevenLabs: Hope - upbeat and clear)
  - **3. Paul**: Fleet Manager (ElevenLabs: Mark - ConvoAI)
  - **4. "Sam"**: The AI Business Coach (ElevenLabs: Cassidy)
- Gen audio: Fake Meeting 03

“Speaker ID” Audio ML Project

Process — (demo) Speaker ID Engine

Time Rank	Speaker	% Speaking	Time Speaking
0	Claire: Lead CS Rep	**38.66%**	28.50
1	Paul: Fleet Manager	**29.84%**	22.00
2	"Sam": AI Business Coach	**24.98%**	18.42
3	Charlie: Business Owner	**23.26%**	17.15
4	[Unknown]	**18.90%**	13.93

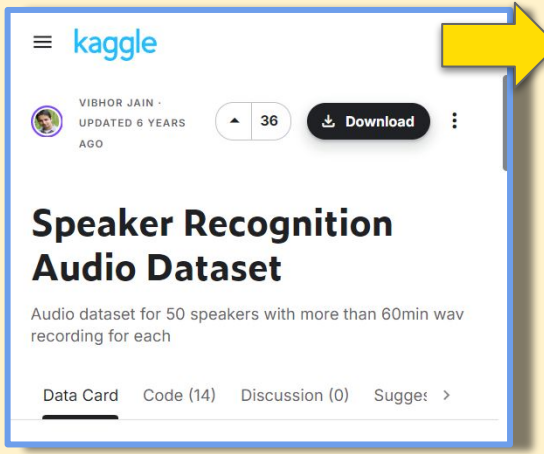
“Speaker ID” Audio ML Project

Solution — A Speaker Identification Engine

- Short enrollment recordings (30–90s per person) to learn new “voiceprints” — practical during the speaker’s first explanation of their job role
- Run future meetings through this pipeline:
  - Splits the audio into speaker segments (existing technology)
  - Matches each segment to the closest enrolled person
- Output: **A transcript labeled with real names, not “Speaker 1/2/3”**

“Speaker ID” Audio ML Project

Process — (high-level) How it Works



1. **Train Base Model**: Built voice pattern model using 50 hrs of Kaggle speech recordings from 50 people
2. **Enroll New Voices**: Capture 30-90 seconds of each known speaker’s voice
3. **Analyze Meeting Audio**: Diarize meeting audio → segment “who spoke when”
4. **Identification**: Match each segment to enrolled voices (or mark “Unknown”)
5. **Output**: Timeline “Justin” talking, **tied to a file on their running job goals, wins, and roadblocks**

“Speaker ID” Audio ML Project

Early Results & Next Step Roadmap

- **Next steps**: Improve accuracy on <2 second “Unknown” clips:
  - **ASR+LLMs**: Guess who’s talking based on sentence context
  - **top-k nearest enrolled speakers + margin**: Make approximate guesses
  - Build this into a full front-end Chat interface in a page on my Shopify website
- **Product for AI consulting** (applying for a job right now with this use case) **and internal meeting visibility**
- **Let’s connect on LinkedIn: https://www.linkedin.com/in/justin-anderson-product-manager/**

“Speaker ID” Audio ML Project

asdf

- asdf

“Speaker ID” Audio ML Project

Impact — Why This Matters

1. Action items may be misattributed
2. Teams lose context when reviewing “Speaker 00” meeting notes
3. **AI-automated business consulting isn’t possible until this is fixed**