

Note 2 — Make the World Sound Alive: Music & Voice with AI

1) What you'll learn

- How AI represents audio as **spectrograms** (pictures of sound).
- Prompting music and voice logically (purpose, genre, emotion, structure).
- Academic concepts: **sampling rate, Fourier transform, vocoder, TTS (text-to-speech), prosody, phonemes, alignment, evaluation.**
- Multilingual narration while preserving emotion and pacing.

2) Sound as pictures (spectrograms 101)

- **Spectrogram:** Time on the x-axis, frequency (pitch) on the y-axis, and darkness/brightness showing loudness.
- **Fourier Transform (idea):** A way to break complex sounds into simple waves. The computer analyzes these to understand and recreate audio.
- **Sampling rate:** How often sound is measured per second (e.g., 16,000 or 44,100 samples/sec). Higher rates can keep more detail.

3) The music pipeline (simplified)

- Your **prompt** → AI plans the **melody/harmony/rhythm** in spectrogram form → a **vocoder** (like WaveNet-style or others) turns that spectrogram into a waveform you can hear.
- **Structure matters:** Telling the model how the piece should develop over time (intro → build → finish) guides the energy.

4) Voice pipeline (simplified TTS)

- **Text normalization:** Expand numbers/dates ("2025" → "twenty twenty-five").
- **Phonemes:** Small sound units (like /k/, /aa/). Converting text to phonemes helps accurate pronunciation.
- **Prosody:** Rhythm, stress, and intonation—the musicality of speech.

- **Alignment:** Matching phonemes to the right time positions so speech sounds natural and synced to video.
- **Vocoder:** Final step converting features into clean audio.

5) Prompting templates

Music prompt: Purpose + Genre/Instruments + Emotion + Structure + Length.

“45-second Indian classical fusion with sarod, soft choir, gentle dholak. Solemn→hopeful arc: 0–10s sarod solo, 10–25s choir fades in, 25–45s steady dholak pulse. Clear sound; no distortion.”

Voice prompt: Audience + Tone + Length + Key words to emphasize + Pace.

“Narration for Grade 8 science on the water cycle. Calm, warm tone, ~70 words. Emphasize ‘evaporation,’ ‘condensation,’ ‘precipitation.’ Moderate pace; clear pauses after each term.”

Multilingual request:

*“Re-voice the same script in Hindi and Tamil with the **same prosody and pacing** as the English version. Keep technical words consistent.”*

6) Academic-style evaluation (kid-friendly)

- **Intelligibility:** Can you clearly understand words?
- **Naturalness:** Does it sound like a real person?
- **Emotion match:** Happy when it should be, calm when needed.
- **Timing:** Does music swell at the right moments? Are pauses helpful?

7) Interactive Activity (10–12 minutes)

Choose one topic: “Parts of a cell” or “Solar cooker basics.” 1) Write a 35–60 word script in English. Mark **emphasis words** with * (e.g., *cell membrane*). 2) Draft a music prompt with structure (intro → build → close). 3) (If tools available) Generate. If not, swap prompts and peer-review using the evaluation checklist.

8) Troubleshooting & fixes

- **Music too loud over voice** → “Lower music –6 dB; sidechain compress under voice.”

- **Muddy mix** → “Reduce low frequencies; increase clarity for 2–4 kHz (voice area).”
- **Pronunciation errors** → Add phonetic hints: “Srinivasan (Shree-nee-vaa-sun).”
- **Emotion mismatch** → Specify “gentle, hopeful; avoid dramatic drums.”

9) Ethics and good sense

- Don’t copy famous songs or voices without permission.
- Use respectful language and avoid mocking accents.
- Add a small credit line: “Audio created with AI for education.”
- Privacy: Do not share classmates’ voices without consent.

10) Mini-project (20–30 minutes)

Create a **soundtrack pack** for one school topic: - One 45-second background music. - One 45-60 word narration (English). - Optional: Hindi or Tamil version with the **same prosody**. - Submit prompts + audio files (or a written “expected sound” if generation isn’t possible).

11) Reflection

- One audio concept I liked (spectrogram/vocoder/prosody): _____
- A fix that improved clarity: _____
- How I made it inclusive for classmates: _____