*2026 AI Skills Series*

# DAY 7

## AI Audio & Voice Generation

*Your Personal Recording Studio, Powered by AI*

Student eBook & Reference Guide

90-Minute Virtual Classroom | Complete with Workbook & Assignments

### How to Use This eBook

Welcome to Day 7. You have now mastered the visual side of AI creation — prompting, writing, research, images, presentations, and video. Today you add the final sensory dimension: sound. The difference between a good video and a great one is often the audio. A professional voice-over, clean sound quality, and the right background music can transform any AI-generated clip into a polished production.

Today's session covers the three pillars of AI audio: Voice Generation (narration that sounds human), Audio Enhancement (turning bad recordings into studio quality), and Sound Design (original music and effects generated on demand).

| 🎙️ VOICE | 🔊 ENHANCE | 🎵 COMPOSE |
|---|---|---|
| Generate and clone voices | Clean and upgrade recordings | Generate original music & SFX |

# Table of Contents

Chapter 1: The Audio Layer — Why Sound Completes the Picture

Chapter 2: Your Audio Toolkit — Tools for Every Sound Need

Chapter 3: Pillar 1 — Voice Generation with ElevenLabs

Chapter 4: The Vocal Persona Formula — Performance Style Prompting

Chapter 5: Pillar 2 — Audio Enhancement with Adobe Podcast

Chapter 6: Pillar 3 — Music & Sound Design with Suno and Udio

Chapter 7: Practice Drills — Voice, Enhancement, and Music

Chapter 8: Advanced Techniques — Voice Cloning, SFX, and Multilingual Audio

Chapter 9: Combining Audio with Video — The Day 6 + Day 7 Integration

Chapter 10: Ethics of AI Audio — Consent, Cloning, and Authenticity

Chapter 11: Workbook — In-Session Exercises

Chapter 12: Day 7 Assignment — Your Complete Audio Production

Chapter 13: References & Further Learning

# Chapter 1: The Audio Layer — Why Sound Completes the Picture

## 1.1 The Most Underestimated Production Element

Film schools teach a principle that surprises most students: audiences will forgive poor video quality far more readily than poor audio quality. A slightly blurry image is tolerable. A robotic voice, background noise, or a jarring music mismatch makes people stop watching within seconds.

In 2026, every professional AI content creator understands that audio is not a secondary consideration — it is the foundation of emotional impact. The same AI-generated video clip will feel completely different depending on the voice, the music, and the sound design layered over it.

> 🎧 **The 50/50 Rule of Production Quality**
> Professional media producers say that audio accounts for 50% of the perceived quality of any video. This is why Hollywood spends as much on sound as on visual effects. Your AI video from Day 6 is visually impressive. Today you add the other 50%.

## 1.2 The Three Pillars of AI Audio

Day 7 is structured around three distinct capabilities, each solving a different audio problem:

| Pillar | The Problem It Solves | Primary Tool |
|---|---|---|
| 🎙️ **Voice Generation** | You need a professional narrator for a video but don't want to record yourself, or you need a voice in a language you don't speak | ElevenLabs — elevenlabs.io |
| 🔊 **Audio Enhancement** | You have a recording but it sounds amateurish — background noise, echo, thin sound quality from a phone mic | Adobe Podcast Enhance — podcast.adobe.com/enhance |
| 🎵 **Sound Design** | You need background music that matches your brand and video perfectly — but you cannot afford a composer or a music licence | Suno — suno.com \| Udio — udio.com |

## 1.3 What You Will Produce in This Session

- A professional-quality voice-over narration — using a styled vocal persona prompt
- A studio-quality audio recording enhanced from a phone recording
- An original piece of background music generated for your specific content

- A complete audio layer ready to merge with your Day 6 video
- A cloned version of your own voice (optional but transformative)

# Chapter 2: Your Audio Toolkit

| Tool | Best For | Access | Key Strength |
|---|---|---|---|
| **ElevenLabs** | Ultra-realistic voice narration and voice cloning | Free tier — elevenlabs.io | Industry-leading voice quality with emotion, whispers, emphasis. Instant voice clone from 60 seconds of audio |
| **Adobe Podcast(Enhance)** | Fixing bad audio recordings instantly | Free — podcast.adobe.com/enhance | One-click 'studio quality' from phone recordings. Removes noise, echo, and room reverb automatically |
| **Suno v4.5** | Generating full original songs with lyrics and instruments | Free tier — suno.com | Text-to-full-song. 4-minute tracks with vocals, instruments, and professional mixing from a one-sentence description |
| **Udio** | Custom music, specific genres, instrumental-only tracks | Free tier — udio.com | More granular control than Suno. Better for instrumental background tracks with custom tempo and mood |
| **Murf AI** | Professional voice-over with presenter avatars | Free trial — murf.ai | Studio-quality narration with visual presenter. Integrates voice + talking avatar in one tool |
| **ElevenLabsText-to-SFX** | Sound effects generated from text descriptions | Built into ElevenLabs — elevenlabs.io | 'Sound of a door creaking open in an old castle' → realistic SFX in seconds. Massive time-saver for video production |

> ⭐ **Session Tool Recommendations**
>
> Use ElevenLabs for Drills 1 and 2 (voice generation and cloning). Use Adobe Podcast Enhance for Drill 3 (audio cleanup). Use Suno for Drill 4 (music generation). All have free tiers that are fully sufficient for today's exercises. No payment required.

# Chapter 3: Pillar 1 — Voice Generation with ElevenLabs

## 3.1 Why Default AI Voices Sound Robotic

When most people first use a text-to-speech tool, they paste their text, click generate, and receive audio that sounds like a GPS navigation system reading a shopping list. Flat. Monotone. Robotic. Every sentence sounds the same. There is no breath, no emotion, no variation in pace.

This is not a technology limitation — it is a prompting limitation. ElevenLabs' voices are capable of remarkable nuance: warmth, urgency, whispers, dramatic pauses, excitement, gravity. But the AI defaults to neutral unless you tell it what you want. The Vocal Persona Formula is how you unlock this capability.

## 3.2 ElevenLabs Voice Library

ElevenLabs maintains a library of over 3,000 voices categorised by gender, age, accent, and use case. For professional content, the most useful categories are:

- Narration voices — warm, confident, authoritative; ideal for documentary, e-learning, and training videos
- News / Corporate voices — crisp, clear, professional; ideal for business communications and presentations
- Storytelling voices — expressive, varied, emotional; ideal for creative content and brand storytelling
- Conversational voices — natural, relaxed, warm; ideal for podcasts, social media, and informal training
- Promotional voices — energetic, bright, forward-leaning; ideal for marketing content and product videos

## 3.3 Setting Up ElevenLabs

1. Go to elevenlabs.io and create a free account.
2. From the dashboard, click 'Text to Speech.'
3. Paste your script into the text box.
4. Click 'Voice Library' to browse and select a voice.
5. Adjust Voice Settings (see Chapter 3.4) before generating.
6. Click 'Generate' and listen to the preview.
7. Download as MP3 for use in CapCut, Gamma, or any video/presentation tool.

## 3.4 ElevenLabs Voice Settings — The Controls That Matter

| Setting | What It Does | Practical Guidance |
|---|---|---|

| | | |
|---|---|---|
| **Stability** | Controls how consistently the voice stays 'in character' | High = steady and reliable. Low = more expressive, takes emotional risks. Lower it to 40-50% for storytelling or promotional content |
| **Similarity Boost** | How closely the output matches the reference voice | Keep at 75-85% for a natural balance between character and flexibility |
| **Style Exaggeration** | How dramatically the voice performs style cues in the script | 0% = subtle. 50%+ = theatrical. Use sparingly unless you want a very expressive performance |
| **Speaker Boost** | Enhances the clarity and presence of the voice | Toggle on for noisy environments or when clarity is critical (e.g. instructional content) |

# Chapter 4: The Vocal Persona Formula

## 4.1 The Problem with Pasting Text

Most people use AI voice tools like a printer: paste the text, press generate, receive audio. The result is technically accurate but emotionally flat. The voice reads the words correctly but does not feel them — and audiences hear the difference immediately.

Professional voice-over directors give their talent detailed performance direction: 'Start warm and welcoming, build to confident urgency in the middle section, close with a soft, inspirational tone. Take a full breath pause after the question in paragraph two.' You can give ElevenLabs the same instructions using the Vocal Persona Formula.

## 4.2 The Vocal Persona Formula

| Formula Element | What to Specify | Example Values |
|---|---|---|
| **Tone** | The emotional quality of the voice | *"Warm and storytelling" \| "Authoritative and confident" \| "Energetic and upbeat" \| "Calm and reassuring"* |
| **Pacing** | The speed and rhythm of delivery | *"Slow and deliberate" \| "Conversational pace" \| "Brisk and energetic" \| "Varied — fast in the examples, slow on key points"* |
| **Pauses** | Where to add dramatic or natural breathing space | *"Short pause after the opening question" \| "Breathe after each key point" \| "Hold 2 beats before the conclusion"* |
| **Emphasis** | Which words should carry extra weight | *"Stress the word 'free' in the third sentence" \| "The phrase 'right now' should feel urgent"* |
| **Emotional Arc** | How the feeling changes across the script | *"Start curious and questioning, build to confident, close with inspiring warmth"* |

## 4.3 The Vocal Persona Prompt — Assembled

BASIC APPROACH: **No Performance Direction — Generic Result**

```
Hello, and welcome to this training module on artificial intelligence in the workplace.Today we will explore three key applications of AI that can immediately improve your productivity.
```

The above delivers the words correctly. But it sounds like a text-to-speech disclaimer at the end of a radio advert.

**VOCAL PERSONA APPROACH: Full Performance Direction**

*[Deliver this script in a warm, storytelling voice with slow, deliberate pacing.][Add a short pause for 1 second after the opening hook.][Build confidence gradually — by the third sentence, the voice should feel inspiring and authoritative.][Stress the word 'three' in the final sentence.]Hello — and welcome. [PAUSE]The way you work is about to change completely. Not in years. Not in months.Right now, today, there are three AI tools that can give you back three hours of your working week — starting tomorrow.*

💡 **Square Bracket Performance Instructions**

In ElevenLabs, place performance directions inside square brackets [like this] throughout your script. The AI reads these as instructions, not as spoken text. This allows you to place emotion and pause instructions exactly where they are needed — not just at the top of the prompt.

## 4.4 Phonetic Spelling — When AI Gets Names Wrong

AI voice tools occasionally mispronounce uncommon words, names, and African-language terms. Use phonetic respelling to fix this:

| Written Word | Common AI Mispronunciation | Phonetic Fix |
|---|---|---|
| **Gemini** | GEM-in-ee | **"Jem-in-eye"** |
| **Abuja** | Ab-OO-ja | **"Ah-BOO-jah"** |
| **Seun** | Seen | **"Shay-OON"** |
| **Naira** | NAY-ra | **"NAH-eer-ah"** |
| **Okafor** | Oh-KAH-for | **"Oh-KAH-faw"** |
| **SaamSoft** | SAAM-soft | **"SAHM-sawft"** |

# Chapter 5: Pillar 2 — Audio Enhancement with Adobe Podcast

## 5.1 The Studio Quality Hack

Adobe Podcast Enhance is one of the most immediately impressive AI tools in this entire series. You upload any audio recording — recorded on a cheap phone in an echoey kitchen, a car, or a noisy office — and it outputs the same recording with professional studio quality.

The AI does not just reduce noise. It actively reconstructs the audio signal, removing room reverb, cleaning background noise, and enriching the voice frequency to produce the warmth and clarity of a recording made in a professional studio booth.

## 5.2 How to Use Adobe Podcast Enhance

8. Open podcast.adobe.com/enhance in your browser — no account required for basic use.
9. Click 'Enhance Speech.'
10. Upload your audio file (MP3, WAV, M4A, or MP4 with audio).
11. Click 'Enhance' and wait 30-90 seconds depending on file length.
12. Listen to the before/after comparison using Adobe's split player.
13. Download the enhanced file as WAV.

> 🏠 **The Room Acoustics Secret**
> The best recording environments for phone recording (before enhancement) are: inside a wardrobe full of clothes, inside a car with the windows up, under a heavy duvet, or in a small bathroom. These environments naturally reduce echo and room noise — giving Adobe Enhance cleaner input material to work with. The combination of the right recording environment + AI enhancement produces results indistinguishable from a proper studio recording.

## 5.3 Other Enhancement Capabilities

- Adobe Podcast also has a 'Studio' feature for live recording — it applies real-time noise cancellation and enhancement as you speak
- For podcast episodes, it can remove filler words ('um', 'uh', 'er') automatically — a separate feature called 'Remove Filler Words'
- Krisp.ai is a real-time noise cancellation tool that works inside video calls — it removes background noise from your microphone in Zoom, Teams, or Google Meet
- LALAL.AI can separate voice from background music in an existing recording — useful if you need to extract a voice-over from a mixed track

# Chapter 6: Pillar 3 — Music & Sound Design with Suno and Udio

## 6.1 The End of Royalty Headaches

Using music in content has always been a legal minefield. YouTube Content ID, Instagram music restrictions, licensing fees, and copyright strikes have made it genuinely difficult to add music to professional video content without either paying significant fees or using generic, often low-quality royalty-free tracks.

In 2026, AI music generation eliminates this problem entirely. With Suno or Udio, you describe the music you want, and an original piece is generated — unique, in exactly the style you request, with no licensing concerns. The music is yours.

## 6.2 Suno v4.5 — Full Song Generation

Suno generates complete songs with lyrics, melody, instrumentation, and professional mixing from a single text description. The quality in 2026 is remarkable — tracks that are genuinely listenable, with appropriate genre conventions and emotional character.

### Basic Music Prompt

BASIC SUNO PROMPT: **Too Vague**

```
"Background music for a video."
```

Result: generic, could be anything, does not serve your specific content.

### Professional Music Prompt

**PROFESSIONAL SUNO PROMPT: Full Description**

```
A lo-fi, chill hip-hop track for a travel vlog about West Africa.Acoustic guitar
lead melody, soft brushed drums, warm bass.No vocals — purely instrumental.Mood:
peaceful, inspiring, slightly nostalgic — like watching a sunset over the
Sahara.Tempo: slow to moderate, unhurried.Length: 3-4 minutes with a natural
fade-out.
```

## 6.3 Suno Style Tags — Quick Reference

You can also use genre and style tags in Suno to quickly narrow the output. Combine multiple tags for precision:

| Genre / Style | Best Content Use | Add These Tags For... |
|---|---|---|
| **lo-fi hip-hop** | Study content, tutorials, background learning | *"no vocals, soft drums, warm" for pure ambience* |
| **corporate / inspiring** | Business presentations, pitch videos | *"uplifting, piano-led, builds to triumphant" for pitch climax* |
| **cinematic orchestral** | Documentary, brand films, dramatic scenes | *"strings, no vocals, epic build" for maximum impact* |
| **Afrobeat / Afropop** | African market content, cultural celebrations | *"West African percussion, joyful, energetic" for vibrant content* |
| **ambient / meditation** | Wellness, calm explainer videos, e-learning | *"drone pads, no melody, serene" for focused attention* |
| **jazz / bossa nova** | Luxury, lifestyle, food and travel content | *"brushed drums, double bass, cool" for sophisticated feel* |

## 6.4 Text-to-Sound Effects (SFX) — ElevenLabs

The ElevenLabs SFX feature generates any sound effect you can describe in text. This is transformative for video production:

**SFX PROMPT EXAMPLES: Sound Effects from Text**

```
"The sound of a futuristic door sliding open with a hydraulic hiss and electronic tone.""Heavy rain falling on a tin roof in a West African village at night, distant thunder.""A crowd of about 50 people applauding enthusiastically in an outdoor space.""The sound of a technology startup office – keyboards typing, a notification ping, quiet conversation.""5-second intro jingle: upbeat, modern, tech brand, ascending notes, confident ending."
```

# Chapter 7: Practice Drills

## 7.1 Drill 1 — Voice Narration with the Vocal Persona Formula

Use ElevenLabs. This drill shows you the difference a performance direction makes.

### Step 1 — The Bare Paste

14. Go to elevenlabs.io → Text to Speech.
15. Select any narration voice from the Voice Library.
16. Paste only this script — no instructions:

**BARE SCRIPT: No Performance Direction**

```
In 2026, artificial intelligence has changed the way we work, create, and communicate.The tools available today are more powerful than anything seen in previous decades.For professionals who embrace these tools, the opportunity is extraordinary.
```

17. Click Generate. Listen. Note how it sounds.

### Step 2 — The Vocal Persona Upgrade

18. Now add full performance direction to the same script:

**VOCAL PERSONA SCRIPT: Full Performance Direction**

```
[Deliver in a warm, authoritative storytelling voice. Start with quiet gravitas and build steadily.][Add a 1-second pause after the first sentence.][Stress the word 'extraordinary' in the final sentence — deliver it slowly and with conviction.]In 2026, artificial intelligence has changed the way we work, create, and communicate. [PAUSE]The tools available today are more powerful — more transformative — than anything seen in previous decades.For professionals who embrace these tools — who truly learn them — the opportunity is extraordinary.
```

Generate both versions and compare. The difference demonstrates exactly what the Vocal Persona Formula does.

## 7.2 Drill 2 — The Studio Quality Hack

Use Adobe Podcast Enhance. You need a short audio recording for this drill.

19. Record yourself speaking for 30-60 seconds using your phone's voice recorder app. Do not go to a quiet room — record in a normal, slightly echoey environment (kitchen, office, corridor).

20. Go to podcast.adobe.com/enhance.
21. Upload your recording.
22. Click 'Enhance Speech.'
23. Listen to the before/after split comparison. Screenshot or note the quality difference.
24. Download the enhanced file.

> 📱 **Can't Record Right Now?**
> If you cannot record live during the session, use any low-quality audio file you have — a phone voice note, a WhatsApp voice message, or a recording from a previous meeting. Adobe Enhance works on any audio with a human voice. The results are most dramatic with recordings that have significant background noise or room echo.

## 7.3 Drill 3 — Generate Original Background Music

Use Suno (suno.com) for this drill.

25. Go to suno.com and create a free account.
26. Click 'Create' → 'Custom Mode' (for more control) or just the main text field.
27. Enter this prompt:

**SUNO PROMPT: Background Music for a Professional Video**

```
A lo-fi, chill hip-hop instrumental track for a travel vlog about West
Africa.Acoustic guitar melody, soft brushed drums, warm bass groove.No vocals —
purely instrumental.Mood: peaceful, inspiring, subtly nostalgic.Tempo: slow and
unhurried.Fade out naturally at the end.
```

28. Generate 2 variations and choose your favourite.
29. Download as MP3.

Then try this shorter version for a 5-second intro jingle:

**JINGLE PROMPT: 5-Second Brand Intro**

```
A 5-second upbeat tech brand intro jingle.Short ascending notes, confident and
modern.Like a startup app notification sound — but more musical and
professional.Clean, minimal, memorable.
```

# Chapter 8: Advanced Techniques

## 8.1 Instant Voice Cloning

ElevenLabs' Instant Voice Cloning feature creates a digital replica of any voice from just 60 seconds of clean audio. Once cloned, this voice can narrate any text — including text in 29 languages — with your vocal tone, character, and personality.

### Step-by-Step Voice Cloning

30. Go to elevenlabs.io and navigate to 'Voice Lab' → 'Add a Generative or Cloned Voice.'
31. Select 'Instant Voice Clone.'
32. Upload a 60-second recording of yourself speaking clearly — no background noise, no music.
33. Name your cloned voice.
34. Go to Text to Speech, select your cloned voice, and type any script.
35. Generate — hear yourself speaking text you never actually recorded.
36. Try the multilingual button: switch language to French, Spanish, or Swahili — the AI speaks in your voice in that language.

> ⭐ **The 60-Second Recording Standard**
> For best cloning results, record yourself speaking naturally for 60 seconds. Not reading — speaking as you would in a conversation. Use a quiet room. No music. No background noise. The richer and more natural the source recording, the more expressive and accurate the clone. A noisy or flat recording produces a flat, noisy clone.

## 8.2 SSML — Speech Synthesis Markup Language

SSML (Speech Synthesis Markup Language) is a professional standard for controlling AI voice output with precision. ElevenLabs supports a subset of SSML tags:

- Pause: <break time="1.5s"/> — inserts a 1.5 second pause at that exact point
- Emphasis: <emphasis level="strong">extraordinary</emphasis> — stresses a specific word
- Rate: <prosody rate="slow">This is delivered slowly</prosody> — controls local pacing
- Whisper effect: <amazon:effect name="whispered">Delivered as a whisper</amazon:effect>

These tags can be embedded directly in your script text alongside the square bracket performance instructions for maximum control.

## 8.3 Stem Separation — Isolating Instruments

LALAL.AI and Spleeter (open-source) can separate any piece of music into its individual stems: vocal, drums, bass, melody, and other. This is useful when you have an existing piece of music and want to use only the instrumental version, or you want to create a remix. These tools are more advanced and are referenced here as a 'further exploration' option.

## 8.4 Real-Time Voice Enhancement for Video Calls

For professionals who present or conduct meetings online, Krisp.ai is a standalone real-time noise cancellation app that works inside any video call platform — Zoom, Google Meet, Microsoft Teams. It:

- Removes all background noise from your microphone in real time
- Cancels echo and room reverb on your outgoing audio
- Suppresses background voices (other people in the room) without affecting your voice
- Has a free tier that covers most professional use cases

# Chapter 9: The Day 6 + Day 7 Integration

## 9.1 Adding Audio to Your Video — The Complete Workflow

This chapter combines the video clips you generated in Day 6 with the audio you produce today. The result is a complete audio-visual production — created entirely with AI tools and a free video editor.

| Step | Action | How |
|---|---|---|
| 1 | **Generate narration** | Write a 1-3 sentence script for your Day 6 video clip. Apply Vocal Persona Formula. Generate in ElevenLabs. Download as MP3. |
| 2 | **Generate background music** | Describe the mood and style of your video in Suno. Generate an instrumental track. Download as MP3. |
| 3 | **Generate 5-second intro jingle** | Use the jingle prompt from Chapter 7.3 in Suno. Keep it short and memorable. |
| 4 | **Open CapCut** | Import your Day 6 video clip as the base. Create a new project at capcut.com or in the CapCut app. |
| 5 | **Add audio layers** | Drag your narration MP3 onto the audio track below the video. Drag your music MP3 onto a second audio track. The music should play under the narration at reduced volume (around 20-30%). |
| 6 | **Add the intro jingle** | Place the 5-second jingle at the very start of the timeline, before the main content begins. |
| 7 | **Add captions (optional)** | CapCut has auto-caption: click the narration track → 'Auto Captions.' AI transcribes the voice and places captions precisely. |
| 8 | **Export as MP4** | File → Export → MP4 at 1080p. Your complete AI-produced audio-visual piece is ready to share. |

🎯 **Audio Mixing Levels — The Professional Balance**

In any video with both narration and background music, use these approximate volume levels for a professional mix: Narration: 100% (or -6 dB). Background music: 20-25% (or -20 dB). Sound effects: 60-70% (or -10 dB). The narration should always be clearly intelligible — music should be felt, not heard competing with the voice.

# Chapter 10: Ethics of AI Audio

## 10.1 Voice Cloning Consent — Absolute Requirement

Voice cloning is among the most powerful and most easily misused capabilities in AI audio. The ability to make a voice say anything opens obvious avenues for fraud, impersonation, misinformation, and abuse.

- You may clone your OWN voice freely for any personal or professional use.
- You may clone another person's voice ONLY with their explicit, specific, written consent.
- You may NEVER clone the voice of a celebrity, politician, public figure, or any private individual without consent.
- Using a cloned voice to impersonate someone for deception, fraud, or character assassination is criminal in most jurisdictions as of 2025-26.

## 10.2 AI Audio in Journalism and Official Communications

The growing sophistication of AI voice cloning has created a verification crisis in media and public life. In 2025-26:

- AI-generated audio of political figures saying things they never said has already influenced elections in multiple countries.
- Business email compromise scams using cloned CEO voices to authorise fraudulent wire transfers have increased dramatically.
- Most responsible media organisations now require audio verification for any audio evidence presented in journalism.

> ⚠️ **The Disclosure Obligation**
> Any AI-generated voice content that could reasonably be mistaken for a real person's voice must be clearly disclosed as AI-generated in its description, caption, or accompanying material. This applies to marketing content, social media posts, training materials, and any published communication. Concealing AI voice generation when it matters to the listener is deceptive, regardless of the intent.

## 10.3 Copyright and AI Music

The legal status of AI-generated music is still actively contested in 2026. Practical guidance:

- Music generated entirely by AI tools like Suno or Udio is generally considered to have no copyright protection in the EU and US (as of 2026) — you cannot enforce copyright on it, but others generally cannot either.
- Some AI music tools (including Suno) have updated terms that allow commercial use of generated tracks. Read the current Terms of Service of each tool before commercial deployment.
- AI music that closely imitates a specific existing artist's style may face legal challenge under developing AI music case law. Describe the style generically rather than naming specific artists.

- For enterprise use, Soundful, Artlist, and Epidemic Sound offer AI-generated music with explicit commercial licensing.

# Chapter 11: Workbook — In-Session Exercises

Complete during the live classroom session.

## Exercise 1: The Vocal Persona Comparison

Generate both versions of the script from Drill 1 in ElevenLabs and answer these questions.

| | |
|---|---|
| 1 | **What tone and pacing did you specify in your Vocal Persona performance direction?**<br>Your answer: _____ |
| 2 | **Describe the most noticeable difference between the basic and Vocal Persona outputs:**<br>Your answer: _____ |
| 3 | **Which square bracket performance instruction had the biggest impact on the delivery?**<br>Your answer: _____ |

## Exercise 2: The Studio Quality Hack

| | |
|---|---|
| A | **What environment did you record in for the Adobe Enhance test? Was it noisy or echoey?**<br>Your answer: _____ |
| B | **On a scale of 1-10, how much did the Adobe Enhance output improve the quality of your recording?**<br>Your answer: _____ |

| | |
|---|---|
| C | **Identify a real recording in your work or professional life that would benefit from this enhancement:**<br>Your answer: _____ |

## Exercise 3: Music Generation

Generate your background music track in Suno for your Day 6 video. Document the process.

| | |
|---|---|
| A | **What was your Suno music prompt? Write it in full below:**<br>Your answer: _____ |
| B | **Describe the track Suno generated — tempo, instruments, mood. Does it match your intent?**<br>Your answer: _____ |
| C | **What would you change in the prompt to get a slightly different result?**<br>Your answer: _____ |

# Chapter 12: Day 7 Assignment

### 🎧 Your Complete Audio Production

*Complete before Day 8. Estimated time: 30-45 minutes.*

## The Task

Take one of your Day 6 video clips and add a complete, professional audio layer: narration, background music, and an intro jingle. Export a finished audio-visual piece ready to share.

## Step-by-Step Instructions

37. Write a 1-3 sentence narration script for your chosen Day 6 video clip. Apply the Vocal Persona Formula — write full performance direction using square bracket instructions.
38. Generate the narration in ElevenLabs using a voice that matches your content's tone. Download as MP3.
39. Optional but encouraged: Record 60 seconds of your own voice and create an Instant Voice Clone in ElevenLabs. Use your clone for the narration.
40. Generate a 3-4 minute background instrumental track in Suno using a detailed music description. Download as MP3.
41. Generate a 5-second intro jingle using the jingle prompt from Chapter 7.3.
42. Open CapCut and assemble the complete production following the 8-step workflow from Chapter 9.1.
43. Add auto-captions to the narration using CapCut's caption feature.
44. Export as MP4 at 1080p. Save to your AI Video Portfolio folder.

## Reflection Questions

| | |
|---|---|
| 1 | **How did the Vocal Persona performance instructions change the emotional quality of the narration?**<br>Your answer: _____ |
| 2 | **How long did the full audio production workflow take? What surprised you about the process?**<br>Your answer: _____ |

| | |
|---|---|
| 3 | **How does the complete audio-visual piece (video + narration + music) compare to the silent video alone?**<br>Your answer: _____ |
| 4 | **What professional context will you use this complete audio-visual production workflow for first?**<br>Your answer: _____ |

# Chapter 13: References & Further Learning

## 13.1 Core References


- Betker, A. et al. (2023). Improving Image Generation with Better Captions. OpenAI Technical Report. (Underpins ElevenLabs voice architecture principles)
- Adobe (2024). Adobe Podcast: AI-Powered Audio Enhancement Technology. Available at: podcast.adobe.com/about
- ElevenLabs (2024). ElevenLabs Multilingual v2 Technical Overview. Available at: elevenlabs.io/docs
- Agostinelli, A. et al. (2023). MusicLM: Generating Music From Text. Google Research. Available at: arxiv.org/abs/2301.11325
- IFPI (2024). Global Music Report 2024: AI and Music Industry Implications. Available at: ifpi.org
- Partnership on AI (2024). Responsible Practices for AI-Generated Voice. Available at: partnershiponai.org


## 13.2 Tools to Explore

- ElevenLabs — elevenlabs.io (voice generation and cloning — start here)
- Adobe Podcast Enhance — podcast.adobe.com/enhance (free, no account needed)
- Suno v4.5 — suno.com (full song generation)
- Udio — udio.com (instrumental tracks with more granular control)
- Murf AI — murf.ai (voice + presenter avatar in one tool)
- Krisp.ai — krisp.ai (real-time noise cancellation for video calls)
- CapCut — capcut.com (free video editor for assembling the final production)

## 13.3 Further Learning

- ElevenLabs Documentation — docs.elevenlabs.io (comprehensive voice settings guide)
- Suno Discord Community — discord.gg/suno (prompt sharing and generation tips)
- Audio Engineering Society — aes.org (professional audio standards and AI audio research)

**End of Day 7 eBook**

*Day 8 — AI for Automation & Workflows: Let AI Handle the Repetitive Work*

2026 AI Learning Roadmap | Created for beginners, designed for results.