www.sunrisetechnology.it https://youtu.be/JsGD5VaBdeE

Operational Guide: From Original Music to AI Video

In brief

Modern multimedia production makes it possible to turn a raw musical sketch - such as a piano track recorded on a smartphone - into a complete orchestral soundtrack with a cinematic video clip. This process leverages the synergy between traditional audio software (e.g. Audacity), AI platforms for sound arrangement (e.g. Suno AI, Udio), language models for storyboarding (e.g. ChatGPT, Claude, Gemini) and multimodal visual generation engines (e.g. Midjourney, Runway, Luma, Seedance).

Multimodal Pipeline Diagram (Music-to-Video)

Phase	Tools Used	Process Description
1. Audio Restoration	Audacity, Adobe Podcast AI	Background noise cleanup, peak normalization, equalization and removal of unwanted frequencies from the mono/piano track.
2. AI Rearrangement	Suno AI, Udio, BandLab	Audio-to-Audio extension: adding violins, violas, a string section, percussion and orchestral dynamics while keeping the piano line.
3. Storyboard & Concept	ChatGPT / Claude / Gemini	Analysis of the sonic mood, drafting of the narrative concept, scene breakdown with timing and generation of image prompts.
4. Visual Generation	Midjourney, Runway, Luma, Gemini, Qwen, Seedance	Creation of high-resolution master frames and their 4K video animation with camera and motion control.
5. Editing & Sync	CapCut, DaVinci Resolve. Adobe Premiere	Assembly of the animated clips, metronomic sync on bar changes, color grading and final mastering.

Operational Phases in Detail

Phase 1: Preparing and Cleaning the Music Track

Recording a piece in a home environment often leads to acoustic imperfections. Before moving on to generative AI, the following editing steps are applied:

- **Noise Reduction (Denoise):** Sampling the background hiss in Audacity and selective removal.
- **AI Enhancement:** Running the track through speech/audio enhancement tools to rebuild lost harmonics and stabilize the RMS volume.

Phase 2: Orchestral Rearrangement and Expansion with AI

By uploading the cleaned file to specialized AI engines (e.g. the Upload Audio function on Suno or Udio):

- **Style Prompting:** Entering genre cues such as 'cinematic orchestral crescendo, emotional violin solo, cello section, epic percussion, piano lead'.
- **Audio-to-Audio Continuity:** The AI extends the piece while keeping the original key and progressively bringing out the strings and percussion up to the final climax.

Phase 3: Structuring the Storyboard and Prompt Design

The final audio is analyzed by an LLM to generate the narrative structure of the video. Below is a sample table of an operational storyboard produced by AI:

Scene & Timing	Shot & Movement	Image Prompt (Midjourney, GPT)	Video Prompt (Runway / Luma / Seedance)
Scene 1 (00:00 - 00:08)	Macro Close-Up Slow zoom in	A glowing ancient musical ember inside a dark stone sphere, cinematic warm light, 8k --ar 16:9	Slow camera push-in, embers floating gently, soft flicker effect
Scene 2 (00:08 - 00:20)	Long Shot Horizontal pan	Majestic concert hall made of ice and golden glowing violin strings, cinematic lighting --ar 16:9	Slow pan left to right, light rays pulsing with music rhythm
Scene 3 (00:20 - 00:35)	Medium Shot Camera rotation	Violin player made of starlight performing on a cliff under a nebula sky, dramatic atmosphere --ar 16:9	360 orbit camera around character, starlight particles flowing

Production Operational Checklist

Task	Required Action
1. Audio Pre-Processing	Export the track as WAV 24-bit 44.1kHz and remove peaks > -3dB in Audacity.
2. AI Arranging	Test at least 3 prompt variations on Suno/Udio (e.g. by selecting the track with the best string dynamics).
3. Image Consistency	Keep the same seed or character reference for visual generation on Midjourney/Leonardo/Seedance.
4. Final Editing	Cut the video clips to the music's bpm and apply 0.5s crossfade transitions - Librosa.

Technical Glossary

Term	Meaning
Audio-to-Audio	AI technique that takes an input audio file and generatively transforms it while keeping the basic structure.
Stem Separation	Splitting the mixed components (vocals, piano, drums, strings) into separate tracks for processing.
Storyboard LLM	Narrative sequence created by a language model with directions for camera work, scenes and timing.
Image-to-Video	Generating animation and motion from a static guide image.
Multimodal Prompt Engineering	The discipline of writing text instructions optimized for AI audio, images and video.