



Sound Design AI



Oriol Nieto

October 15th, 2026

Conversational AI Reading Group

MILA

Sound Design



How can I add sound?

Traditional Sound Design: Sound Library

- Expensive \$\$\$
- Slow
- Daunting

The screenshot displays the SoundQ application interface, which is a sound library management tool. The main window is divided into several sections:

- Libraries:** A sidebar on the left shows the source of the sounds, including 'Local' and 'Pro Sound Effects'.
- File List:** A central table lists the audio files with columns for Filename, Description, Category, Subcategory, Duration, Library, Bit Depth, and Sample Rate. The files are organized by category, such as 'Ambience' and 'Animals'.
- File Info:** A panel on the right provides detailed information about the selected file, including its name, description, category, subcategory, duration, library, file type, bit depth, sample rate, path, and artist.
- Transport and Controls:** At the bottom, there are controls for transport (play, stop, etc.), volume, speed, and a waveform display.

The interface is designed to be user-friendly, allowing users to browse and manage their sound library efficiently.

Text-to-SFX

Prompt

For the best output, use short, specific prompts focused on one element of the sound.

A sci-fi spaceship speeding past

Voice performance (optional)

Use your voice to control the timing and energy by imitating the desired sound. [Learn more.](#)

Record your voice

From device

Record voice

Generate



0000:01:0000:30:00

Upload media

00:00:0000:00:0500:1000:1500:2000:2500:30

spaceship.mp4

spaceship.wav

Add audio track

Clear all

Adobe

Creativity & Design

Adobe Firefly

Features

Mobile

Compare Plans

Go to Firefly

Adobe Firefly

Make some noise with AI sound effect generator.

Imagine any sound effect and create it with Generate Sound Effects. Describe the effect, upload reference audio, or act it out into your mic — then easily add your high-quality effect to any video.

Generate sound effects

Rocket launching into space

Video

Audio

Create showstopping sound effects, fast.

Amazing audio with AI sound effects.

drum roll - 1

6.7

drum roll - 2

6.7

drum roll - 3

6.7

drum roll - 4

6.7

GenSFX: text-to-SFX

- Cost effective \$
- Fast
- Easy
- BUT: timing

Prompt: “a gigantic snow avalanche is rolling down a glacier”

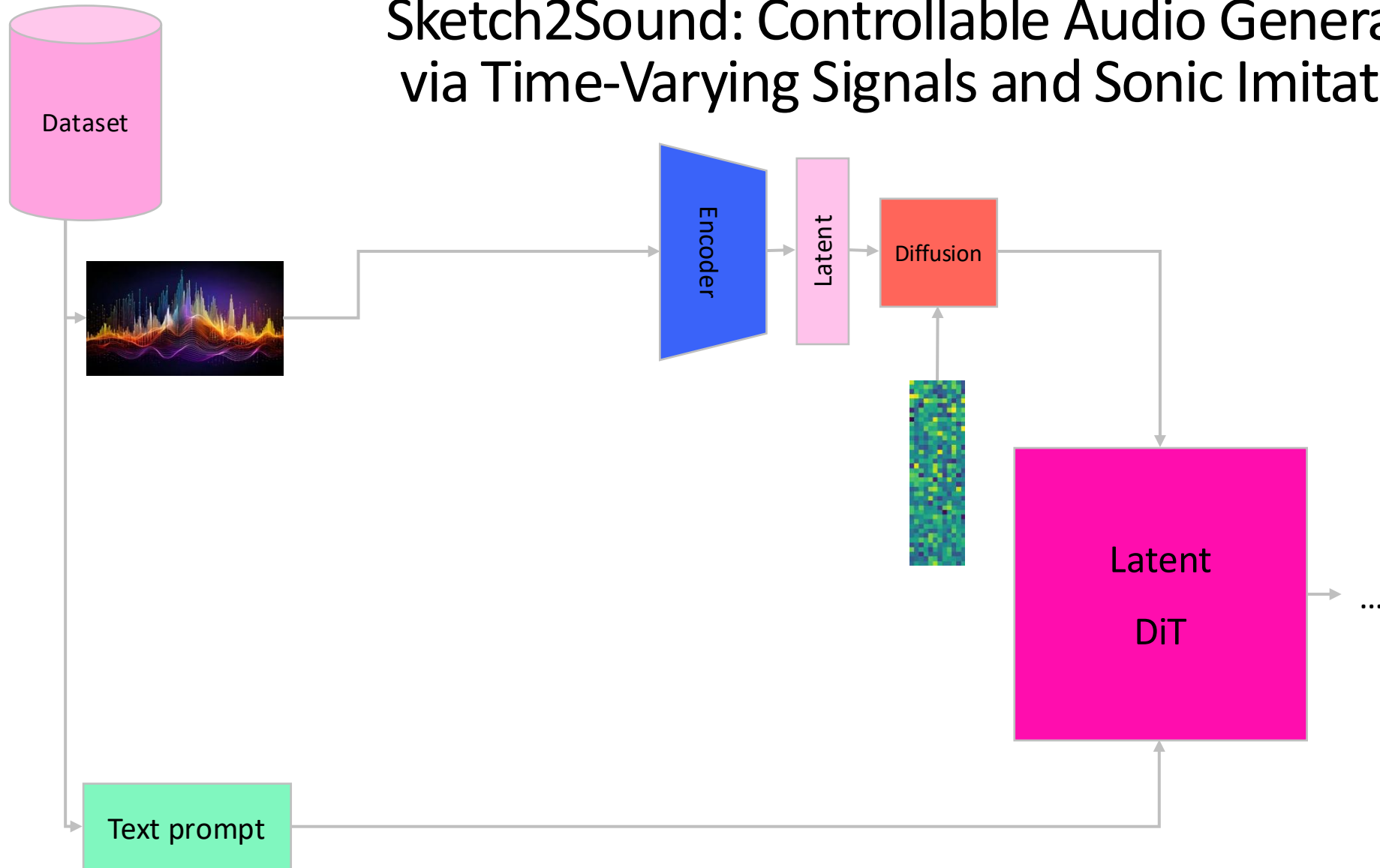


Voice-to-SFX

“Sketch2Sound: Controllable Audio Generation via Time-Varying Signals and Sonic Imitations”, Flores García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 50th International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Hyderabad, India, 2025

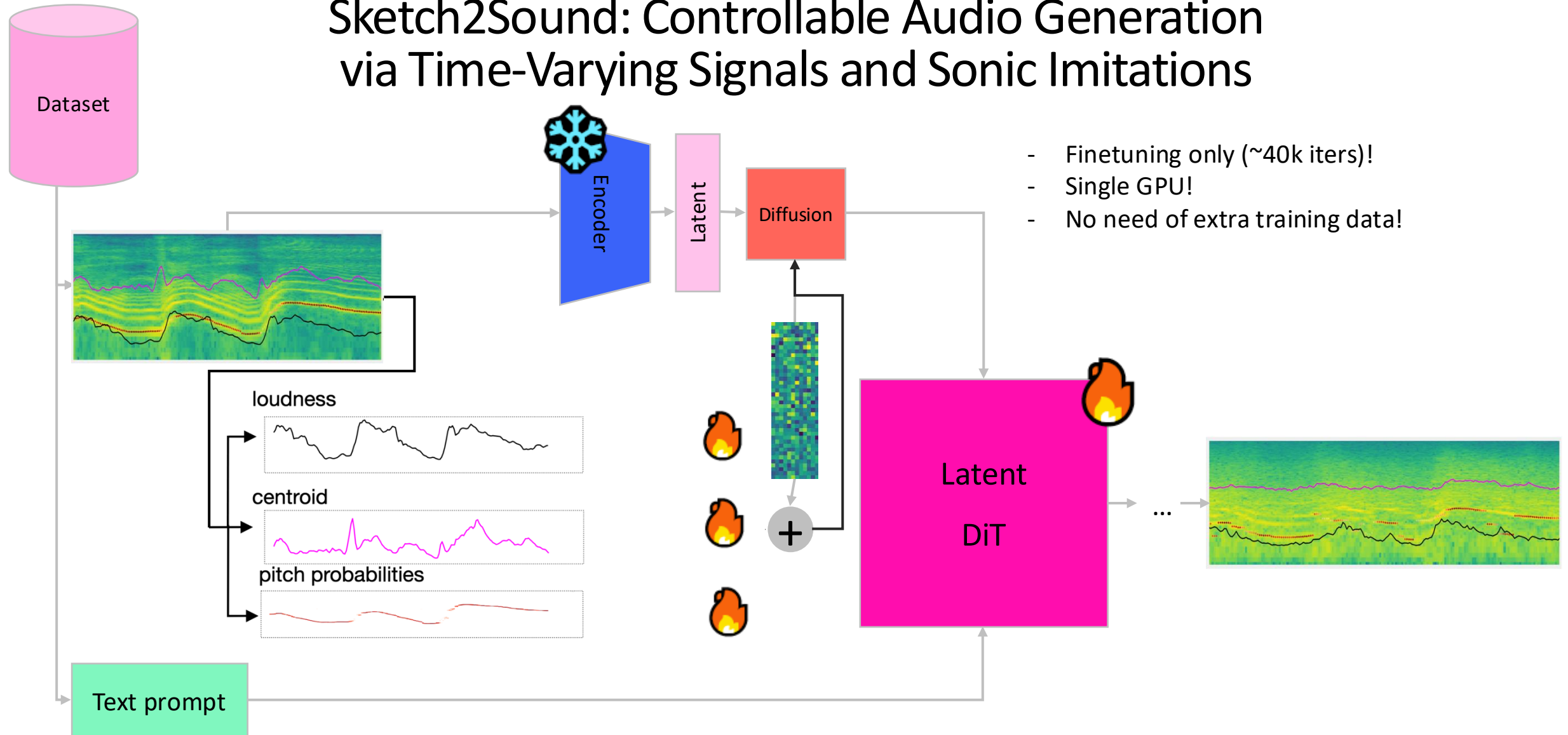
Voice-to-SFX

Sketch2Sound: Controllable Audio Generation via Time-Varying Signals and Sonic Imitations



Voice-to-SFX

Sketch2Sound: Controllable Audio Generation via Time-Varying Signals and Sonic Imitations



“Sketch2Sound: Controllable Audio Generation via Time-Varying Signals and Sonic Imitations”, Flores García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 50th International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Hyderabad, India, 2025





input (sonic imitation)

text prompt: lion roaring



6:46

SOS



0:00020:010:020:030:1019

Countryside ambience cicadas

Body armor rustle

Cloth rustle

Fight grunts

Fight grunt high pitch

Sword who

Sword hit

Sword ring whoosh

Sword hits

Bi

Cinematic hits

Ciner

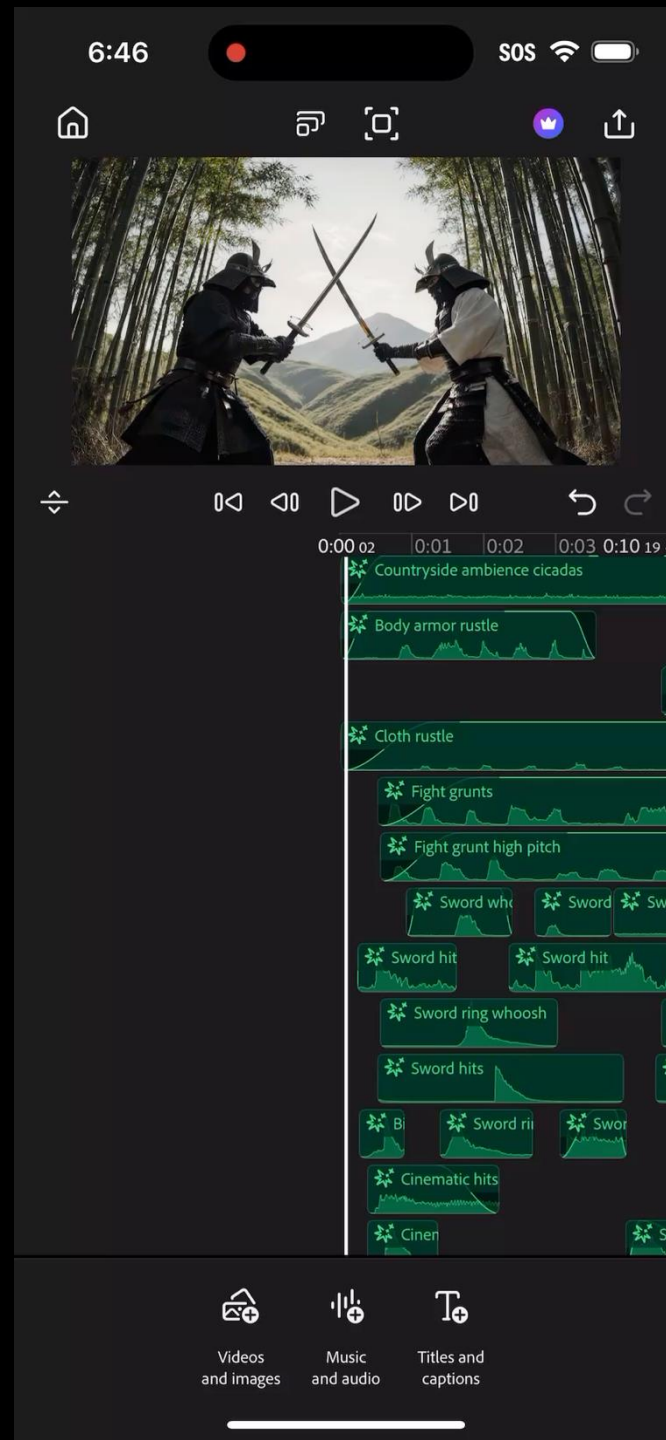
Videos and images

Music and audio

Titles and captions

GenSFX: voice-to-SFX

- Cost effective \$
- Fast
- Expressive and FUN
- Control over timing & energy
- BUT: takes practice

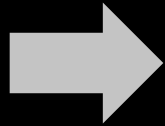


Video-to-SFX

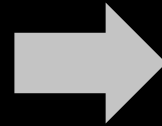
MultiFoley

“Video-Guided Foley Sound Generation with Multimodal Controls”, Chen, Z., Seetharaman, P., Russell, B., Nieto, O., Bourgin, D., Owens, A., Salamon, J., The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR). Nashville, TN, USA, 2025.

Video-to-SFX: MultiFoley

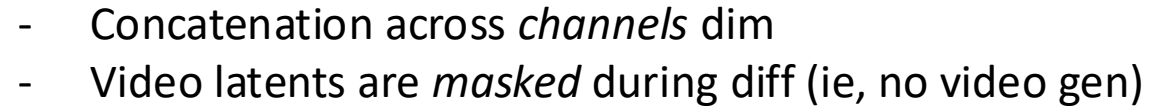


MultiFoley



Prompt: skateboard, wheels spinning



A pink cylinder with the word "Dataset" written in black text in the center.

Video-to-SFX: MultiFoley



Video-to-SFX: MultiFoley



Prompt: *“skateboard, wheels spinning”*

Video-to-SFX: MultiFoley



Prompt: ***"typewriter"***

Video-to-SFX: MultiFoley

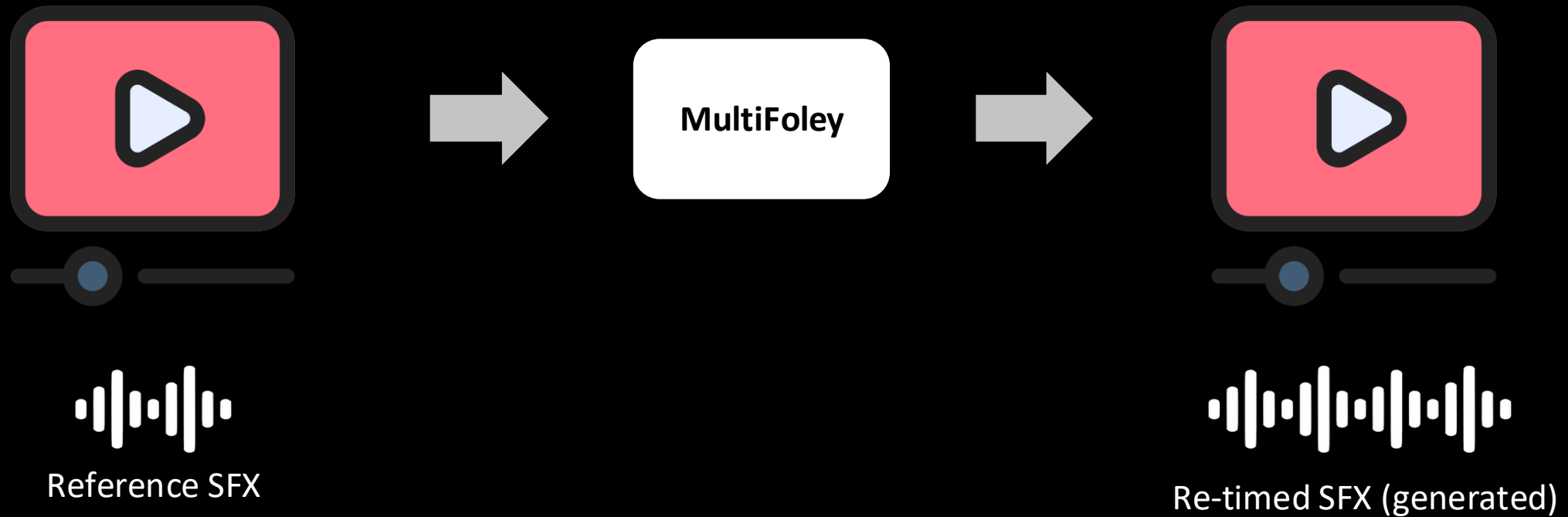


Prompt: *“lion roaring”*

Video-to-SFX: MultiFoley



Video-to-SFX: MultiFoley



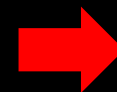
Video-to-SFX

Input: silent video

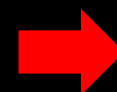


Input: reference SFX

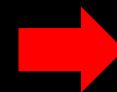
Reference SFX 1



Reference SFX 2



Reference SFX 3



Output: synced SFX



Running MultiFoley multiple times



Source: (no clips) Lumetri Scopes Effect Controls Audio

Program: video-1-2025-12-02T09-58-37

Source: video-1-2025-12-02T09-58-37 video-1-2025-12-02T09-58-37

Video

fx Motion

Position960.0540.0

Scale100.0

Scale Width100.0

Uniform Scale

Rotation0.0

Anchor Point960.0540.0

Anti-flicker Filter0.00

fx Opacity

Opacity100.0 %

Blend ModeNormal

fx Time Remapping

Speed100.00%

00:00:02:02

Fit

1/2

00:00:10:00

Project: samurai_video2sfx Media Browser Libraries Inf

1 of 5 items selected

Name	Frame Rate
video-1-2025-12-02T09-58-37	24.00 fps
video-1-2025-12-02T09-58-37.mp4	24.00 fps
video-1-2025-12-02T10-01-36.mp4	24.00 fps
video-2-2025-12-02T09-58-41.mp4	24.00 fps
video-3-2025-12-02T10-09-50.mp4	24.00 fps

video-1-2025-12-02T09-58-37

00:00:02:02

V2

V1 V1 Video 1

A1 A1 Audio 1

A2 Audio 2

A3 Audio 3

A4 Audio 4

video-1-2025-12-02T09-58-37.mp4

video-1-2025-12-02T09-58-37.mp4

video-1-2025-12-02T10-01-36.mp4

video-3-2025-12-02T10-09-50.mp4

video-2-2025-12-02T09-58-41.mp4

Lumetri Color

Essential Sound

Browse Edit

No Selection

Preset:

Select clips with Tag:

Dialogue

Music

SFX

Ambience

Adobe

© 2026 Adobe. All © 2026 Adobe. All Rights Reserved.



Sound Stager

Full Agentic Sound Design

“SoundStager: Interactive Design of Story-Driven GenAI Soundscapes for Video”, Yoo, S., Hernandez-Sebastian, A., Seetharaman, P., Salamon, J., Nieto, O., Truong, A., Proc. of the ACM Conference on Human Factors in Computing Systems (CHI). Barcelona, Spain, 2026

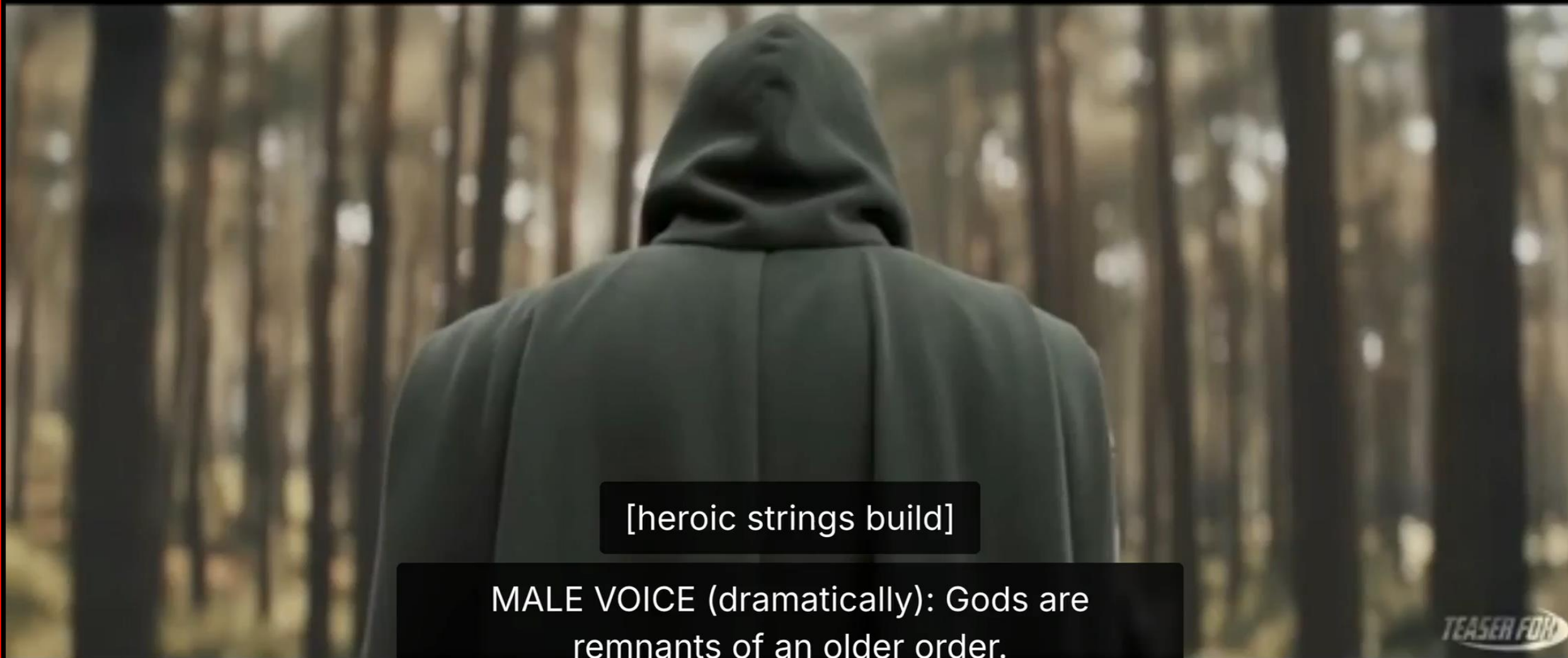


ProjectSoundStager
#ProjectSoundStager
ProjectSoundStager
ProjectSoundStager
ProjectSoundStager

ProjectSoundStager
#ProjectSoundStager
ProjectSoundStager
ProjectSoundStager
ProjectSoundStager

TAC: Timestamped Audio Captioner

“TAC: Timestamped Audio Captioner”, Sonal Kumar, Prem Seetharaman, Ke Chen, Oriol Nieto, Jiaqi Su, Zhepei Wang, Rithesh Kumar, Dinesh Manocha, Nicholas J. Bryan, Zeyu Jin, Justin Salamon, NeurIPS, 2026

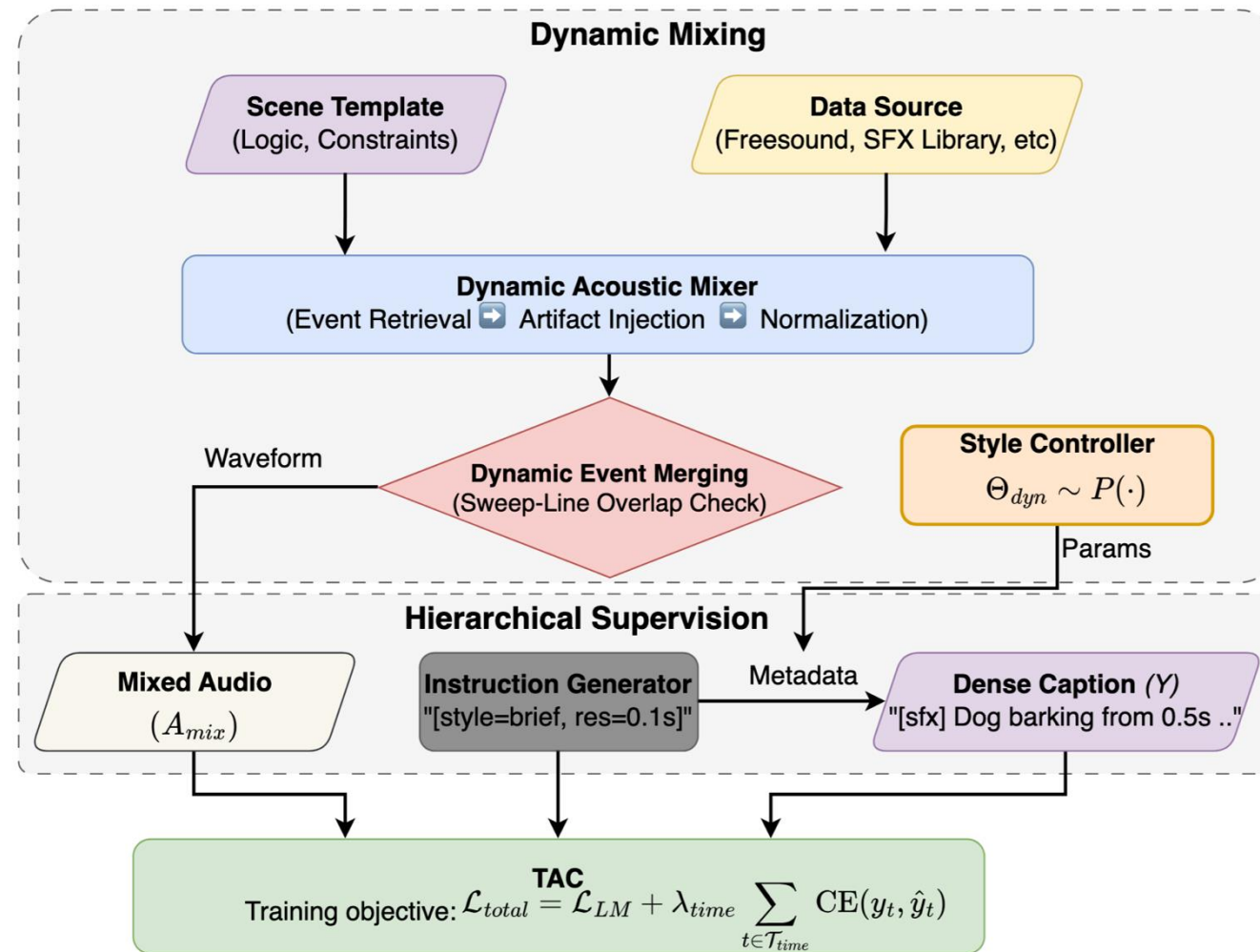


[heroic strings build]

MALE VOICE (dramatically): Gods are
remnants of an older order.

TEASER FOR

TAC: Timestamped Audio Captioner

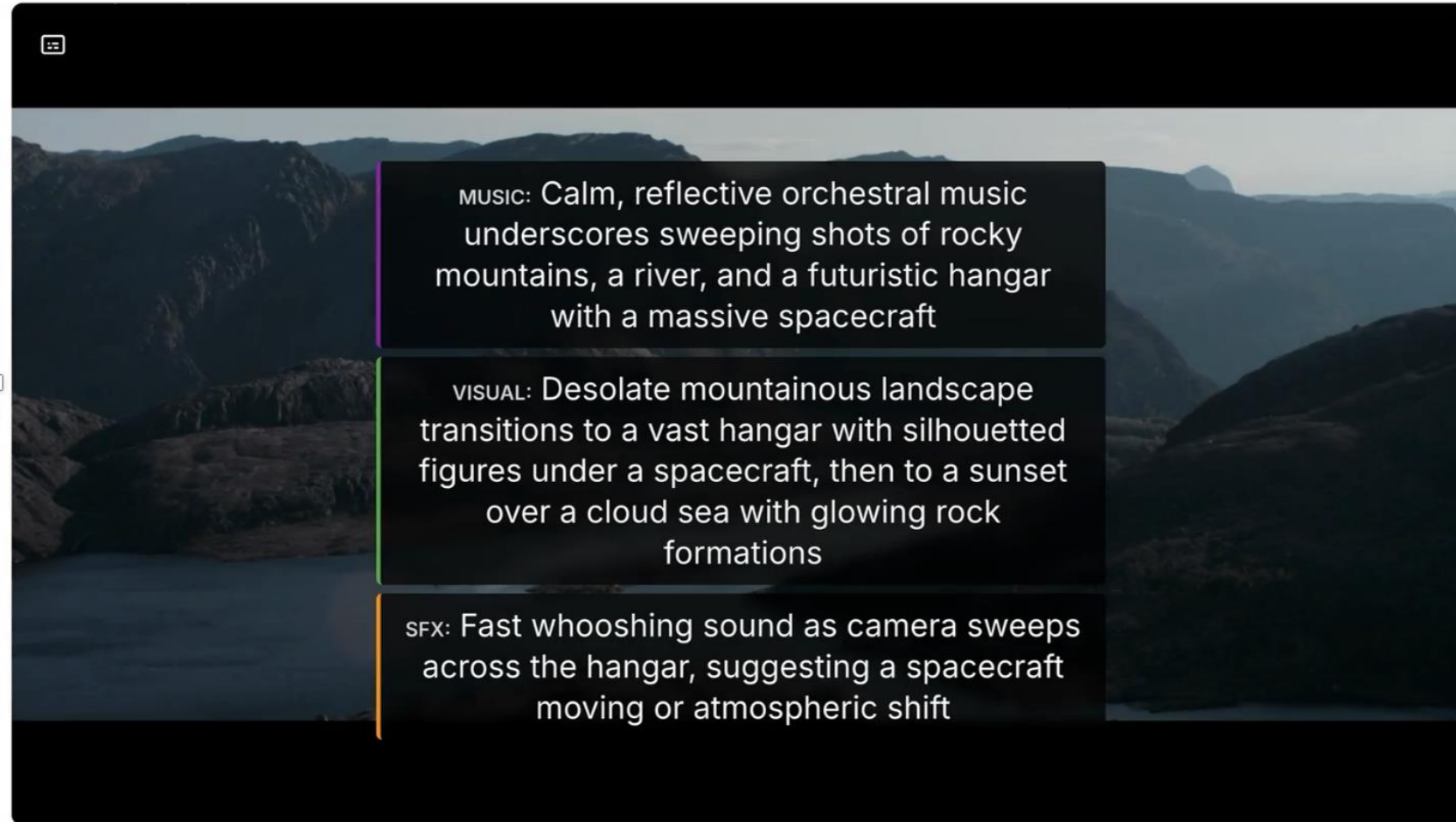


- Finetune Qwen2-Audio
- Only 5k iters
- SOTA

“TAC: Timestamped Audio Captioner”, Sonal Kumar, Prem Seetharaman, Ke Chen, Oriol Nieto, Jiaqi Su, Zhepei Wang, Rithesh Kumar, Dinesh Manocha, Nicholas J. Bryan, Zeyu Jin, Justin Salamon, NeurIPS, 2026

Audio Captioner + Video Captioner

Caption Results



MUSIC: Calm, reflective orchestral music underscores sweeping shots of rocky mountains, a river, and a futuristic hangar with a massive spacecraft

VISUAL: Desolate mountainous landscape transitions to a vast hangar with silhouetted figures under a spacecraft, then to a sunset over a cloud sea with glowing rock formations

SFX: Fast whooshing sound as camera sweeps across the hangar, suggesting a spacecraft moving or atmospheric shift

Events (107)

Q Search...

VISUAL **MUSIC** **SPEECH** **SFX** **MISC**

BACKGROUND

Calm, reflective orchestral music underscores sweeping shots of rocky mountains, a river, and a futuristic hangar with a massive spacecraft

00:20.00 - 00:30.00 **VISUAL**

Desolate mountainous landscape transitions to a vast hangar with silhouetted figures under a spacecraft, then to a sunset over a cloud sea with glowing rock formations

00:20.19 - 00:22.50 **SFX**

Fast whooshing sound as camera sweeps across the hangar, suggesting a spacecraft moving or atmospheric shift

00:22.60 - 00:23.39 **SPEECH**

Bearded man in tactical gear speaks positively, standing under the spacecraft with the younger man

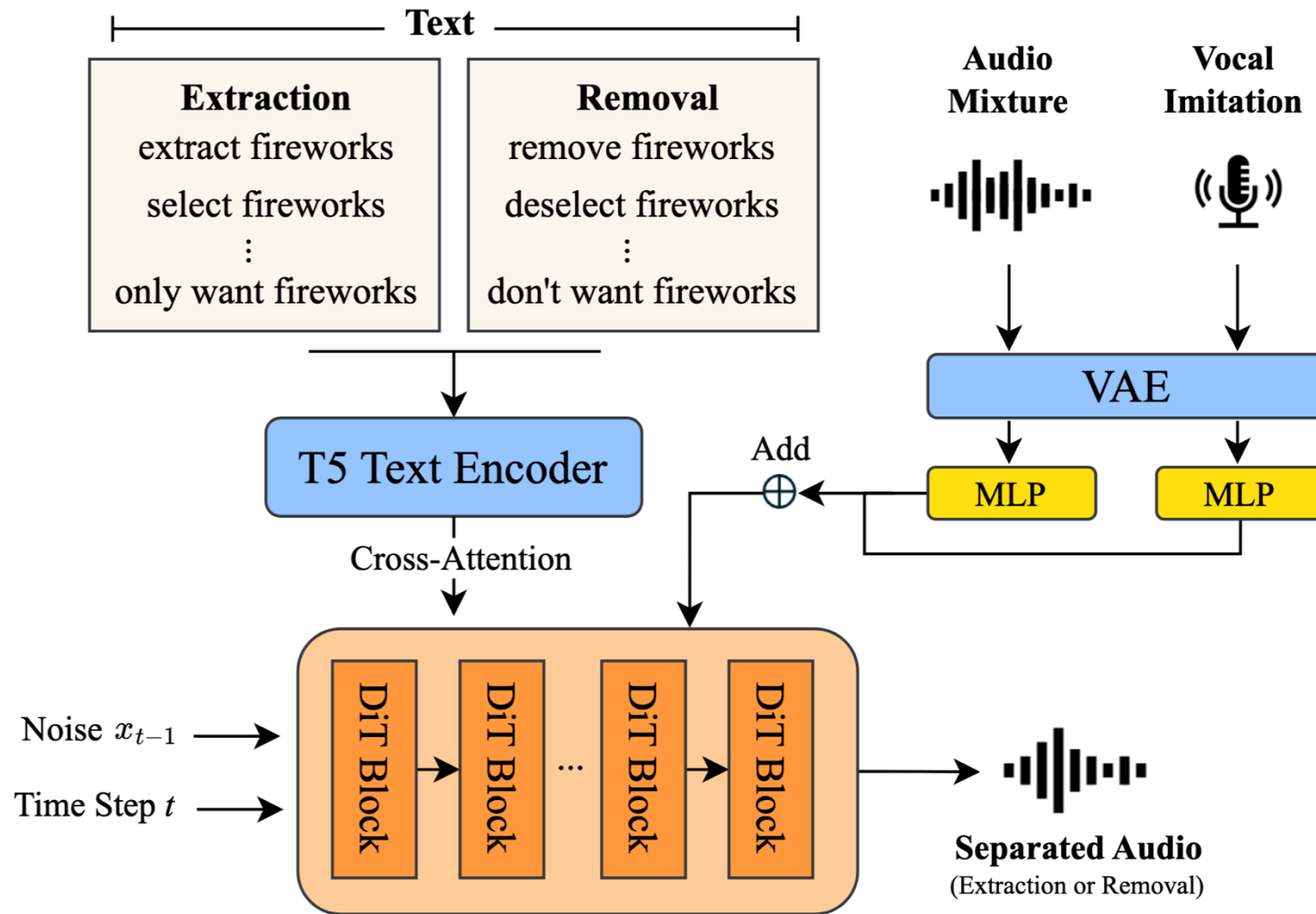
"So you're going tomorrow?"

PromptSep

Sound Separation via Prompts

“PromptSep: Generative Audio Separation Via Multimodal Prompting”, Wen, Y., Chen, K., Seetharaman, P., Nieto, O., Su, J., Kumar, R., Kim, M., Smaragdis, P., Jin, Z., Salamon, J., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

PromptSep: Sound Separation via Prompts

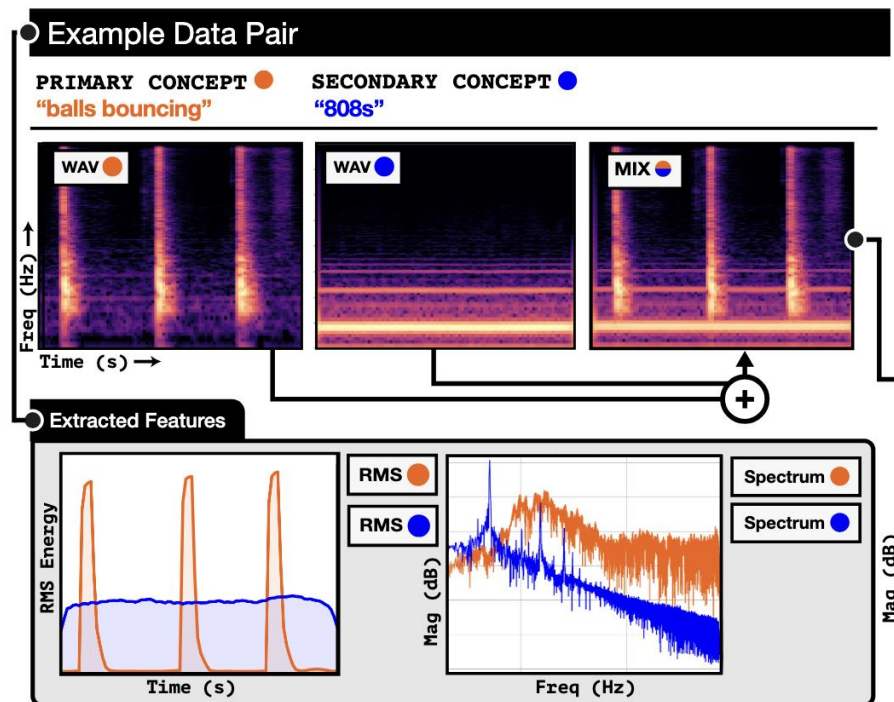


"PromptSep: Generative Audio Separation Via Multimodal Prompting" Wen, Y., Chen, K., Seetharaman, P., Nieto, O., Su, J., Kumar, R., Kim, M., Smaragdis, P., Jin, Z., Salamon, J., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

Audio Generative Morphing

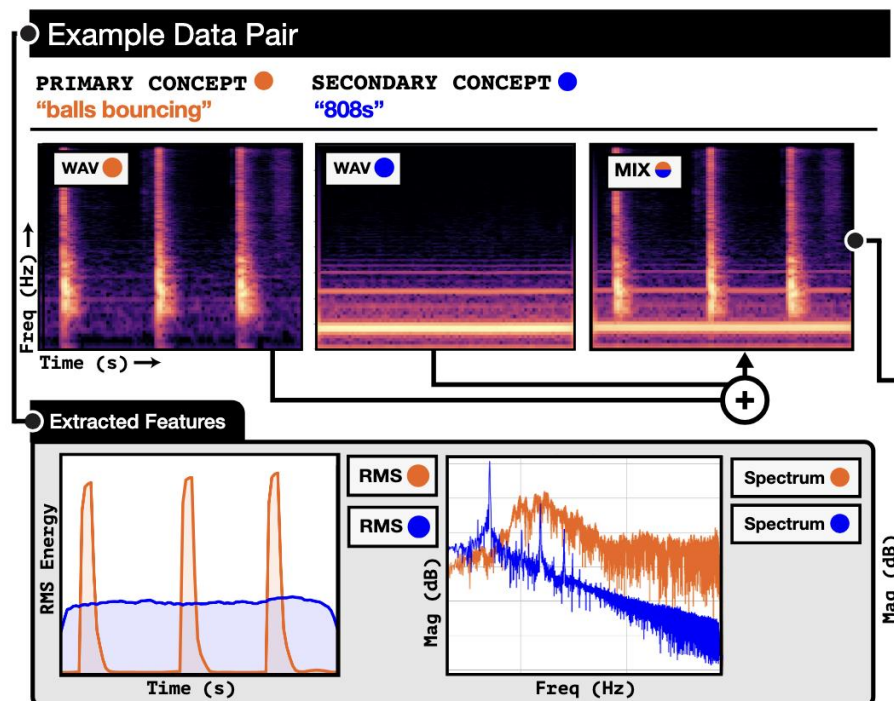
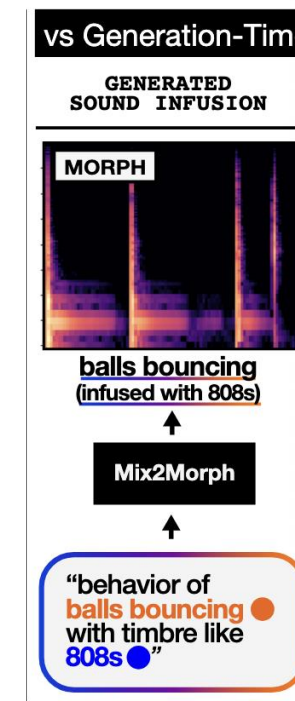
“Mix2Morph: Learning Sound Morphing From Noisy Mixes”, Chu, A., Flores-García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026

Audio Generative Morphing



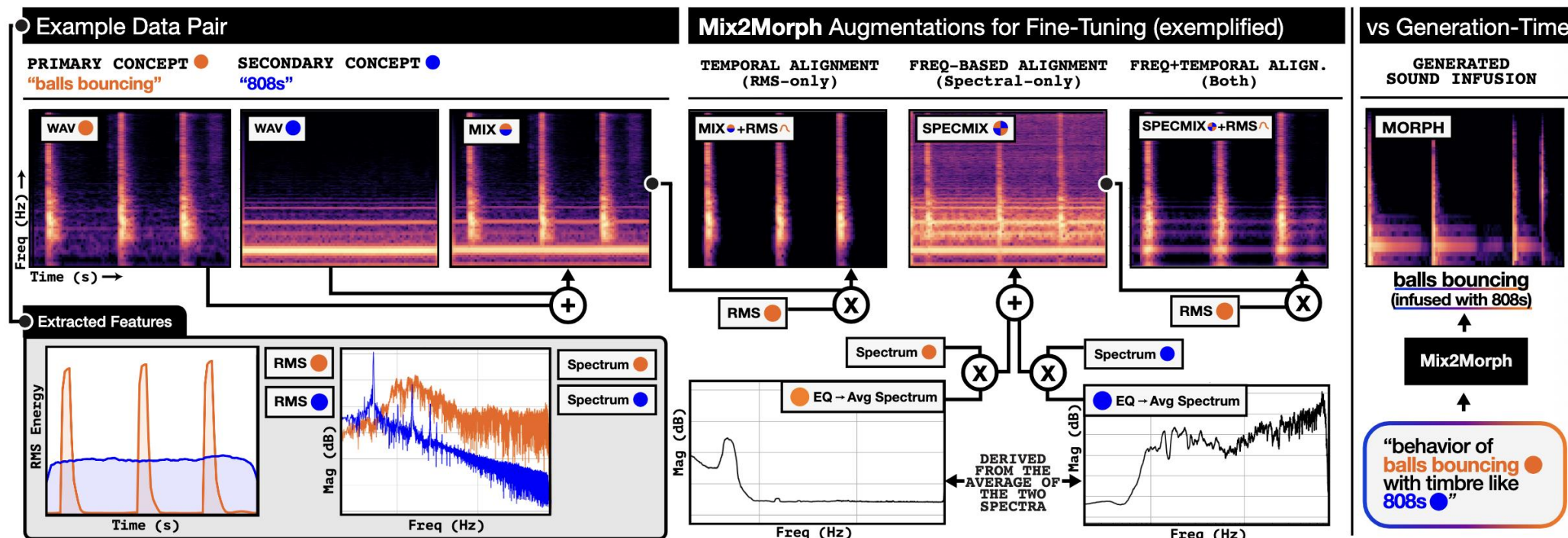
"Mix2Morph: Learning Sound Morphing From Noisy Mixes", Chu, A., Flores-García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026

Audio Generative Morphing



"Mix2Morph: Learning Sound Morphing From Noisy Mixes", Chu, A., Flores-García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026

Audio Generative Morphing



"Mix2Morph: Learning Sound Morphing From Noisy Mixes", Chu, A., Flores-García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026

Audio Generative **Morphing**

Door creaking + Zombie Groan



Tennis Ball + Xylophone Hit



Monster Growling + Motorcycle Rewing



Sneak Peek

FoleyForge: Audio Generative Editing

Input



"I want this to sound like a sci-fi gun"

Output



Sneak Peek

FoleyForge: Audio Generative Editing

Input



“Make it sound more like wood”

Output



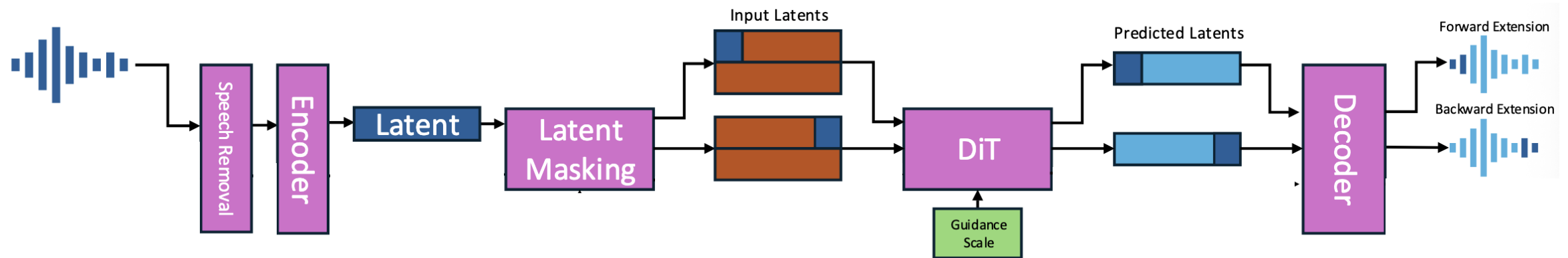
Audio Generative Extend

“Generative Audio Extension and Morphing”, Seetharaman, P.,* Nieto, O.,* Salamon, J.,
Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

Audio Generative Extend

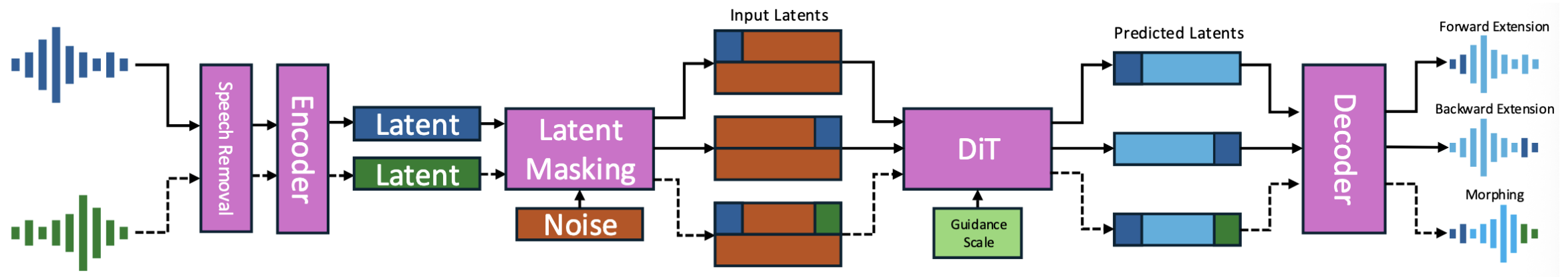


Audio Generative Extend



“Generative Audio Extension and Morphing”, Seetharaman, P.,* Nieto, O.,* Salamon, J.,
Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

Audio Generative Morphing



“Generative Audio Extension and Morphing”, Seetharaman, P.,* Nieto, O.,* Salamon, J.,
Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

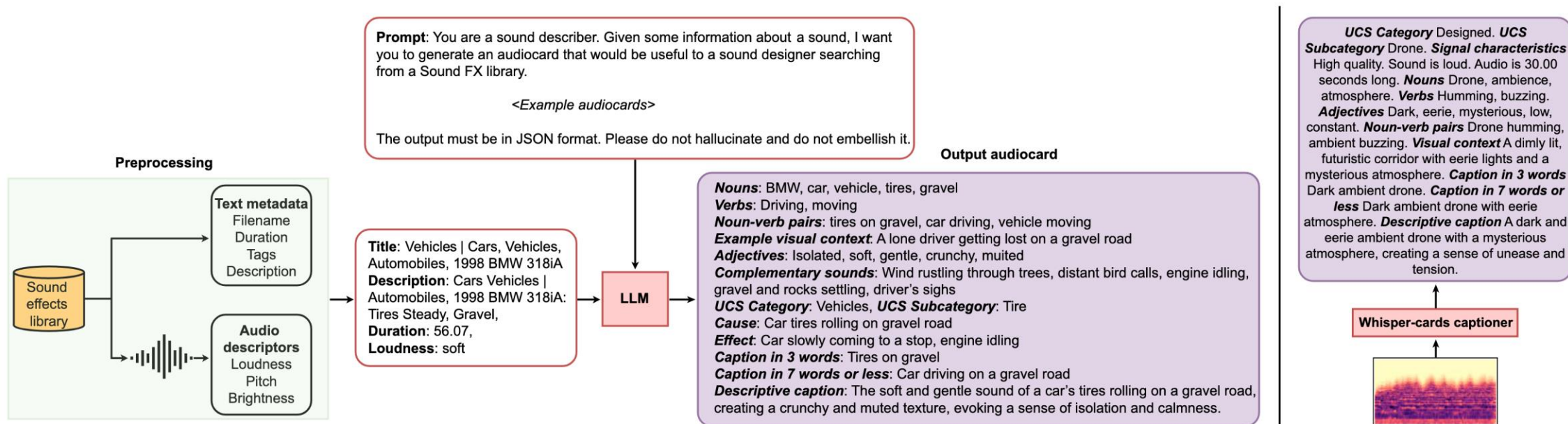
Audio Generative Extend



Audio Cards

“Audiocards: Structured Metadata Improves Audio Language Models For Sound Design”, Sripathi Sridhar, Prem Seetharaman, Oriol Nieto, Mark Cartwright, Justin Salamon, Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

Audio Cards



“Audiocards: Structured Metadata Improves Audio Language Models For Sound Design”, Sripathi Sridhar, Prem Seetharaman, Oriol Nieto, Mark Cartwright, Justin Salamon, Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

SODA technologies: Recap

- Text-to-SFX
- Voice-to-SFX (ICASSP25)
- Video-to-SFX: MultiFoley (CVPR25)
- Sound Stager (Agentic Sound Design) (CHI26)
- TAC: Audio Captioner (NeurIPS 2026)
- PromptSep (prompt to extract sound) (ICASSP26)
- Mix2Morph (ICASSP26)
- Audio Generative Extend (ICASSP26)
- AudioCards (ICASSP26)

References

- **“Sketch2Sound: Controllable Audio Generation via Time-Varying Signals and Sonic Imitations”**, Flores García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 50th International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Hyderabad, India, 2025.
- **“Video-Guided Foley Sound Generation with Multimodal Controls”**, Chen, Z., Seetharaman, P., Russell, B., Nieto, O., Bourgin, D., Owens, A., Salamon, J., The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR). Nashville, TN, USA, 2025.
- **“SoundStager: Interactive Design of Story-Driven GenAI Soundscapes for Video”**, Yoo, S., Hernandez-Sebastian, A., Seetharaman, P., Salamon, J., Nieto, O., Truong, A., Proc. of the ACM Conference on Human Factors in Computing Systems (CHI). Barcelona, Spain, 2026.
- **“TAC: Timestamped Audio Captioner”**, Sonal Kumar, Prem Seetharaman, Ke Chen, Oriol Nieto, Jiaqi Su, Zhepei Wang, Rithesh Kumar, Dinesh Manocha, Nicholas J. Bryan, Zeyu Jin, Justin Salamon, arXiv preprint, 2026.
- **“PromptSep: Generative Audio Separation Via Multimodal Prompting”**, Wen, Y., Chen, K., Seetharaman, P., Nieto, O., Su, J., Kumar, R., Kim, M., Smaragdis, P., Jin, Z., Salamon, J., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.
- **“Mix2Morph: Learning Sound Morphing From Noisy Mixes”**, Chu, A., Flores-García, H., Nieto, O., Salamon, J., Pardo, B., Seetharaman, P., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.
- **“Generative Audio Extension and Morphing”**, Seetharaman, P.,* Nieto, O.,* Salamon, J., Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.
- **“AudioCards: Structured Metadata Improves Audio Language Models For Sound Design”**, Sripathi Sridhar, Prem Seetharaman, Oriol Nieto, Mark Cartwright, Justin Salamon, Proc. of the 51st International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Barcelona, Spain, 2026.

SOund Design AI



Justin Salamon



Oriol Nieto



Prem Seetharaman



Hugo Flores Garcia



Johannes Imort



Annie Chu



Giovana Morais



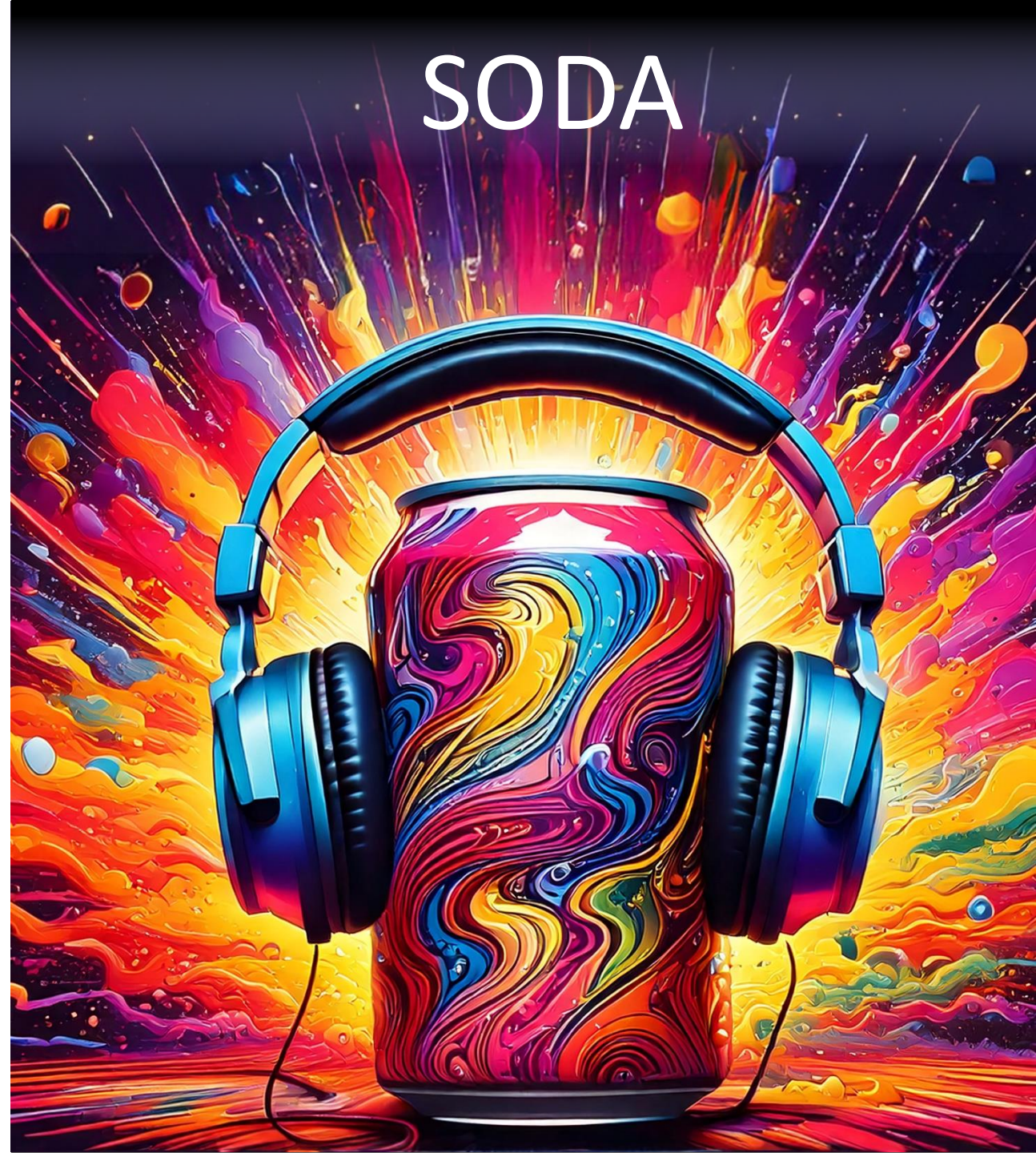
Sripathi Sridhar



Yutong Wen



SODA



SODA technologies

- Text-to-SFX
- Voice-to-SFX (ICASSP25)
- Video-to-SFX: MultiFoley (CVPR25)
- Sound Stager (Agentic Sound Design) (CHI26)
- TAC: Audio Captioner
- PromptSep (prompt to extract sound) (ICASSP26)
- Mix2Morph (ICASSP26)
- Audio Generative Extend (ICASSP26)
- AudioCards (ICASSP26)

