HARP 3.0: Generalizing I/O and API Support for Machine Learning in Digital Audio Workstations

Frank Cwitkowitz[1,*] **Christodoulos Benetatos**[1,*] **Qixin Deng**[2]

Huiran Yu[1] **Nathan Pruyne**[2] **Patrick O'Reilly**[2]

Hugo Flores Garcia[2] **Zhiyao Duan**[1] **Bryan Pardo**[2]

[1]University of Rochester
Rochester, NY, USA

[2]Northwestern University
Evanston, IL, USA

Abstract

HARP is a framework that integrates deep learning models into digital audio workstations (DAWs) through **h**osted, **a**synchronous, **r**emote **p**rocessing, allowing users to access state-of-the-art models without leaving their production environment. In HARP 3.0, we introduce generalized input–output handling for audio, MIDI, and text; expand API compatibility beyond Gradio to include providers such as Stability AI; and add workflow features including a media clipboard for storing and reusing processed files. These improvements make HARP 3.0 a more versatile and extensible platform for creative audio applications. HARP is open source and available for download at `https://github.com/TEAMuP-dev/HARP`.

1 Introduction

Deep learning has enabled a wide range of advances in audio and music technology, powering tasks such as music generation [Evans et al., 2025], transcription [Simon et al., 2022], source separation [Rouard et al., 2023], etc. While these models offer new tools for composition, production, and analysis, they are typically released by researchers as code repositories or web demos intended for other researchers. Musicians and producers often lack the expertise needed to install or debug such projects, and this usage paradigm is disconnected from their workflows in DAWs (*e.g.*, Logic Pro). As a result, access to cutting-edge deep learning models remains limited, despite their creative potential.

Previous efforts to integrate deep learning into music workflows include custom DAW builds [Garcia et al., 2021] or plugin-based systems such as NeuTone [Mitcheltree et al., 2025] and NeuralMidiFx [Haki et al., 2023]. While these solutions have demonstrated the potential for deploying models in DAWs, they are limited by computational resource and real-time constraints as well as platform restrictions. Our previous work introduced HARP 1.0, an Audio Random Access[1] plugin that provided users access to powerful, GPU-powered models through a uniform DAW interface, by routing audio through external Gradio endpoints hosted on platforms like Hugging Face. HARP 2.0 extended this framework to support MIDI-based models and time-stamped labeling, and was rebuilt as an external sample editor to support cross-platform use [Benetatos et al., 2024].

Despite these advances, earlier versions of HARP were limited to fixed input–output (I/O) modalities and single-step interactions with Gradio endpoints, restricting the ability to support multi-modal

[*]Equal contribution.
[1]`https://github.com/Celemony/ARA_SDK`

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Music.

models, alternative providers, and iterative workflows. In HARP 3.0, we address these limitations through three key extensions. First, we introduce generalized I/O handling, allowing for arbitrary combinations of audio, MIDI, and text, to support broader deep learning models and applications. Second, we take an initial step towards generalized API support, extending compatibility beyond Gradio to include Stability AI's text-to-audio and audio-to-audio generation models[2]. Finally, we introduce other new features, including a media clipboard enabling users to store and reuse processed files across sessions and route them back into the DAW, thereby facilitating multi-step interactions.

2 Generalized Inputs & Outputs

Many models for general audio or music, such as those for source separation [Rouard et al., 2023], rely on multiple (and sometimes optional) inputs or outputs. Some models, such as those for music transcription [Simon et al., 2022] or MIDI synthesis [Wu et al., 2021], also feature multi-modal processing. Other models, such as those for text-based music generation [Evans et al., 2025], don't use audio or MIDI inputs at all. HARP was originally built to support only single-track audio-to-audio or MIDI-to-MIDI models as a consequence of adopting the ARA standard. Since HARP was redesigned as an external sample editor in version 2.0, these restrictions were no longer in place. As such, in HARP 3.0 we have generalized the I/O to support all of the above interactions.

3 Generalized API Support

In previous versions, HARP was limited to Gradio endpoints hosted locally or through Hugging Face Spaces[3]. While this paradigm is extremely flexible and offers GPU acceleration, it neglects models that have already been deployed by alternate providers. In HARP 3.0, we take the first step towards generalized API support, enabling integration with external services that expose standard web APIs. This extension allows users to access models beyond Gradio, *i.e.*, Stable Audio's text-to-audio and audio-to-audio generation models, while broadening API compatibility, improving access to faster compute, and lowering the barrier for third-party providers to integrate their models with HARP.

4 Media Clipboard & Other Features

In HARP 3.0, we introduce various features that improve workflow and accessibility. A collapsible media clipboard with the ability to hold audio or MIDI tracks has been added to the main window. With this clipboard, users can now store any potential inputs or intermediate results as they switch between models. A new send-to-DAW feature has also been incorporated to grant users more control with respect to deciding when and how to overwrite DAW-linked tracks. Information defined by model developers for controls and I/O is now displayed in the instruction box when hovering over respective elements. Finally, there have been various other small bug fixes and efficiency improvements.

5 Future Work

Moving forward, our team has several envisioned improvements and objectives to further increase the utility and usability of HARP. In terms of major features, next we plan to add tabs for models, such that multiple models can be open at a time. We also plan to add the ability to save and open entire HARP projects, in order to keep track of open media files, models, parameter settings, and more. These features will allow users to establish their own unique workflows that can aid in their creative ventures. Another goal is to foster a community around HARP, connecting researchers in audio and music machine learning with musicians and producers who seek to discover and use cutting-edge research in their work. This will involve encouraging researchers in communities such as the International Society for Music Information Retrieval (ISMIR) to deploy their models to HARP, as well as outreach to musicians and producers to assist them in adopting the technology.

[2] `https://stableaudio.com/`

[3] `https://huggingface.co/spaces`

Acknowledgments and Disclosure of Funding

This work is supported by National Science Foundation (NSF) grants No. 2222129 and 2222369.

References

Christodoulos Benetatos, Frank Cwitkowitz, Nathan Pruyne, Hugo Flores Garcia, Patrick O'Reilly, Zhiyao Duan, and Bryan Pardo. HARP 2.0: Expanding hosted, asynchronous, remote processing for deep learning in the DAW. *ISMIR Late Breaking Demo Papers*, 2024.

Zach Evans, Julian D Parker, CJ Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. Stable audio open. In *Proceedings of ICASSP*, 2025.

Hugo Flores Garcia, Aldo Aguilar, Ethan Manilow, Dmitry Vedenko, and Bryan Pardo. Deep learning tools for Audacity: Helping researchers expand the artist's toolkit. In *NeurIPS Workshop on Machine Learning for Creativity and Design*, 2021.

Behzad Haki, Julian Lenz, and Sergi Jorda. NeuralMidiFx: A wrapper template for deploying neural networks as VST3 plugins. In *Proceedings of AI Music Creativity*, 2023.

Christopher Mitcheltree, Bogdan Teleaga, Andrew Fyfe, Naotake Masuda, Matthias Schäfer, Alfie Bradic, and Nao Tokui. Neutone SDK: An open source framework for neural audio processing. In *Proceedings of AES International Conference on Artificial Intelligence and Machine Learning for Audio*, 2025.

Simon Rouard, Francisco Massa, and Alexandre Défossez. Hybrid transformers for music source separation. In *Proceedings of ICASSP*, 2023.

Ian Simon, Josh Gardner, Curtis Hawthorne, Ethan Manilow, and Jesse Engel. Scaling polyphonic transcription with mixtures of monophonic transcriptions. In *Proceedings of ISMIR*, 2022.

Yusong Wu, Ethan Manilow, Yi Deng, Rigel Swavely, Kyle Kastner, Tim Cooijmans, Aaron Courville, Cheng-Zhi Anna Huang, and Jesse Engel. MIDI-DDSP: Detailed control of musical performance via hierarchical modeling. In *Proceedings of ICLR*, 2021.

A HARP 3.0 Screenshots

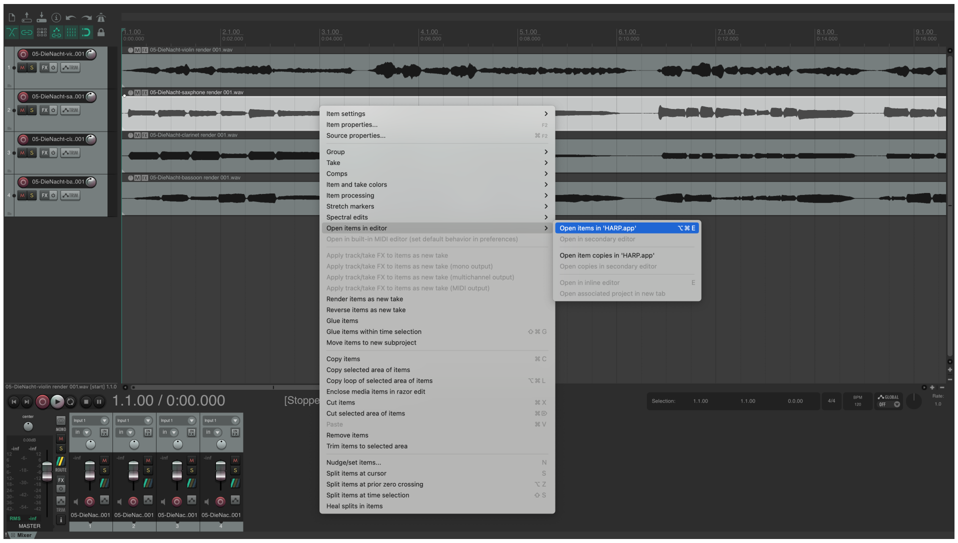

Figure 1: Invoking HARP from REAPER as an external sample editor.

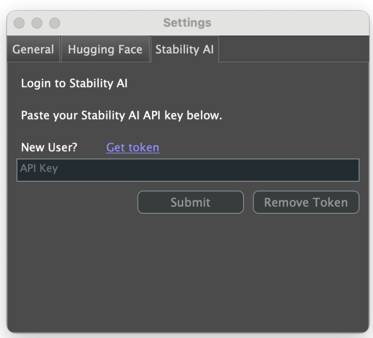

Figure 2: Login window to enter an access token for Stability AI.

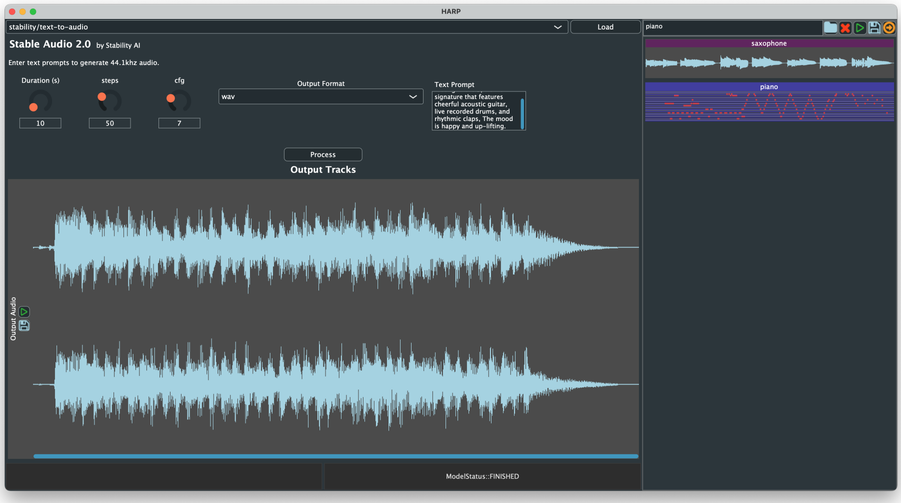

Figure 3: Main HARP window illustrating Stable Audio text-to-audio model and media clipboard containing DAW-linked (indicated by purple banner) track "saxophone" and MIDI track "piano".

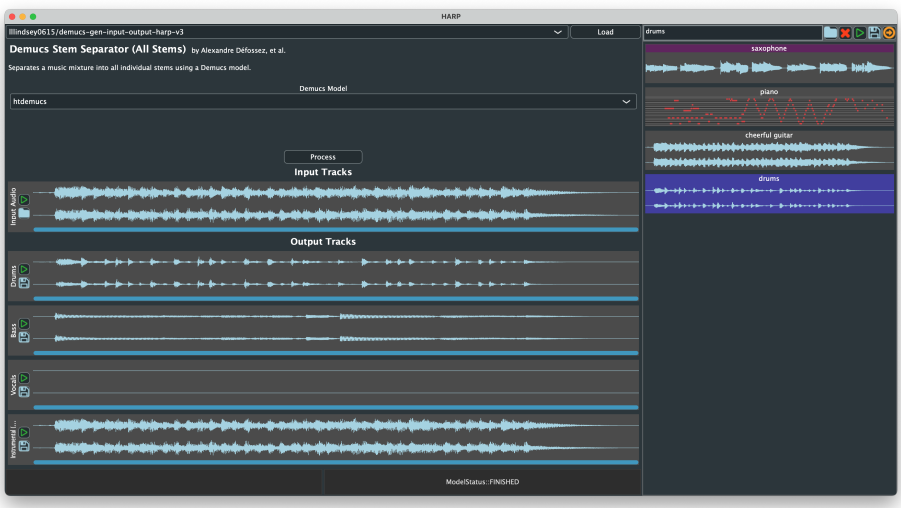

Figure 4: Main HARP window illustrating support for models with arbitrary I/O.

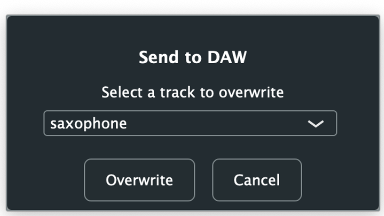

Figure 5: Send-to-DAW window, opened by clicking orange right-most button in media clipboard, with selection for DAW-linked track to overwrite.