

Better music demixing with sliCQT

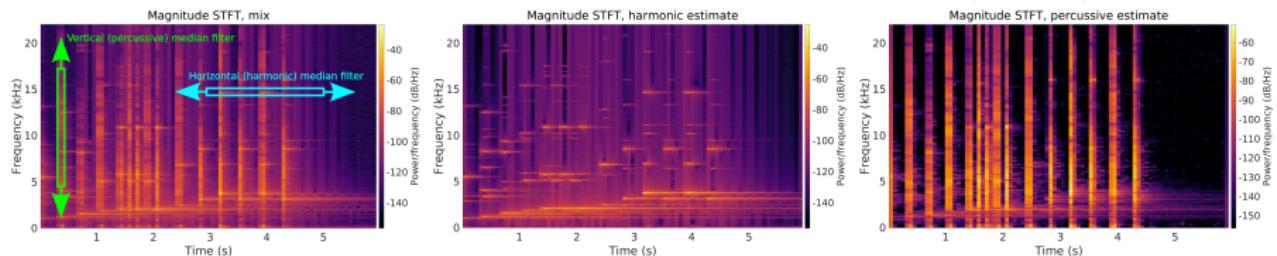
Submission to Cadenza Challenge CAD1

Sevag Hanssian

December 08, 2023

Time-frequency tradeoffs

Median-filtering harmonic/percussive source separation (HPSS)¹



- ① Short window (256) for percussion, long window (4096) for harmonic
- ② Short-time Fourier Transform (STFT) window size matters per-target² in VDBO problems
- ③ In musical and auditory contexts, frequency resolution should increase from high to low frequencies (vice-versa for time resolution)³
- ④ CQT⁴ uses long windows in low frequencies and short windows in high frequencies for the 12-tone Western pitch scale

¹ Derry Fitzgerald. 2010; Jonathan Driedger et al. 2014.

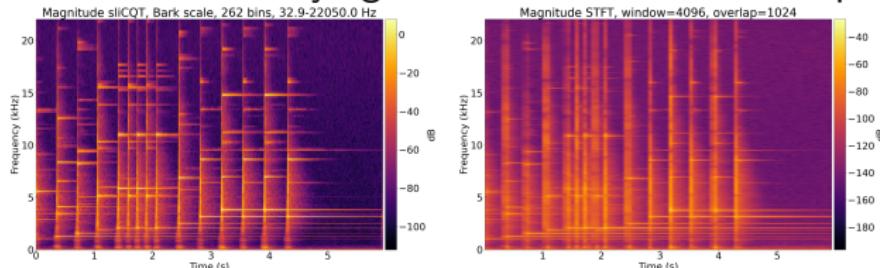
² Ilya Kavalerov et al. IEEE, 2019.

³ Christian Schörkhuber et al. 2012; Monika Dörfler. PhD thesis. 2002.

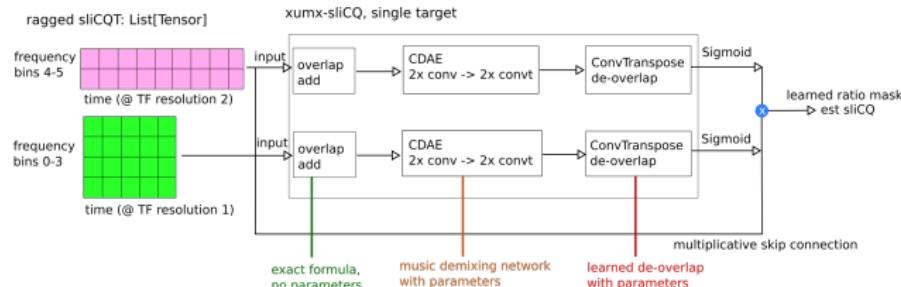
⁴ Judith Brown. (1991).

xumx-sliCQ v1 @ MDX 2021

- ① sliCQ Transform⁵ is an STFT-like transform with **perfect inverse** that uses time-varying windows for nonuniform spectral analysis



- ② Bark scale (262 bins 32.9–22050 Hz) sliCQT + convolutional denoising autoencoder (CDAE) architecture⁶ to achieve 3.6 dB SDR

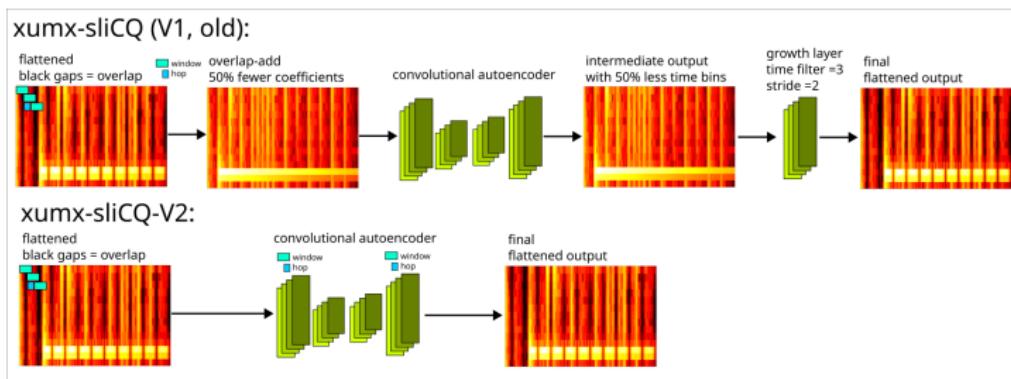


⁵ Gino Angelo Velasco et al. 2011; Nicki Holighaus et al. (2013).

⁶ Emad M. Grais et al. 2017.

xumx-sliCQ v2 @ CAD1 2023

- ① Bark scale may have some benefits for human listeners
- ② Focused solely on VDBO demixing problem
- ③ Better handling of overlap-add, mask sum loss, differentiable Wiener filtering, and complex MSE⁷: github.com/sevagh/xumx-sliCQ



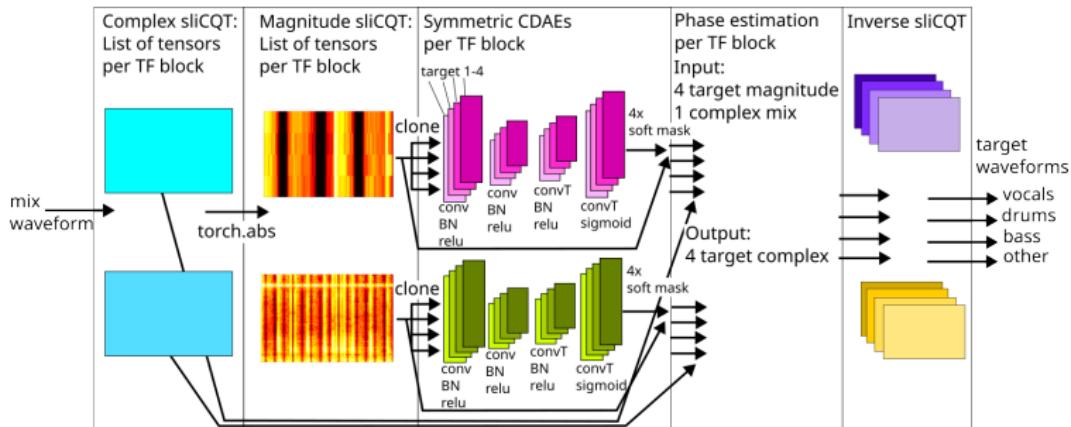
$$x_{\text{mix}} = x_v + x_d + x_b + x_o$$

$$|X|_{\text{mix}} = M_v |X|_{\text{mix}} + M_d |X|_{\text{mix}} + M_b |X|_{\text{mix}} + M_o |X|_{\text{mix}}$$

$$\rightarrow 1 = M_v + M_d + M_b + M_o$$

⁷ Chin-Yun Yu et al. *arXiv preprint arXiv:2112.03752* (2021).

xumx-sliCQT v2: results



- ① 4.4 dB SDR up from 3.6
- ② HAAQI score: mean of 0.094 vs. 0.255 of Baseline 1 (demucs)
- ③ BAQ score: mean of 41.84 vs. 41.40 of Baseline 1 (demucs)
- ④ More efficient (using bfloat16 for faster training and inference, etc.)
- ⑤ Weights are 60 MB
- ⑥ Fast realtime variant with 4.0 dB SDR using causal convolutions

New demixing-related project

Aim of these systems are to improve listening experience for those with different hearing.⁸ VDBO models are not very accessible (inscrutable Python errors, need >64GB RAM, etc.)

<https://freemusicdemixer.com>

Optimized C++ inference for UMX + Demucs, compiled to WebAssembly, running in the web, client-side on your browser, under 4 GB of memory

free-music-demixer
Split songs, demix music, and separate stems with our AI-based tool: free, private, and unlimited use directly in your browser

[Home](#) [About](#) [Blog](#) [Sponsors](#) [Source code](#)

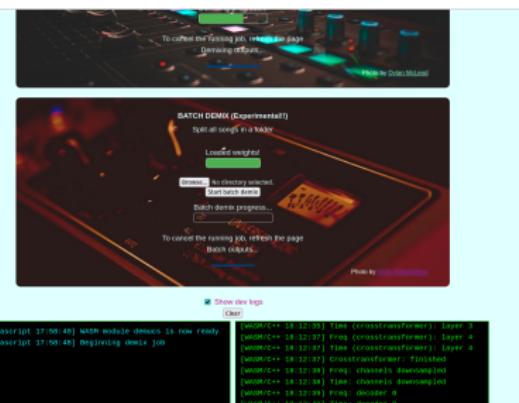
Free AI-powered music demixer

In music demixing, music source separation, or song splitting, AI models are used to separate the different instruments from a music recording into stems. This web application allows you to demix or split your music files, free and with no usage limits since it runs on your computer! ⓘ
Load a song to deconstruct it into bass, drums, vocals, melody, and karaoke using a near-state-of-the-art AI model. Open [UMX](#) with the [UMX](#), pretrained weights. This site is created and maintained by Sevag H.



Runs locally in your browser!

Unlike similar projects, it's free to use and doesn't store your data. All processing is done in your browser, and your files are never uploaded anywhere to fully respect user privacy. It runs well on computers and very slowly on smartphones; user beware.



To cancel the running job, refresh the page
Photo by [Sevag H.](#)

BATCH DEMIX (Experimental)
Split all songs in a folder
Loaded weightfile
Demos: No directory selected
Start batch demo
Batch demo progress...
To cancel the running job, refresh the page
Batch outputs...
Photo by [Sevag H.](#)

Show dev logs
 Clear

```
[JavaScript] 17:00:48 WASM module demucs is now ready
[JavaScript] 17:00:48 Beginning demix job
[JavaScript] 17:00:48 (17:00:48) Time (17:00:47.997777ms), Layer 0
[JavaScript] 17:00:48 (17:00:48) Free (constrainttransformer), layer 0
[JavaScript] 17:00:48 (17:00:48) Time (17:00:48.000000ms)
[JavaScript] 17:00:48 (17:00:48) constrainttransformer: finished
[JavaScript] 17:00:48 (17:00:48) Time (17:00:48.000000ms)
[JavaScript] 17:00:48 (17:00:48) constrainttransformer: uninitialised
[JavaScript] 17:00:48 (17:00:48) Free: demucs is uninitialised
[JavaScript] 17:00:48 (17:00:48) Time (17:00:48.000000ms)
[JavaScript] 17:00:48 (17:00:48) Free: demucs is initialised
[JavaScript] 17:00:48 (17:00:48) Time (17:00:48.000000ms)
```

⁸<http://cadenzachallenge.org/about>