MDX Workshop @ ISMIR • 12 November 2021

Deep learning for automatic mixing

Challenges and next steps

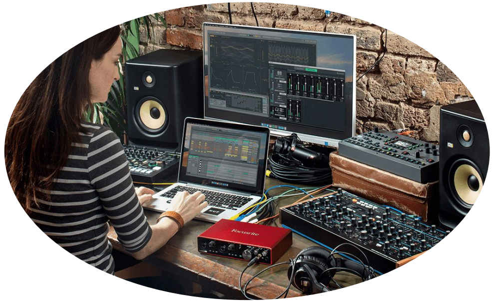
Christian J. Steinmetz
c.j.steinmetz@qmul.ac.uk
Centre for Digital Music, Queen Mary University of London

Queen Mary University of London

AI + MUSIC

UK Research and Innovation

More people are creating audio & music content than ever before



Over 60,000 new songs on Spotify every day

More than 500,000 active podcast shows

31 million YouTube channels producing audio/visual content

Demand for audio editing/creation tools is increasing,

especially tools that elevate the quality of **budget productions** or expedite workflow for professionals.

What is mixing?

Combine elements to create a cohesive mixture that serves the composition.

Combine can mean:
Leveling
Panning
Equalization
Compression
Limiting
Reverb/Delay
Creative effects...

It can be **really** complex!

LOGIC SESSION BREAKDOWN: "Moon River"

1,002,614 views • Streamed live on Dec 10, 2019

31K 312 SHARE SAVE ...

Jacob Collier

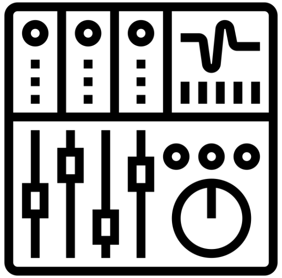
921K subscribers

SUBSCRIBE

Jacob Collier breaks down the logic session of his GRAMMY winning arrangement of Moon River.

Automatic mixing

Does mixing have to be so hard?

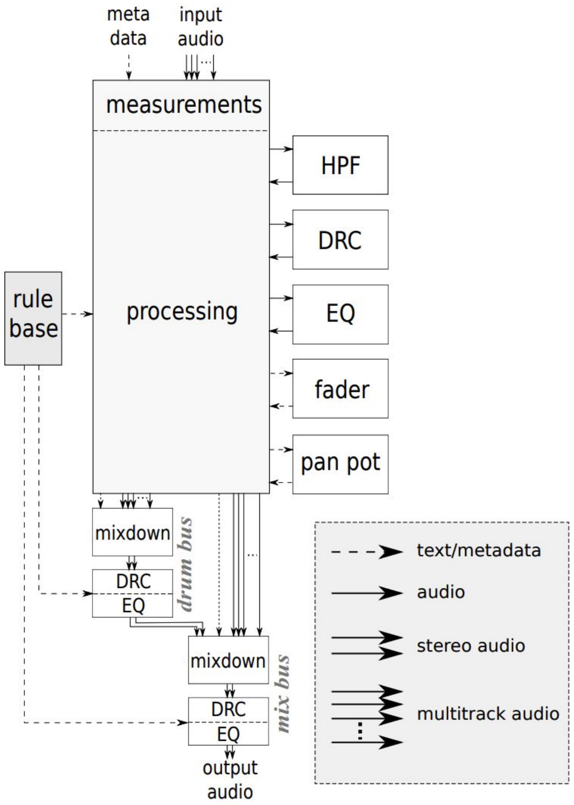


Expert systems

Design a set of rules based to create a mix based on analysis of the inputs.

Pro: Explainable decisions

Con: Often lacks sufficient complexity



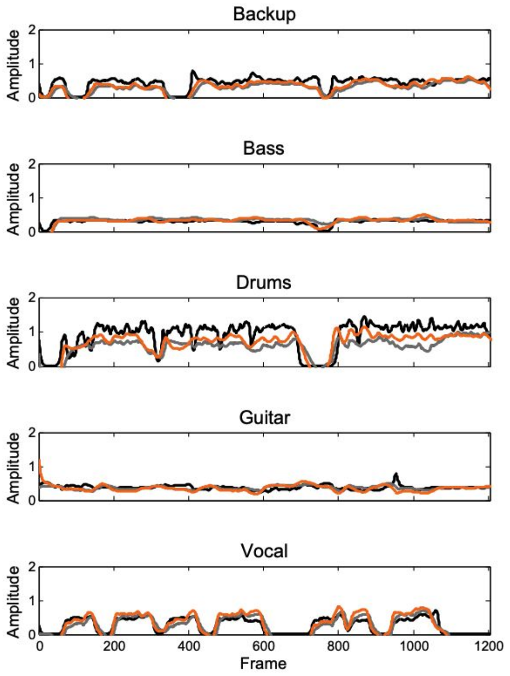
A knowledge-engineered autonomous mixing system
Brecht De Man, Joshua D. Reiss AES 2013

Machine Learning*

Learn to create a mix by leveraging parametric data collected from pros.

Pro: Greater model flexibility

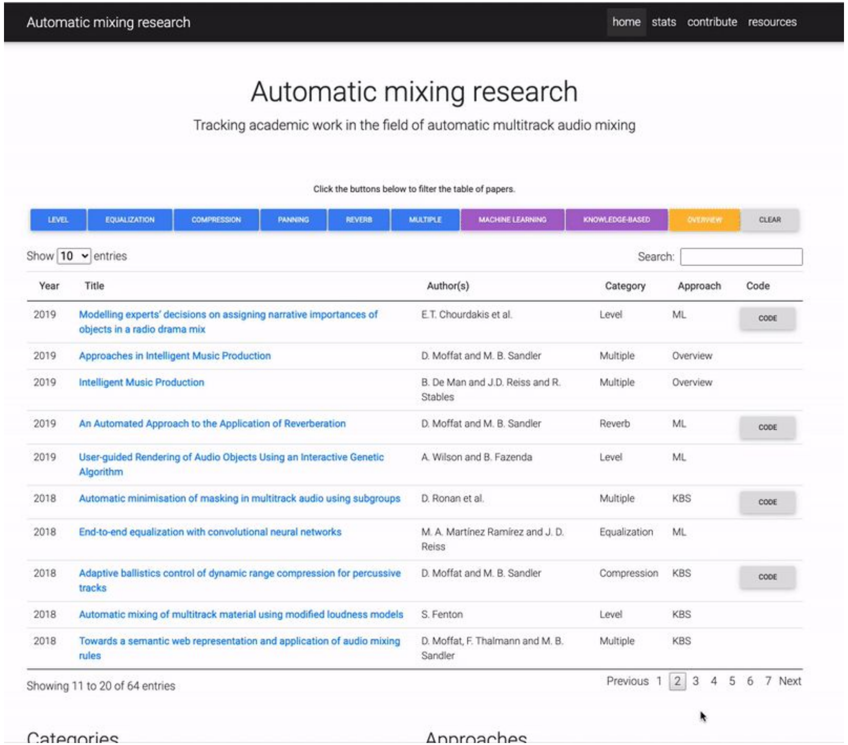
Con: Requires data (parametric)



Analysis of acoustic features for automated multitrack mixing
Jeffrey J. Scott. Youngmoo E. Kim ISMIR 2011

*Approaches that use classical machine learning techniques

More works on automatic mixing research


Automatic mixing research
home stats contribute resources

Automatic mixing research

Tracking academic work in the field of automatic multitrack audio mixing

Click the buttons below to filter the table of papers.

LEVEL EQUALIZATION COMPRESSION PANNING REVERB MULTIPLE MACHINE LEARNING KNOWLEDGE-BASED OVERVIEW CLEAR

Show 10 entries Search:

Year	Title	Author(s)	Category	Approach	Code
2019	Modelling experts' decisions on assigning narrative importances of objects in a radio drama mix	E.T. Chourdakis et al.	Level	ML	CODE
2019	Approaches in Intelligent Music Production	D. Moffat and M. B. Sandler	Multiple	Overview	
2019	Intelligent Music Production	B. De Man and J.D. Reiss and R. Stables	Multiple	Overview	
2019	An Automated Approach to the Application of Reverberation	D. Moffat and M. B. Sandler	Reverb	ML	CODE
2019	User-guided Rendering of Audio Objects Using an Interactive Genetic Algorithm	A. Wilson and B. Fazenda	Level	ML	
2018	Automatic minimisation of masking in multitrack audio using subgroups	D. Ronan et al.	Multiple	KBS	CODE
2018	End-to-end equalization with convolutional neural networks	M. A. Martínez Ramírez and J. D. Reiss	Equalization	ML	
2018	Adaptive ballistics control of dynamic range compression for percussive tracks	D. Moffat and M. B. Sandler	Compression	KBS	CODE
2018	Automatic mixing of multitrack material using modified loudness models	S. Fenton	Level	KBS	
2018	Towards a semantic web representation and application of audio mixing rules	D. Moffat, F. Thalmann and M. B. Sandler	Multiple	KBS	

Showing 11 to 20 of 64 entries

Previous 1 2 3 4 5 6 7 Next

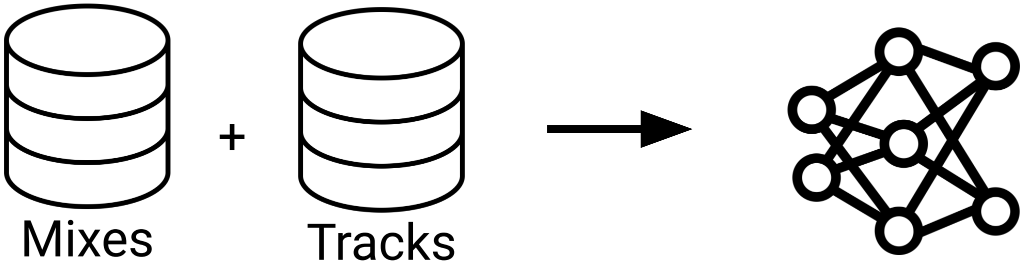
Categories Approaches

Searchable/filterable table of relevant papers and stats

https://csteinmetz1.github.io/AutomaticMixingPapers

Deep learning for automatic mixing

The promise of deep learning



*Can we **learn** to produce mixes directly from tracks and mixes without any knowledge of the parameters used?*

Many challenges

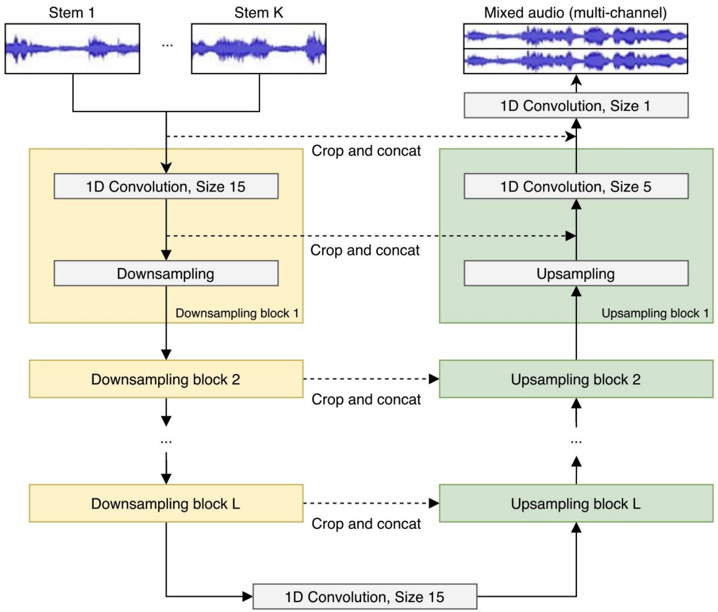
1. **Limited training data**	We need the original tracks and good mixes.
2. **Evaluation of mixes**	What makes a good mix? According to who?
3. **Highly variable inputs**	No consistent size and structure to inputs.
4. **High-fidelity required**	High sampling rates and no artifacts.
5. **User interaction**	Audio engineers need to tweak the output.

Great time to innovate on these hard problems!

Approaches

How can we apply deep learning?

Fully end-to-end approach



Contributions

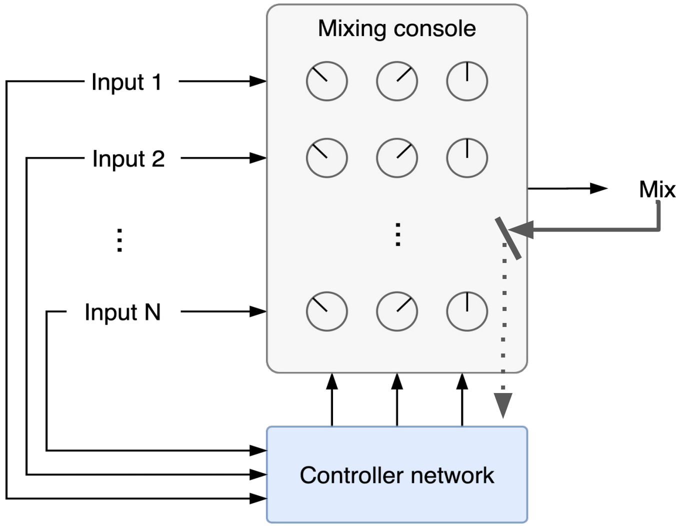
- Model problem as “reverse” source separation
- Produce high quality drum mixes
- Architecture requires no domain knowledge

Limitations

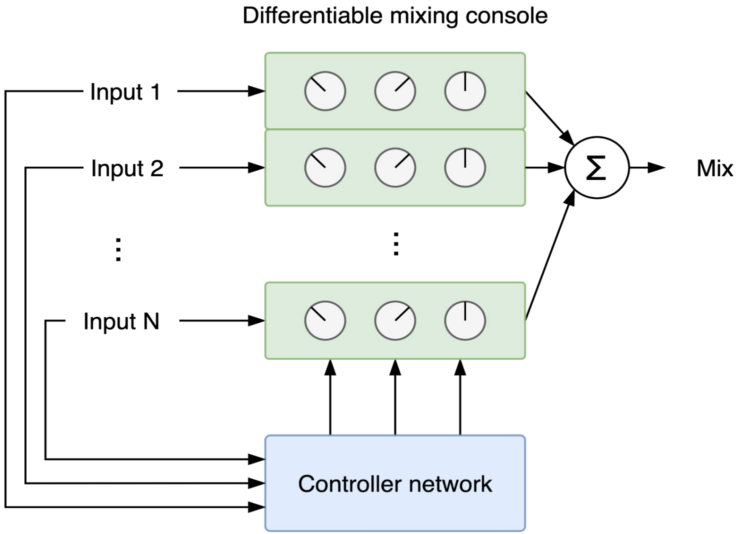
- Cannot adapt to variable track count
- Not permutation invariant to input tracks
- User cannot adjust the final result

A deep learning approach to intelligent drum mixing with the Wave-U-Net
Marco A Martínez Ramírez, Daniel Stoller, David Moffat JAES 2021

How can we learn to control audio effects?



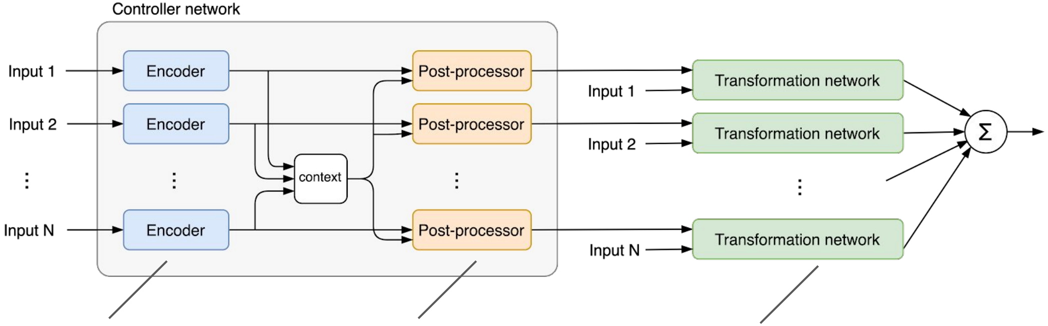
Unfortunately, the mixing console is not differentiable



Can we construct a differentiable mixing console?

Neural audio effects (Proxies)

Weight sharing across all input tracks



Analyze inputs

Predict effect parameters

Diff. proxy model processes audio

Contributions

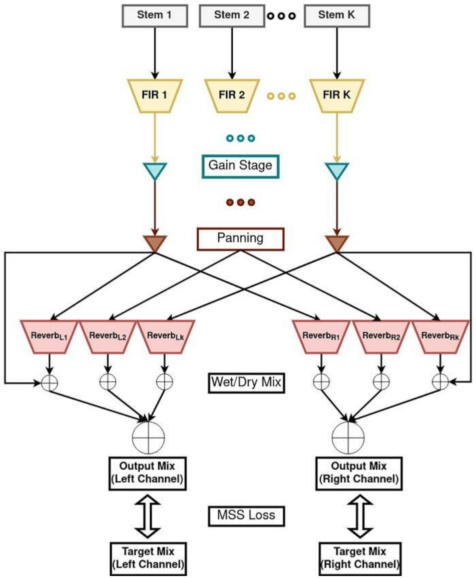
- Scale to any number of tracks
- Permutation invariant w.r.t input tracks
- Human readable control parameters
- Enforces no input taxonomy

Limitations

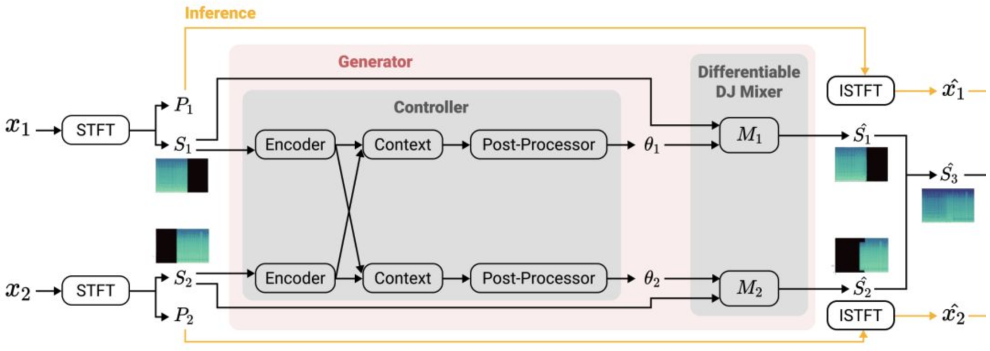
- Requires pre-training of effect networks
- Training requires significant compute

Automatic multitrack mixing with a differentiable mixing console of neural audio effects
Christian J. Steinmetz, Jordi Pons, Santiago Pascual, Joan Serrà ICASSP 2021

Autodiff (DDSP) effects

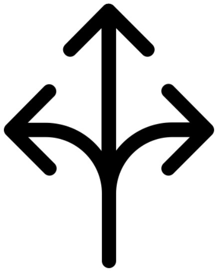


Reverse engineering of a recording mix with differentiable digital signal processing
Joseph T. Colonel and Joshua Reiss   JASA 2021

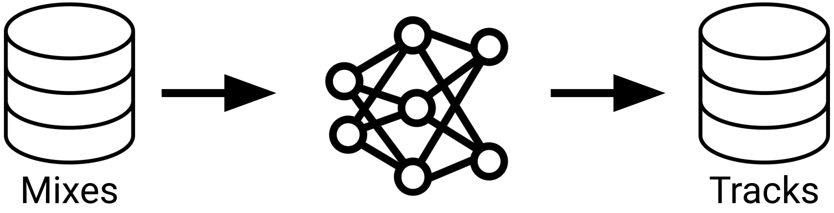


Automatic DJ Transitions with Differentiable Audio Effects and Generative Adversarial Networks
Chen et al.   arXiv 2021

Source separation for automatic mixing



Source separation for mixing data generation



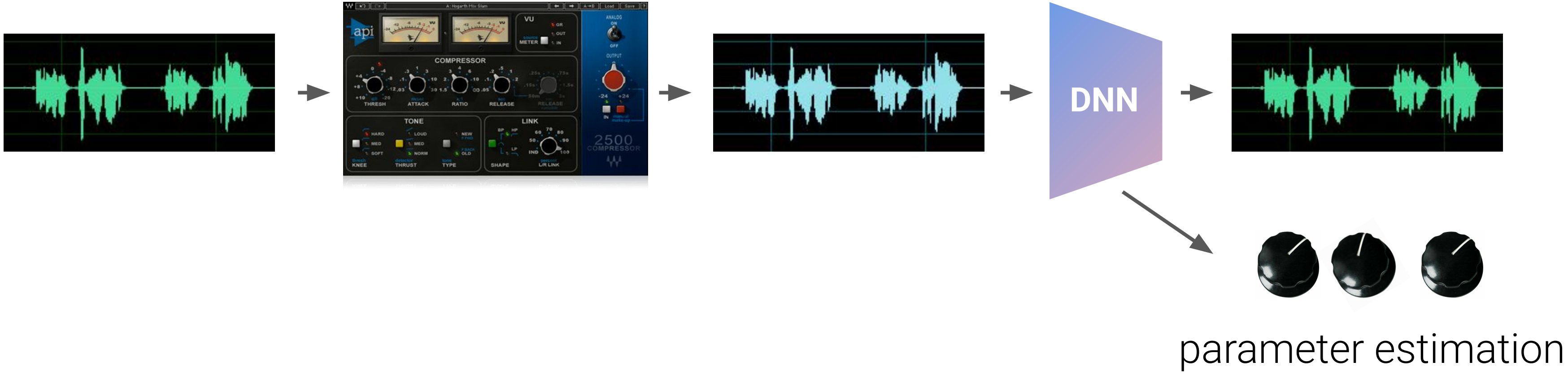
We have many more mixes than we do multitracks

Can we use source separation to leverage the large collections of produced audio datasets that are available (FMA, MSD, MTG-Jamendo, etc)?

... But, current source separation algorithms do not consider audio effects...

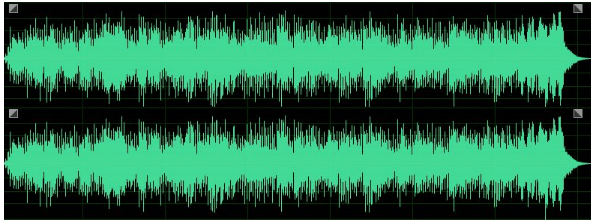
Estimating the loudness balance of musical mixtures using audio source separation
Dominic Ward, Hagen Wierstorf, Russell Mason, Mark Plumbley, Christopher Hummersone WIMP 2017

Audio effect removal (and parameter estimation)

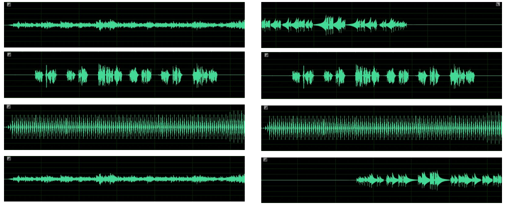


Guitar effects recognition and parameter estimation with convolutional neural networks
Marco Comunità, Dan Stowell, Joshua D Reiss JAES 2021

🏆 **Holy grail:** Fully deconstructing a mix

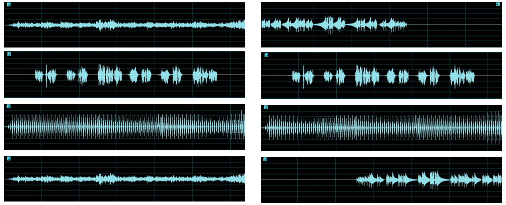


1. Detailed separation



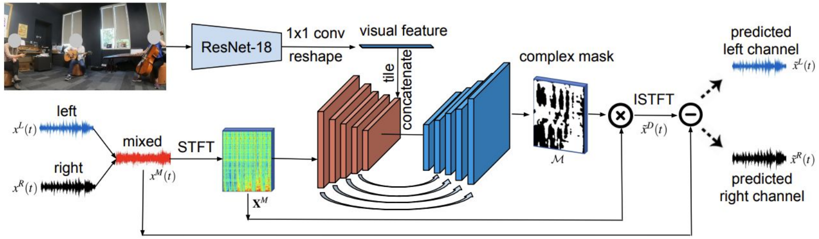
2. Effect parameter estimation

3. Effect removal

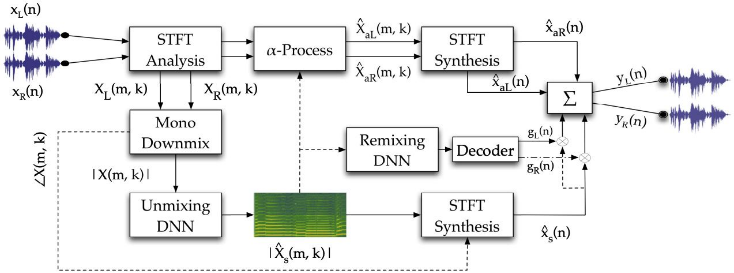


Automatic upmixing

Can we separate music sources and automatically generate spatial audio content?

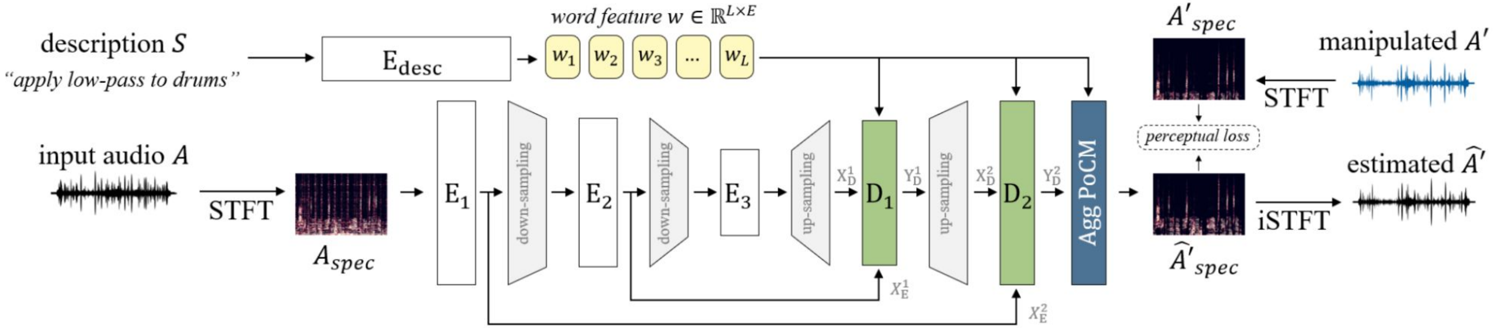


2.5D Visual Sound
Ruohan Gao, Kristen Grauman CVPR 2019



New sonorities for jazz recordings: Separation and mixing using deep neural networks
Stylianos Ioannis Mimilakis, et al. WIMP 2016

Audio source manipulation



AMSS-Net: Audio manipulation on user-specified sources with textual queries
Choi et al. ACM Multimedia 2021

Music source separation is a subset of general audio manipulation!

Audio source manipulation

What could future audio manipulation interfaces look like?

Recap

Recap

- Significant demand for audio tools that make things easier
- Research on **deep learning** for audio production is **very** fresh
- Potential research directions
    a. Advancing DL for automatic mixing
        i. Data generation via source separation
        ii. More differentiable audio effects
        iii. Scalable architectures for multichannel audio
    b. Music source separation with audio effect removal
    c. Separation for automatic upmixing (e.g. stereo -> Dolby Atmos)
    d. General audio source manipulation

References

A knowledge-engineered autonomous mixing system
Brecht De Man, Joshua D. Reiss
AES 2013

Analysis of acoustic features for automated multitrack mixing
Jeffrey J. Scott, Youngmoo E. Kim
ISMIR 2011

A deep learning approach to intelligent drum mixing with the Wave-U-Net
Marco A Martínez Ramírez, Daniel Stoller, David Moffat
JAES 2021

Automatic multitrack mixing with a differentiable mixing console of neural audio effects
Christian J. Steinmetz, Jordi Pons, Santiago Pascual, Joan Serrà
ICASSP 2021

Reverse engineering of a recording mix with differentiable digital signal processing
Joseph T. Colonel, Joshua Reiss
JASA 2021

Automatic DJ Transitions with Differentiable Audio Effects and Generative Adversarial Networks
Chen et al.
arXiv 2021

Estimating the loudness balance of musical mixtures using audio source separation
Ward et al.
WIMP 2017

Guitar effects recognition and parameter estimation with convolutional neural networks
Marco Comunità, Dan Stowell, Joshua D Reiss
JAES 2021

2.5D Visual Sound
Ruohan Gao, Kristen Grauman
CVPR 2019

New sonorities for jazz recordings: Separation and mixing using deep neural networks
Stylianos Ioannis Mimilakis, et al.
WIMP 2016

AMSS-Net: Audio manipulation on user-specified sources with textual queries
Choi et al.
ACM Multimedia 2021

Attribution

Videos

- Analog Mixing (SSL Console) GoPro POV - "Get Out Of Bed" by Magician's Nephew
  From Daniel Duskin

- LOGIC SESSION BREAKDOWN: "Moon River"
  From Jacob Collier

Icons

- mixing board by Eucalyp from the Noun Project
- arrows by Ker'is from the Noun Project
- History by The Icon Z from the Noun Project
- Head by ainul muttaqin from the Noun Project
- Brain by achmad mulyana from the Noun Project
- Data by arjuazka from the Noun Project
- Neural Network by Ian Rahmadi Kurniawan from the Noun Project

MDX Workshop @ ISMIR • 12 November 2021

Deep learning for automatic mixing

Challenges and next steps

Christian J. Steinmetz
c.j.steinmetz@qmul.ac.uk
Centre for Digital Music, Queen Mary University of London

Queen Mary University of London
AI + MUSIC
UK Research and Innovation