INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

Musically Informed Audio Decomposition

Meinard Müller and Jonathan Driedger

International Audio Laboratories Erlangen
meinard.mueller@audiolabs-erlangen.de

Oberseminar
Lehrstuhl für Complex and Intelligent Systems
Universität Passau
18.01.2017

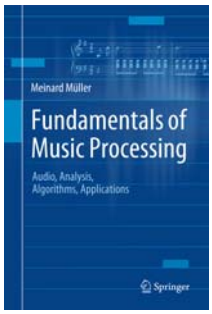
FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Meinard Müller

- 2001 PhD, Bonn University
- 2002/2003 Postdoc, Keio University, Japan
- 2007 Habilitation, Bonn University
  "*Information Retrieval for Music and Motion*"
- 2007-2012 Senior Researcher
  Max-Planck Institut für Informatik, Saarland
- 2012: Professor
  Semantic Audio Processing
  Universität Erlangen-Nürnberg

Book: Fundamentals of Music Processing



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

International Audio Laboratories Erlangen

Fraunhofer IIS

mp3

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

AUDIO LABS

Research Group
Semantic Audio Signal Processing

- Thomas Prätzlich
- Christof Weiß
- Stefan Balke
- Christian Dittmar
- Patricio López-Serrano
- Frank Zalkow





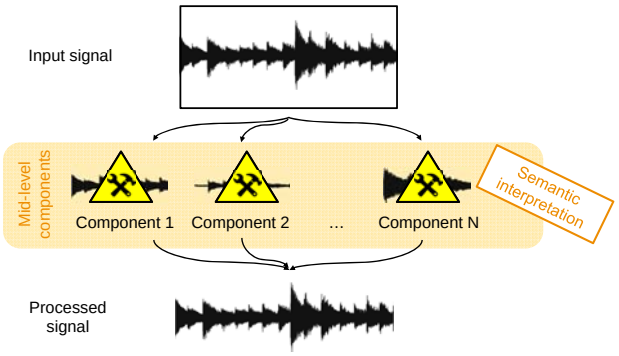

Research Group
Semantic Audio Signal Processing

- Jonathan Driedger

PhD Thesis
Processing Music Signals Using Audio Decomposition Techniques
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
May 2016

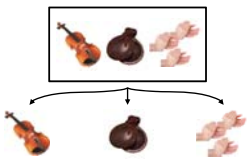
Music Signal Processing

Why Audio Decomposition?



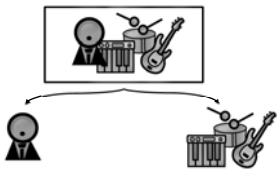
Structure of this Talk

- **Harmonic-percussive-residual decomposition**



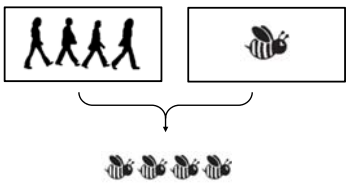
Structure of this Talk

- Harmonic-percussive-residual decomposition
- **Singing voice extraction**

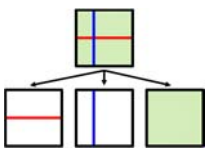


Structure of this Talk

- Harmonic-percussive-residual decomposition
- Singing voice extraction
- **Audio mosaicing**

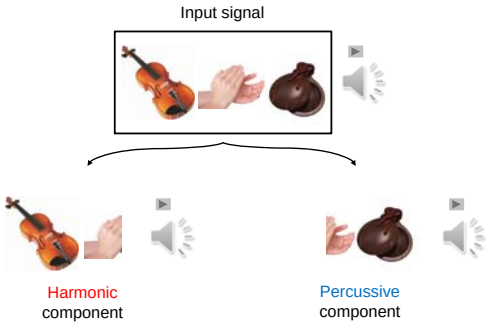


Harmonic-Percussive-Residual Decomposition



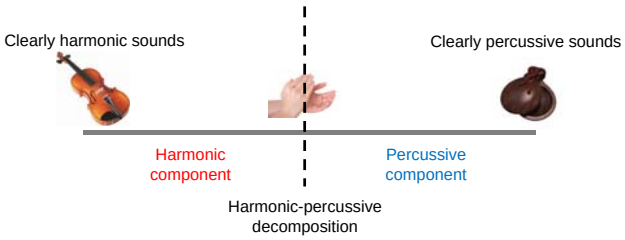
Harmonic-Percussive Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]



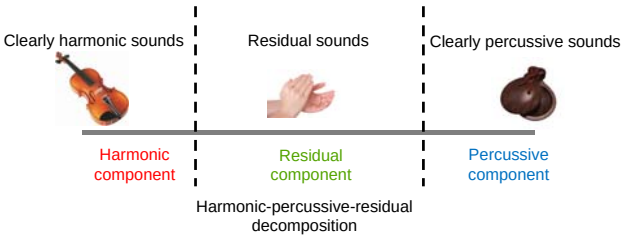
Harmonic-Percussive Decomposition

The "harmonic-percussive scale":



Harmonic-Percussive-Residual Decomposition

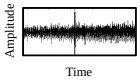
The "harmonic-percussive scale":



Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

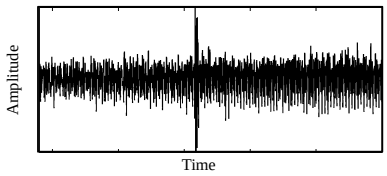
Audio



Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

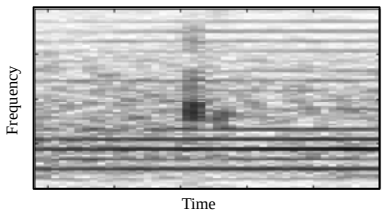
Audio



Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

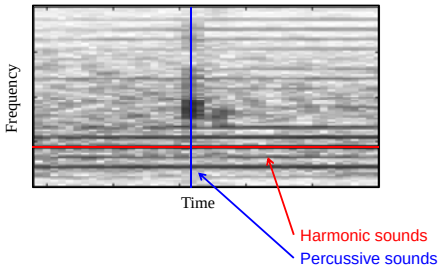
Spectrogram



Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

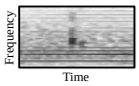
Spectrogram



Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

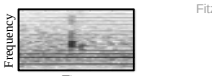
Spectrogram



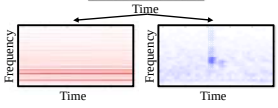
Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

Spectrogram



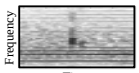
Horizontally and vertically enhanced spectrograms



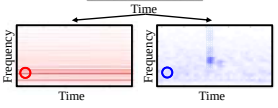
Harmonic-Percussive-Residual Decomposition

[Ono et al. ISMIR 2008, Fitzgerald DAFx 2010]

Spectrogram



Horizontally and vertically enhanced spectrograms

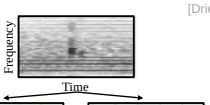


○ > ○ ➔ Harmonic component

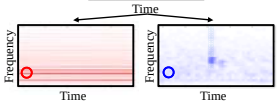
Harmonic-Percussive-Residual Decomposition

[Driedger et al. ISMIR 2014]

Spectrogram

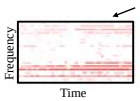


Horizontally and vertically enhanced spectrograms



Separation factor $\beta \geq 1$

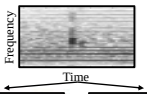
○ > β • ○ ➔ Harmonic component



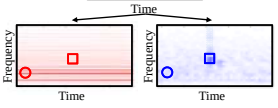
Harmonic-Percussive-Residual Decomposition

[Driedger et al. ISMIR 2014]

Spectrogram



Horizontally and vertically enhanced spectrograms



Separation factor $\beta \geq 1$

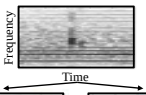
○ > β • ○ ➔ Harmonic component

β • □ < □ ➔ Percussive component

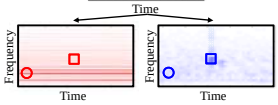
Harmonic-Percussive-Residual Decomposition

[Driedger et al. ISMIR 2014]

Spectrogram



Horizontally and vertically enhanced spectrograms



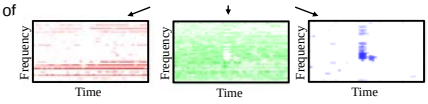
Separation factor $\beta \geq 1$

$\circ > \beta \bullet \circ$ ➔ Harmonic component

$\beta \bullet \square <$ ➔ Percussive component

otherwise ➔ Residual component

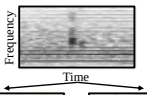
Spectrograms of the harmonic, residual and percussive components



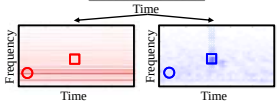
Harmonic-Percussive-Residual Decomposition

[Driedger et al. ISMIR 2014]

Spectrogram



Horizontally and vertically enhanced spectrograms



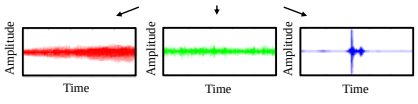
Separation factor $\beta \geq 1$

$\circ > \beta \bullet \circ$ ➔ Harmonic component

$\beta \bullet \square <$ ➔ Percussive component

otherwise ➔ Residual component

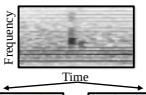
Harmonic, residual and percussive components



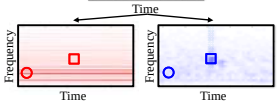
Harmonic-Percussive-Residual Decomposition

[Driedger et al. ISMIR 2014]

Spectrogram



Horizontally and vertically enhanced spectrograms



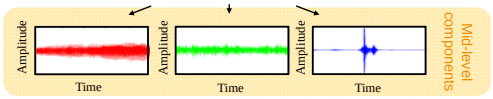
Separation factor $\beta \geq 1$

$\circ > \beta \bullet \circ$ ➔ Harmonic component

$\beta \bullet \square <$ ➔ Percussive component

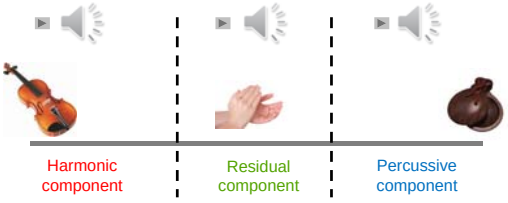
otherwise ➔ Residual component

Harmonic, residual and percussive components



Harmonic-Percussive-Residual Decomposition

Mixture

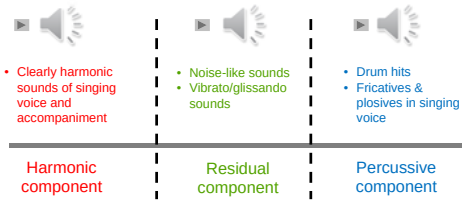


Harmonic component | Residual component | Percussive component

$\beta = 2$

Harmonic-Percussive-Residual Decomposition

Mixture:

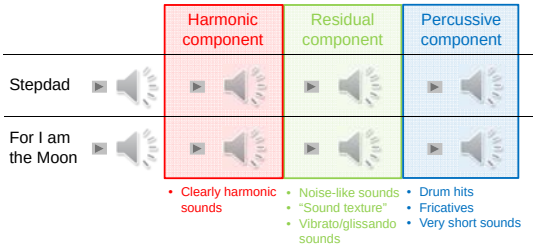


- Clearly harmonic sounds of singing voice and accompaniment
- Noise-like sounds
- Vibrato/glissando sounds
- Drum hits
- Fricatives & plosives in singing voice

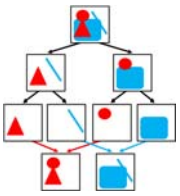
Harmonic component | Residual component | Percussive component

$\beta = 2$

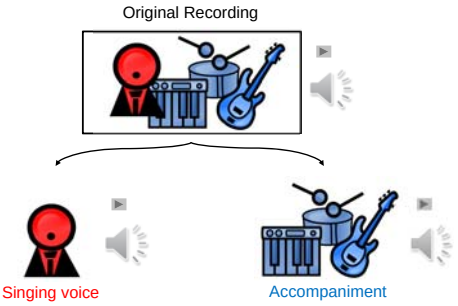
Additional Examples



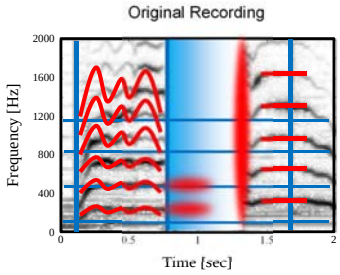
Singing Voice Extraction



Singing Voice Extraction

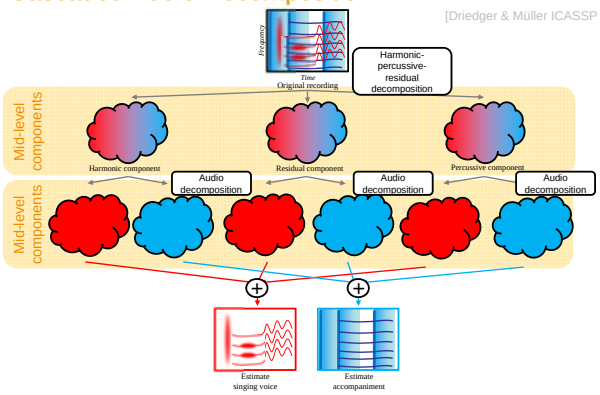


Singing Voice Extraction



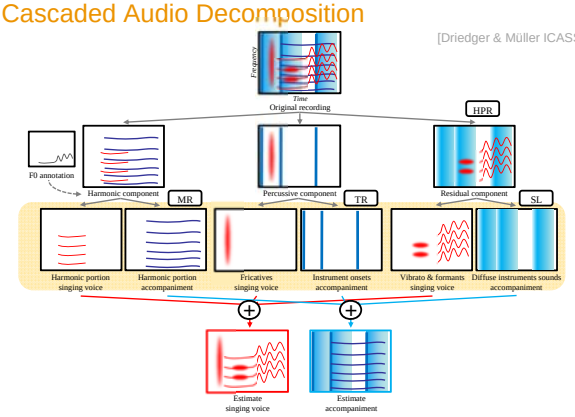
Cascaded Audio Decomposition

[Driedger & Müller ICASSP 2015]

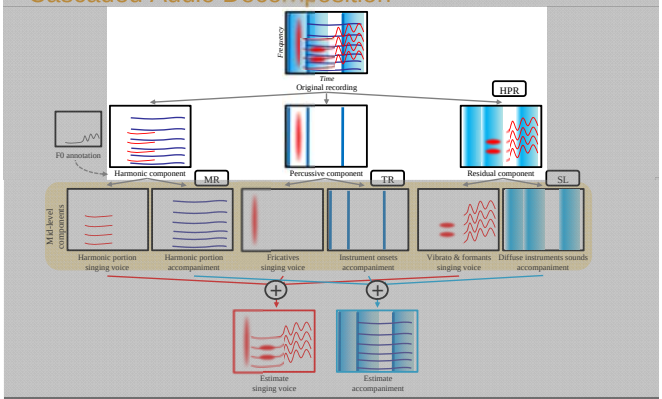


Cascaded Audio Decomposition

[Driedger & Müller ICASSP 2014]

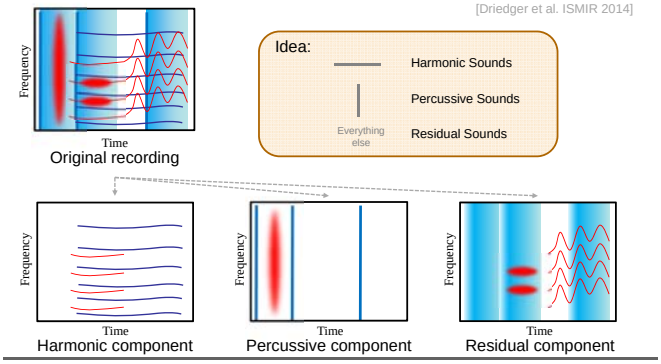


Cascaded Audio Decomposition

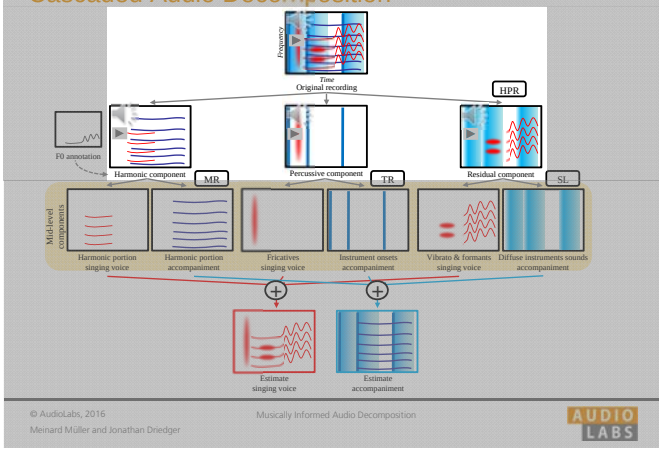


HPR – Harmonic-Percussive-Residual Decomposition

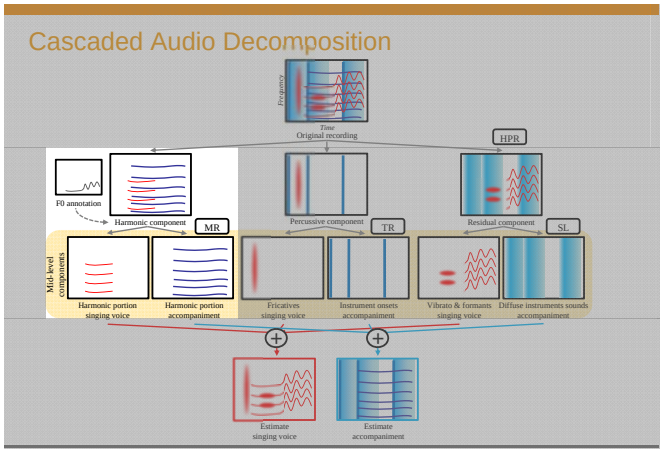
[Tachibana et al. IEEE-TASLP 2013]
[Jeong & Lee IEEE-SPL 2014]
[Driedger et al. ISMIR 2014]



Cascaded Audio Decomposition

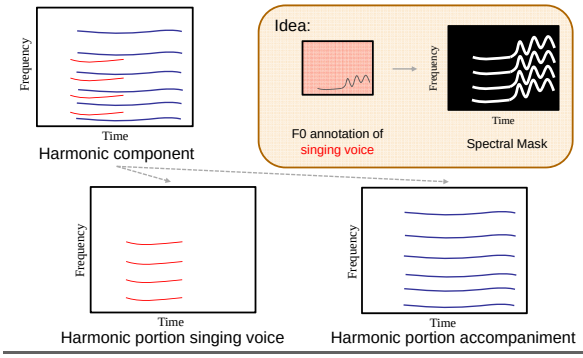


Cascaded Audio Decomposition

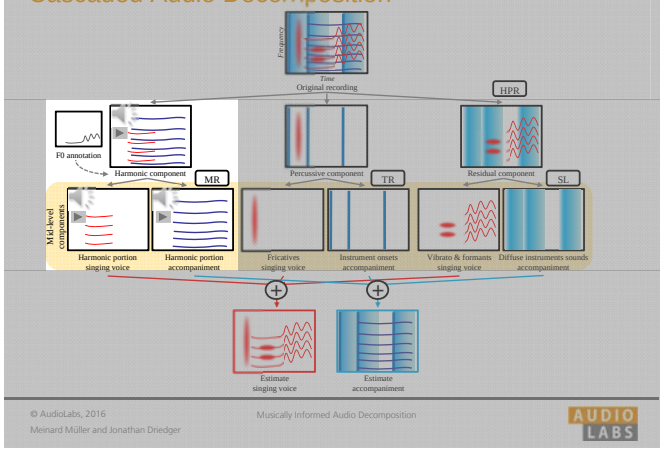


MR – Melody/Residual Decomposition

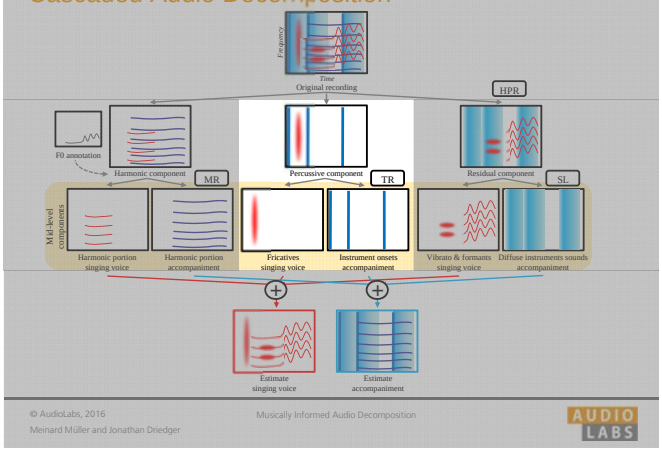
[Virtanen et al. ISCA 2008]
[Salamon & Gómez IEEE-TASLP 2012]



Cascaded Audio Decomposition

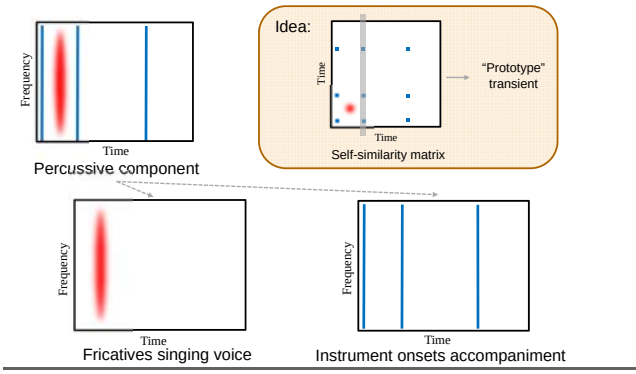


Cascaded Audio Decomposition

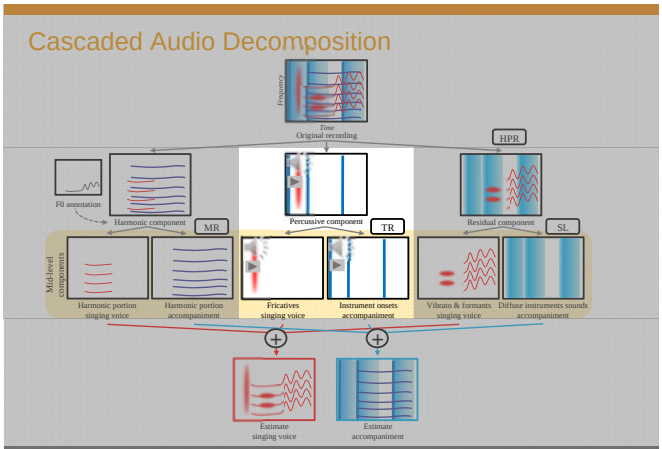


TR – Transient/Residual Decomposition

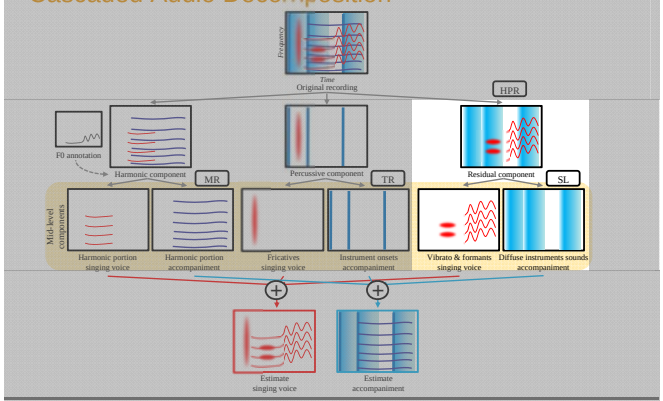
[Talmon et al. IEEE-TASLP 2011]



Cascaded Audio Decomposition

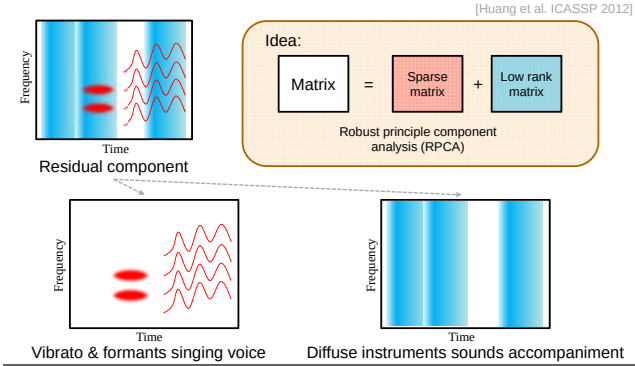


Cascaded Audio Decomposition

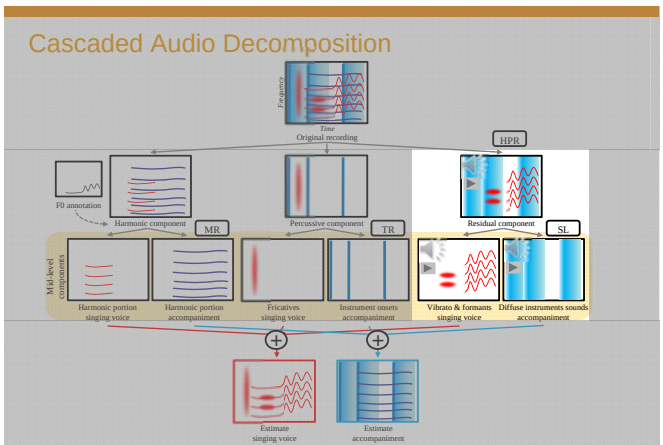


SL – Sparse/Low Rank Decomposition

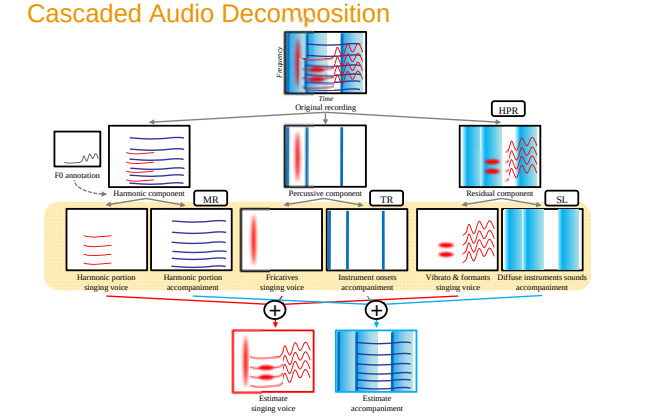
[Candès et al. JACM 2011]
[Huang et al. ICASSP 2012]



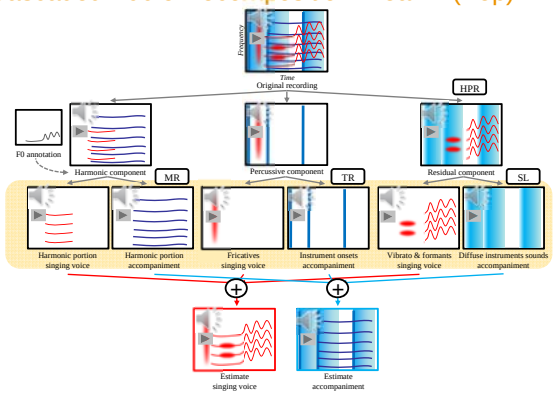
Cascaded Audio Decomposition



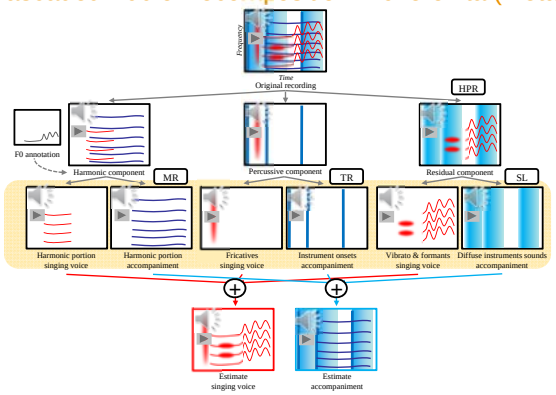
Cascaded Audio Decomposition



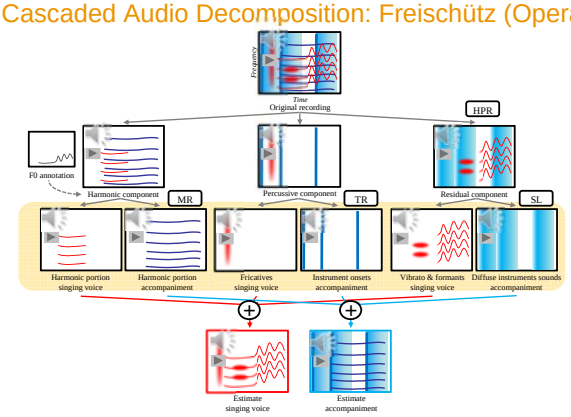
Cascaded Audio Decomposition: Bearlin (Pop)



Cascaded Audio Decomposition: Acheronita (Metal)

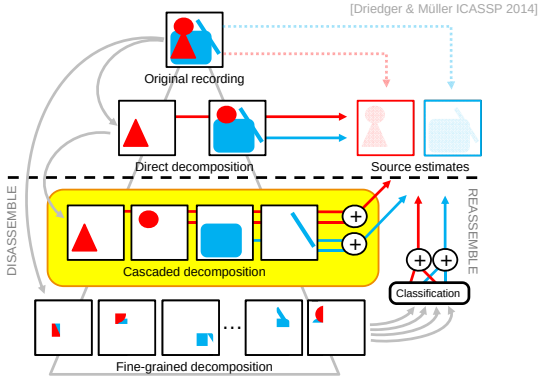


Cascaded Audio Decomposition: Freischütz (Opera)



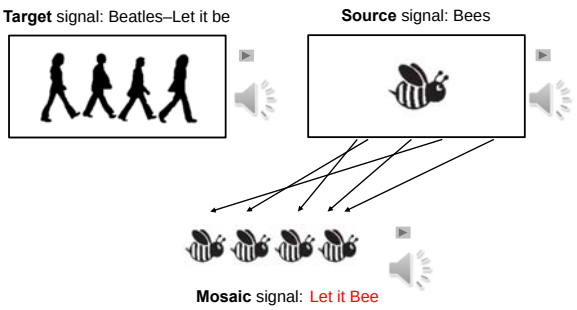
Approaches to Singing Voice Extraction

[Driedger & Müller ICASSP 2014]



Audio Mosaicing

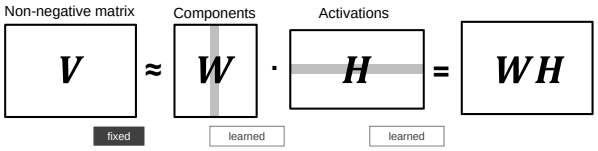
Audio Mosaicing



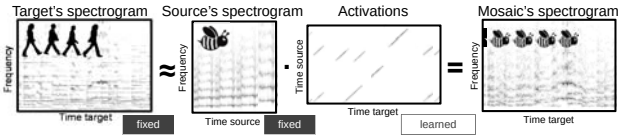
NMF-Inspired Audio Mosaicing

Non-negative matrix factorization (NMF)

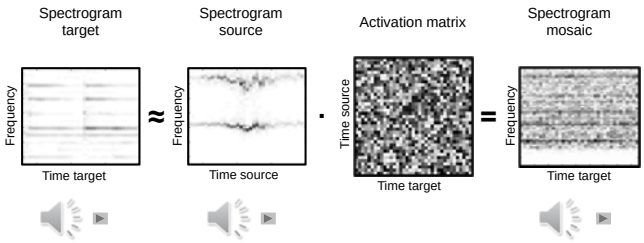
[Driedger et al. ISMIR 2015]



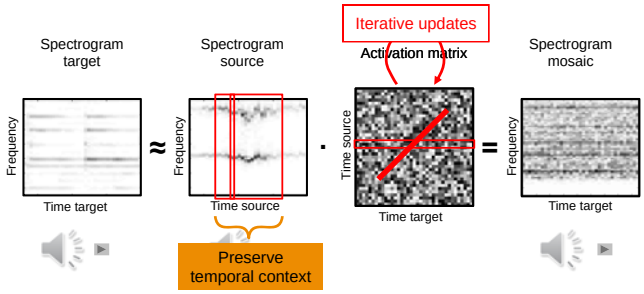
Proposed audio mosaicing approach



Basic NMF-Inspired Audio Mosaicing



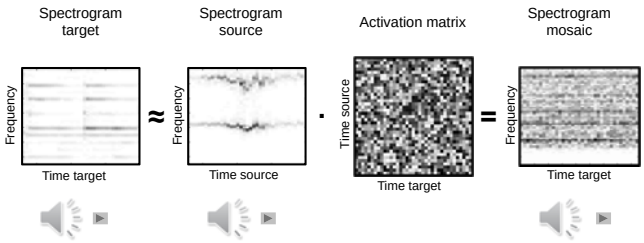
Basic NMF-Inspired Audio Mosaicing



Core idea: support the development of sparse diagonal activation structures

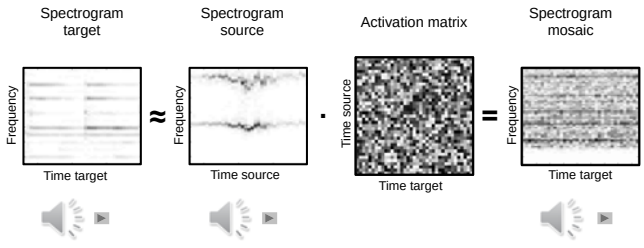
Audio Mosaicing with Extended Set of Update Rules

[Driedger et al. ISMIR 2015]

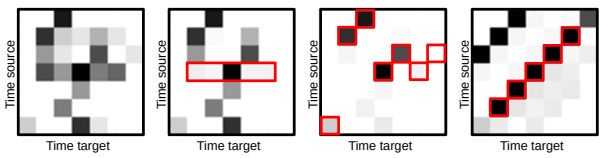


Audio Mosaicing with Extended Set of Update Rules

[Driedger et al. ISMIR 2015]



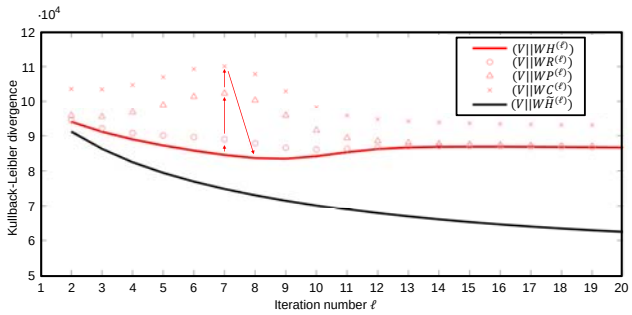
NMF with Extended Set of Update Rules



Activation matrix	Repetition-restricted activation matrix	Polyphony-restricted activation matrix	Continuity-enhanced activation matrix
	Neighborhood constraints	Column constraints	Diagonal smoothing

- Constraints are enforced by additional update rules
- Additional rules are interleaved with standard NMF update rules
- Soft alternative to NMFD

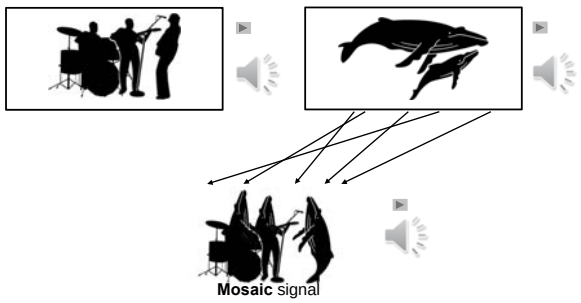
Kullback-Leibler Divergence between Target and Mosaic



Audio Mosaicing

Target signal: Chic–Good times

Source signal: Whales

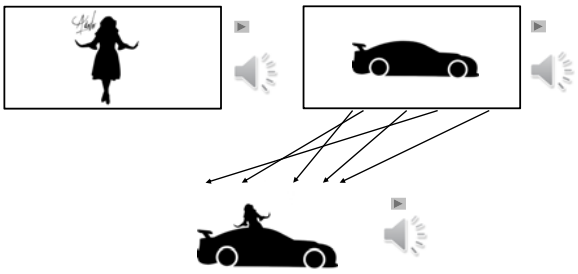


Mosaic signal

Audio Mosaicing

Target signal: Adele–Rolling in the Deep

Source signal: Race car



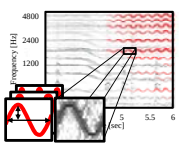
Mosaic signal

Summary: Musically Informed Audio Decomposition

- Harmonic-percussive-residual decomposition
- Singing voice extraction
- Audio mosaicing

Summary: Musically Informed Audio Decomposition

- Harmonic-percussive-residual decomposition
- Singing voice extraction
- Audio mosaicing
- **Vibrato-residual decomposition**



Mix

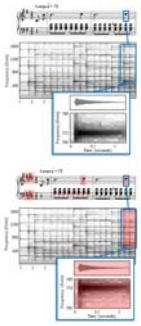
Vibrato

Residual

[Driedger et al., ISMIR 2016]

Summary: Musically Informed Audio Decomposition

- Harmonic-percussive-residual decomposition
- Singing voice extraction
- Audio mosaicing
- Vibrato-residual decomposition
- **Score-informed audio editing**



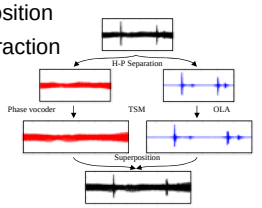
Original

Edited

[Driedger et al. ACM-MM 2013]

Summary: Musically Informed Audio Decomposition

- Harmonic-percussive-residual decomposition
- Singing voice extraction
- Audio mosaicing
- Vibrato-residual decomposition
- Score-informed audio editing
- **Time-scale modification**

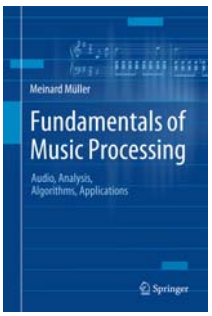


Original

Stretched by a factor of 1.8

[Driedger et al. IEEE-SPL 2014, Driedger & Müller Appl. Sci. 2016]

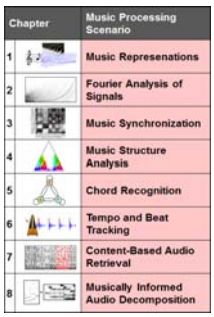
Book: Fundamentals of Music Processing



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

Book: Fundamentals of Music Processing



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

References

[Fitzgerald DAFx 2010] Derry Fitzgerald. Harmonic/percussive separation using medianfiltering. In Proceedings of the International Conference on Digital Audio Effects (DAFx), pages 246–253, Graz, Austria, 2010.

[Tachibana et al. IEEE-TASLP 2013] Hideyuki Tachibana, Nobutaka Ono, and Shigeki Sagayama. Singing voice enhancement in monaural music signals based on two-stage harmonic/percussive sound separation on multiple resolution spectrograms. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(1):228–237, January 2013.

[Driedger et al. ISMIR 2014] Jonathan Driedger, Meinard Müller, and Sascha Disch, "Extending harmonic-percussive separation of audio signals," in Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014.

[Virtanen et al. ISCA 2008] Tuomas Virtanen, Annamaria Mesaros, and Matti Ryynänen, "Combining pitch-based inference and non-negative spectrogram factorization in separating vocals from polyphonic music," in Proceedings of the ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition (SAPA), 2008

[Salamon & Gómez IEEE-TASLP 2012] Justin Salamon and Emilia Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE Transactions on Audio, Speech & Language Processing, vol. 20, no. 6, pp. 1759–770, 2012.

[Talmon et al. IEEE-TASLP 2011] Ronen Talmon, Israel Cohen, and Sharon Gannot, "Transient noise reduction using nonlocal diffusion filters," IEEE Transactions on Audio, Speech & Language Processing, 2011.

[Candès et al. JACM 2011] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," Journal of the ACM, vol. 58, no. 3, pp. 11:1–11:37, June 2011.

[Huang et al. ICASSP 2012] Po-Sen Huang, Scott Deeann Chen, Paris Smaragdis, and Mark Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in Proceedings of ICASSP, 2012.

Acknowledgements

This work has been supported by the German Research Foundation (DFG MU 2686/6-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS.