http://zenodo.org/record/1422385, http://www.ircam.fr, http://abcdj.eu

# Methods and Datasets for DJ-Mix Reverse Engineering

Diemo Schwarz, Dominique Fourer
Ircam Lab, CNRS, Sorbonne Université, Ministère de la Culture, Paris, France
IBISC, Université d'Évry-Val-d'Essonne/Paris-Saclay, Évry, France

ABC_DJ Artist to Business to Business to Consumer Audio Branding System

European Commission

Innovation Union

ircam Centre Pompidou

cnrs
Depuis 80 ans, nos connaissances bâtissent de nouveaux mondes

SORBONNE UNIVERSITÉ

# Collaboration Context: The ABC DJ EU-Project

ABC_DJ Artist to Business to Business to Consumer Audio Branding System

MIR tools for audio branding

automatic, DJ-like playback of playlists in stores

http://abcdj.eu

HearDis! GmbH

The ABC DJ project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 688122.

# Scientific Context:
# Understanding DJ Culture & Practices

Important part of popular music culture

Enables:

- musicological research in popular music
- studies on DJ culture
- computer support of DJing
- automation of DJ mixing

Qualitative accounts exists, but…

# Problem:
# Lack of Annotated Databases of DJ Mixes or DJ Sets

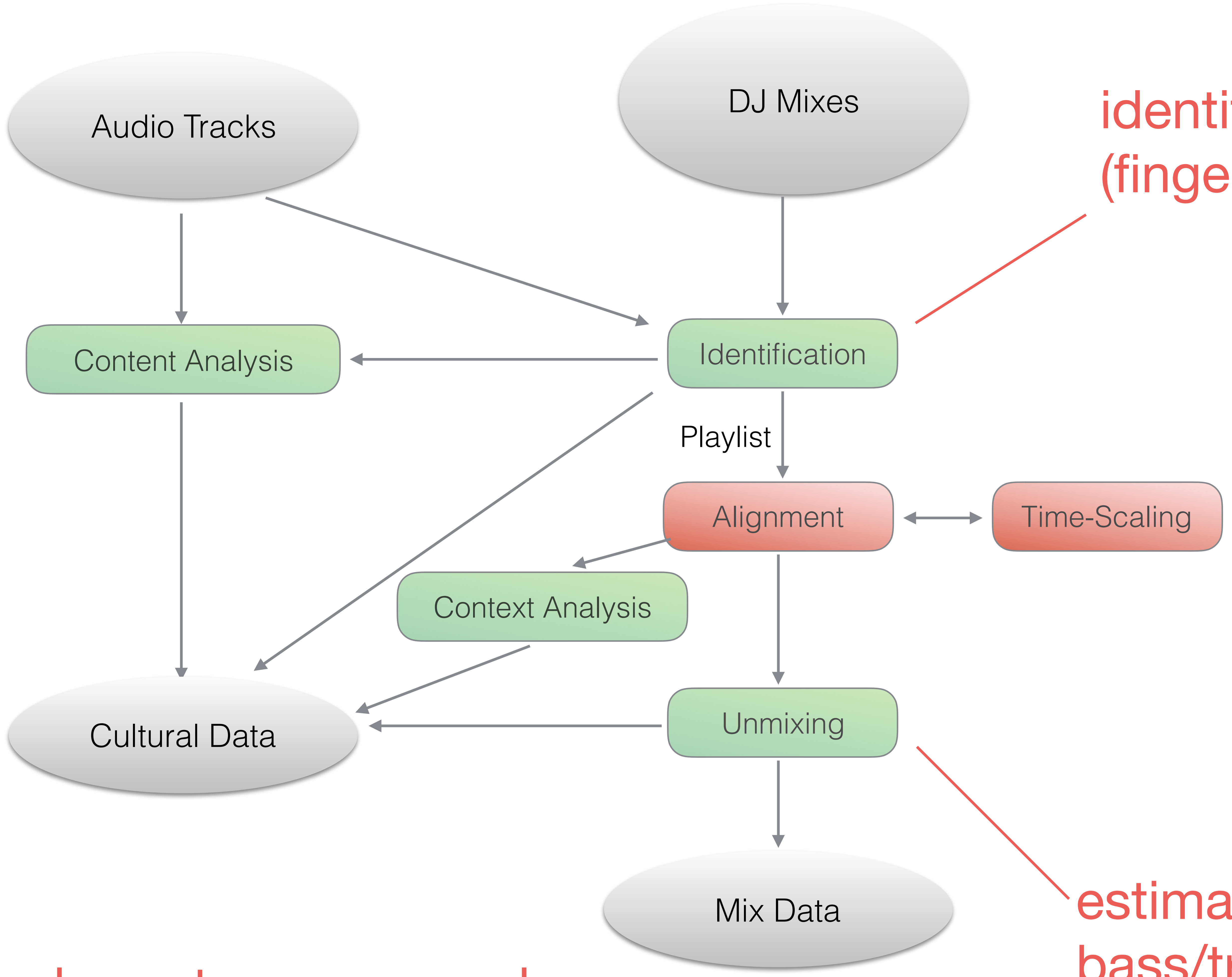
Very large scale availability **(millions)** of DJ mixes, often with tracklist,
e.g. http://www.mixcloud.com, YouTube, podcasts.

**very few annotated databases**

Existing research in studio multi-track mixing and unmixing in DAWs

Existing work on DJ production tools, but **no information retrieval from recorded mixes**

# Needed Components




identify contained tracks (fingerprinting)

get track start and end in mix

determine tempo changes (beat-aligned mixing)

*suggested here*

derive genre and social tags attached to the music → inform about the choices a DJ makes when creating a mix

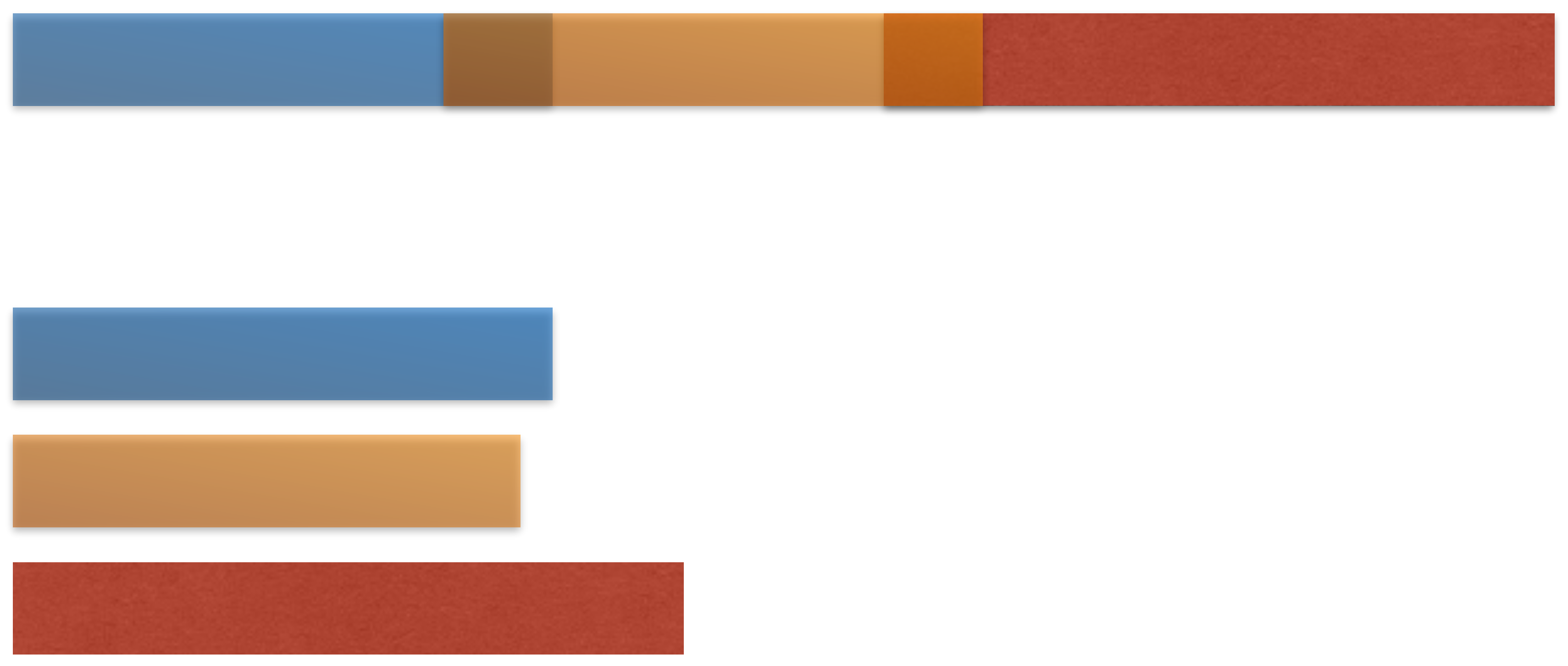
estimate fade curves for volume, bass/treble, and parameters of other effects (compression, echo, etc.)

downstream research enabled by DJ mix annotation

# Proposed Method for DJ Mix Reverse Engineering

**Input**

- recorded DJ mix
- playlist (list of tracks in the mix in correct order)
- audio files of the original tracks



**Five steps**

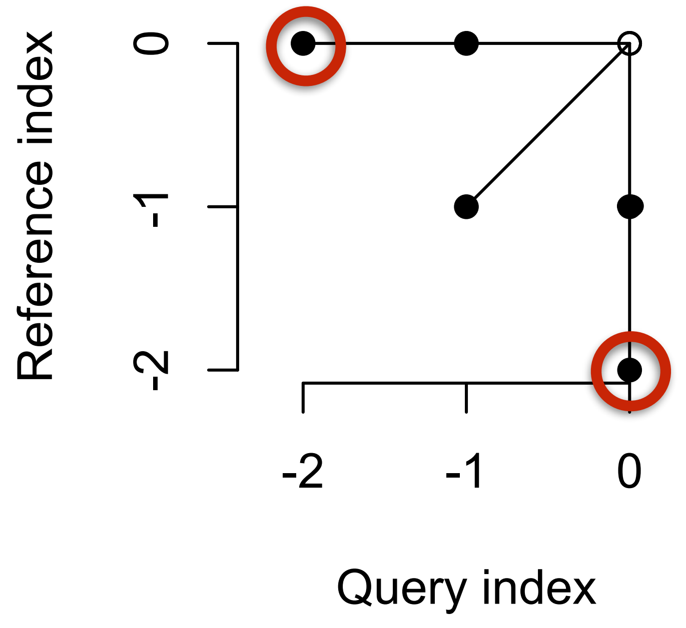
1. rough alignment
2. sample alignment
3. verification by track removal
4. estimation of gain curves
5. estimation of cue regions

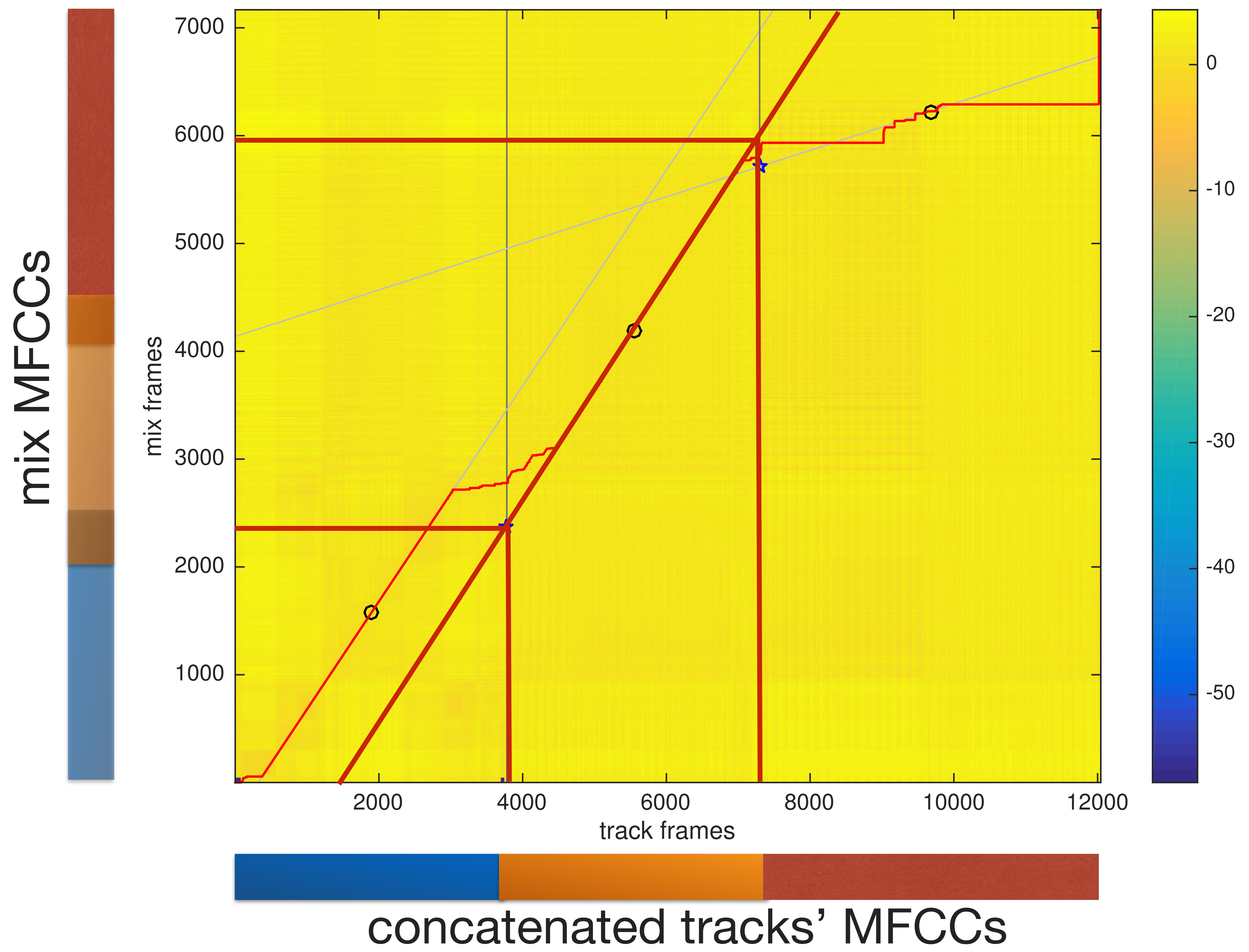
# Step 1: Rough Alignment by DTW

**Dynamic Time Warping alignment** of concatenated MFCCs of tracks with mix

→ relative positioning of the tracks in the mix (intersections)

→ speed factor (slopes of path)



# Step 2: Sample Alignment

Refine alignment to close in to sample precision:

1. **time-scale** source track according to estimated speed factor
2. **search** best sample shift around rough (frame) alignment

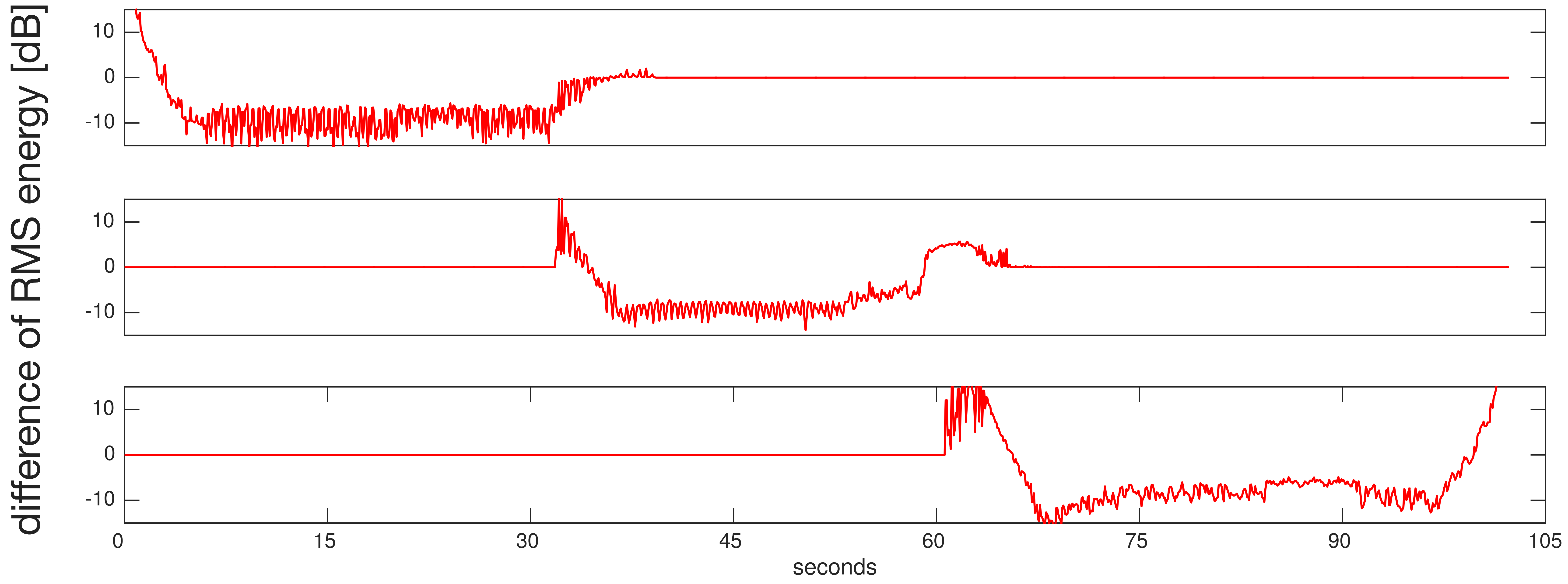
    maximum cross-correlation between mix and track

# Step 3: Verification by Track Removal

Success of sample alignment can be verified by subtracting the aligned and time-scaled track from the mix

→ drop in RMS energy

/!\ Method applicable even when ground truth is unknown or inexact!

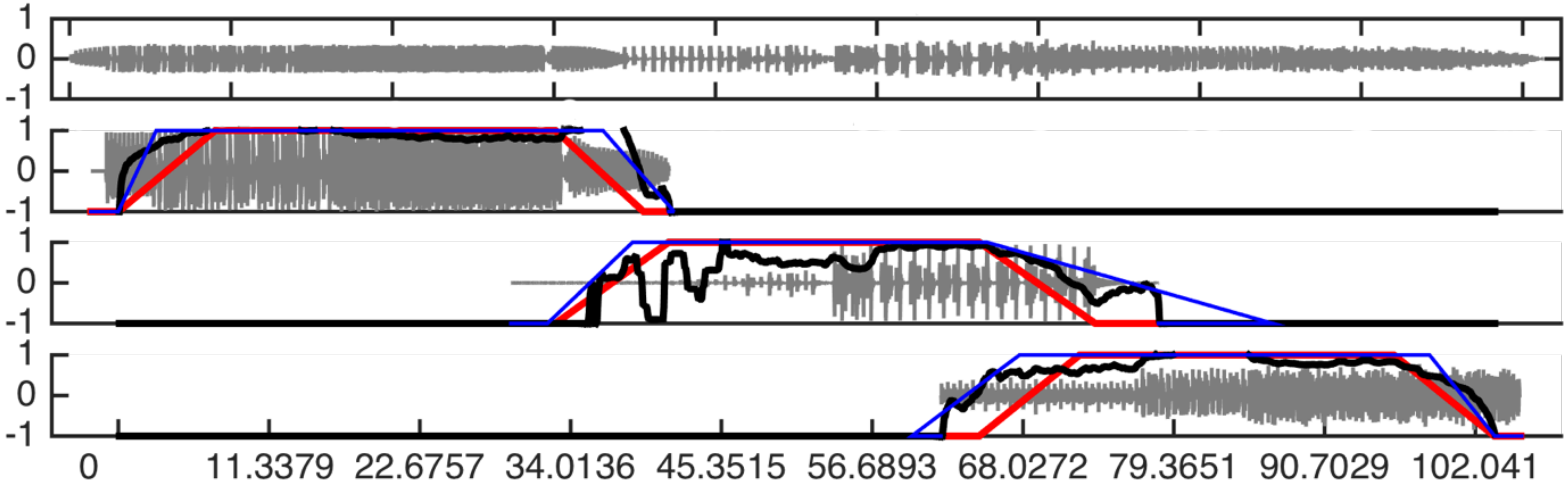


# Step 4: Volume Curve Estimation

Estimate the volume curves $\hat{a}_i$ (**black** lines) applied to each track to obtain the mix

Novel method based on time-frequency representations $X$ (mix) and $S_i$ (track):

$$\hat{a}_i(n) = \begin{cases} \operatorname{median}\left(\frac{|X(n,m')|}{|S_i(n,m')|}\right)_{\forall m' \in \mathbb{M}} & \text{if } \exists m' \text{ s. t. } |S_i(n,m')|^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

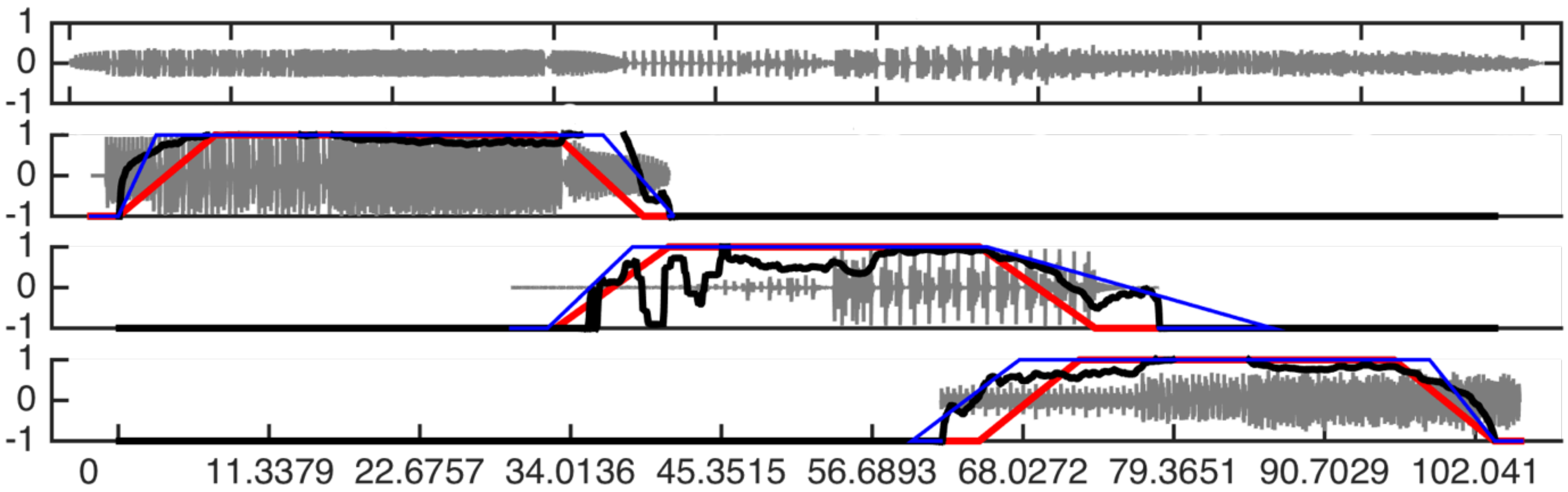


# Step 5: Cue Point Estimation

Cue points are the start and end points of fades

Estimation (**blue** lines) by linear regression of the fade curve *â at* beginning and end (where *â* is between 0 and 70% of its maximum)

Ground truth fade curve in **red**



http://zenodo.org/record/1422385

# The *UnmixDB* Open DJ-Mix Dataset

zenodo

Automatically generated “ecologically valid” beat-synchronous mixes
based on CC-licensed freely available music tracks from net label **http://www.mixotic.net**
curated by Sonnleitner, Arzt & Widmer (2016)

Each mix combines 3 track excerpts of ~40s (start cutting into end on a downbeat)

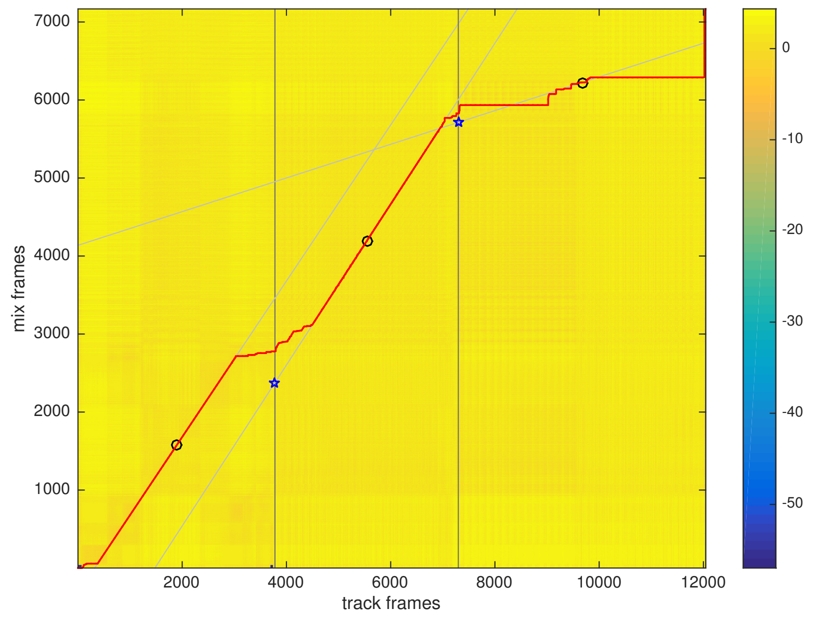
Precise ground truth about the placement of tracks in a mix, fade curves, speed

Mixes generated in 12 variants:
- **4 effects:** no effect, bass boost, dynamics compression, distortion
- **3 time-scaling algorithms:** none, resample, time stretch

6 sets of tracks and mixes, 500 MB – 1 GB, total 4 GB

python source code for mix generation at https://github.com/Ircam-RnD/unmixdb-creation

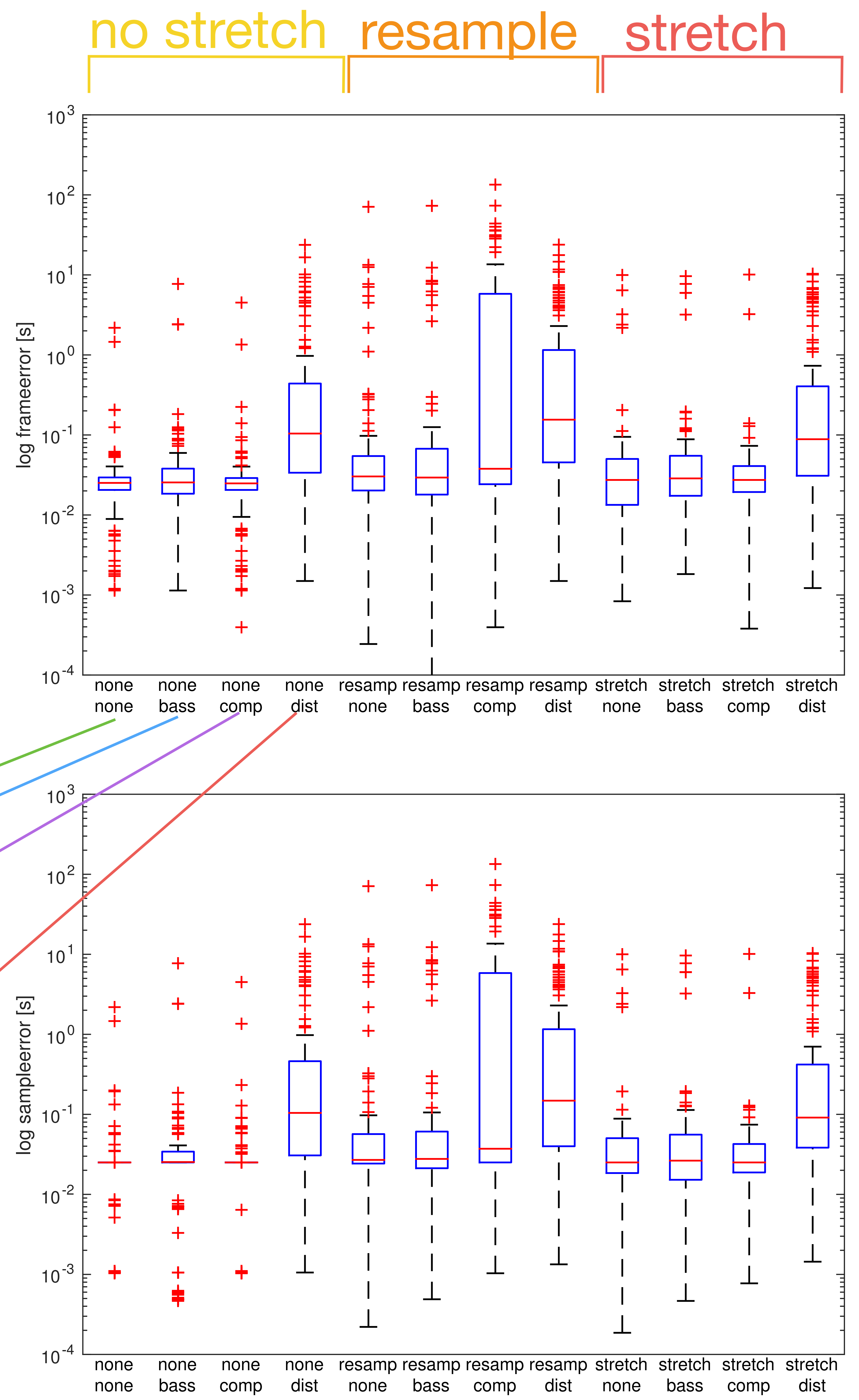
# Evaluation Measures and Results: Alignment



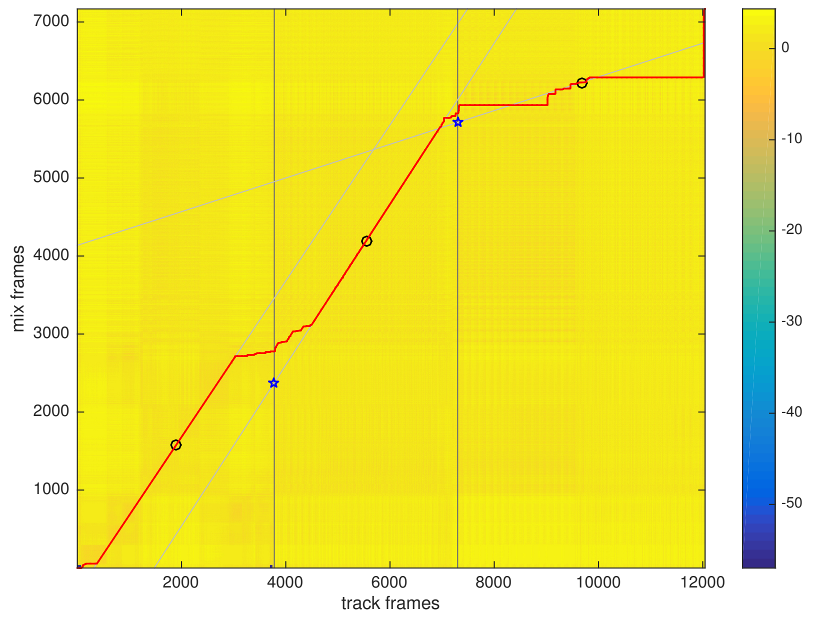

**frame error**: absolute error between ground truth and frame start time from DTW rough alignment (step 1) [s]

**sample error**: absolute error between ground truth and track start time from sample alignment (step 2) [s]

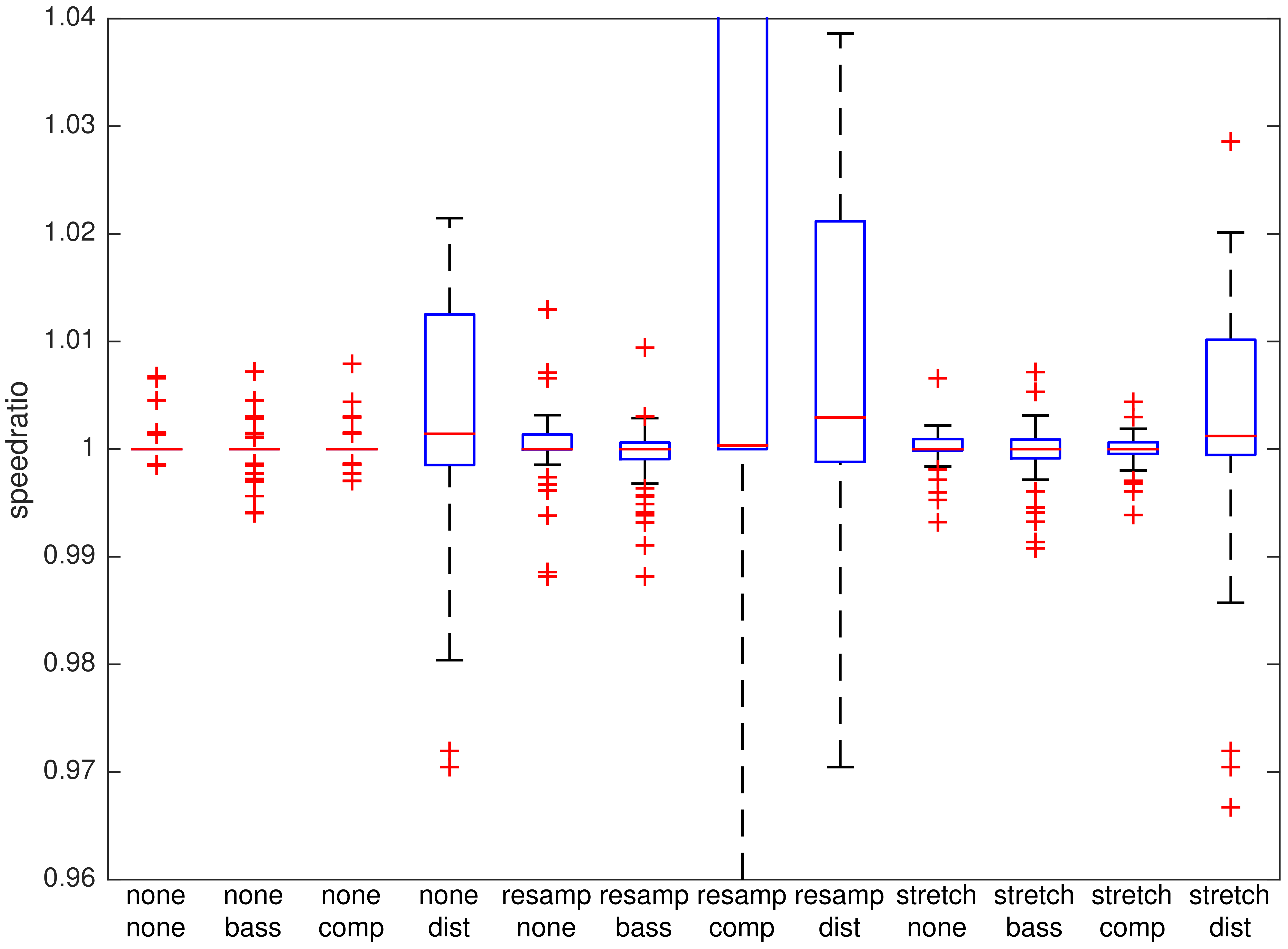
no fx bass compression distortion



# Evaluation Measures and Results: Speed Ratio




**speed ratio**: ratio between ground truth and speed factor estimated by DTW alignment (step 1, ideal value is 1)

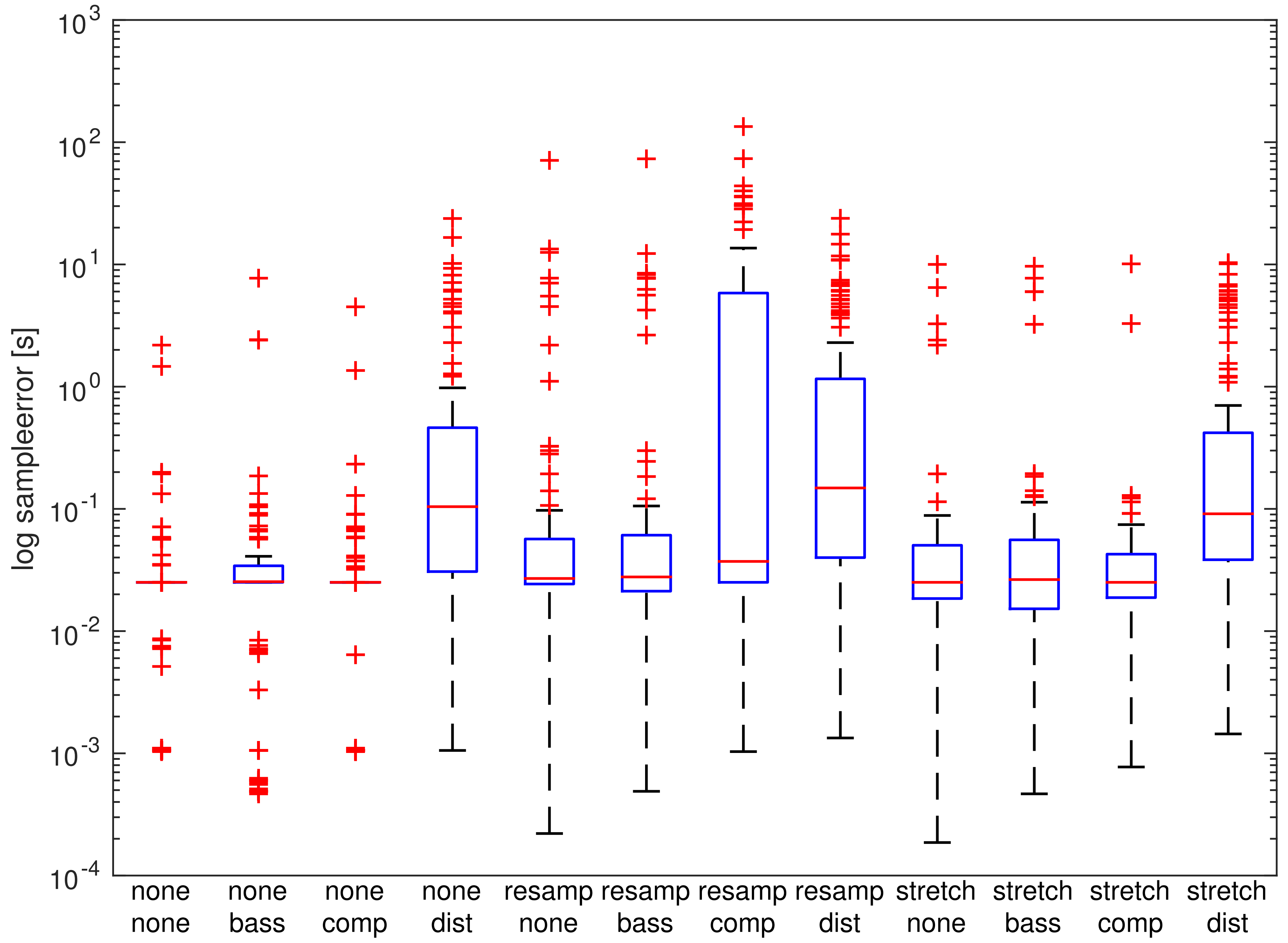


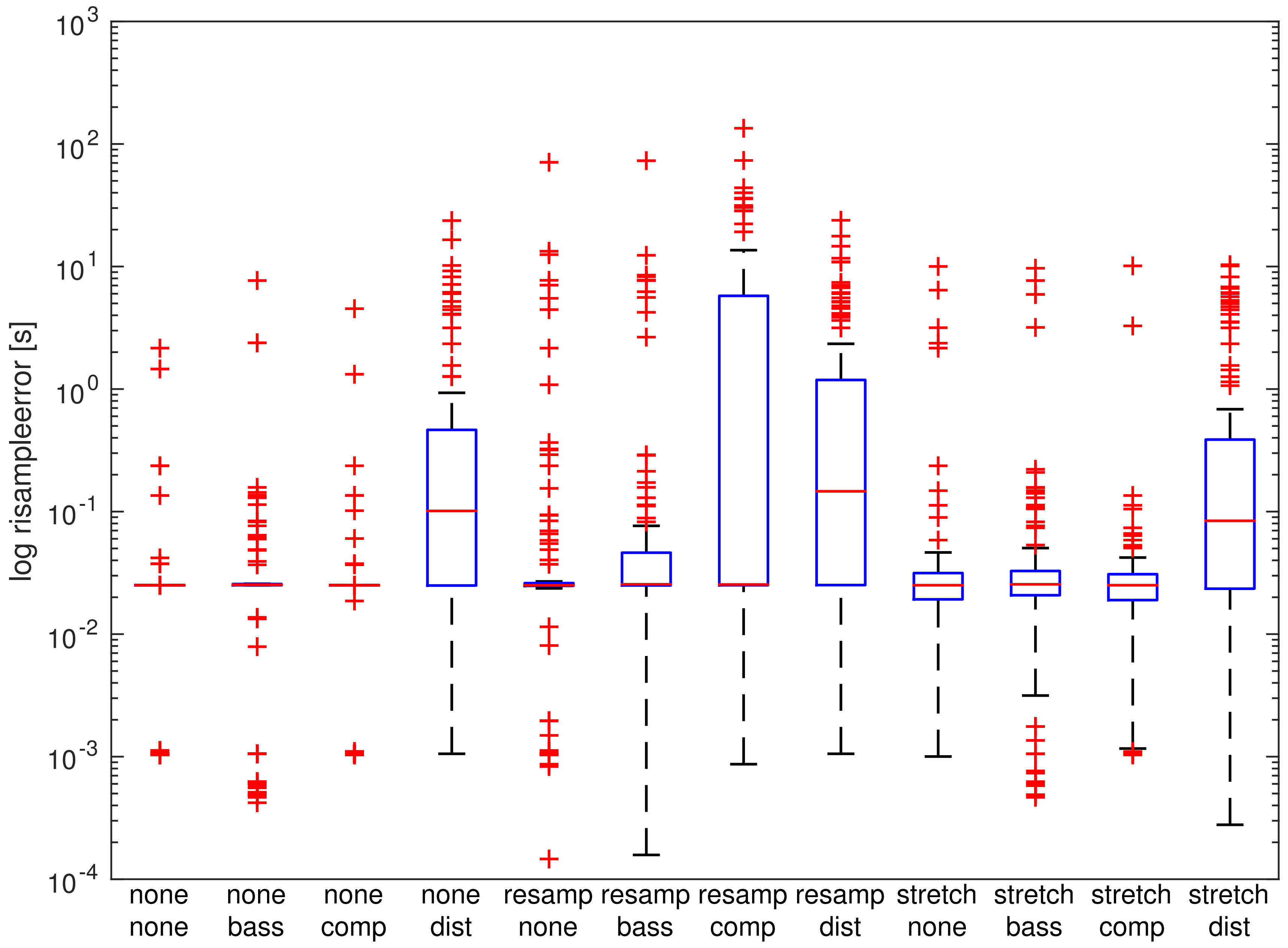
# Reinjecting Ground Truth Speed

High sensitivity of sample alignment and track removal on accuracy of speed estimation from DTW

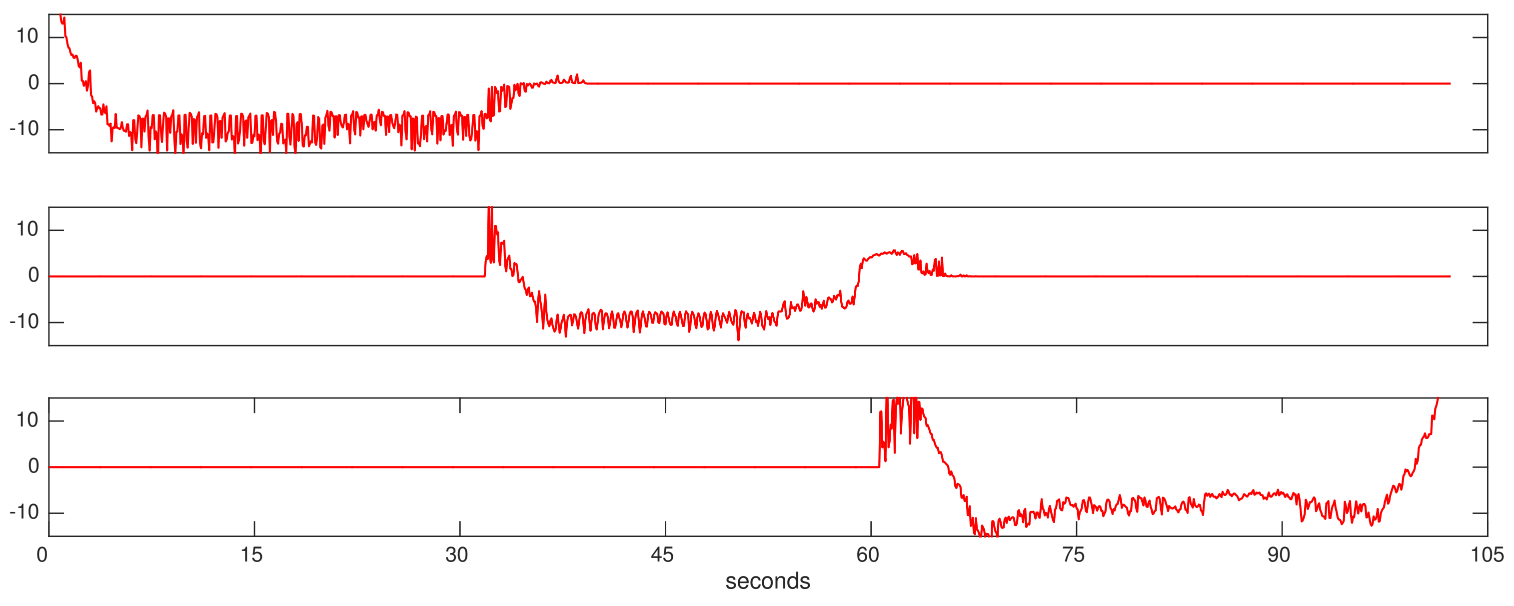
Judge its influence by reinjecting ground truth speed:

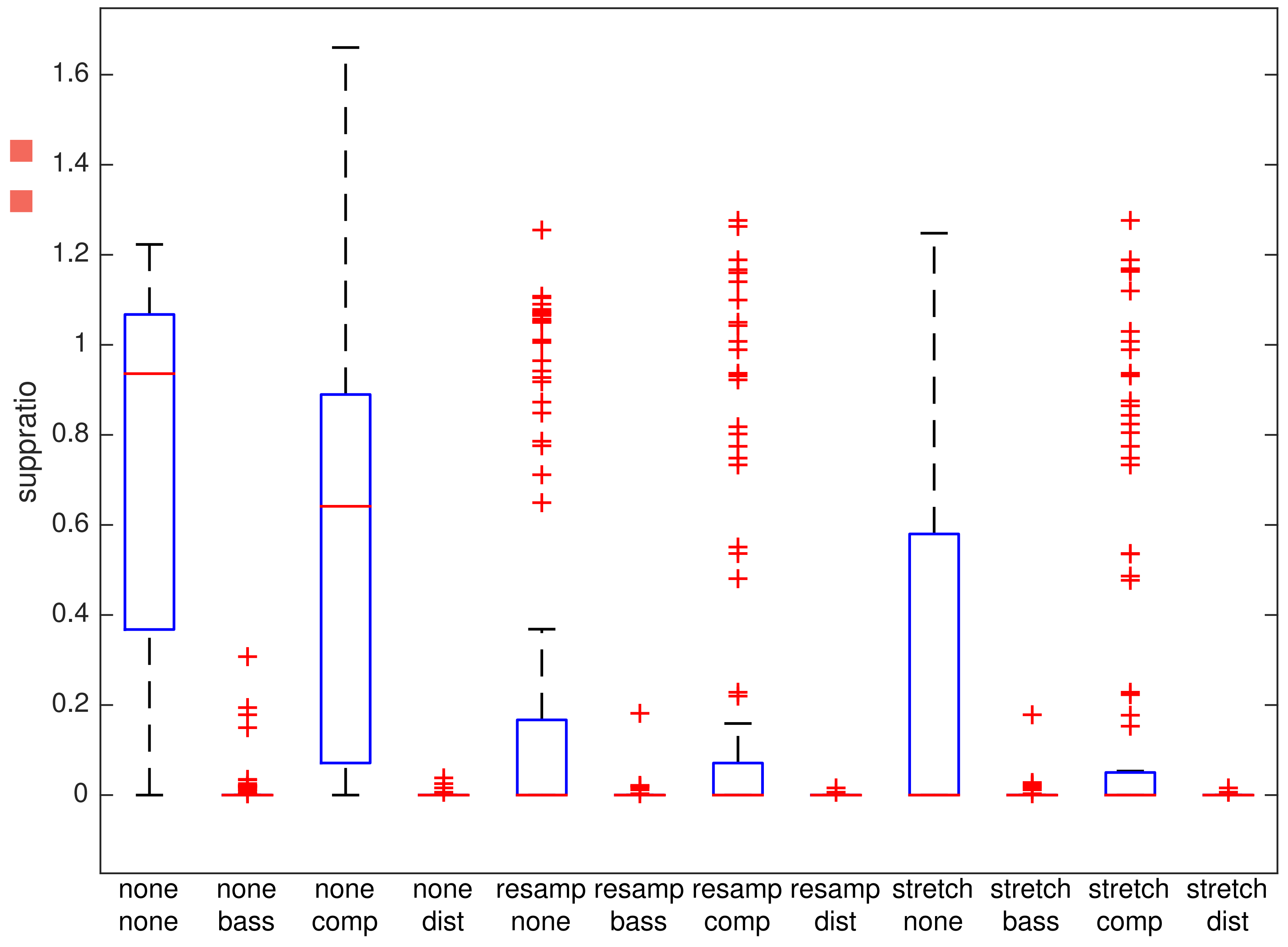
- **top**: sample alignment error
- **bottom**: sample alignment error with ground truth speed reinjected for time-scaling



# Evaluation Measures and Results: Suppression

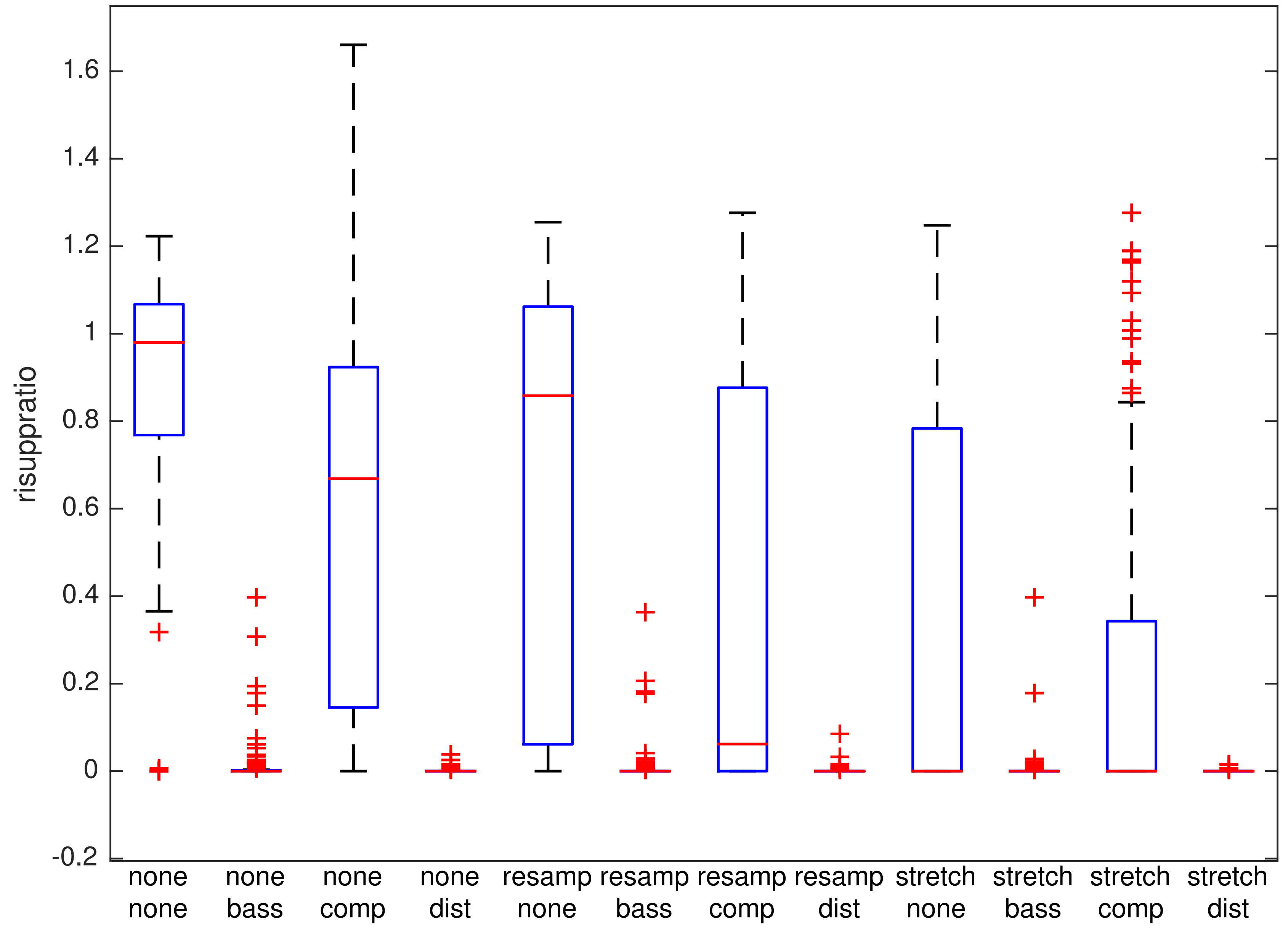


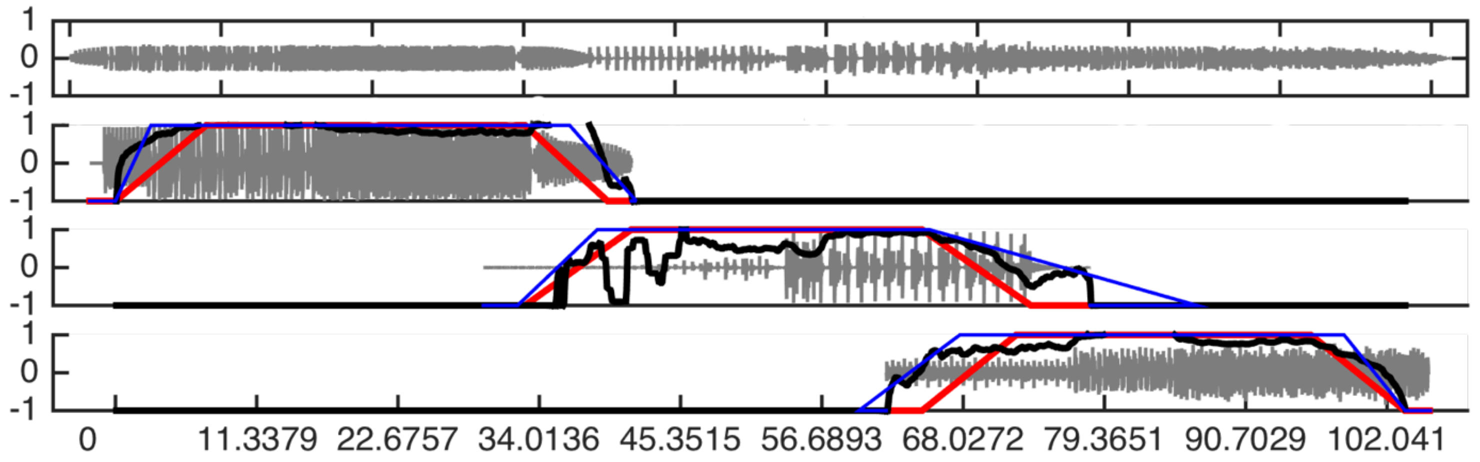

**suppression ratio**: ratio of track time with >15 dB of track removal (step 3, bigger is better)

**top**: with DTW-estimated speed

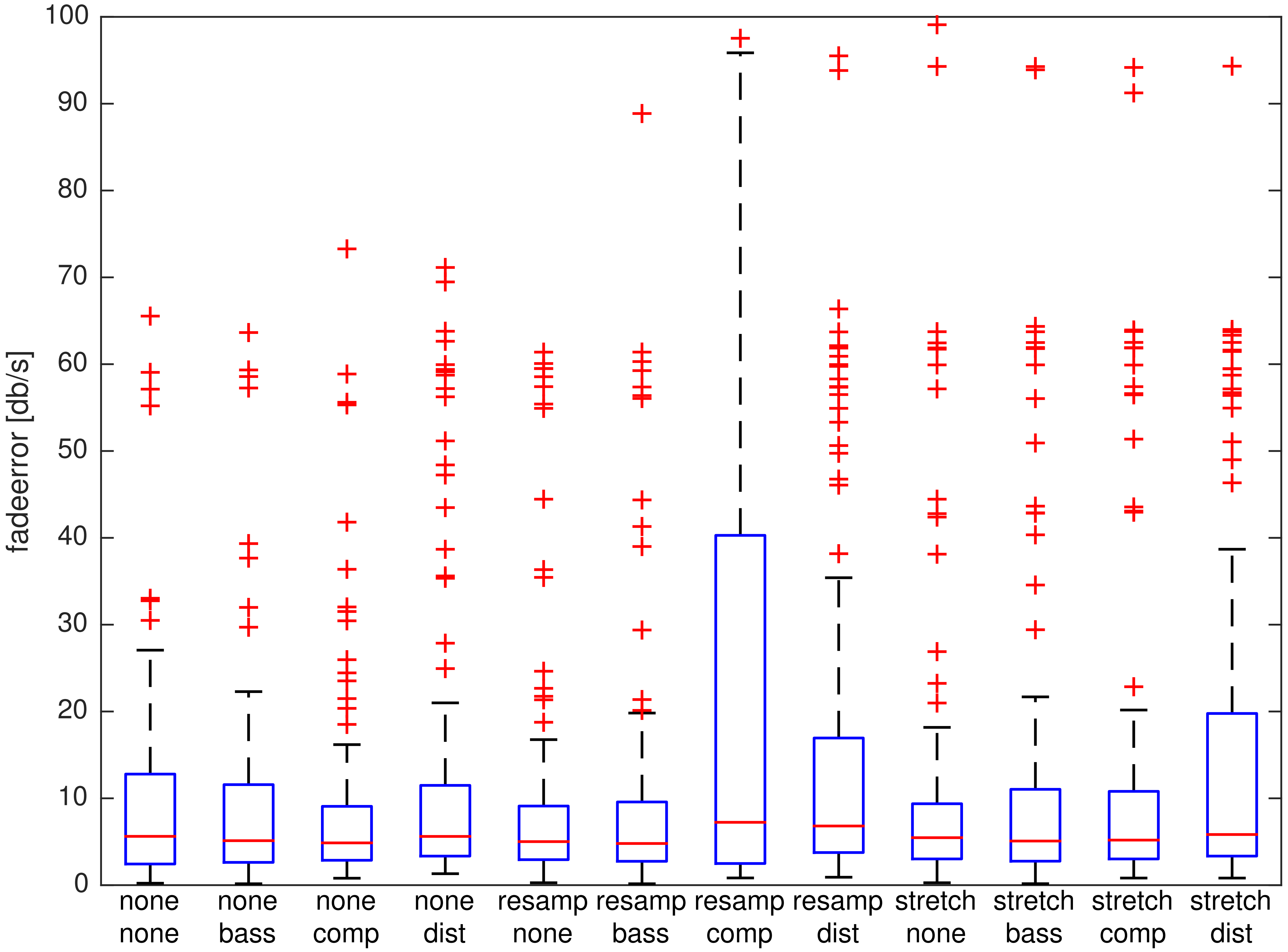
**bottom**: with ground truth speed for time-scaling



# Evaluation Measures and Results: Fade Error



**fade error**: total difference between ground truth and estimated fade curves (steps 4 and 5)



# Conclusion and Future Work

Our DJ-mix reverse engineering method validated on artificial open *UnmixDB* dataset
→ retrieval of rich data from existing real DJ mixes

With some refinements, our method could become robust and precise enough to allow the inversion of EQ and other processing (e.g. compression, echo)

Extend to mixes with non-constant tempo curves, more effects

Close link between alignment, time-scaling, and unmixing hints at a joint and possibly iterative estimation algorithm

Other signal representations (MFCC, spectrum, chroma, scattering transform)?

**beware: IANADJ!**