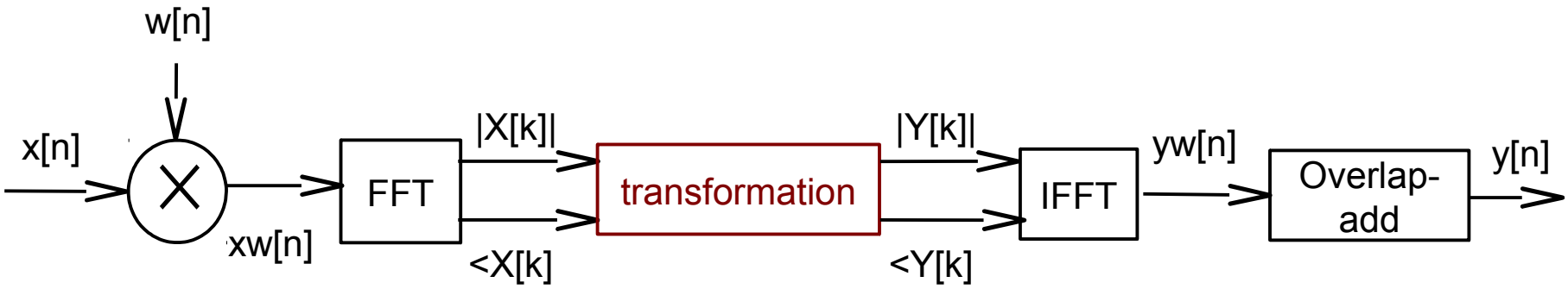
8T1: Spectral-based sound transformations (1 of 2)

Xavier Serra
Universitat Pompeu Fabra, Barcelona
&
Stanford University

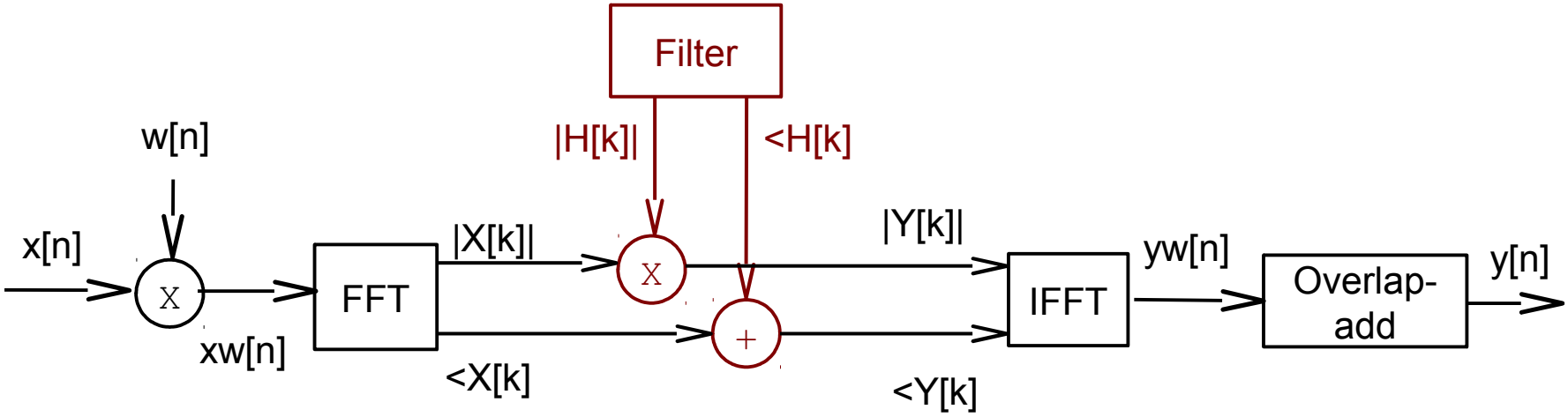
Index

- Short-time Fourier transform
  - Filtering
  - Morphing
- Sinusoidal model
  - Frequency scaling
  - Time scaling

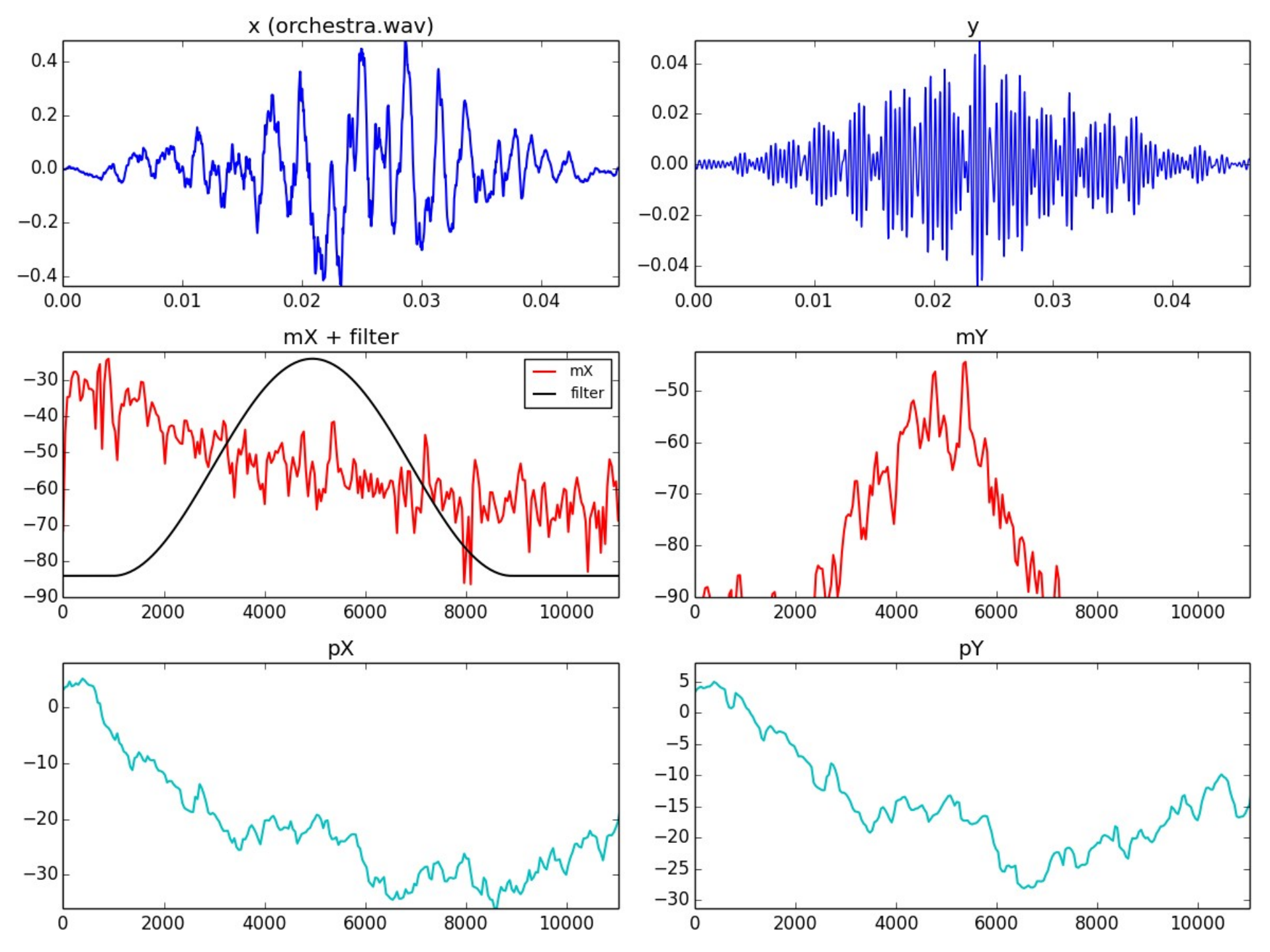
Short-time Fourier transform

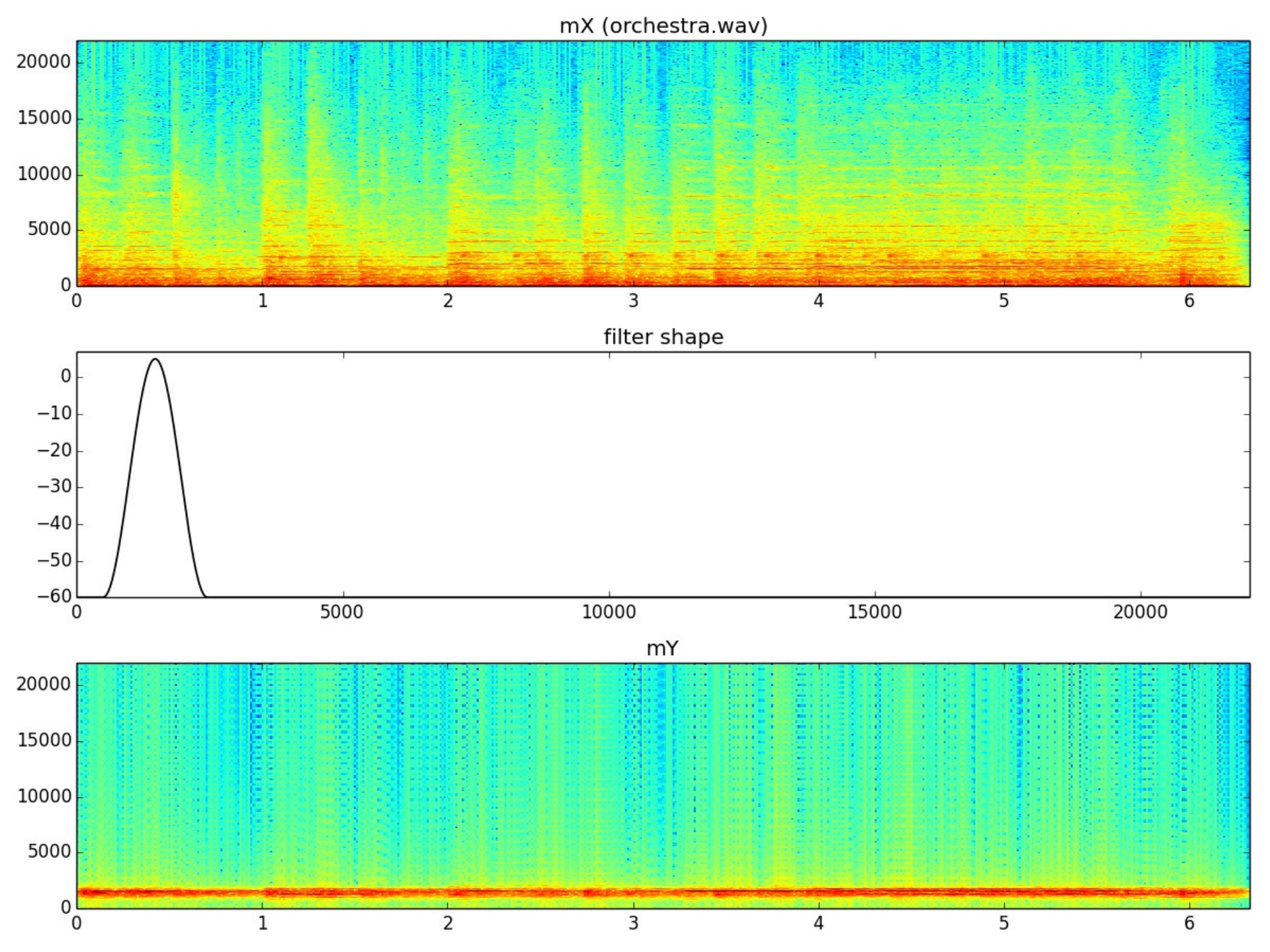


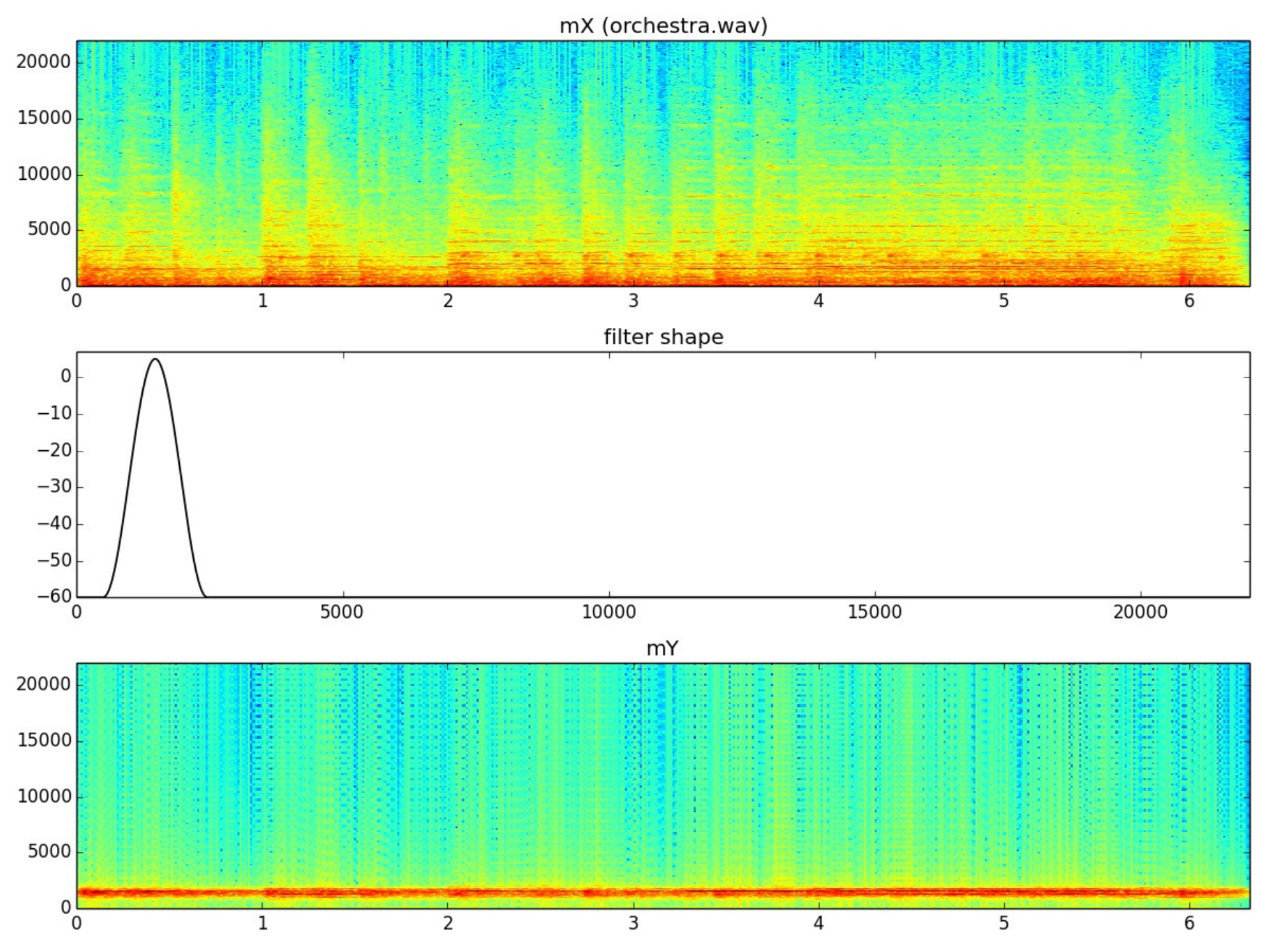
Filtering with STFT

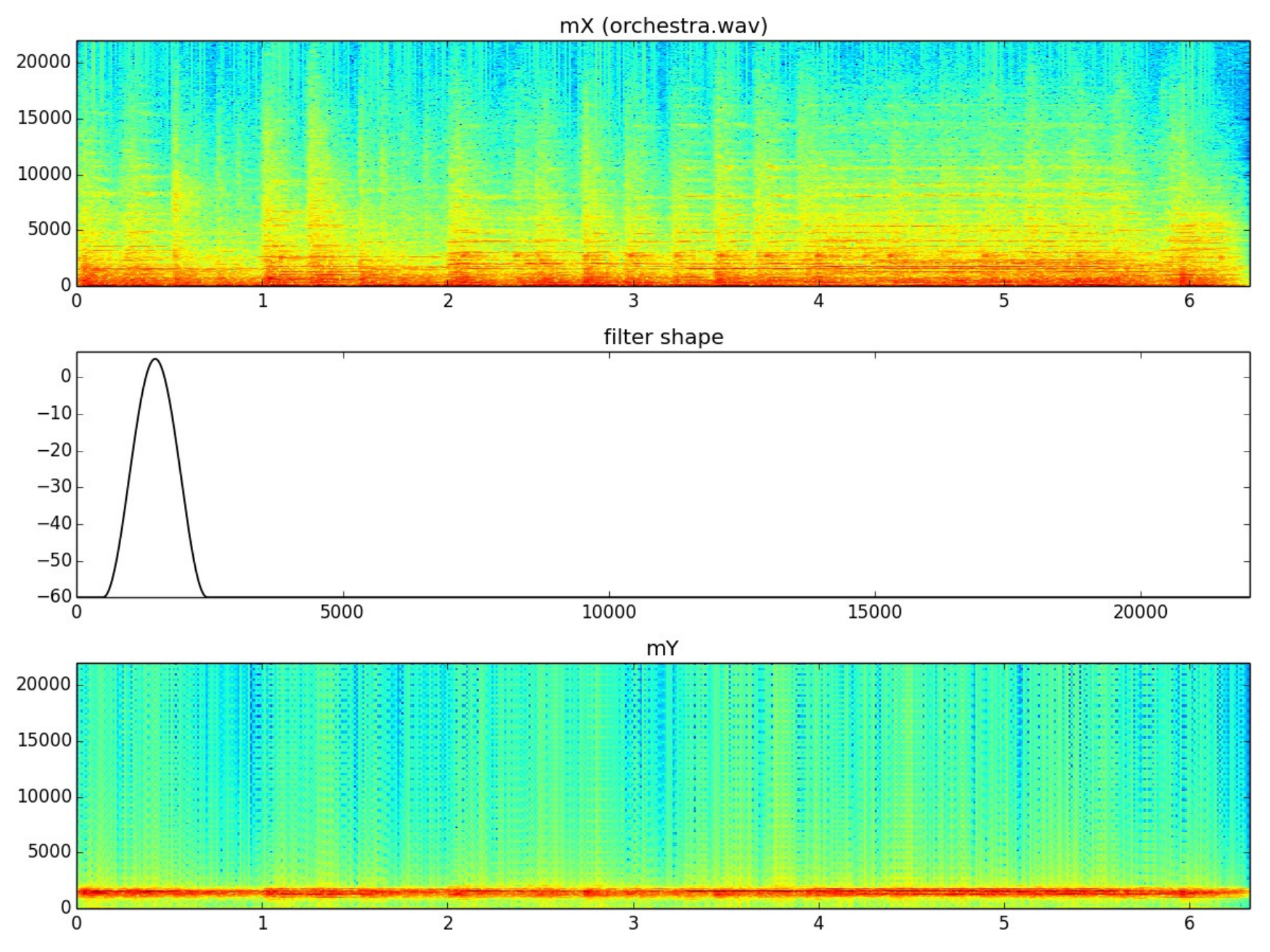


$$Y_l[k] = |H[k]||X_l[k]| e^{j(\measuredangle H[k] + \measuredangle X_l[k])}$$

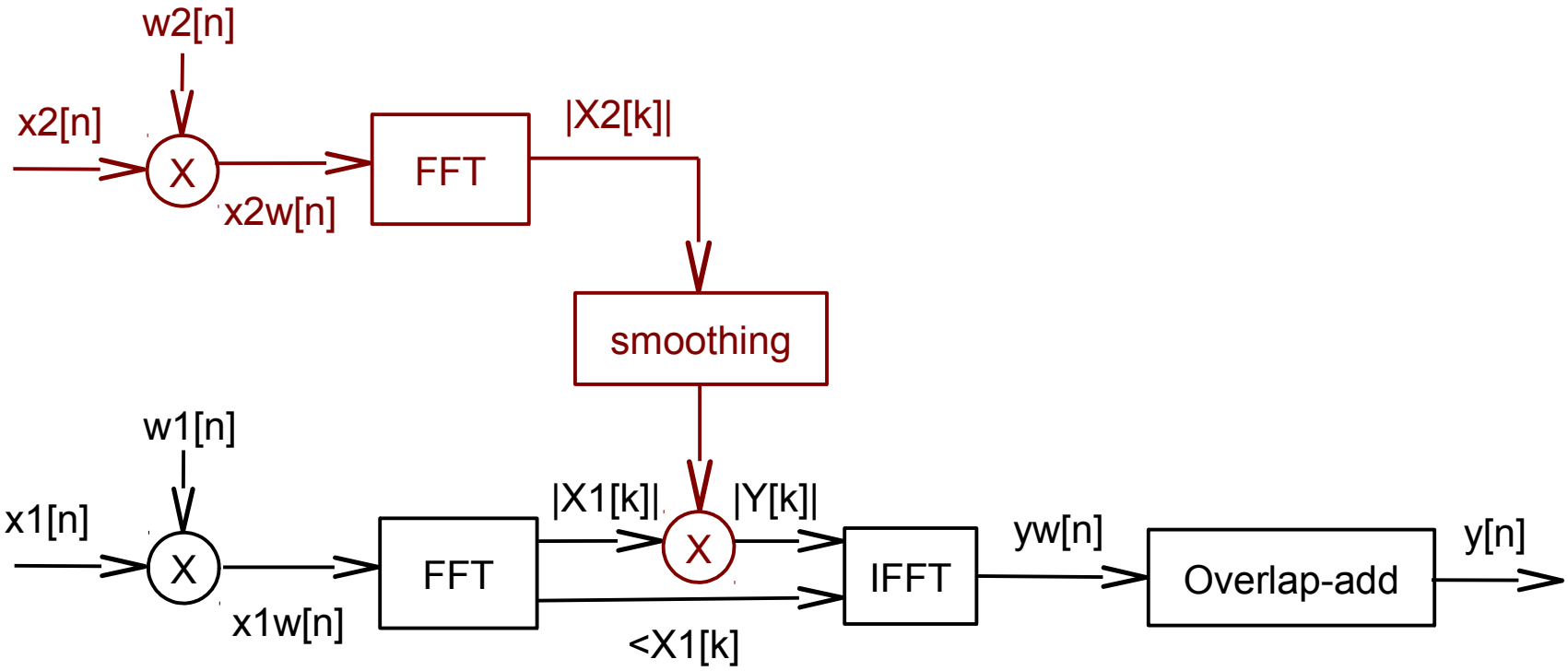

x (orchestra.wav)
y
mX + filter
mX
filter
mY
pX
pY


mX (orchestra.wav)
20000
15000
10000
5000
0
0
1
2
3
4
5
6
filter shape
0
−10
−20
−30
−40
−50
−60
0
5000
10000
15000
20000
mY
20000
15000
10000
5000
0
0
1
2
3
4
5
6

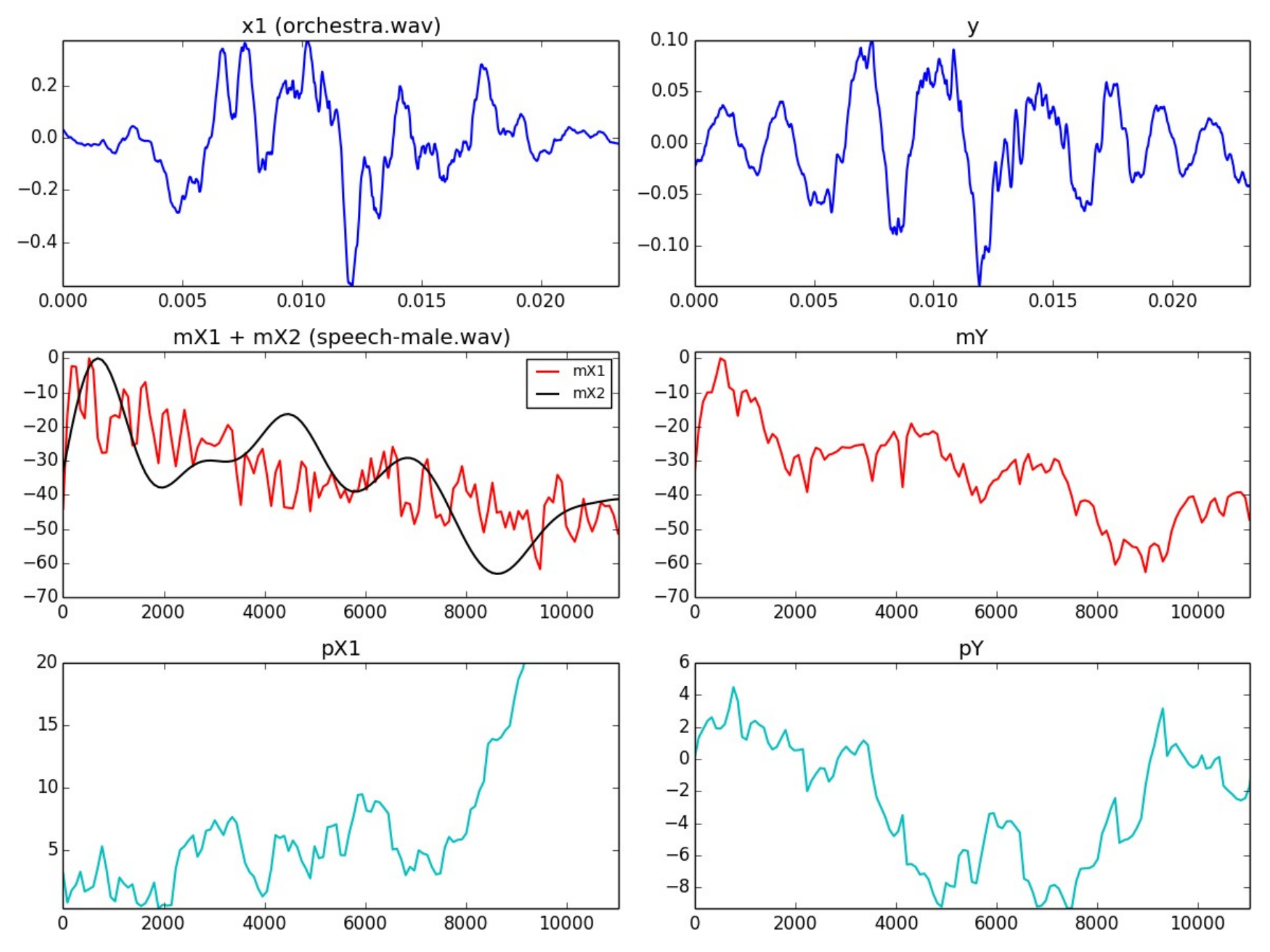

mX (orchestra.wav)
20000
15000
10000
5000
0
0
1
2
3
4
5
6
filter shape
0
−10
−20
−30
−40
−50
−60
0
5000
10000
15000
20000
mY
20000
15000
10000
5000
0
0
1
2
3
4
5
6

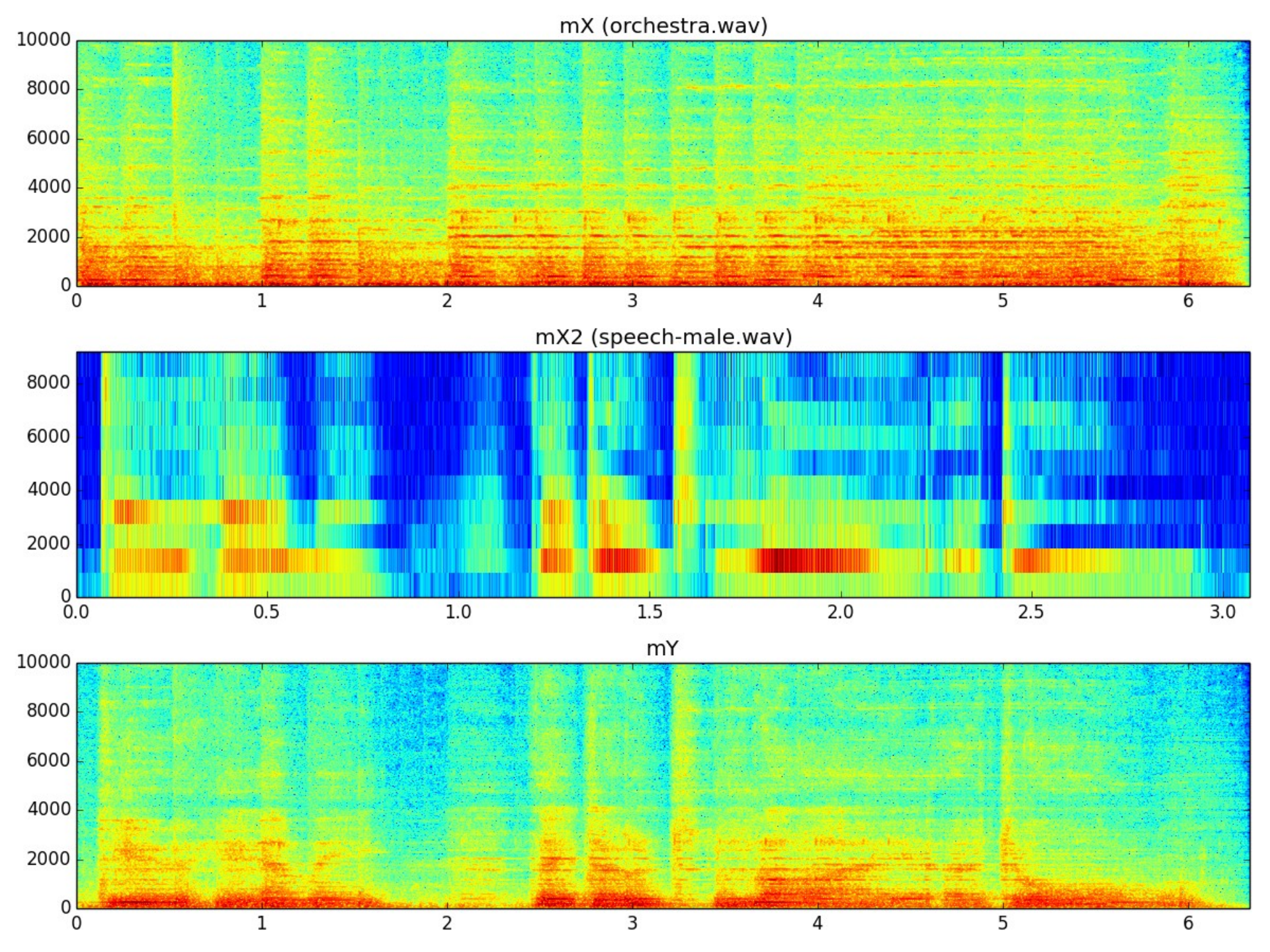

mX (orchestra.wav)
20000
15000
10000
5000
0
0
1
2
3
4
5
6
filter shape
0
−10
−20
−30
−40
−50
−60
0
5000
10000
15000
20000
mY
20000
15000
10000
5000
0
0
1
2
3
4
5
6

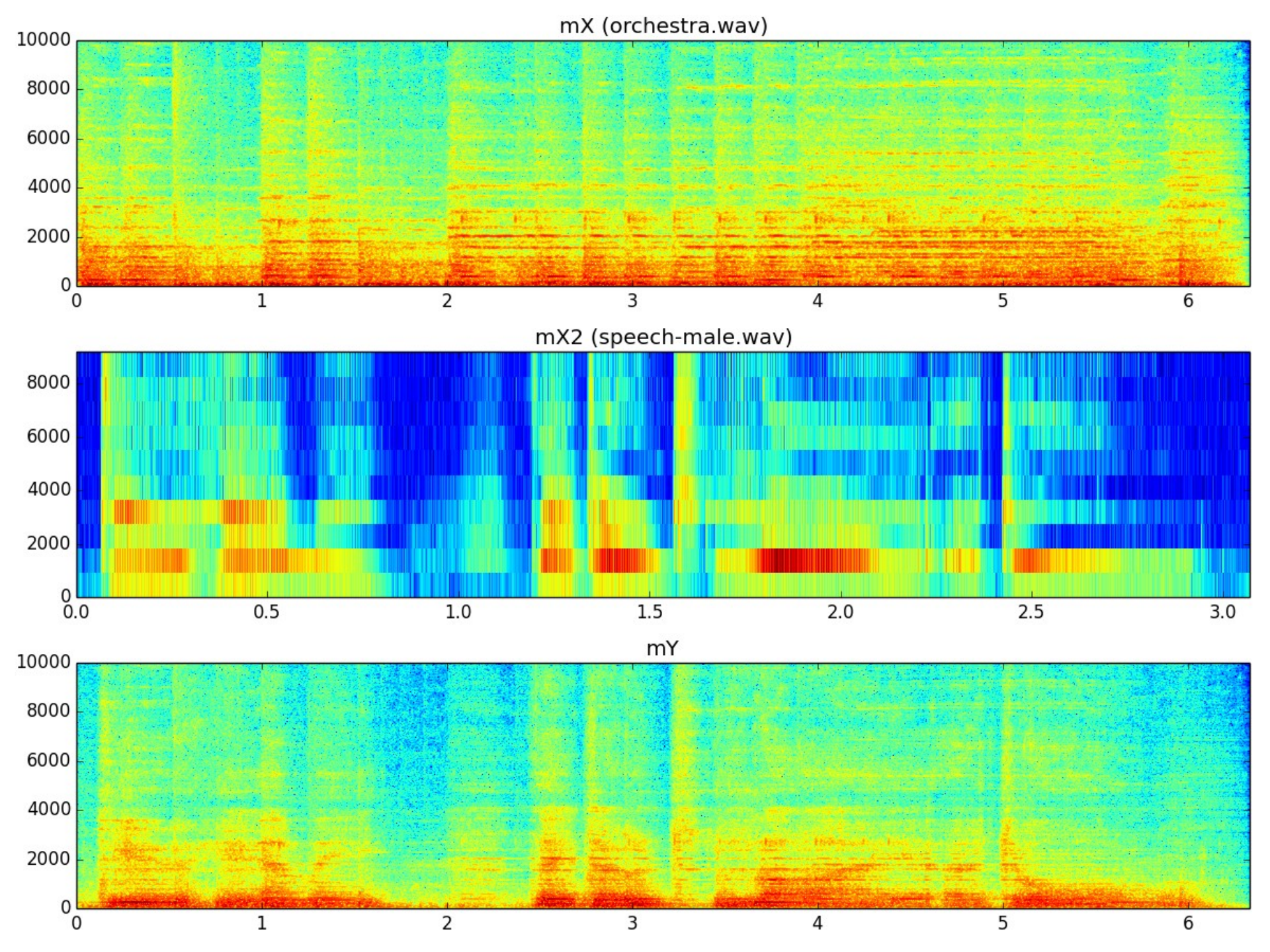
Morphing with STFT

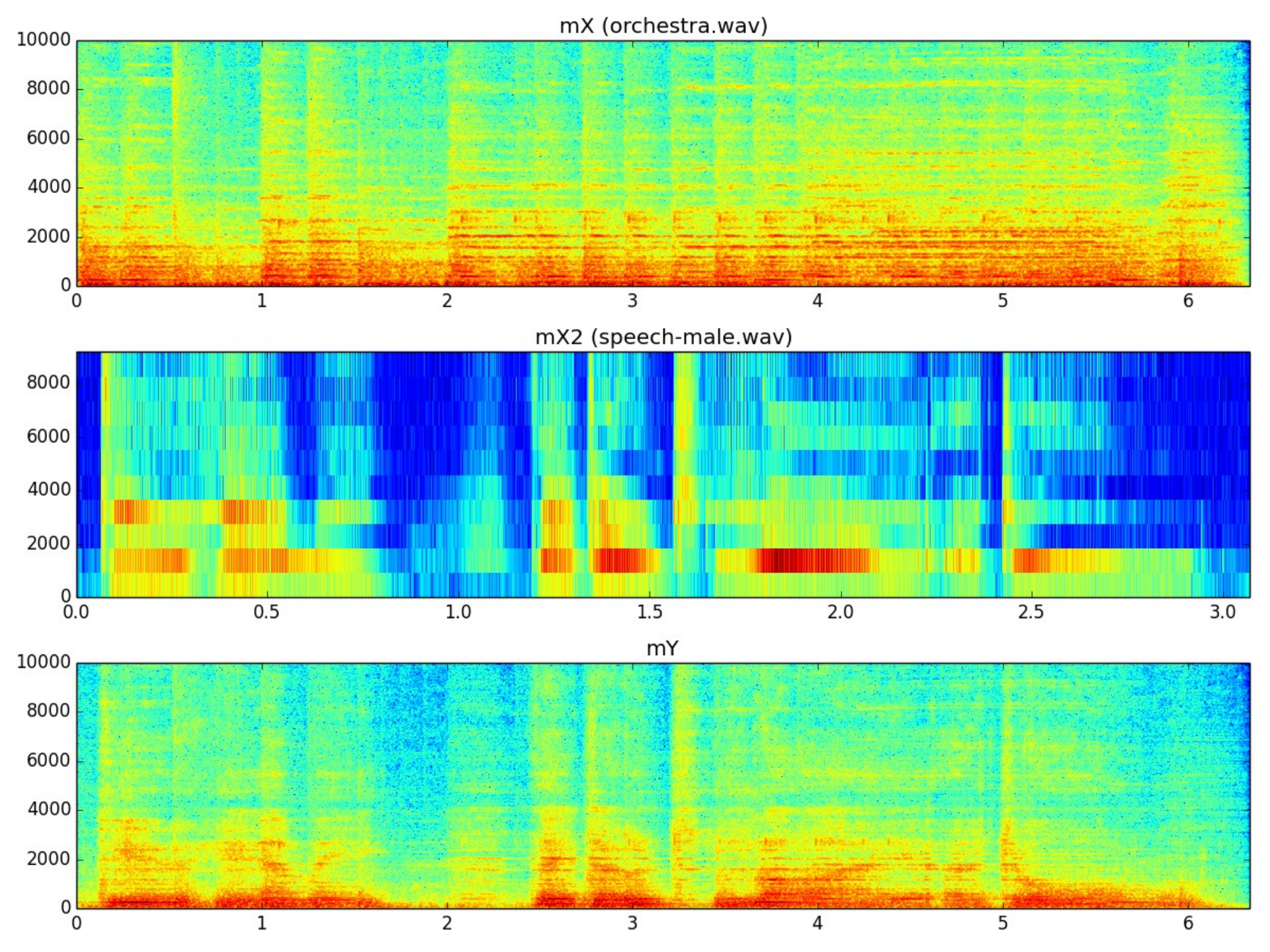


$$Y_l[k] = |X2_l[k]||X1_l[k]|e^{j \measuredangle X1_l[k]}$$

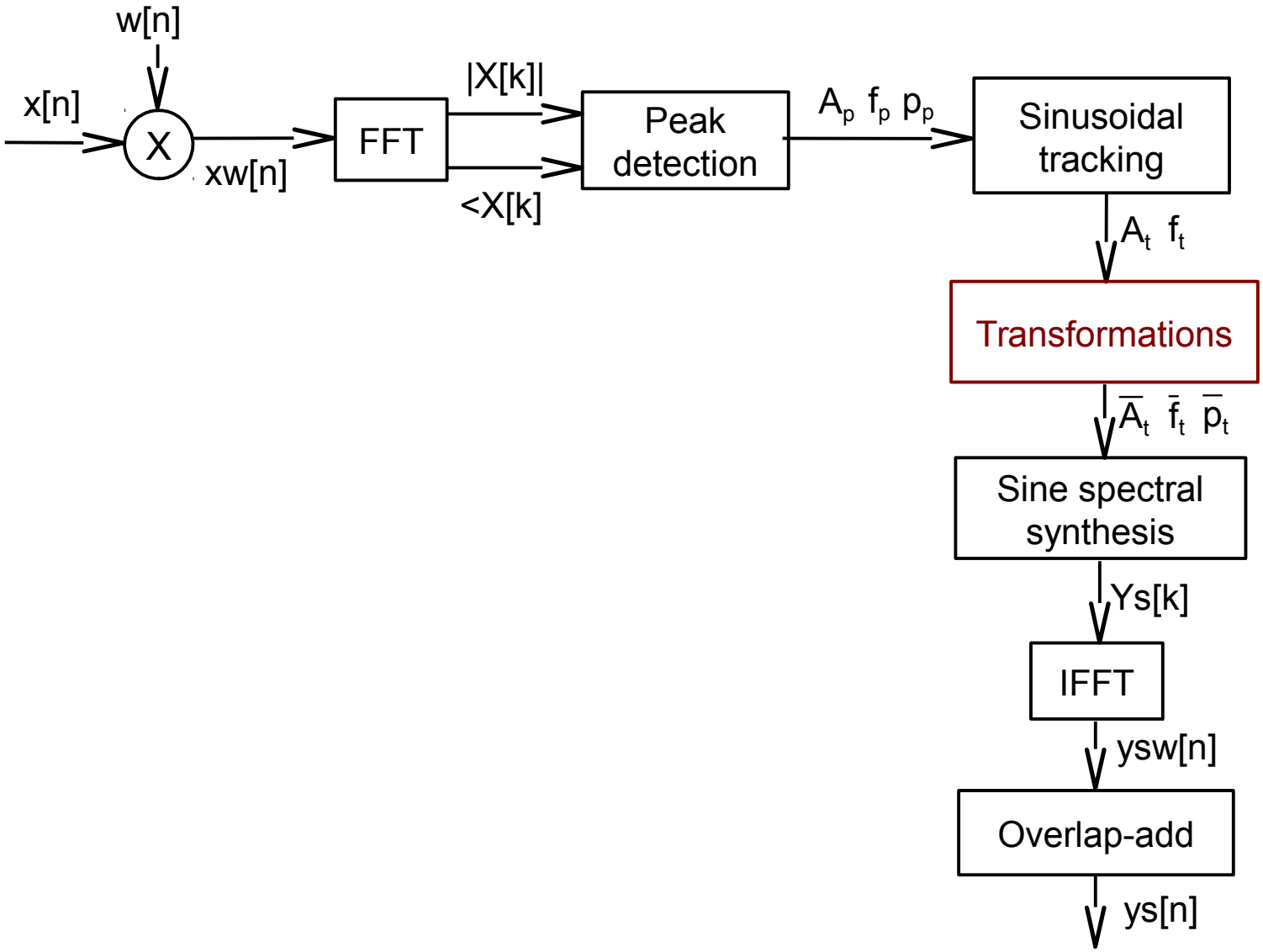

x1 (orchestra.wav)
y
mX1 + mX2 (speech-male.wav)
mX1
mX2
mY
pX1
pY


mX (orchestra.wav)
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
mX2 (speech-male.wav)
8000
6000
4000
2000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
mY
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6


mX (orchestra.wav)
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
mX2 (speech-male.wav)
8000
6000
4000
2000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
mY
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6


mX (orchestra.wav)
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
mX2 (speech-male.wav)
8000
6000
4000
2000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
mY
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6

Sinusoidal model

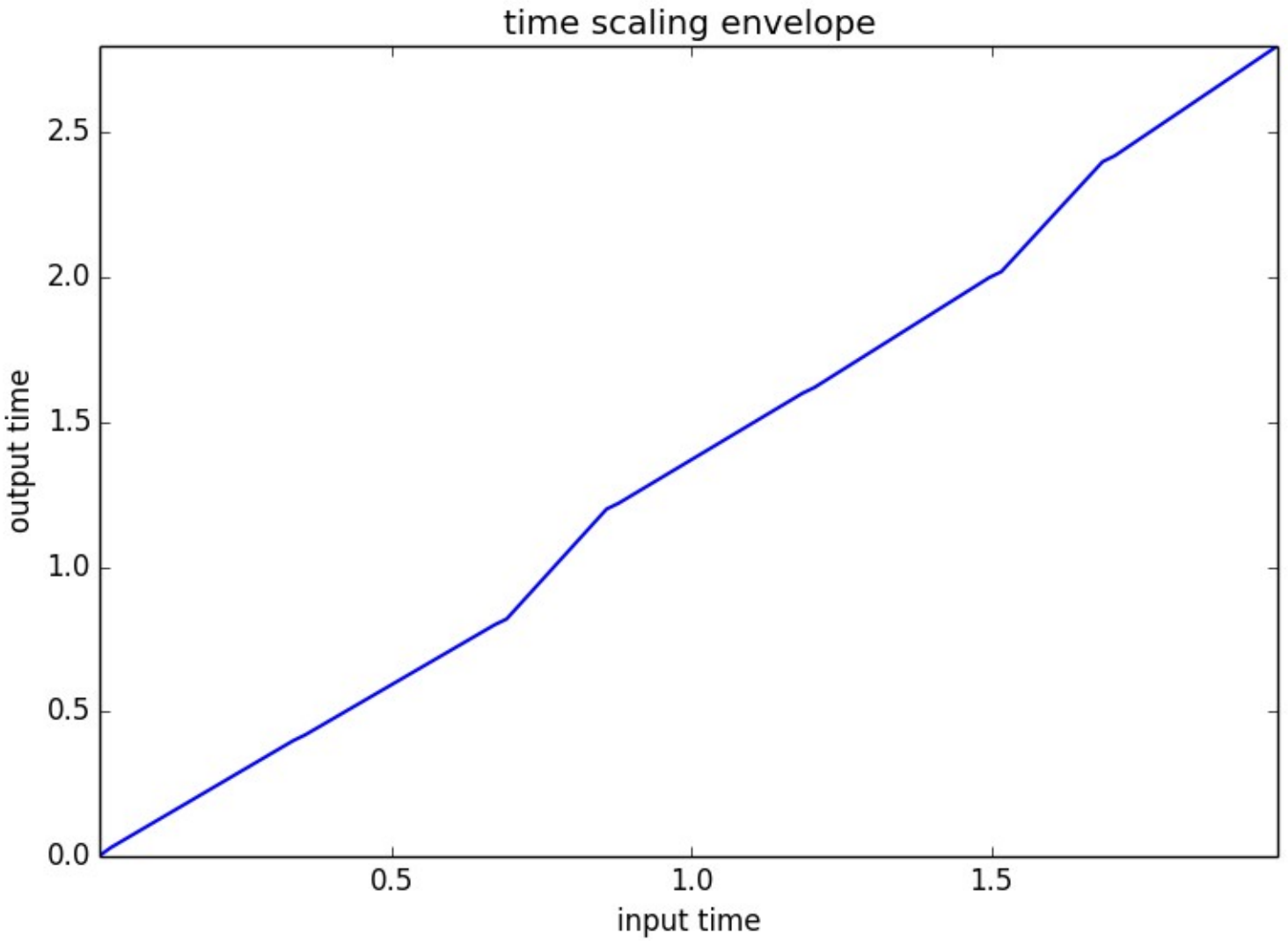


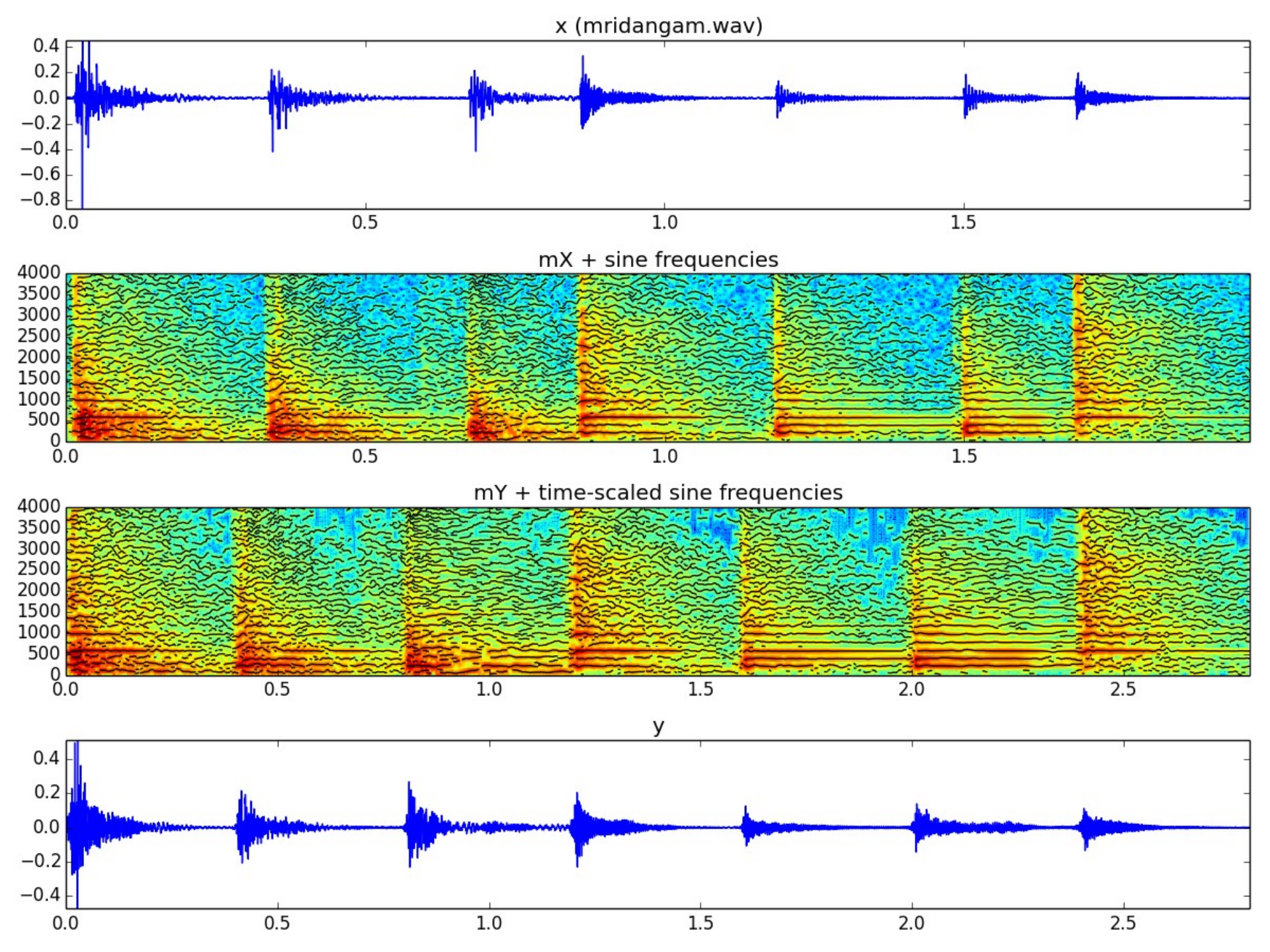
Scaling frequency, amplitude and time

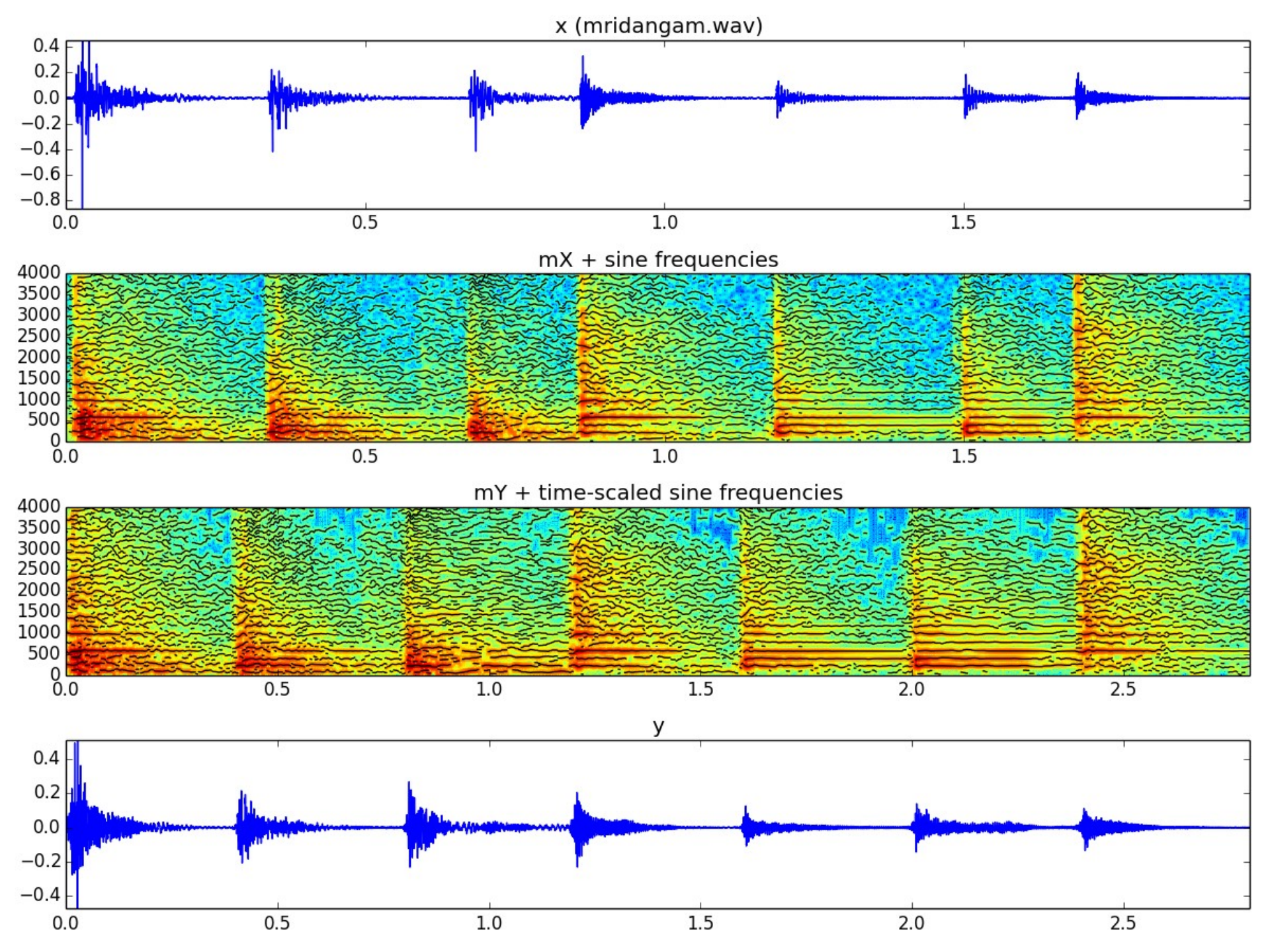
$$\bar{f}_t[q] = sf_t[l] f_t[st_t[l] l]$$
$$\bar{A}_t[q] = sA_t[l] + A_t[st_t[l] l]$$
$$\bar{\varphi}_t[q] = \varphi_t[q-1] + f_t[q]$$

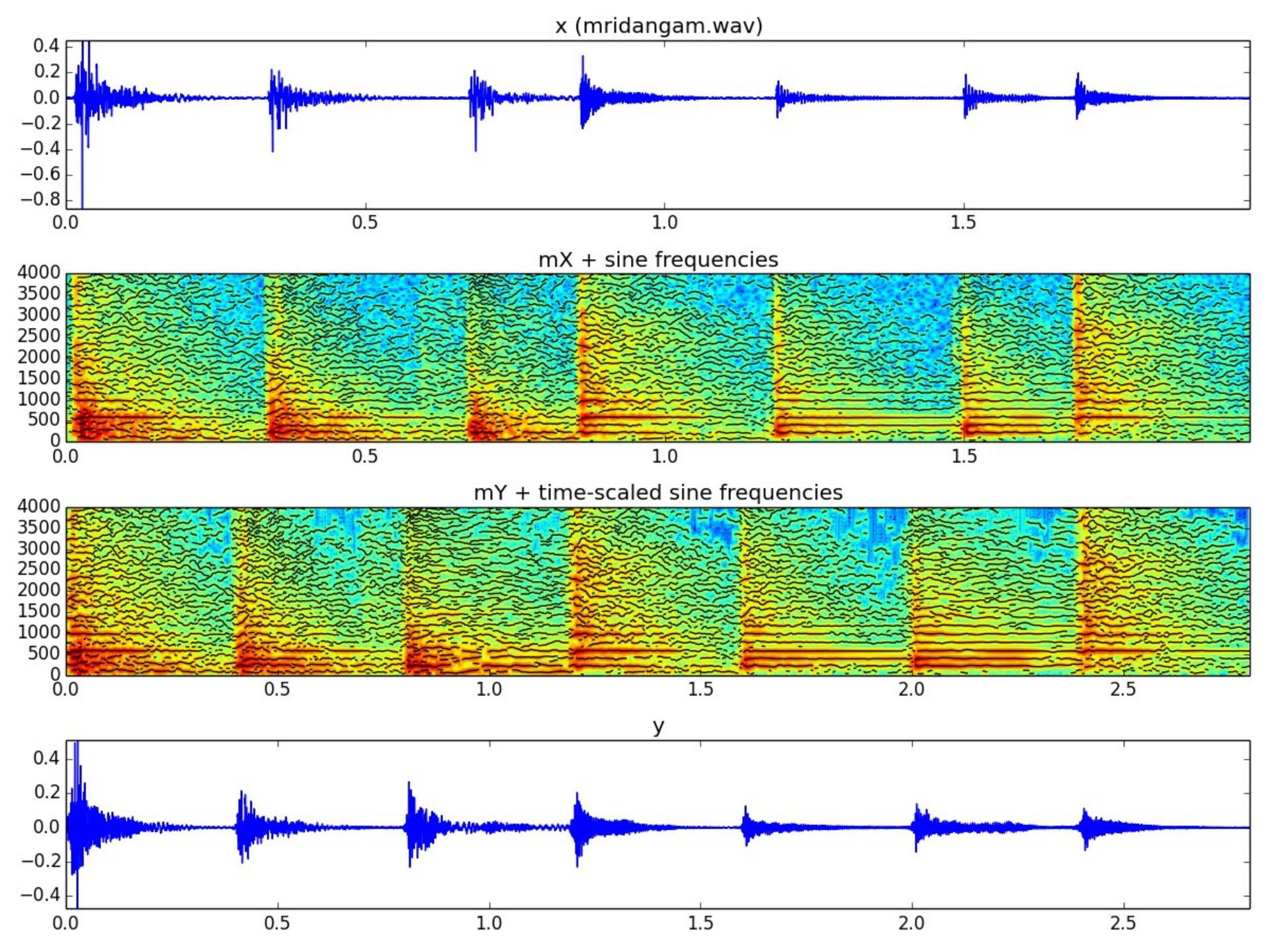
q : output frame index ; l : input frame index ; t : sinusoidal track index
f : input frequency in Hz ; A : input amplitude in dB
sf : scaling frequency ; sA : scaling amplitude ; st : scaling time
$\bar{f}$: output frequency ; $\bar{A}$: output amplitude ; $\bar{\varphi}$: output phase

Time scaling

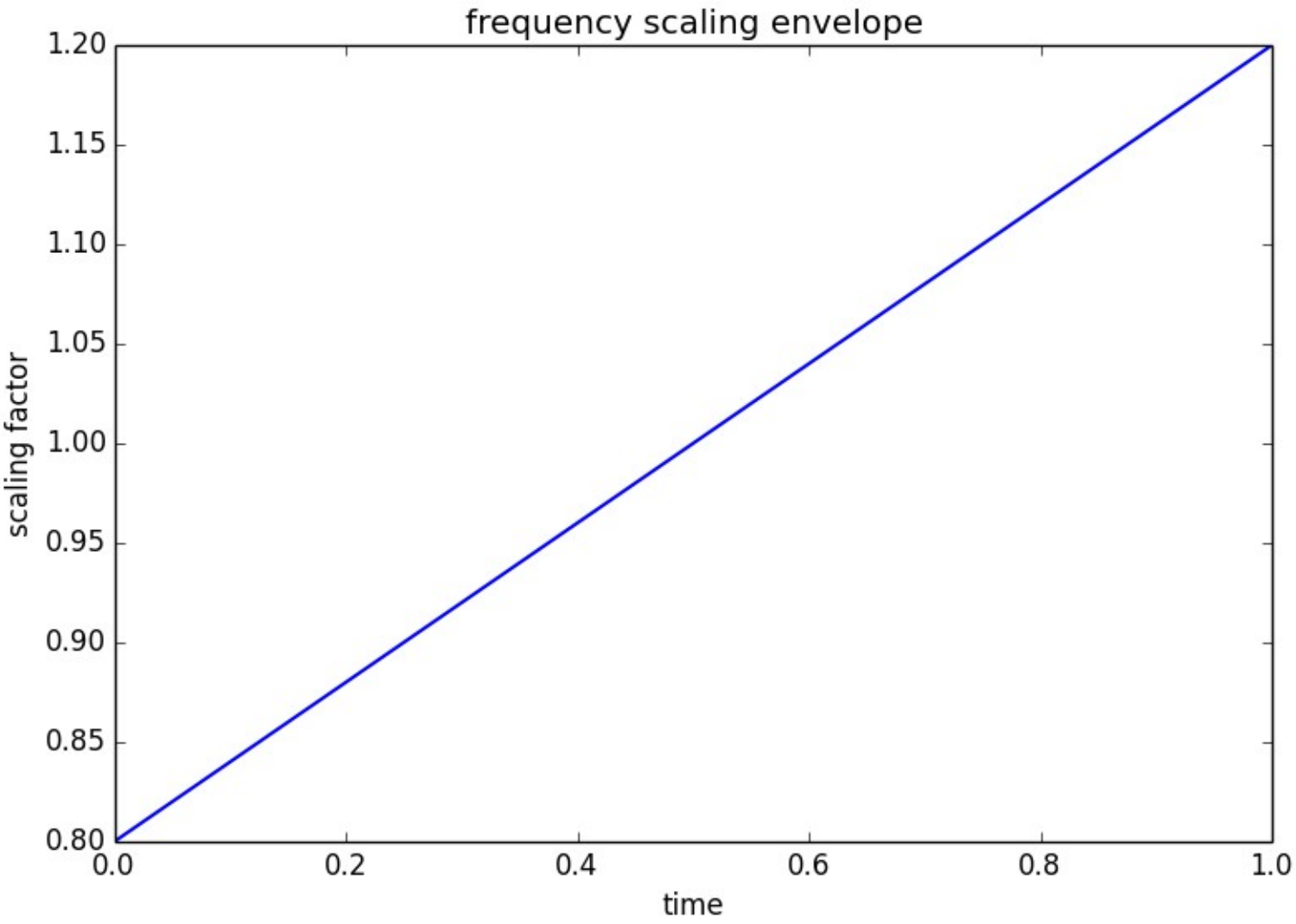


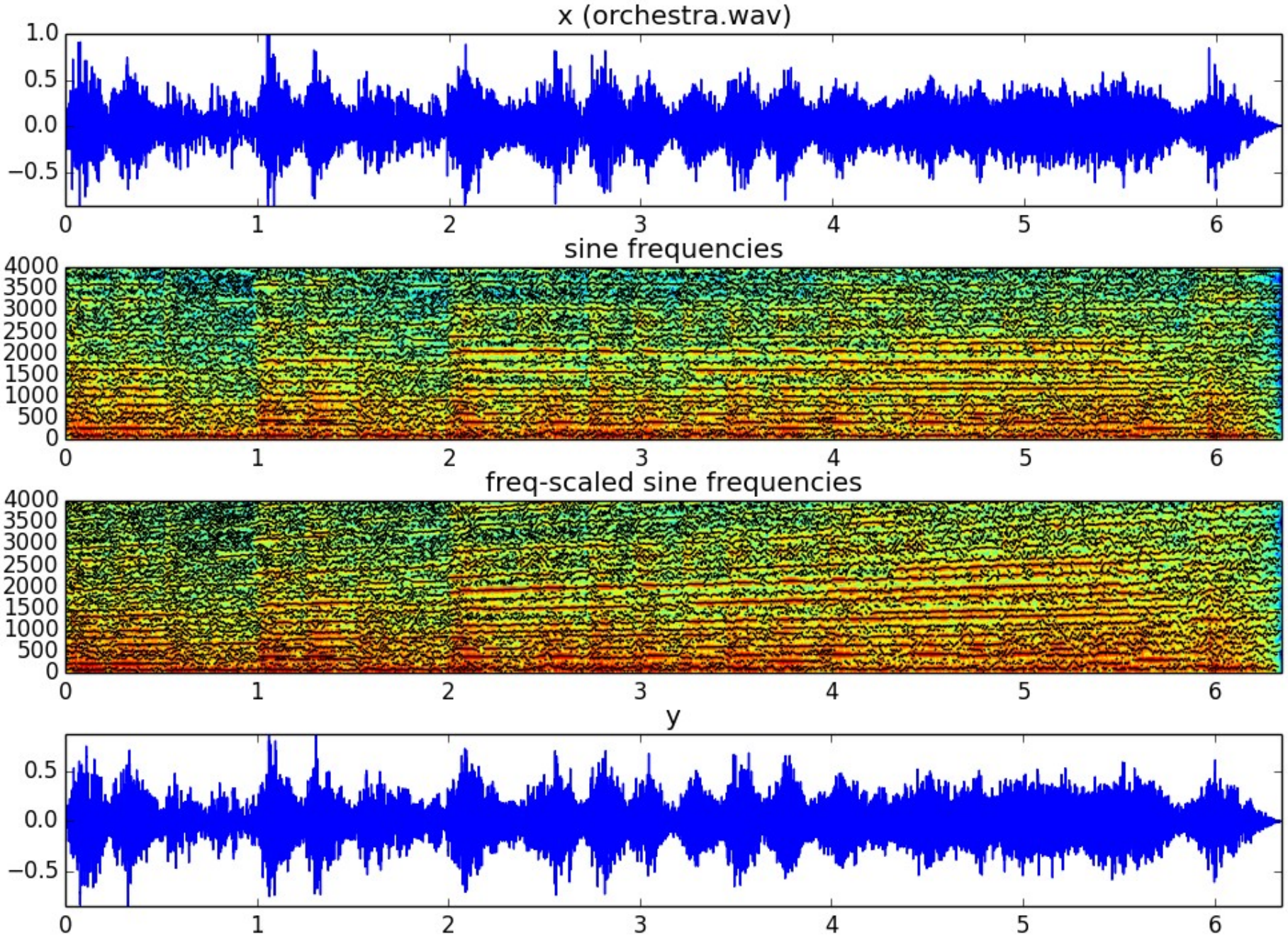

x (mridangam.wav)
mX + sine frequencies
mY + time-scaled sine frequencies
y

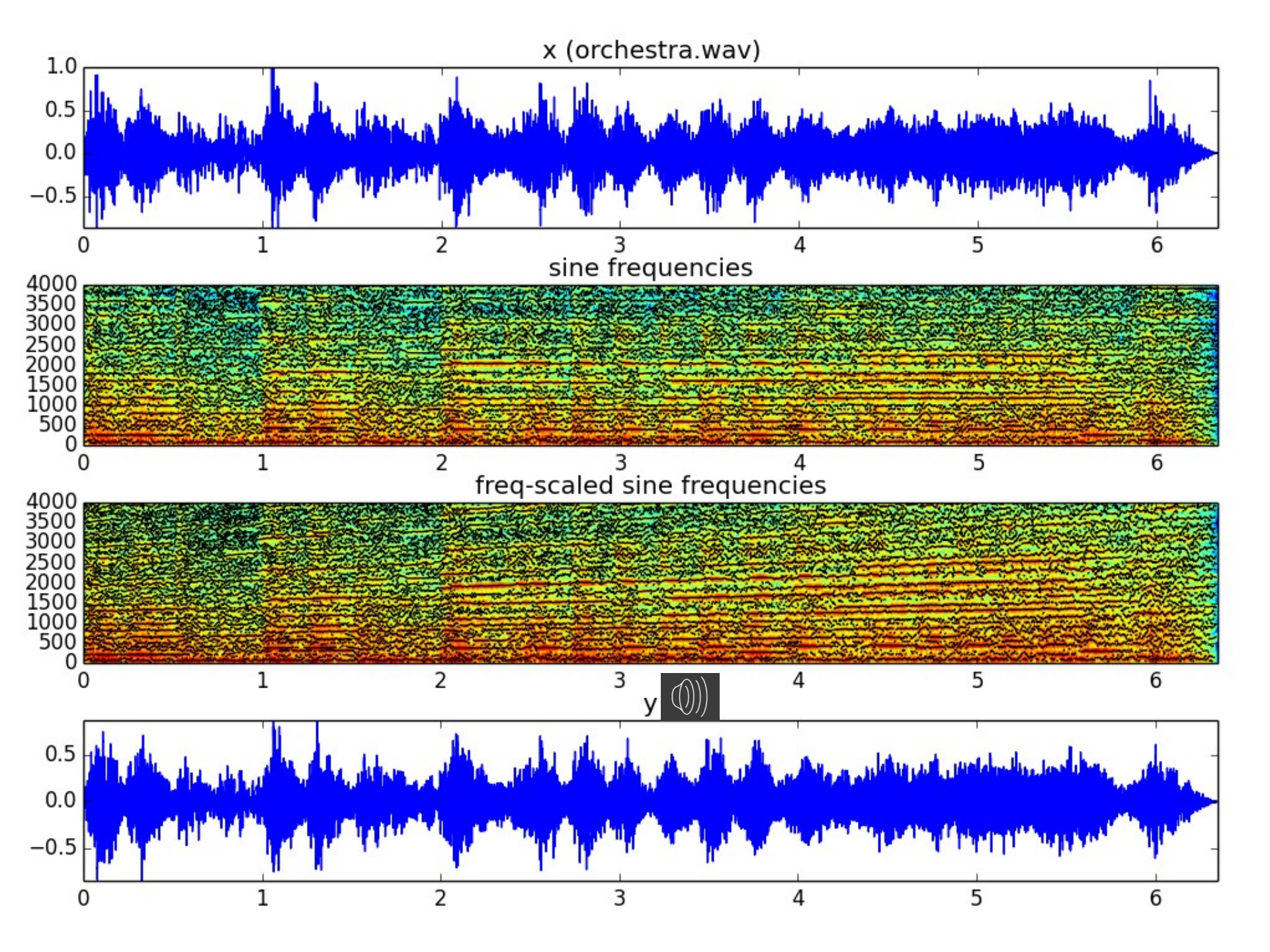

x (mridangam.wav)
mX + sine frequencies
mY + time-scaled sine frequencies
y


x (mridangam.wav)
mX + sine frequencies
mY + time-scaled sine frequencies
y

Frequency scaling




x (orchestra.wav)
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
sine frequencies
4000
3500
3000
2500
2000
1500
1000
500
0
freq-scaled sine frequencies
y


x (orchestra.wav)
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
sine frequencies
4000
3500
3000
2500
2000
1500
1000
500
0
freq-scaled sine frequencies
y

References

- More information on this topic from Wikipedia:
  - http://en.wikipedia.org/wiki/Sound_effects
  - http://en.wikipedia.org/wiki/Equalization_filter
  - http://en.wikipedia.org/wiki/Audio_timescale-pitch_modification
- Sounds: http://www.freesound.org/people/xserra/packs/13038/
- The slides and code are released using the CC Attribution-Noncommercial-Share Alike license or the Affero GPL license and available from https://github.com/MTG/sms-tools

8T1: Spectral-based sound transformations (1 of 2)

Xavier Serra
Universitat Pompeu Fabra, Barcelona
&
Stanford University