Digital Signal Processing for Music

Part 23: Time-stretching and Pitch-shifting

Andrew Beck

Intro

- **Time Stretching**

  Change playback speed/tempo without changing pitch
- **Pitch Shifting**

  Change pitch without changing tempo / playback speed
- **Terms**
  - Time / pitch scaling
  - Time expansion / compression

Applications

- **Beat Matching:**: Align tempo of two or more audio files (mashup)
- **Key lock**: "Align" pitch of two or more audio files (mashup)
- **Pitch / time correction**: Edit intonation, frequency deviation, vibrato, glissando
- Video **frame rate conversion**
- **Sample player / libraries**
- **Sound design**
- **Educational software**: Pitch and timing visualization

Stretch and Pitch Factors

$$s = \frac{t_{output}}{t_{input}} \qquad p = \frac{f_{output}}{f_{input}}$$

Examples:

- *Half speed:* $s = 2$
- *Half pitch:* $p = \frac{1}{2}$
- *Semitone up / down:*

$$p_u = 2^{\frac{1}{12}} = 1.059 \quad p_d = 2^{-\frac{1}{12}} = 0.9439$$

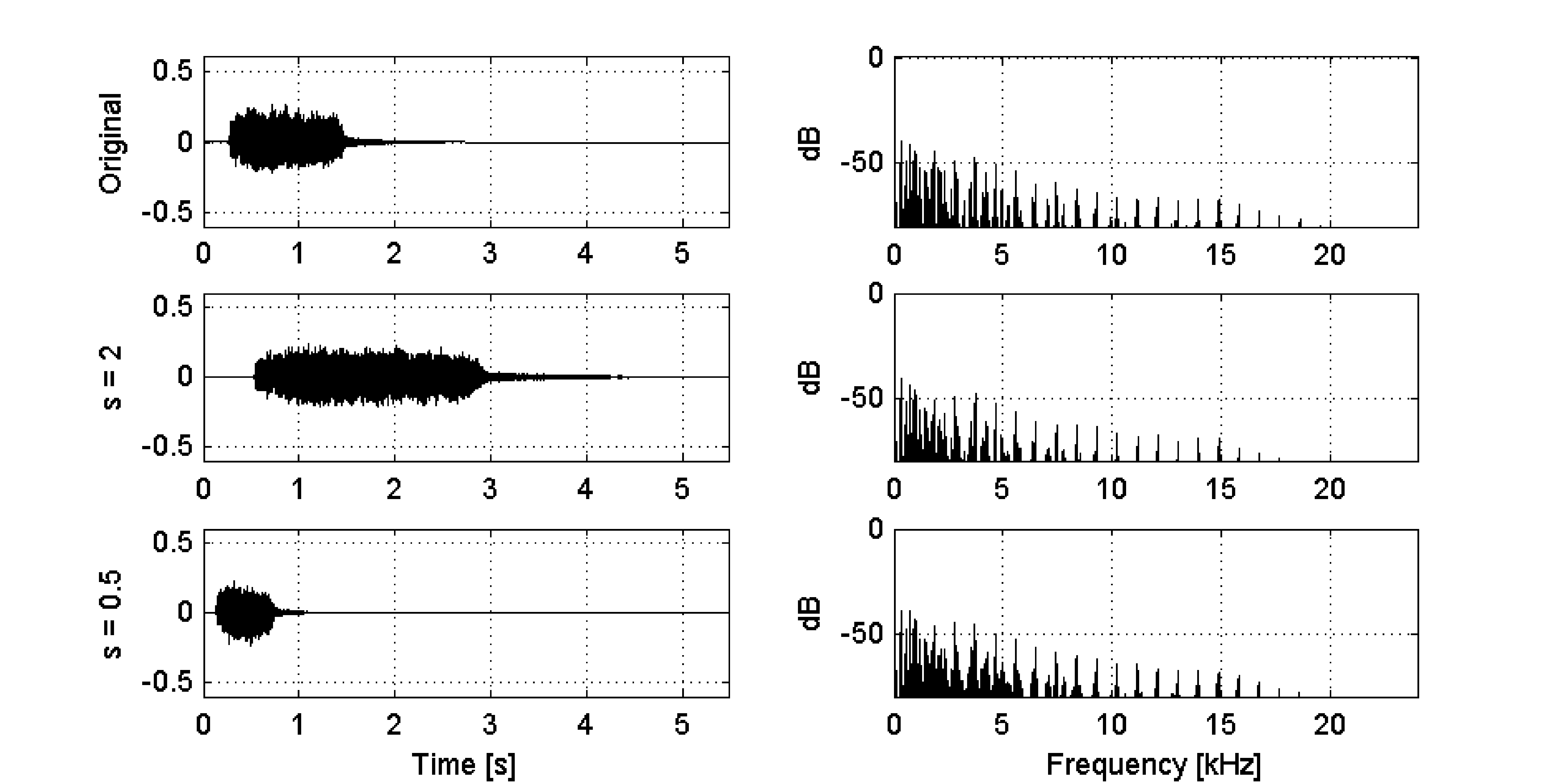
- *100 BPM → 75 BPM:* $s = \frac{4}{3}$

Resampling

- **Traditional**: resampling
  - Change inter-sample 'distance' by interpolation
  - Keep playback sample rate constant
  - Audio example
    - Original: 0:00 / 0:15
    - Resample: 0:00 / 0:11
- Tempo changes results in pitch change (and vice versa)

$$s = \frac{1}{p}$$

Stretching: Effect on Frequency Domain



OLA: Introduction

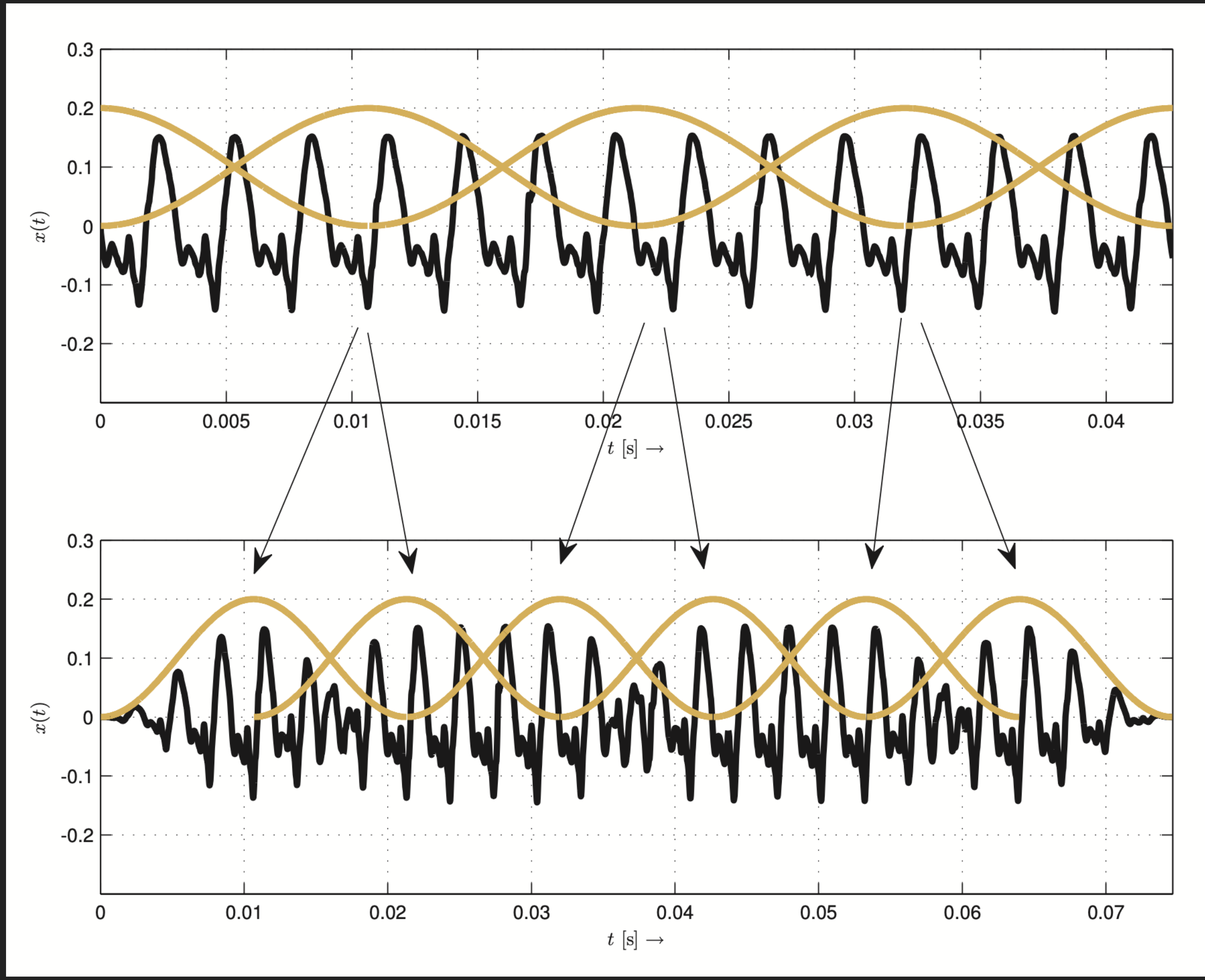
Overlap and add approaches for:

- Granular synthesis
- Time /frequency synthesis and processing
- Time-stretching and pitch-shifting

Time Stretching

Overlap and add

1. **Split input** signal into overlapping blocks
2. **Duplicate or discard blocks** depending on stretch factor



» Original:

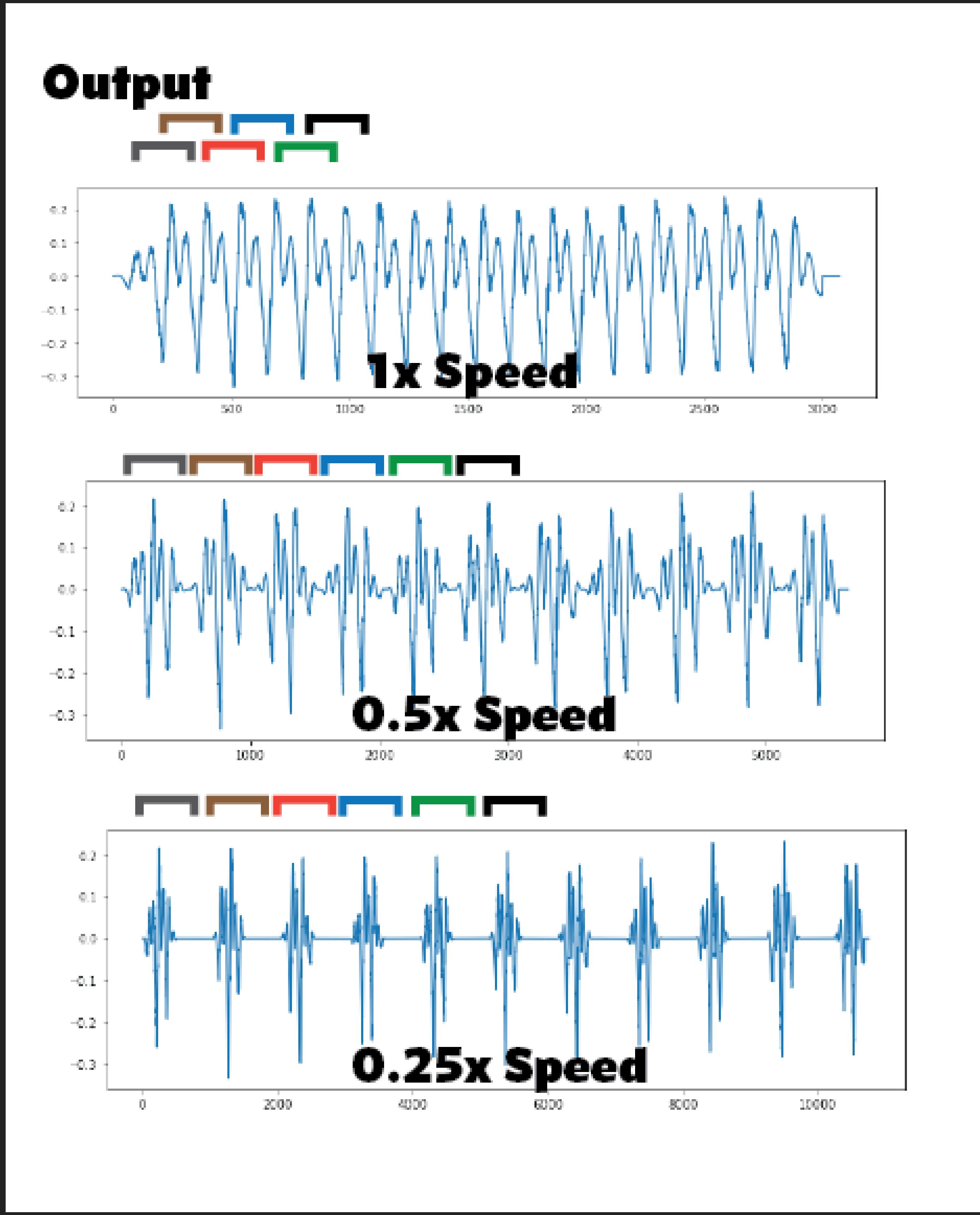
0:00 / 0:15

» $s = \frac{4}{3}$:

0:00 / 0:20

Windowed Grains

Outputing Grains

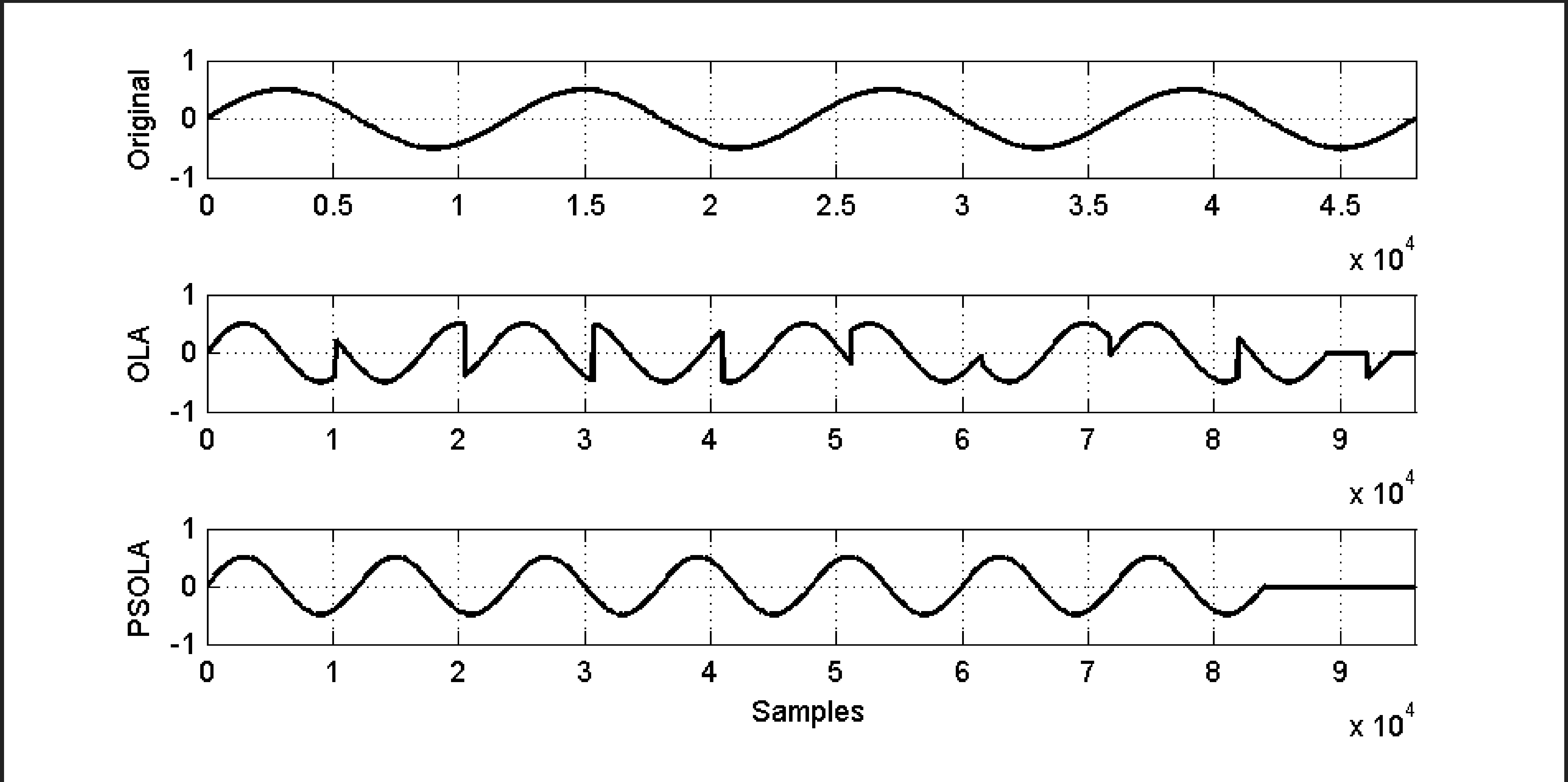


Parameters

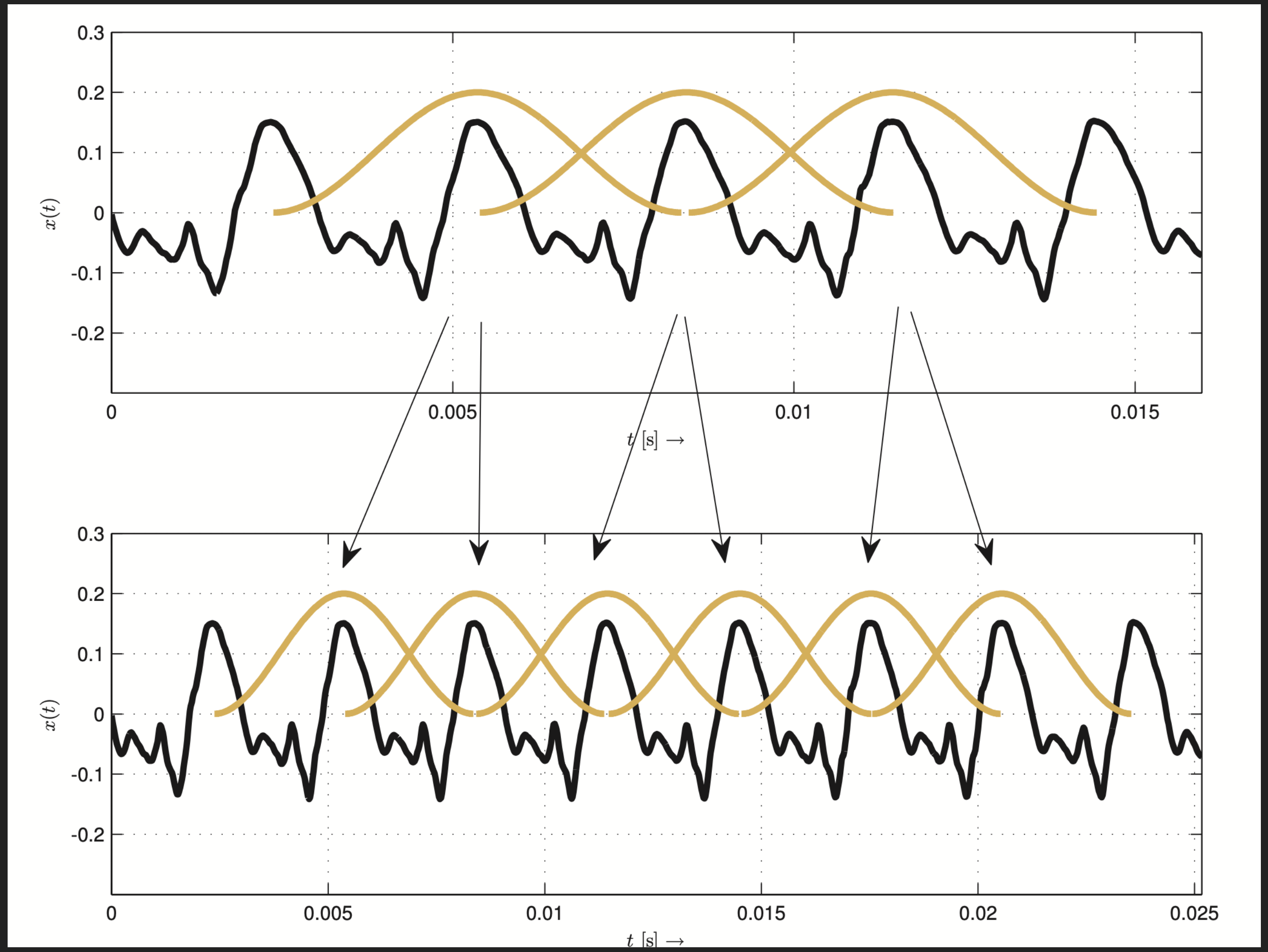
- Grain Size
- Hop Size
- Time Scale
- Pitch Scale
- Time Variance
- Pitch Variance
- Interpolation Method

Pitch Synchronous Overlap and Add

- Use the OLA principle, but
- **Adapt block length** to fundamental period length



PSOLA: Example



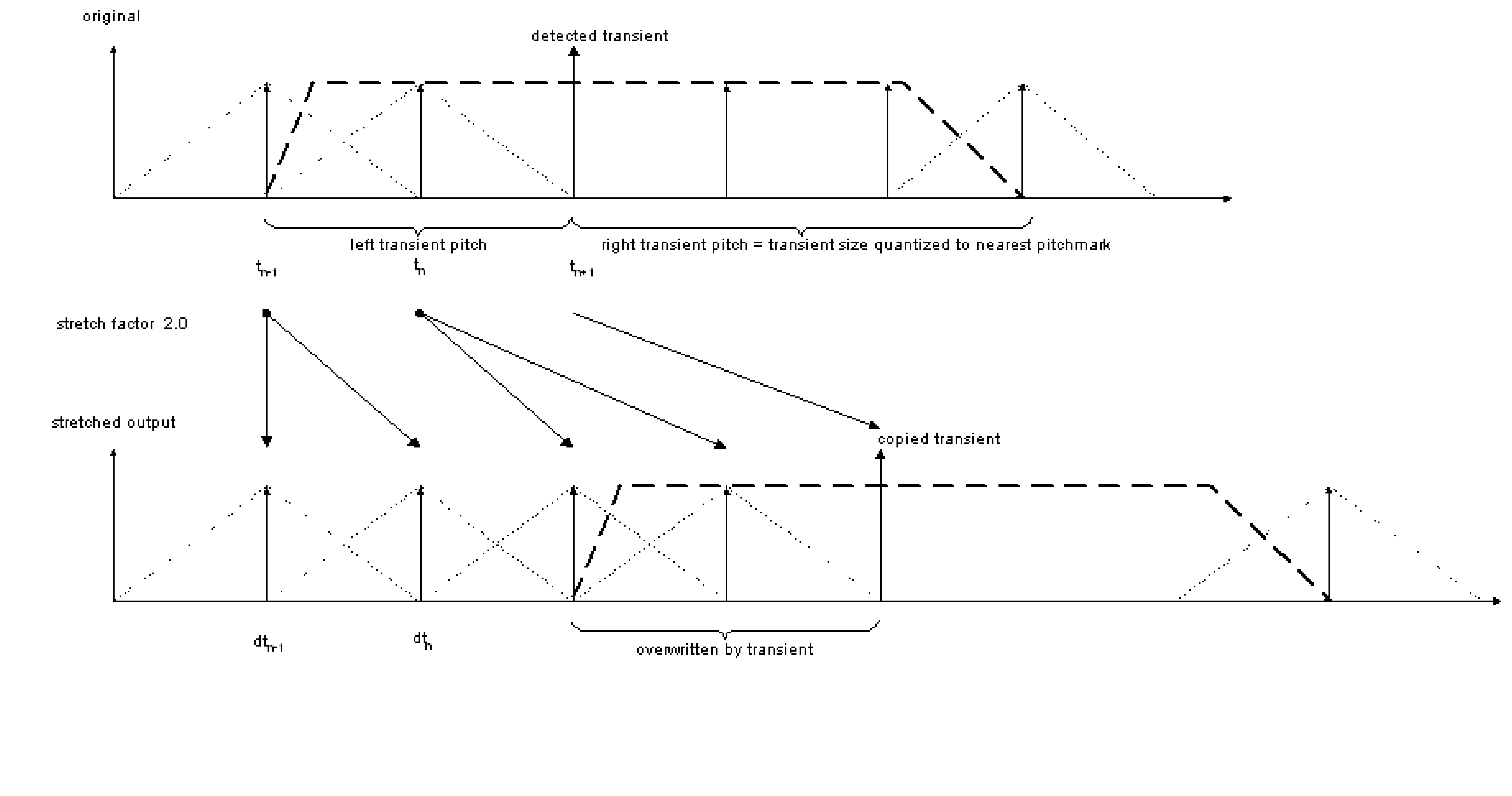
» Original:

0:00 / 0:15

» $s = \frac{4}{3}$:

0:00 / 0:20

PSOLA: Transient Copying



PSOLA: Summary

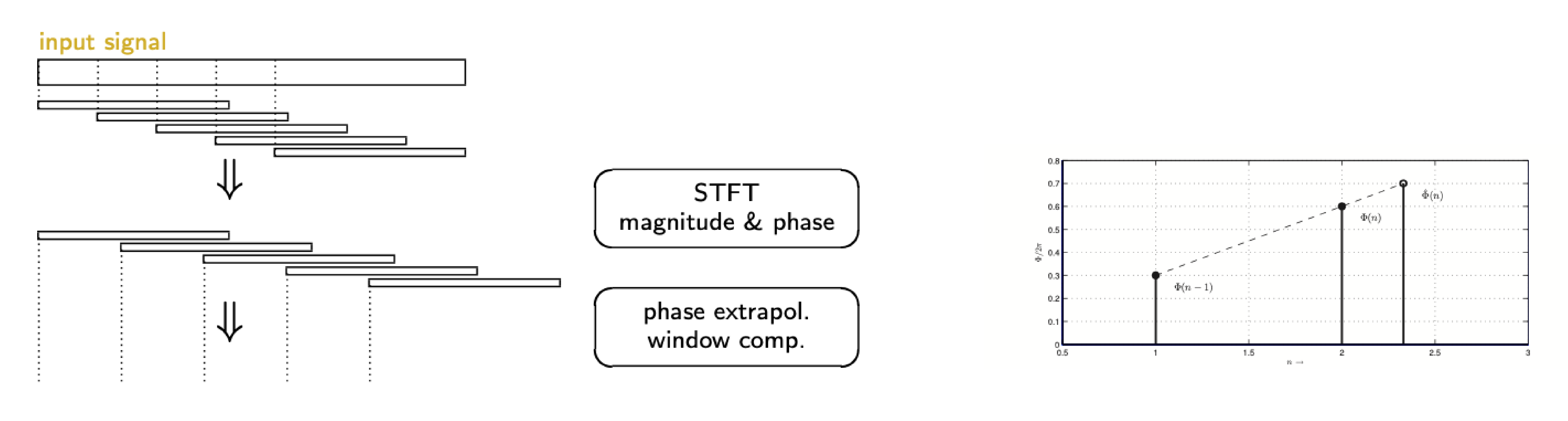
- **Processing Steps**
  - Detect *fundamental frequency* / period length
  - Set *pitch marks*
  - Intelligently *select blocks* to be repeated / discarded
- **Advantages**
  - *High granularity* - Modify audio on period length resolution
  - *High quality*

PSOLA: Summary

- **Problems**
  - Quality depends on *pitch tracking* reliability
  - Quality and timbre depends on *pitch mark positioning*
  - Works only for *monophonic* input signals
    - Polyphonic and noisy segments
    - Reverberation and overlapping tones
  - *Noise, plosives* require special consideration
  - *Copying* artifacts (double transients, timing deviations)
- **Typical Applications**
  - **Standard approach for vocal editing tools**

Phase Vocoder: Frequency Domain OLA

1. **Split input** signal into overlapping blocks
2. Compute **magnitude and phase spectrum** of each block
3. **Change overlap ratio** between blocks depending on stretch factor
4. Keep the magnitude, **adapt the phase per bin** to the blocks new time stamp



Phase Vocoder: Audio Example

- Original: 0:00 / 0:15
- Phase Vocoded: 0:00 / 0:20

Frequency Reassignment: Relation of Phase and Frequency

Phasor Representation:

- Sine value is defined by magnitude and phase
- Decreasing the amplitude → shorter vector
- Increasing the frequency → Increasing speed

Frequency and phase change closely related

- Time for full rotation is period length T with

$$f = \frac{1}{T}$$

- Time for fractional rotation $\Delta\Phi$ is corresponding fraction of period length

$$f = \frac{\Delta\Phi}{\Delta t}$$

Frequency and phase change closely related

- In other words:

$$\Phi(t) = \omega \cdot t$$

$$\Rightarrow \frac{d\Phi(t)}{dt} = \omega = 2\pi f$$

Frequency Reassignment: Principles

Frequency domain

- Instead of using the bin frequency

$$f(k) = k * \frac{f_\mathrm{S}}{\mathcal{K}}$$

- We use the phase of each bin $\Phi(k, n)$
- To compute the frequency from the phase difference of neighboring blocks

$$\omega_\mathrm{I}(k, n) \propto \Phi(k, n) - \Phi(k, n-1)$$

- $\omega_\mathrm{I}(k, n)$ is called **instantaneous frequency** per block per bin

Frequency Reassignment: Scaling Factor

- Instantaneous frequency calculation has to take into account
  - Hop Size: $\mathcal{H}$
  - Sample Rate: f_s

$$\omega_{\mathrm{I}}(k,n) = \frac{\Delta\Phi_{\mathrm{u}}(k,n)}{\mathcal{H}} \cdot f_{\mathrm{S}}$$

- Problem: Phase ambiguity

$$\Phi(k,n) = \Phi(k,n) + j \cdot 2\pi$$

- *Phase unwrapping*

Phase Unwrapping

- Compute unwrapped phase $\Phi_{\mathrm{u}}(k,n)$
  - Estimate unwrapped bin phase

$$\hat{\Phi}(k,n) = \Phi(k,n-1) + \underbrace{2\pi k \cdot \frac{\mathcal{H}}{\mathcal{K}}}_{=\omega_k \cdot \frac{\mathcal{H}}{f_{\mathrm{s}}}}$$

  - Unwrap phase by shifting current phase to estimates range

$$\Phi_{\mathrm{u}}(k,n) = \hat{\Phi}(k,n) + \mathbf{princarg}\left[\Phi(k,n) - \hat{\Phi}(k,n)\right]$$

Phase Unwrapping

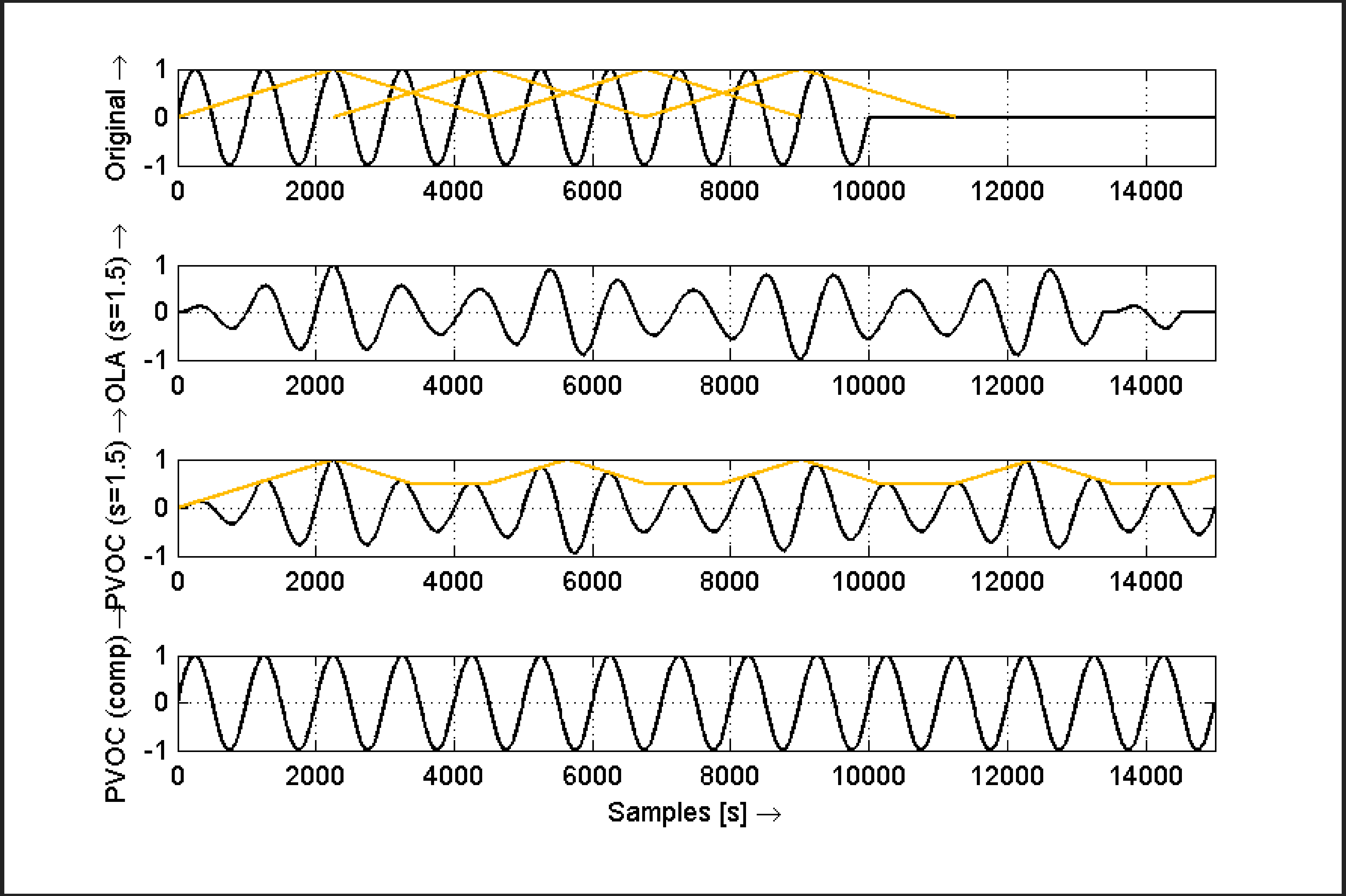
- Compute unwrapped phase difference

$$\begin{aligned}\Delta\Phi_{\mathrm{u}}(k,n) &= \Phi_{\mathrm{u}}(k,n) - \Phi(k,n-1)\\ &= \hat{\Phi}(k,n) + \operatorname{princarg}\left[\Phi(k,n) - \hat{\Phi}(k,n)\right] - \Phi(k,n-1)\\ &= \frac{2\pi k}{\mathcal{K}}\mathcal{H} + \operatorname{princarg}\left[\Phi(k,n) - \Phi(k,n-1) - \frac{2\pi k}{\mathcal{K}}\mathcal{H}\right]\end{aligned}$$

Frequency Reassignment: Problems

- **Overlapping spectral components**
  - Sinusoidal components often overlap (Spectral leakage, several instruments playing the same pitch, ...)
  - Incorrect phase estimates
  - Spectrum should be as sparse as possible, increase STFT length
- **Inaccurate phase unwrapping**

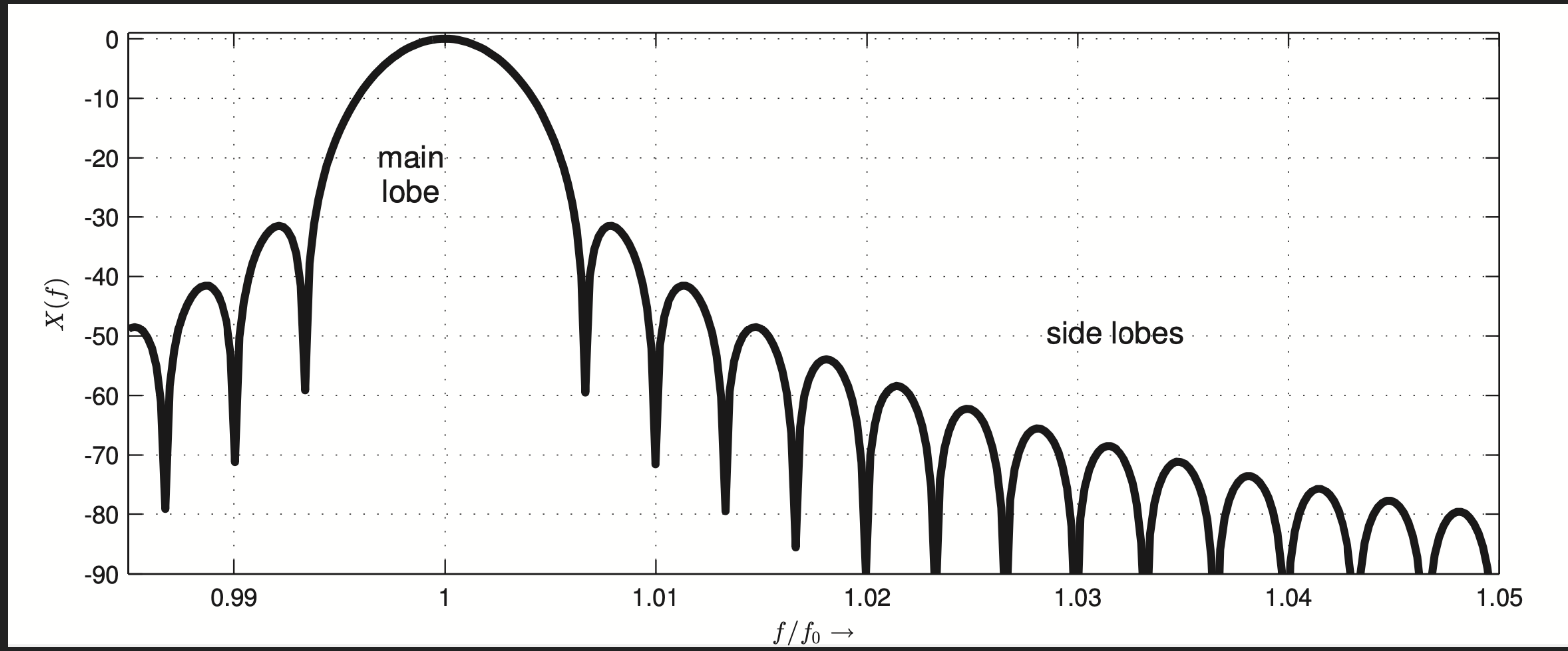
Phase Vocoder Window Compensation

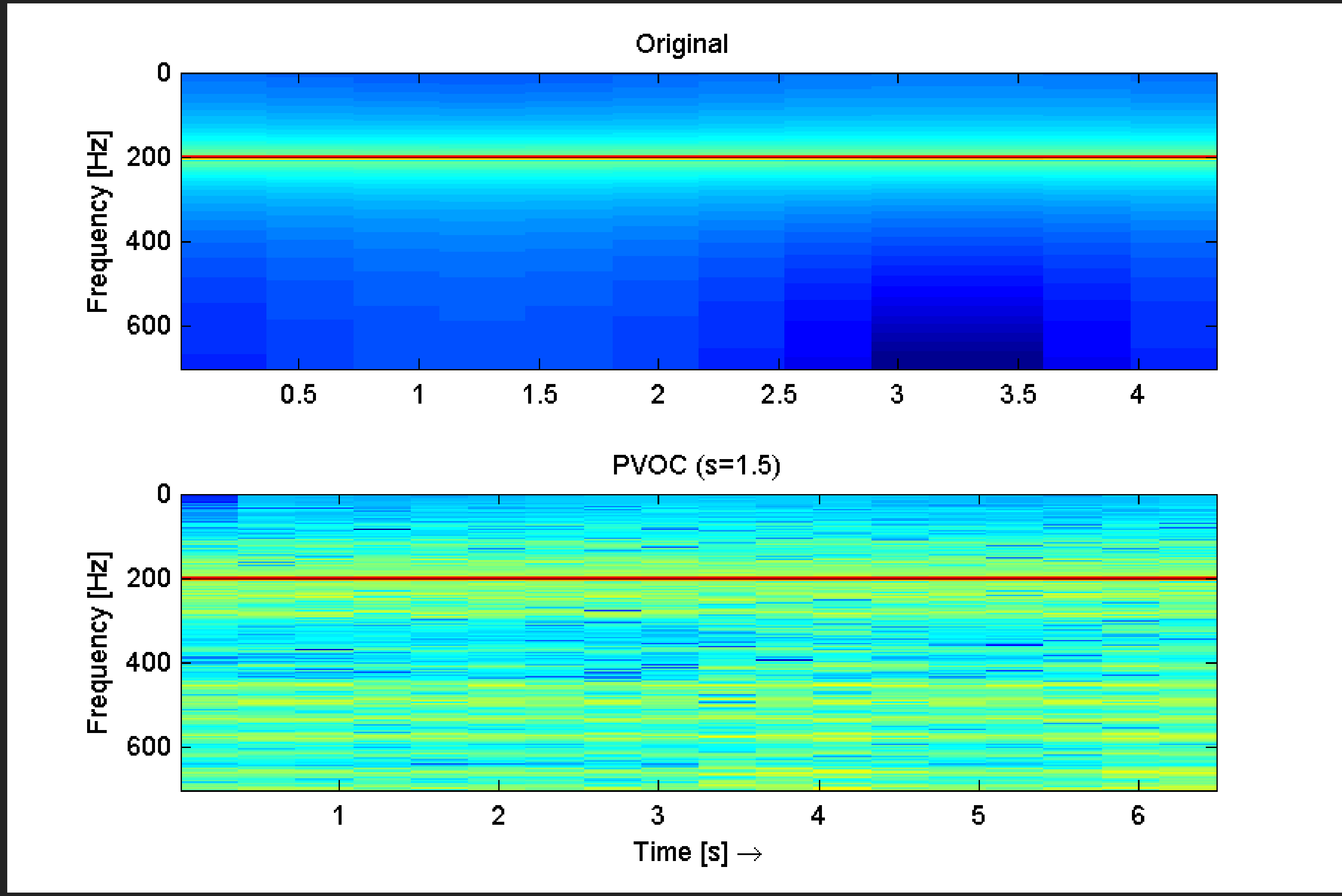


Phase Vocoder - Properties & Artifacts

- **Advantages**
  - Allows *polyphonic input* (assumption: no overlapping harmonics)
  - Absolute *timing stability* (i.e., sample resolution)
- **Disadvantages**
  - *Low granularity* - FFT block size
  - Artifacts: Phasing, Transient smearing / doubling

Phase Vocoder Artifacts: Spectral Leakage



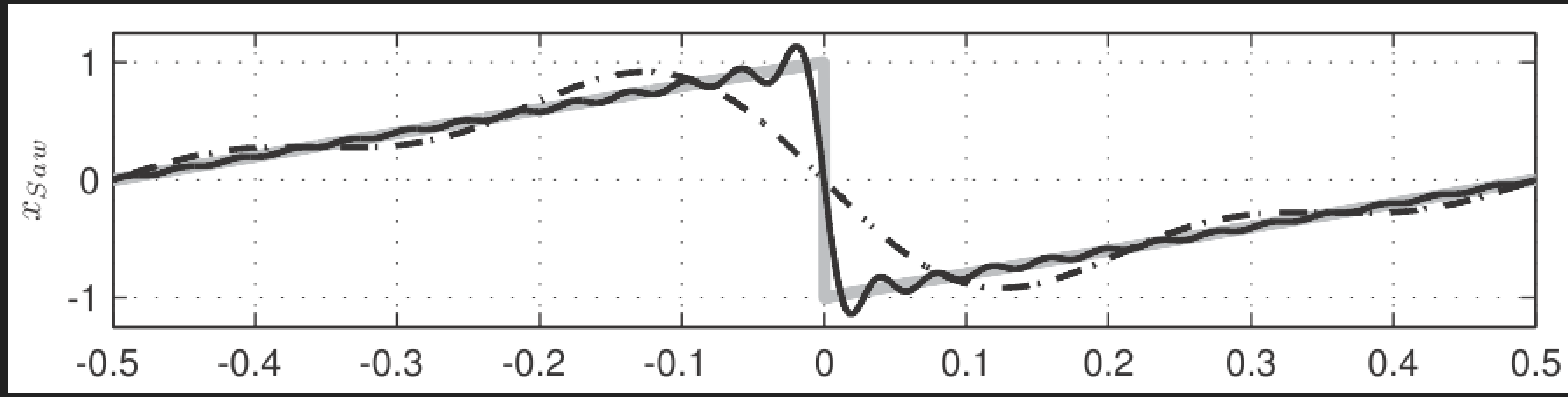


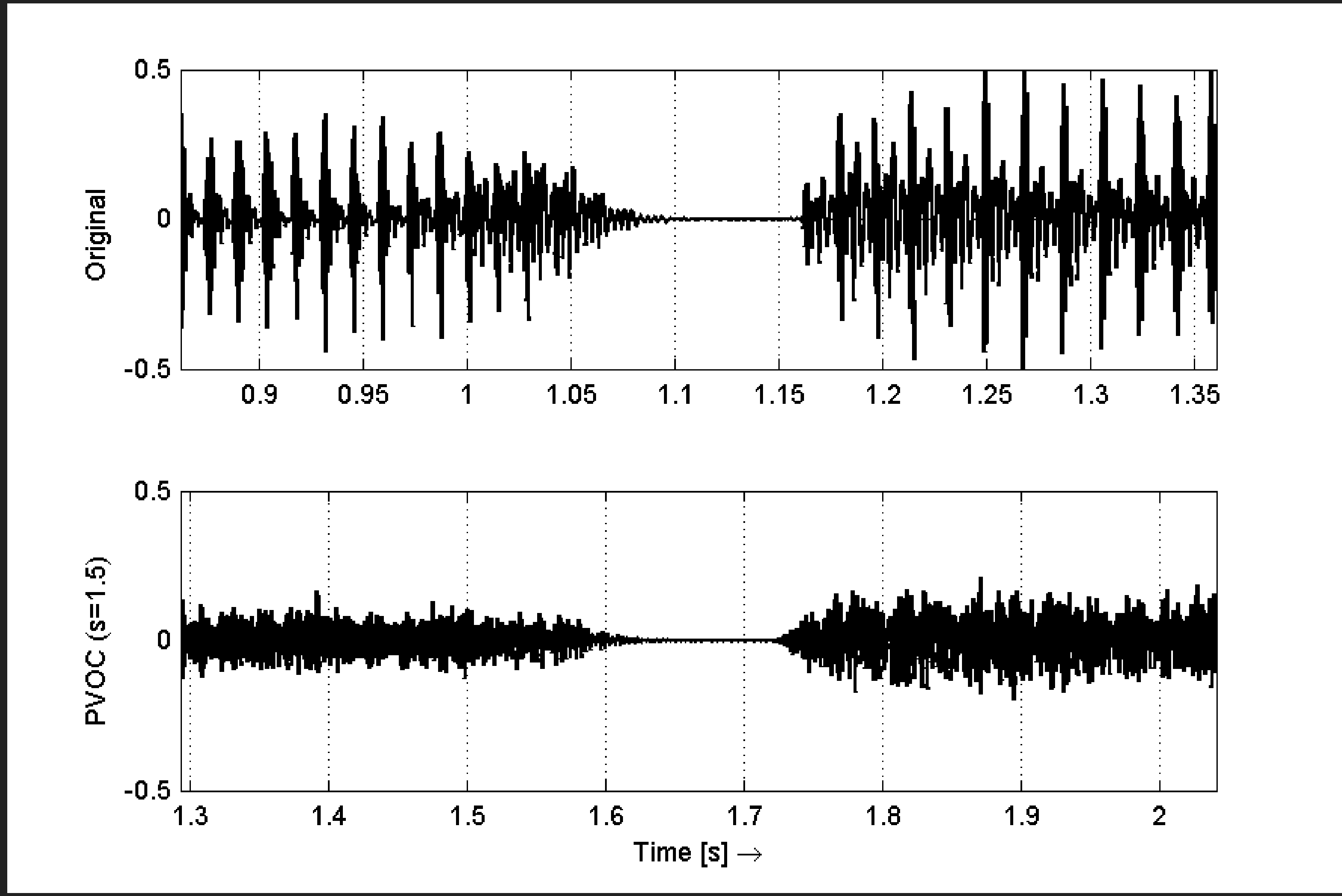
Phase Vocoder Phasing

Use *Frequency Reassignment* for grouping and phase sync

- Original: 0:00 / 0:15
- Phase Vocoder: 0:00 / 0:20
- PV w/ grouped phase: 0:00 / 0:20

Phase Vocoder Artifacts - Unsynced Harmonics





Use *Harmonic Analysis* for Grouping and Phase Sync

- Original: 0:00 / 0:15
- Phase Vocoder: 0:00 / 0:20
- PV w/ synced phase: 0:00 / 0:20

Phase Vocoder Artifacts: Interchannel Phasing

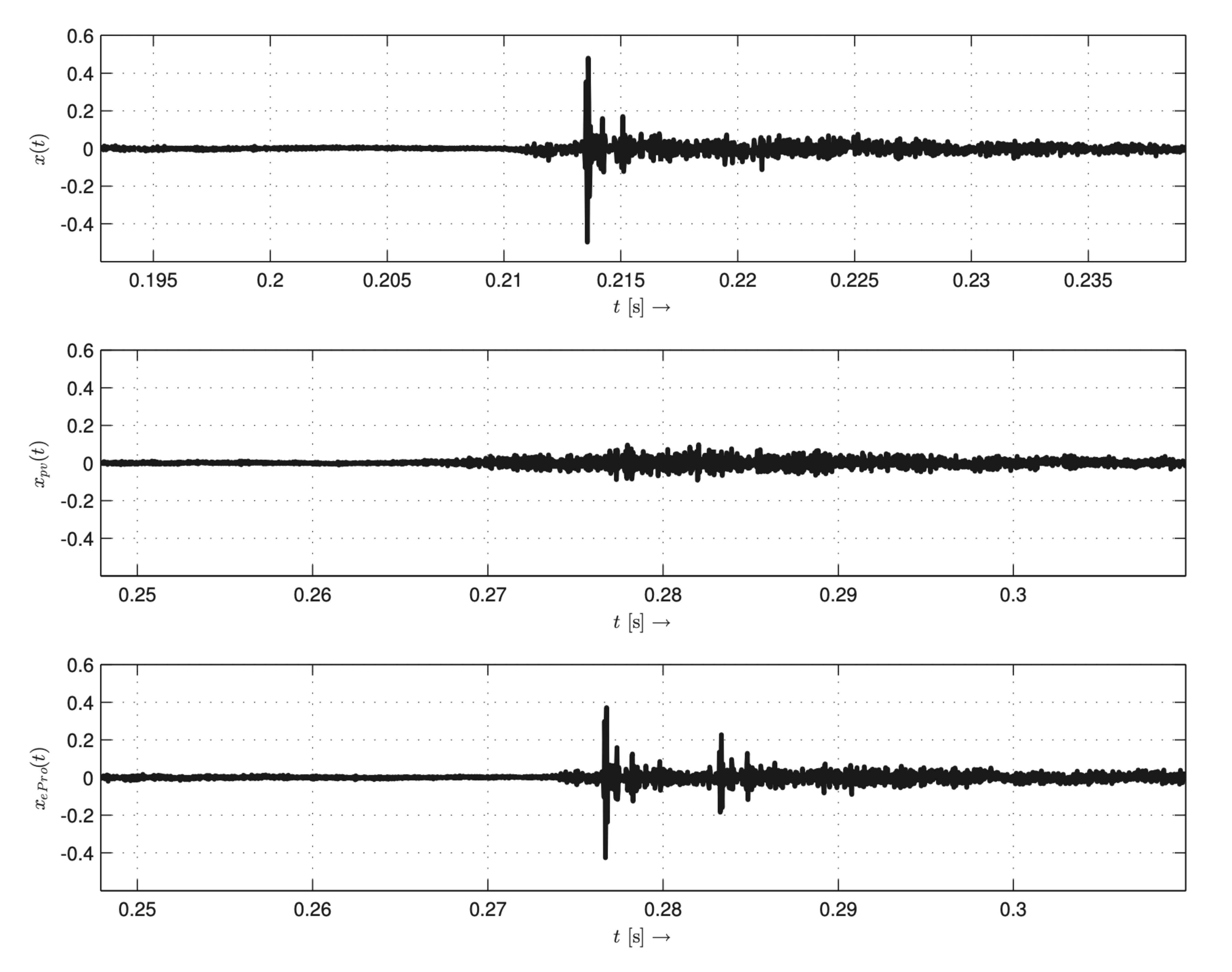
Phase estimation between channels slightly off due to

- Numerical inaccuracies (cumulative!)
- Overlapping frequency components

Change in spatial image

- Original: 0:00 / 0:17
- Phase Vocoder: 0:00 / 0:25

Phase Vocoder Artifacts: Transient Smearing



Transient Smearing Example

- Original: 0:00 / 0:06
- Phase Vocoder: 0:00 / 0:09

Detect transients and *reset phase* per bin

- Original: 0:00 / 0:15
- Phase Vocoder: 0:00 / 0:20
- PV w/ Phase Reset: 0:00 / 0:20

Time Stretching: Inherent Problems

- Stretching the audio data can lead to **"non-natural"** results
- **Examples**
  - Tempo dependent *timing variations*
  - Other performance related aspects may get inappropriate lengths and speed: *vibrato, tremolo, glissando*

Pitch Shifting

- **Definition**

  Change pitch without changing tempo

- **Method**

  Combine stretching and *sample rate conversion* (interpolation)

  Change length with stretching

  Resample to compensate for length difference

- **Implementation**: Differentiate "external" and "internal" parameters
  - *External*: stretch s_e and pitch p_e
  - *External*: stretch s_i and resample r_i

Pitch Shifting: Example

Pitch shift factor $p = \frac{4}{3}$

- *Time stretch* (increase length / decrease tempo) $s = \frac{4}{3}$
- *Resample* (decrease length / increase pitch) $s = \frac{3}{4}$
- OLA: 0:00 / 0:15
- Phase Vocoder: 0:00 / 0:15

Pitch Shifting: Standard Approach Examples

- External: $s_e = 1 \ldots p_e = 2$
  Internal: $s_i = 2 \ldots r_i = \frac{1}{2}$
- External: $s_e = 1 \ldots p_e = \frac{4}{3}$
  Internal: $s_i = \frac{4}{3} \ldots r_i = \frac{3}{4}$
- External: $s_e = \frac{1}{2} \ldots p_e = 2$
  Internal: $s_i = 1 \ldots r_i = \frac{1}{2}$
- External: $s_e = 2 \ldots p_e = 2$
  Internal: $s_i = 4 \ldots r_i = \frac{1}{2}$

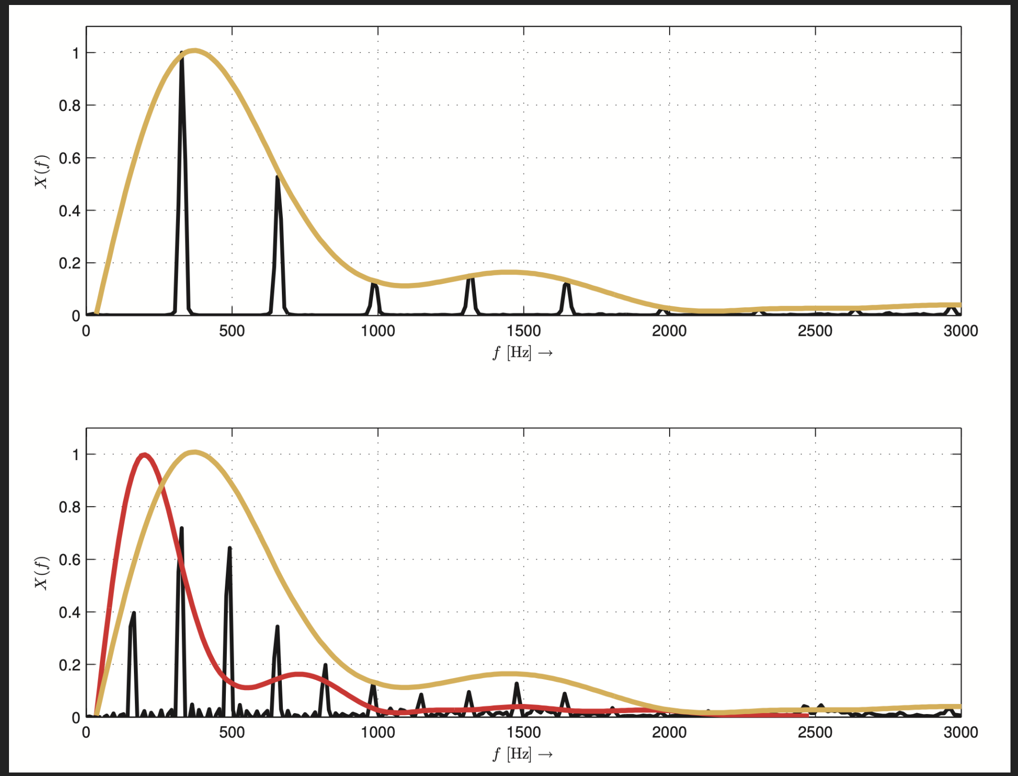
Pitch Shifting: Frequency Domain Approach

- STFT
- Magnitude and phase
- Magnitude and instantaneous frequency
- Resample both magnitude and frequency spectrum according to pitch factor
- Magnitude and phase
- Complex spectrum
- IFFT and OLA

Format Preservation: Time Domain

Formant Preservation: Frequency Domain

- **Idea**
  - Preserve spectral envelope
- **Approach**
  - Measure spectral envelope
  - Apply inverse envelope (whitening)
  - Pitch shift
  - Apply spectral envelope



Spectral Envelope Estimate

- **Approaches**
  - LPC coefficients
  - Spectral maxima
- **Potential issues**
  - *Polyphonic input* audio
    'Superposition' of envelopes
  - *Very high / low pitch factors*: High frequency boost / cut
  - *Estimate resolution*
    - Too coarse → Loss of timbre characteristics
    - Too fine → Impress pitch characteristics (harmonic pattern) on spectrum

Pitch Shifting: Audio Examples

	OLA	PSOLA	PVOC
Original			
Resample			
Formant			

Summary

- Pitch stretching and pitch shifting are largely equivalent algorithms
  - Sample artifacts
  - Same workload
- Monophonic time-stretching with PSOLA-based approaches
  - Easier to solve
  - Has bad artifacts if pitch tracker is off

Summary

- Polyphonic time-stretching with PV-based approaches
  - Complicated due to tradeoffs (e.g., frequency vs time resolution
- General challenges:
  - Noisy and transient signals
  - Resulting timbre changes
  - Perceived naturalness of result
  - Time resolution / accuracy due to blocked processing