Multiresolution Representations of Musical Rhythm & Expectation

Leigh M. Smith

Universiteit van Amsterdam
ILLC / Music Cognition Group

Music Cognition Group

EU Project: *Emergent Cognition through Active Perception (EmCAP)*

(European Commission FP6-IST, contract 013123)

- The study of how cognitive behaviour in artificial systems can emerge through interacting with a musical environment.

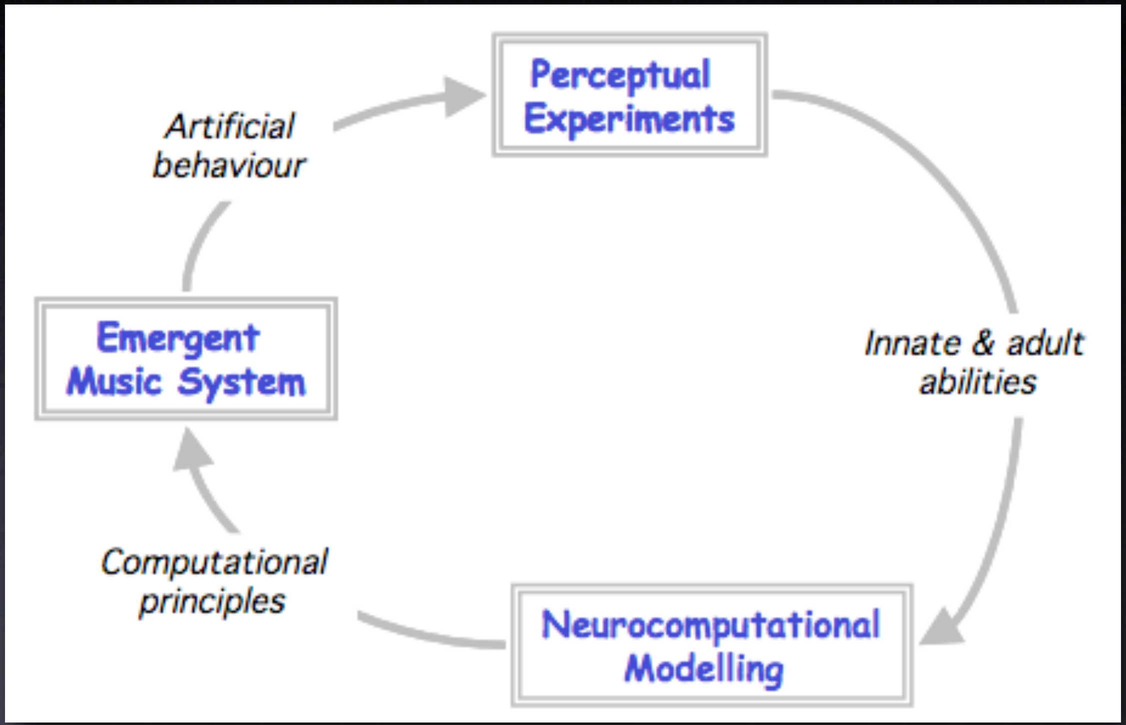
- Neuroimaging innate vs. learned auditory functions.
- Perception of musical form.
- Prefrontal cortical function controlling attention and STM.
- Spectrotemporal response fields in the thalamocortical system.
- Perception and categorisation of rhythmic patterns.
- Active perception, relative pitch and emergence of tonality.
- Interactive music system: *The Music Projector*.

http://www.musiccognition.nl/EmCAP

Music Cognition Group

EU Project: *Emergent Cognition through Active Perception (EmCAP)*

(European Commission FP6-IST, contract 013123)



- Neuroimaging innate vs. learned auditory functions.
- Perception of musical form.
- Prefrontal cortical function controlling attention and STM.
- Spectrotemporal response fields in the thalamocortical system.
- Perception and categorisation of rhythmic patterns.
- Active perception, relative pitch and emergence of tonality.
- Interactive music system: *The Music Projector*.

http://www.musiccognition.nl/EmCAP

EU Project: *Emergent Cognition through Active Perception (EmCAP)*

(European Commission FP6-IST, contract 013123)

- Neuroimaging innate vs. learned auditory functions.
- Perception of musical form.
- Prefrontal cortical function controlling attention and STM.
- Spectrotemporal response fields in the thalamocortical system.
- Perception and categorisation of rhythmic patterns.
- Active perception, relative pitch and emergence of tonality.
- Interactive music system: *The Music Projector*.

Information Society Technologies · University of Plymouth · Universitat Pompeu Fabra · MTA Pszichológiai Kutatóintézet · Universiteit van Amsterdam

http://www.musiccognition.nl/EmCAP

Music
Cognition
Group

Rhythmic Expectation

- Question:
  - What contribution to expectation arises from the temporal structure of the rhythm?

or

  - How much information is actually within the rhythmic signal?

Rhythmic Expectation

- Question:
  - What contribution to expectation arises from the temporal structure of the rhythm?

or

  - How much information is actually within the rhythmic signal?

...some theories of musical rhythm...

Rhythm

Rhythm

- "...the *systematic* patterning of sound in terms of timing, accent, and grouping." (Patel 2008 p.96)

Rhythm

- “...the *systematic* patterning of sound in terms of timing, accent, and grouping.” (Patel 2008 p.96)
  - (Not always periodic patterns)

Rhythm

- "...the *systematic* patterning of sound in terms of timing, accent, and grouping." (Patel 2008 p.96)
  - (Not always periodic patterns)
- Accent sources include: dynamics, melody, harmony, articulation, timbre, onset asynchrony etc.

Rhythm

- "...the *systematic* patterning of sound in terms of timing, accent, and grouping." (Patel 2008 p.96)
  - (Not always periodic patterns)
- Accent sources include: dynamics, melody, harmony, articulation, timbre, onset asynchrony etc.
- Consists of *hierarchical* and *figural* (proximal) temporal structures.

Music
Cognition
Group

Meter

- Meter is expressed in Western music as time-signatures (4/4, 3/4 etc).

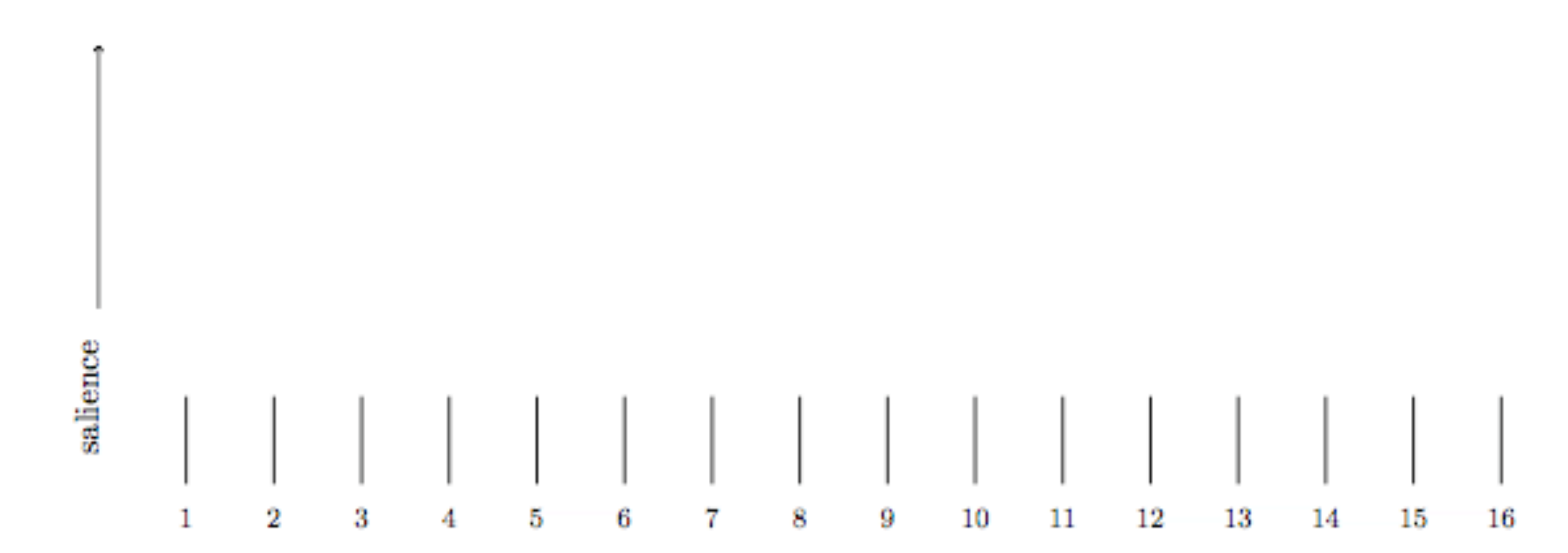
Subdivision of 4/4 (4 beats to the bar):

(Courtesy Olivia Ladinig)

Meter

- Meter is expressed in Western music as time-signatures (4/4, 3/4 etc).

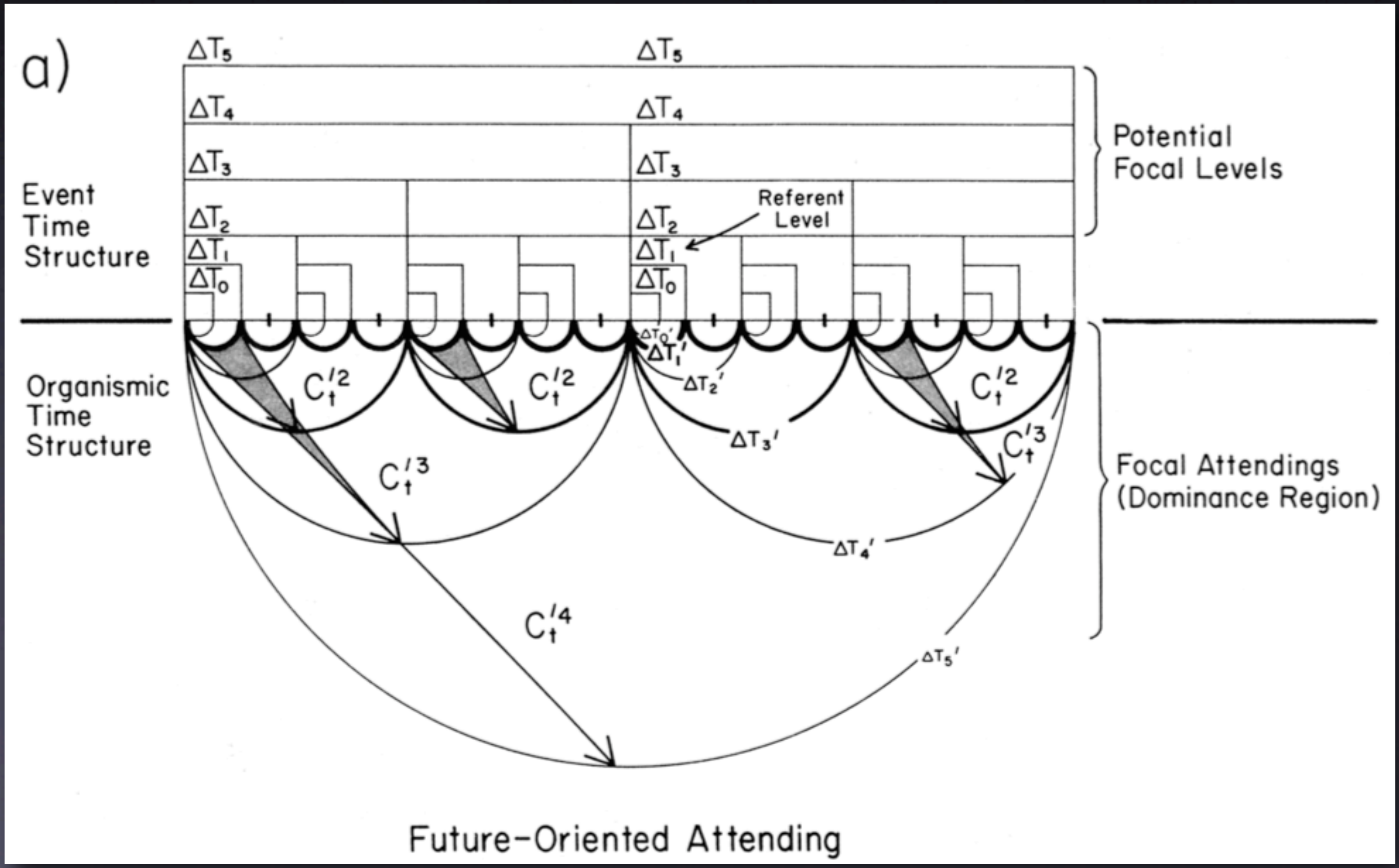
Subdivision of 4/4 (4 beats to the bar):



(Courtesy Olivia Ladinig)

Rhythmic Strata

- Musical rhythm can be considered as composed of a hierarchy of temporal levels or *strata* (Yeston 1976, Lerdahl & Jackendoff 1983, Clarke 1987, Jones & Boltz 1989).



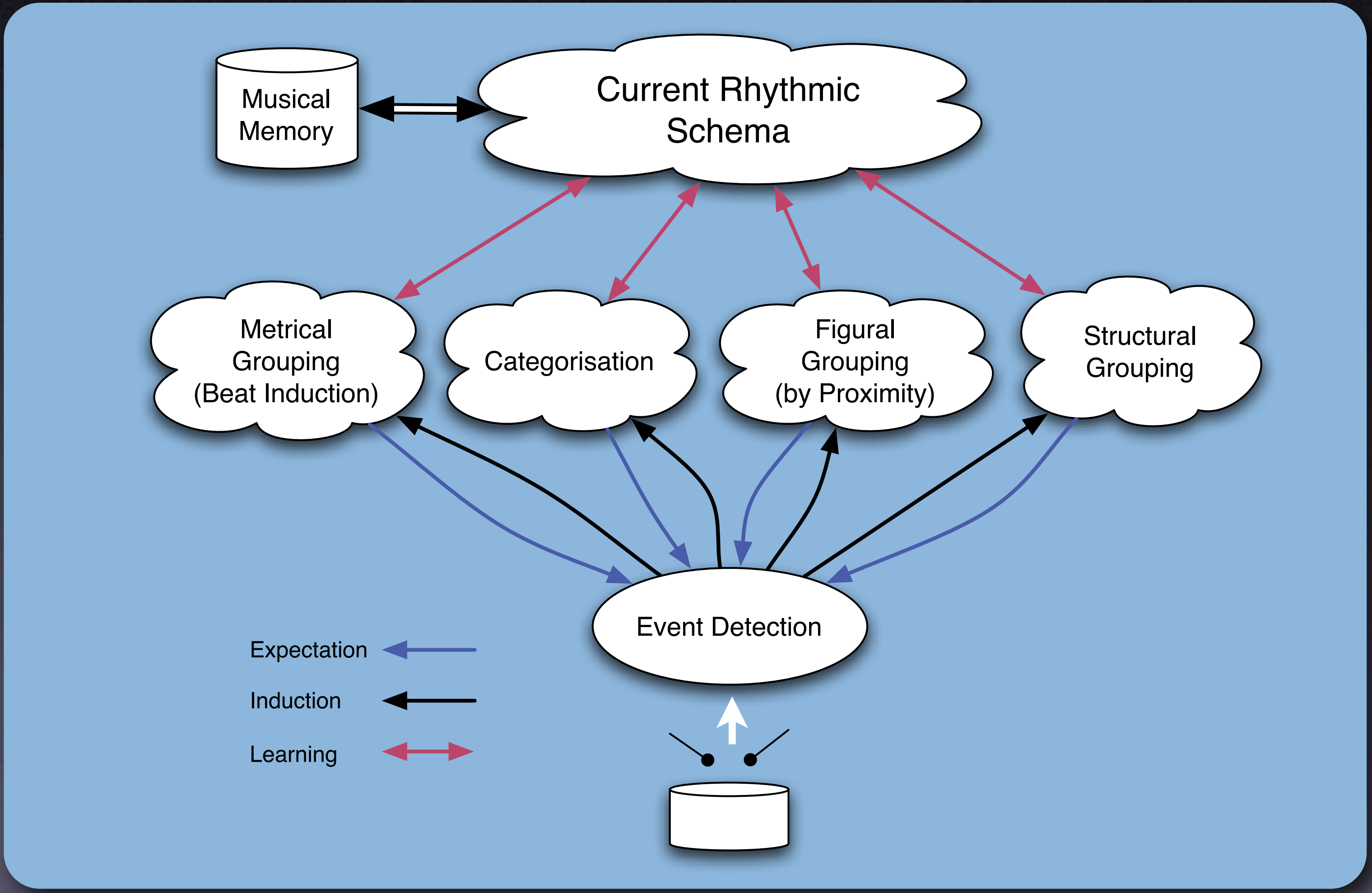
From Jones & Boltz '89

Hierarchical Grouping: Meter

- Meters are argued to arise from the interaction between temporal levels (Yeston 1976).
- Therefore a meter implies two frequencies: the pulse rate and the measure (“bar”) rate.
- The *tactus* is considered as the most salient hierarchical level, consistent with the notated meter, or the foot tapping rate (Desain & Honing 1994).

Active Rhythm Perception

- Viewed as a resonance between top down and bottom-up processes (see e.g Desain & Honing 1999):



Model Requirements

- Accounts for multiple, overlapping, temporal contexts.
- Multiple beat hypotheses.
- Identification of tactus.
- Expressive timing (*tempo rubato*).

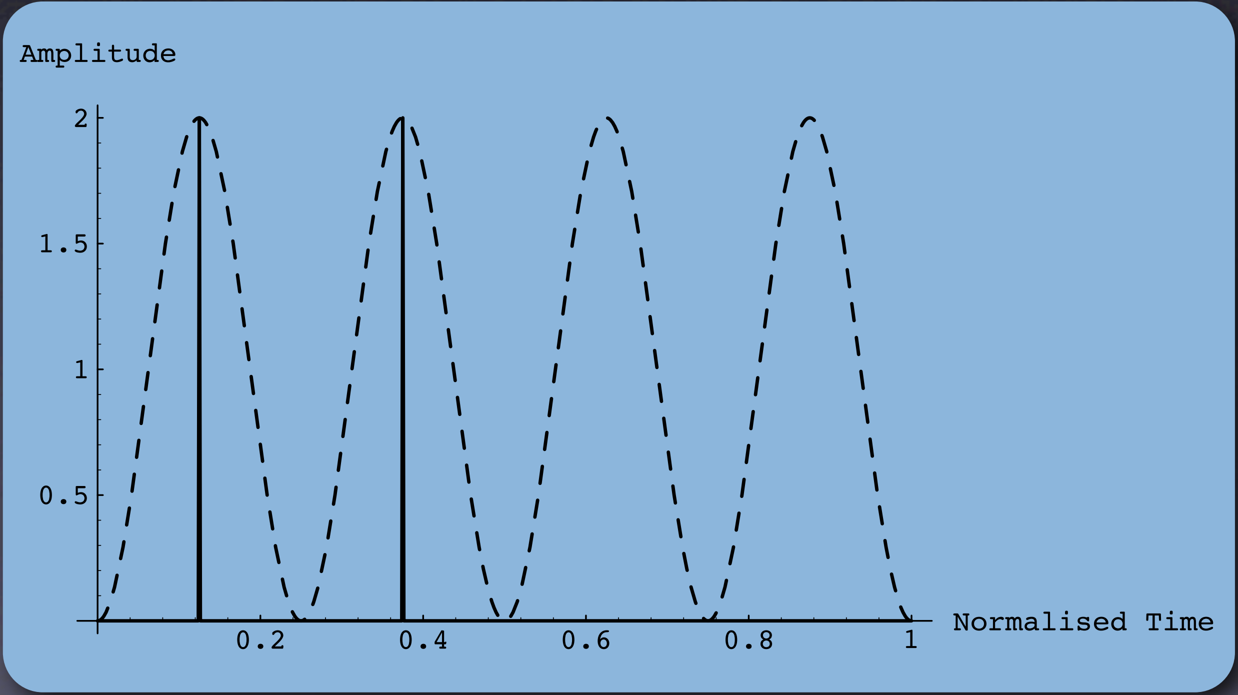
Music Cognition Group

Existing Rhythmic Models

- Parsing metrical grammars (Longuet-Higgins and Lee 1982).
- Forward projection of likelihood (Desain 1992).
- Autocorrelation (Desain & Vos 1990, Brown 1993).
- Oscillator bank entrainment (Toiviainen 1998, Large 1994, Ohya 1994, Miller, Scarborough & Jones 1989).
- Auditory-Motor "Primal Sketch" (Todd 1994, Todd, O'Boyle & Lee 1999) from Sombrero filter banks.

Rhythm as a signal

- **Rhythm models have often implicitly dealt with rhythm as composed of periodic components:**
  - Consider each beat as a *critical sample* of the amplitude envelope, weighted by the peak amplitude.
  - The rhythm analysed is therefore a train of impulses, sampling the rectification of the auditory signal.



Spectrogram (STFT)

Music
Cognition
Group

- The Short Term Fourier Transform has been traditionally used for analysis of time varying signals.
  - Example: Audio analysis...

Spectrogram (STFT)

- The Short Term Fourier Transform has been traditionally used for analysis of time varying signals.
  - Example: Audio analysis...

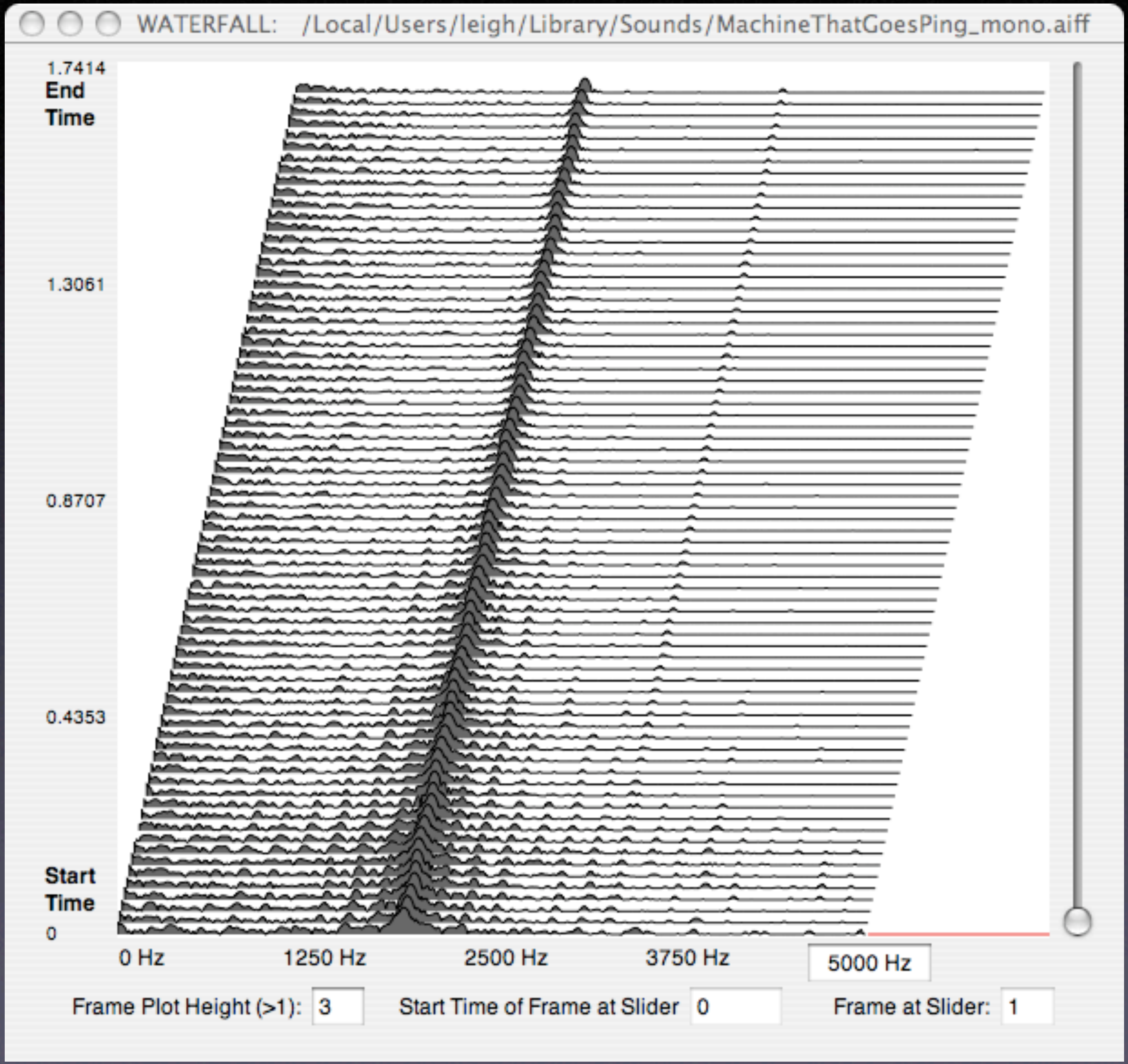
...the
machine that
goes ping...

Spectrogram (STFT)

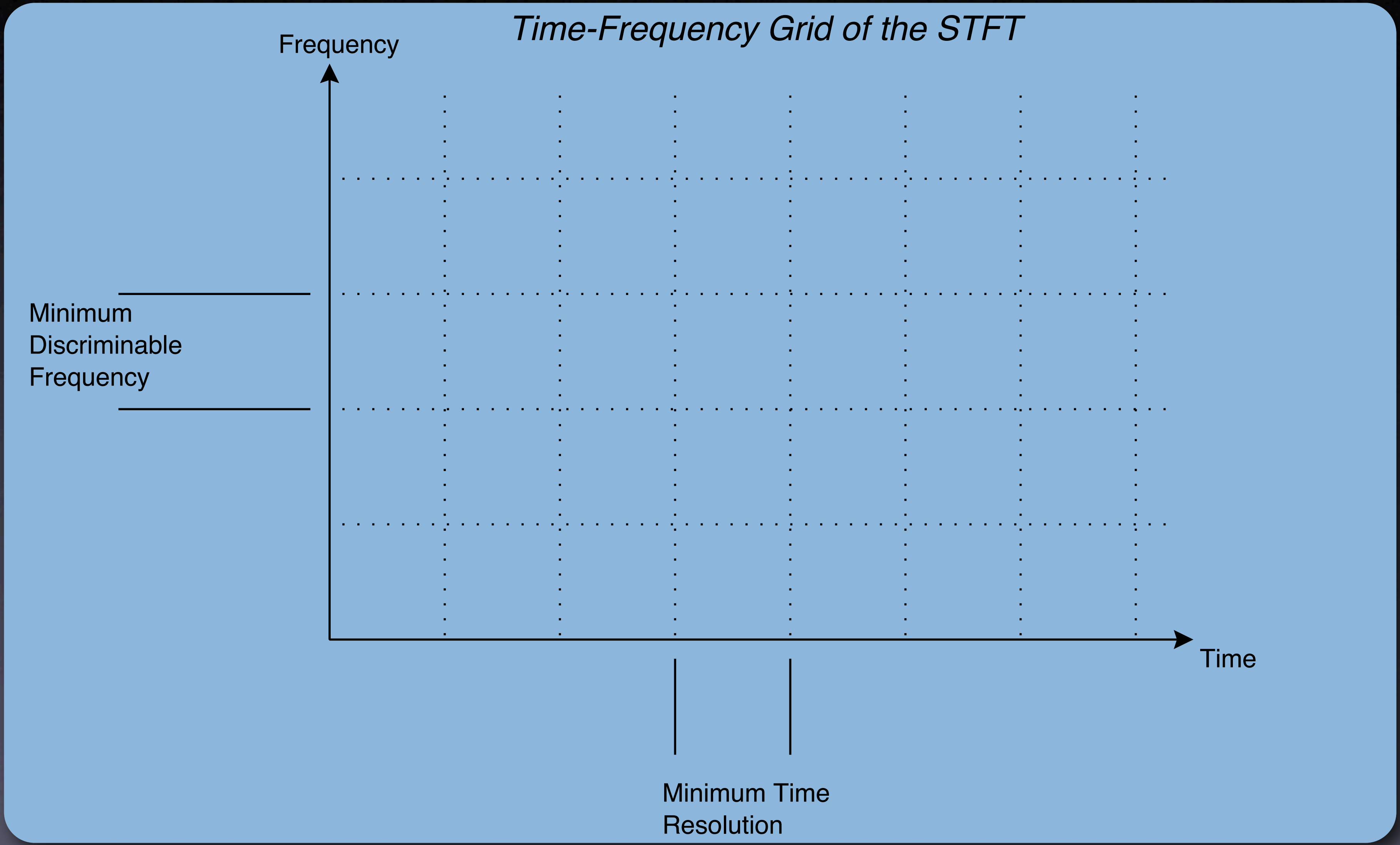
Music
Cognition
Group

Spectrogram (STFT)

Music
Cognition
Group



Short Term Fourier Transform



Wavelet time-frequency analysis

Continuous wavelet transform (CWT) decomposes (**invertibly**) a signal onto scaled and translated instances of a *finite* time "mother function" or "basis".

Wavelet time-frequency analysis

Continuous wavelet transform (CWT) decomposes (**invertibly**) a signal onto scaled and translated instances of a *finite* time "mother function" or "basis".

$$W_s(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau) \cdot \bar{g}(\frac{\tau - b}{a})\, d\tau, \;\; a > 0$$

$$g(t) = e^{-t^2/2} \cdot e^{i\omega_0 t}$$

Wavelet time-frequency analysis

Continuous wavelet transform (CWT) decomposes (**invertibly**) a signal onto scaled and translated instances of a *finite* time "mother function" or "basis".

$$W_s(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau) \cdot \bar{g}(\frac{\tau - b}{a})\, d\tau, \;\; a > 0$$

$$g(t) = e^{-t^2/2} \cdot e^{i\omega_0 t}$$

Wavelet time-frequency analysis

Continuous wavelet transform (CWT) decomposes (**invertibly**) a signal onto scaled and translated instances of a *finite* time "mother function" or "basis".

$$W_s(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau) \cdot \bar{g}(\frac{\tau - b}{a}) d\tau, \;\; a > 0$$

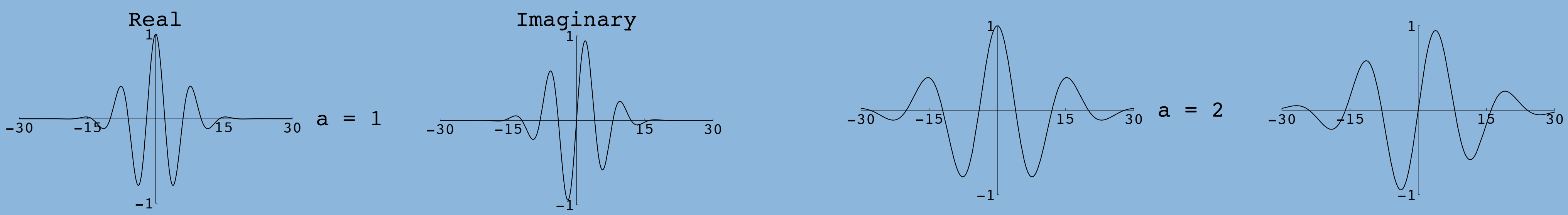
$$g(t) = e^{-t^2/2} \cdot e^{i\omega_0 t}$$

Wavelet time-frequency analysis

Continuous wavelet transform (CWT) decomposes (**invertibly**) a signal onto scaled and translated instances of a *finite* time "mother function" or "basis".

$$W_s(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau) \cdot \bar{g}\left(\frac{\tau - b}{a}\right) d\tau, \quad a > 0$$

$$g(t) = e^{-t^2/2} \cdot e^{i\omega_0 t}$$

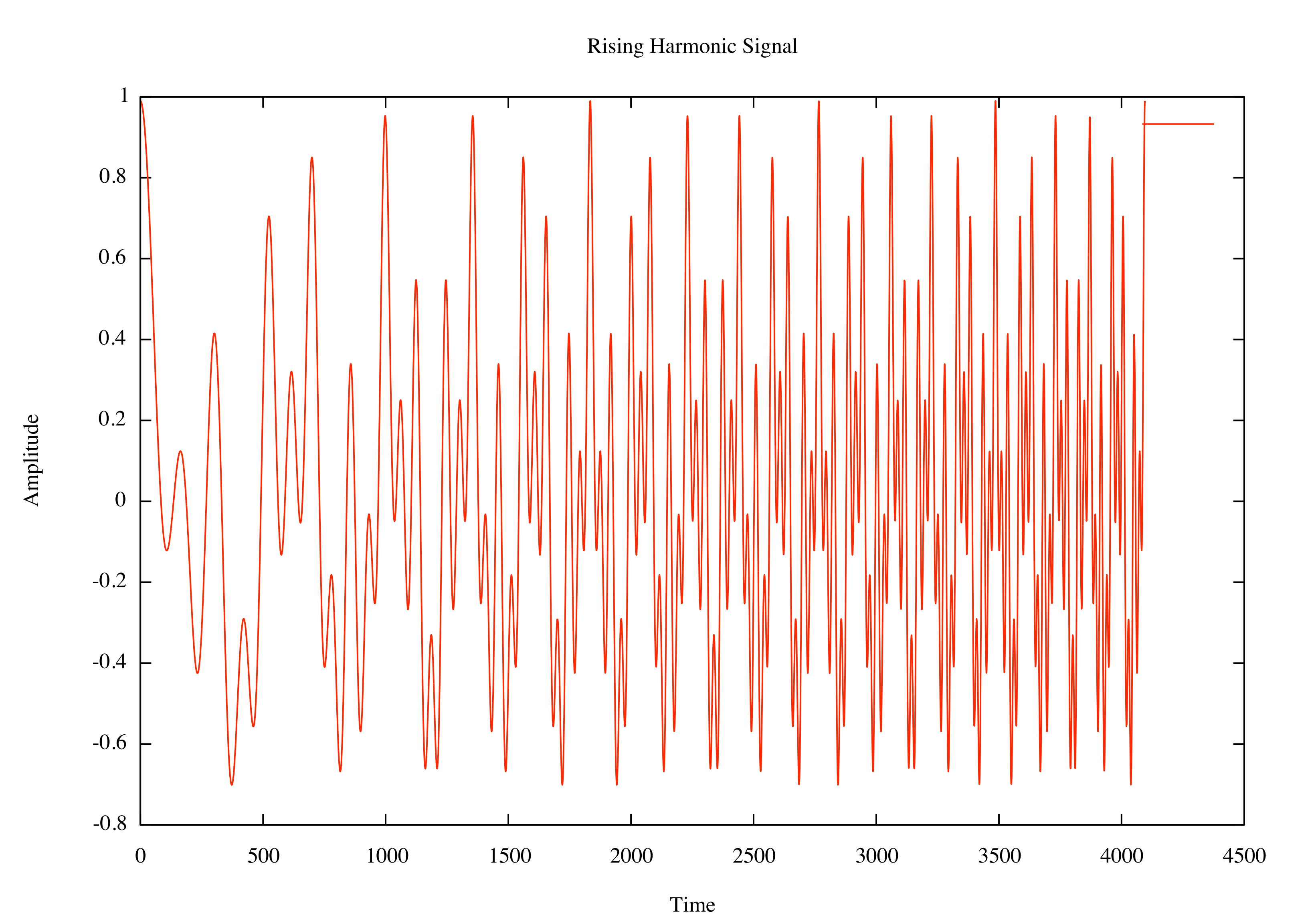


Implementation

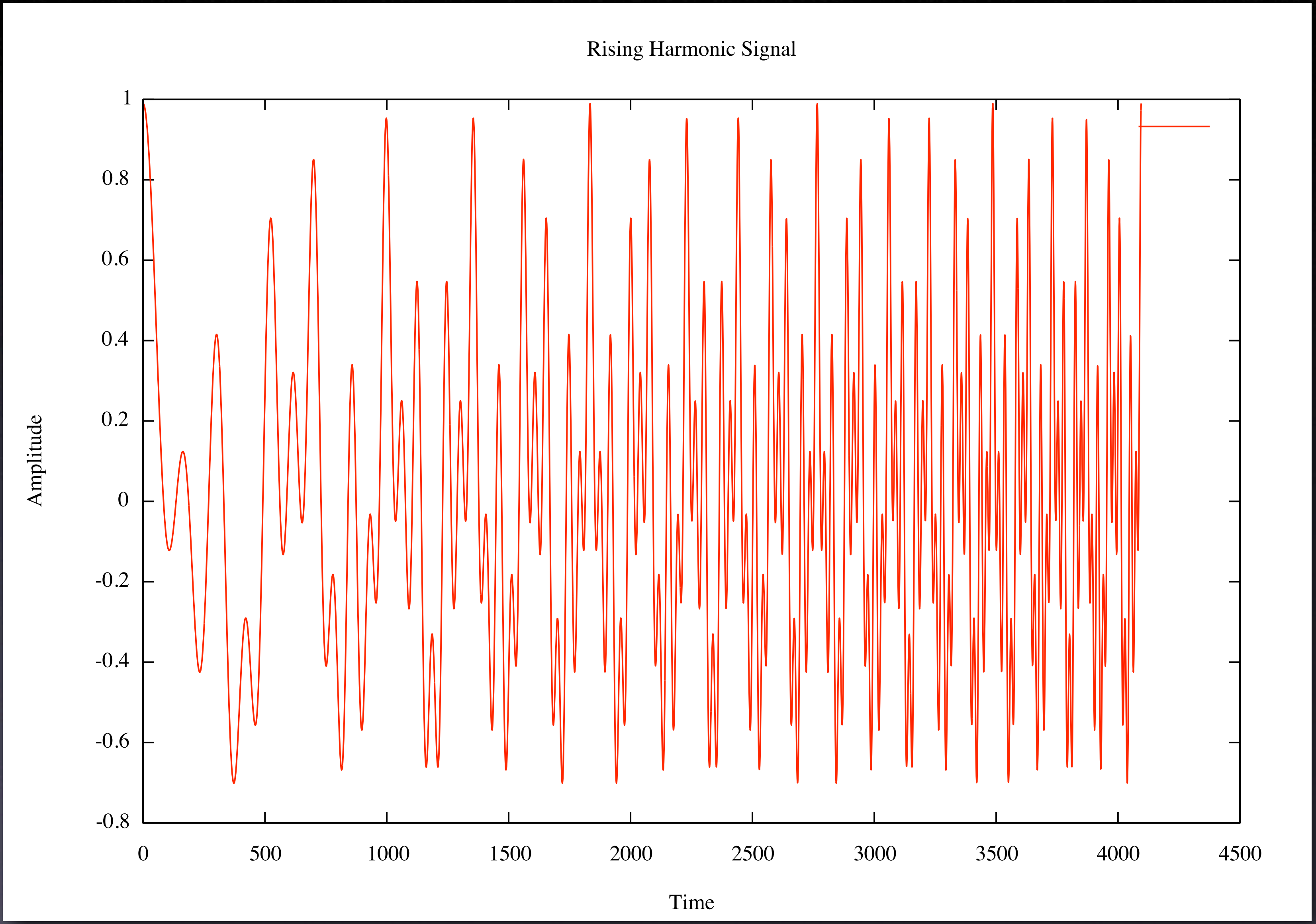
- Implemented as a set of complex value bandpass filters in Fourier domain.
- Scaling produces a “zooming” time window for each frequency “scale”.
- Creates simultaneous time and frequency localisation close to the Heisenberg inequality.

Example: Sinusoidal Signal

Music
Cognition
Group

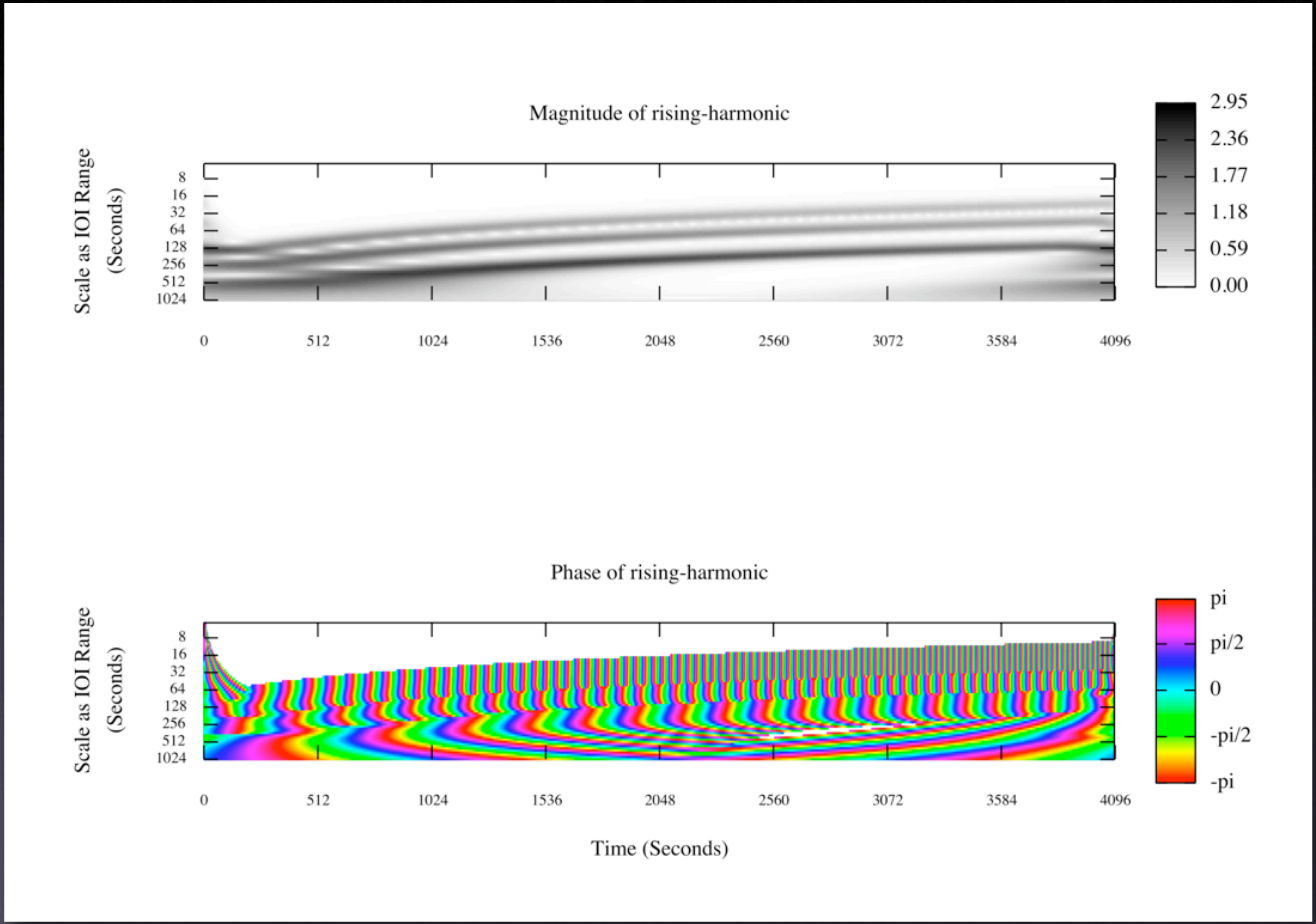


Example: Sinusoidal Signal



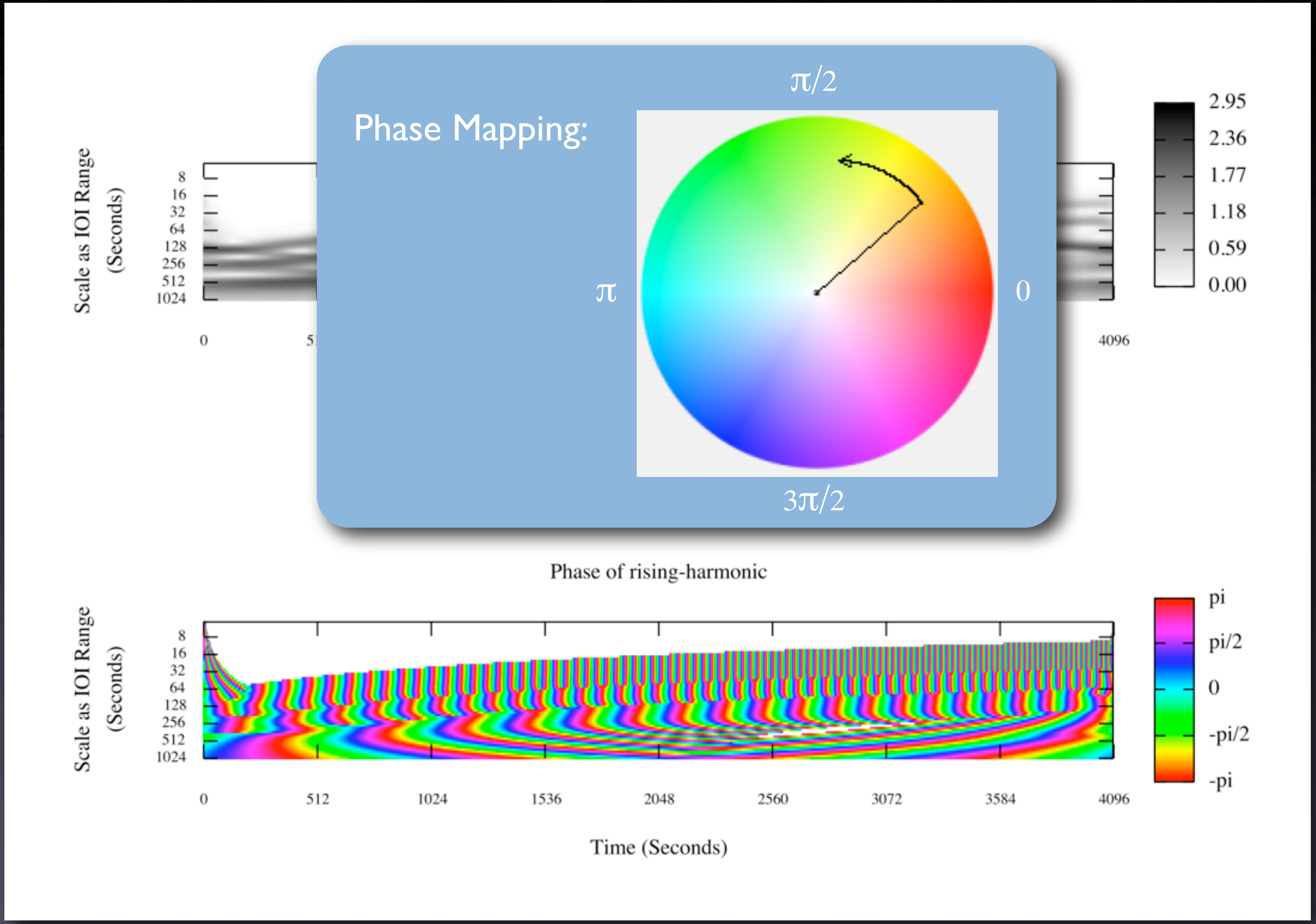
Example: Sinusoidal Signal

Music Cognition Group



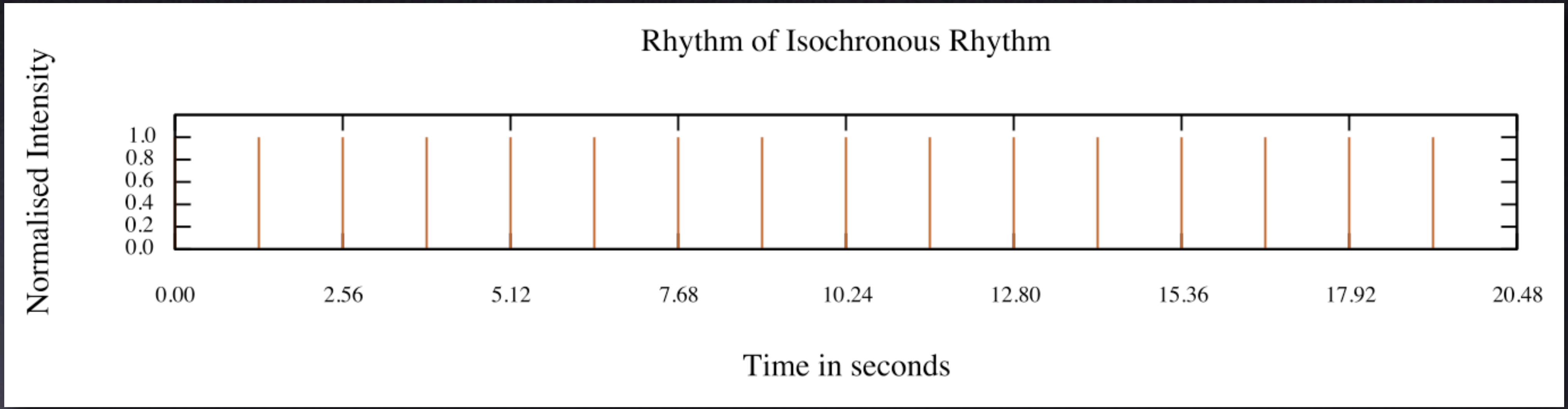
Example: Sinusoidal Signal

Music
Cognition
Group



Example: Simple Rhythm

An isochronous pulse rhythmic signal:



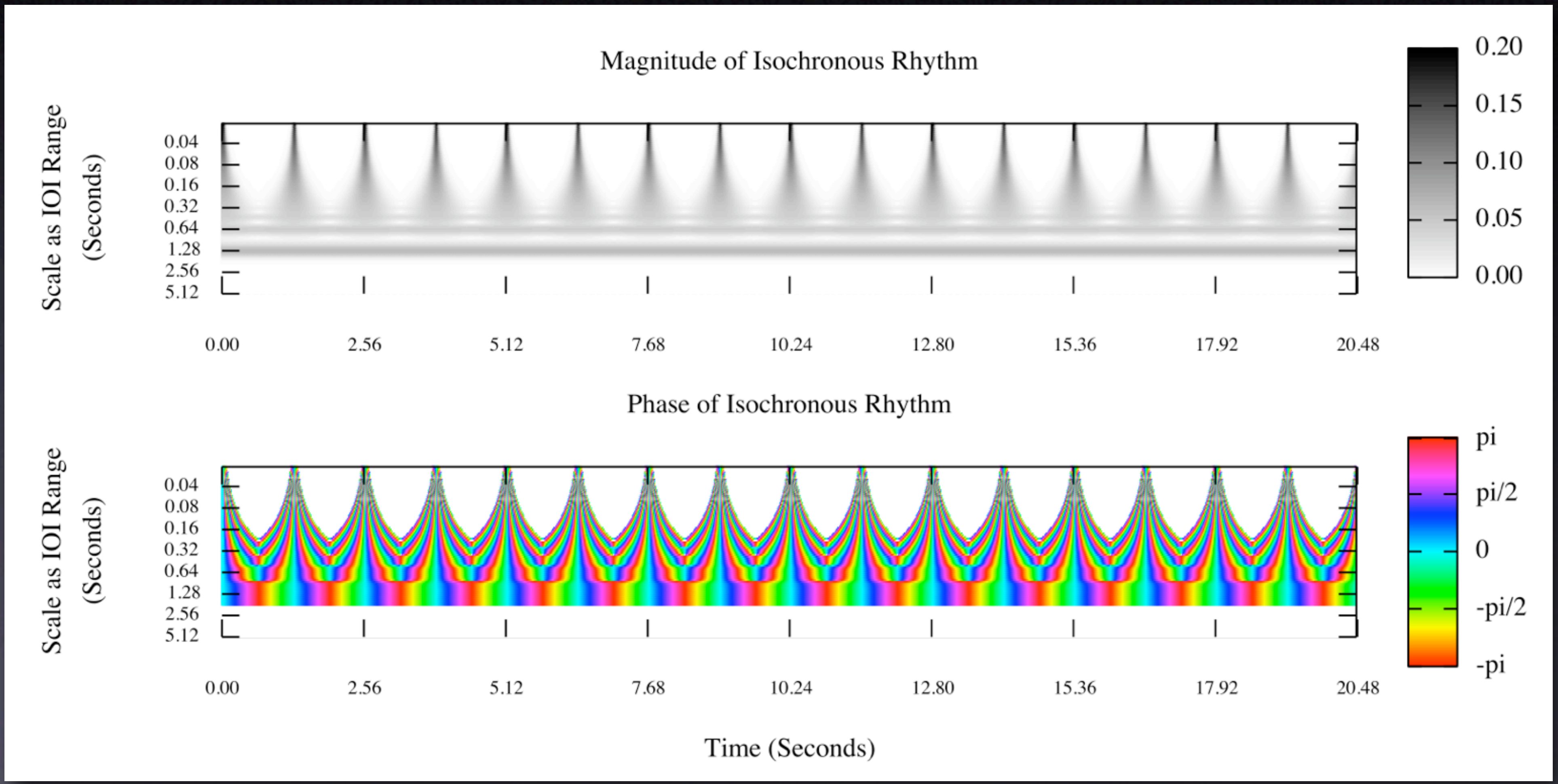
Example: Simple Rhythm

Scalogram and Phasogram of an isochronous pulse rhythmic signal:

Music Cognition Group

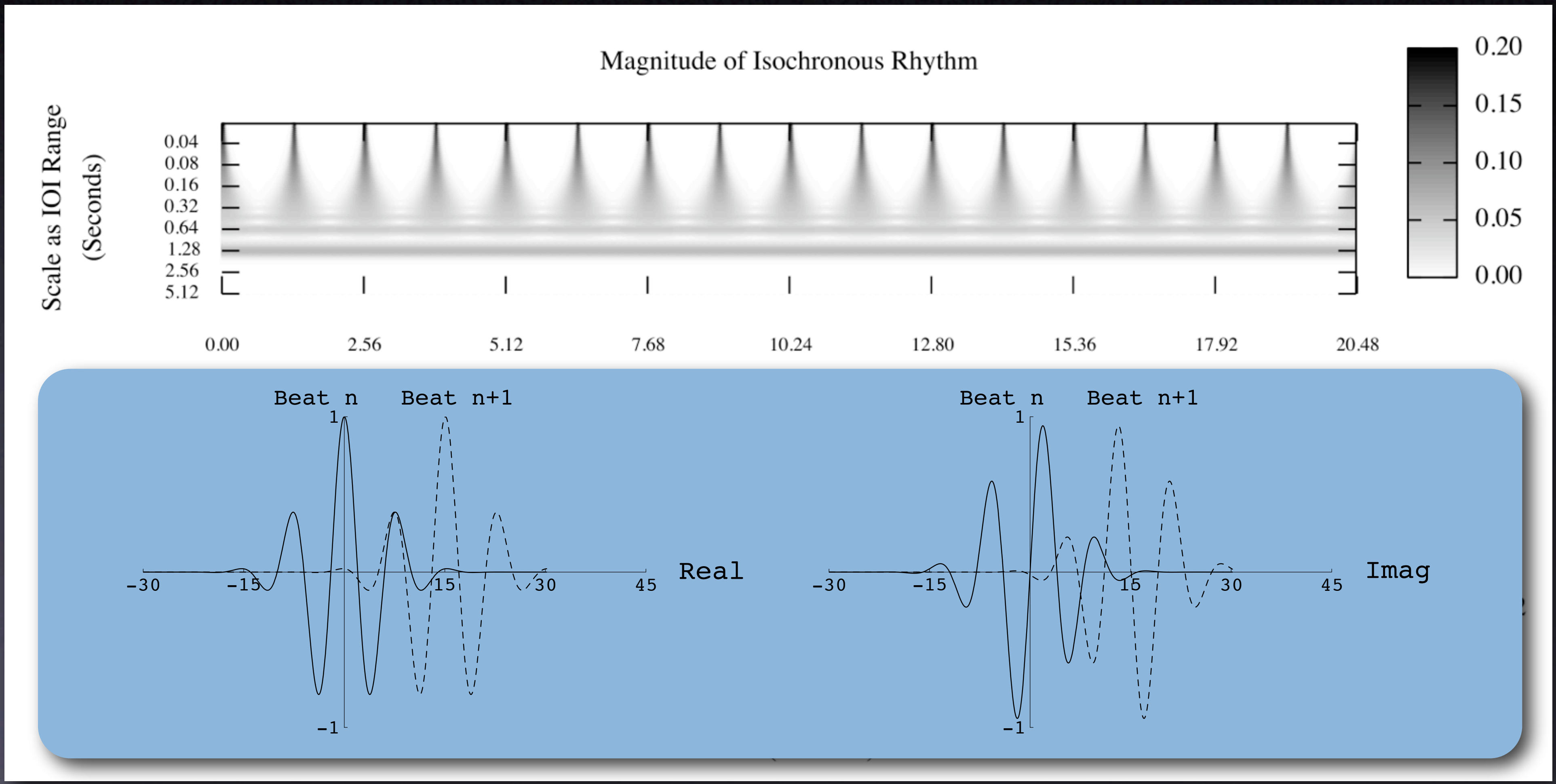
Example: Simple Rhythm

Scalogram and Phasogram of an isochronous pulse rhythmic signal:



Example: Simple Rhythm

Scalogram and Phasogram of an isochronous pulse rhythmic signal:



Music
Cognition
Group

Wavelets for Rhythm

(Smith & Honing in press: Journal of Mathematics & Music 2008)

Music Cognition Group

Wavelets for Rhythm

(Smith & Honing in press: Journal of Mathematics & Music 2008)

- The CWT enables representation of temporal structure in terms of time varying rhythmic frequencies.

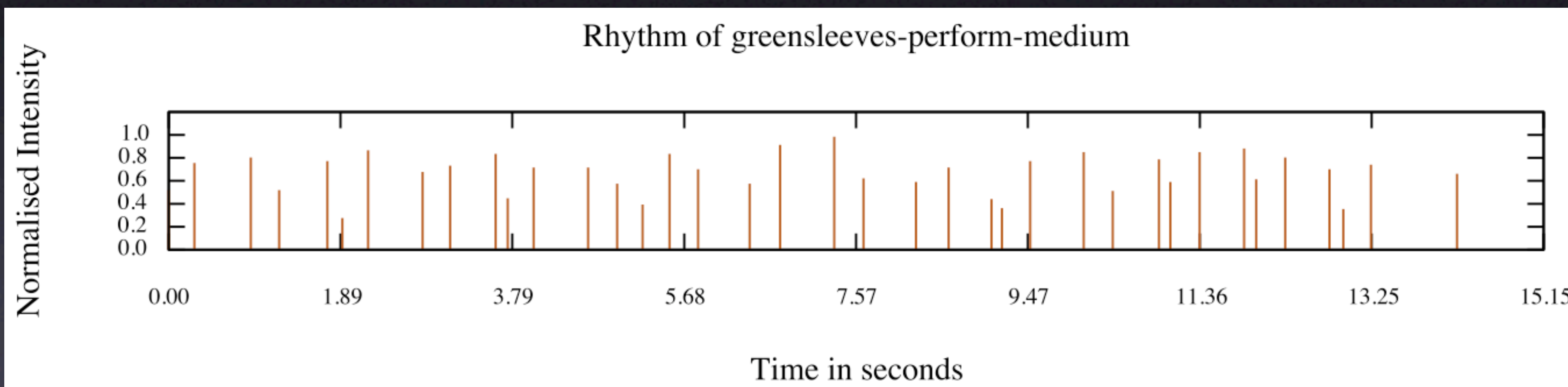
Wavelets for Rhythm

(Smith & Honing in press: Journal of Mathematics & Music 2008)

- The CWT enables representation of temporal structure in terms of time varying rhythmic frequencies.
- Produces magnitude and phase measures which reveal time-frequency *ridges* indicating the frequencies present in the input rhythm signal (collectively a *skeleton,* Tchamitchian & Torrésani '92).

Musical Example

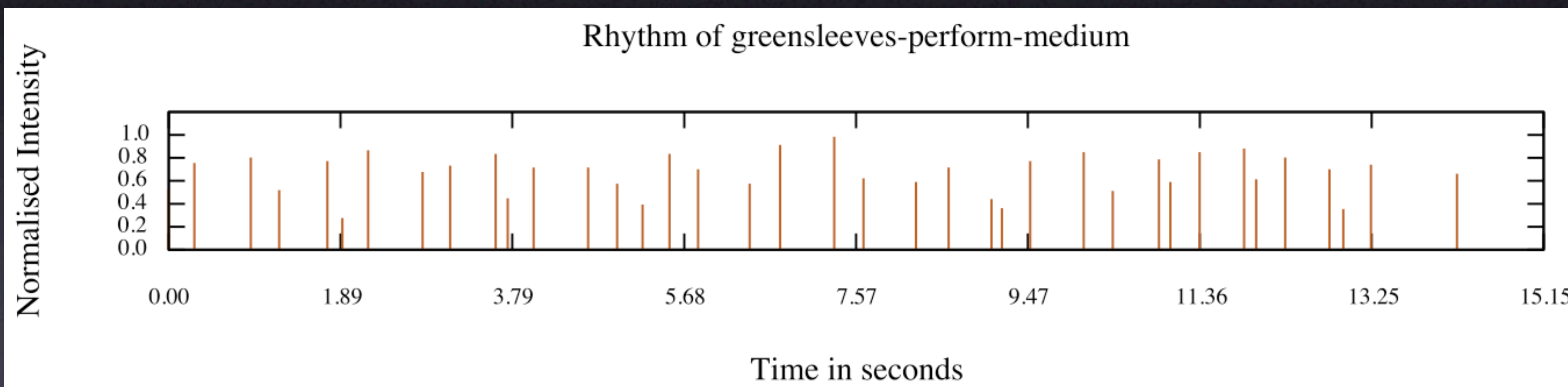
- The rhythm of "Greensleeves"...



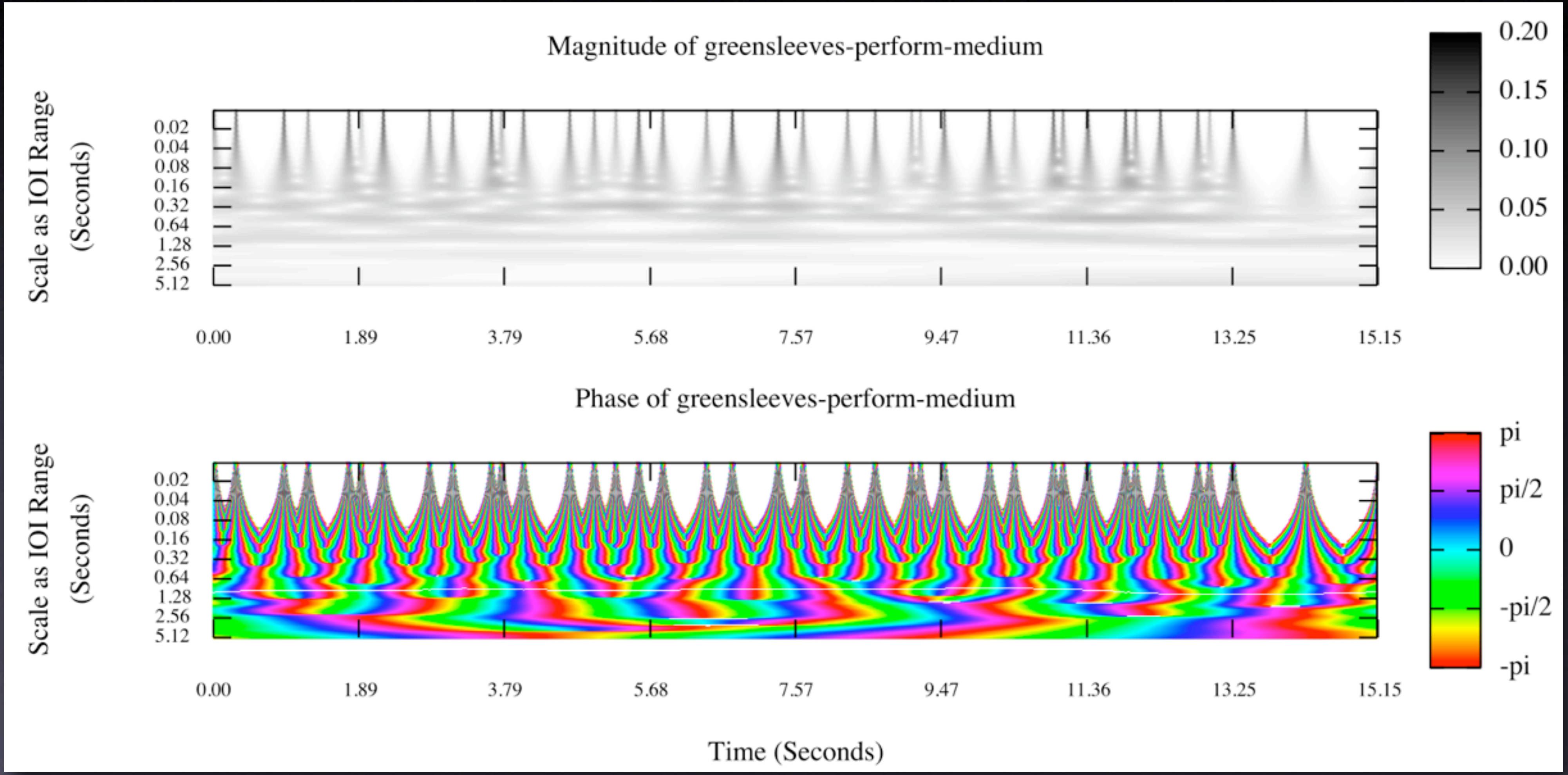
Music
Cognition
Group

Musical Example

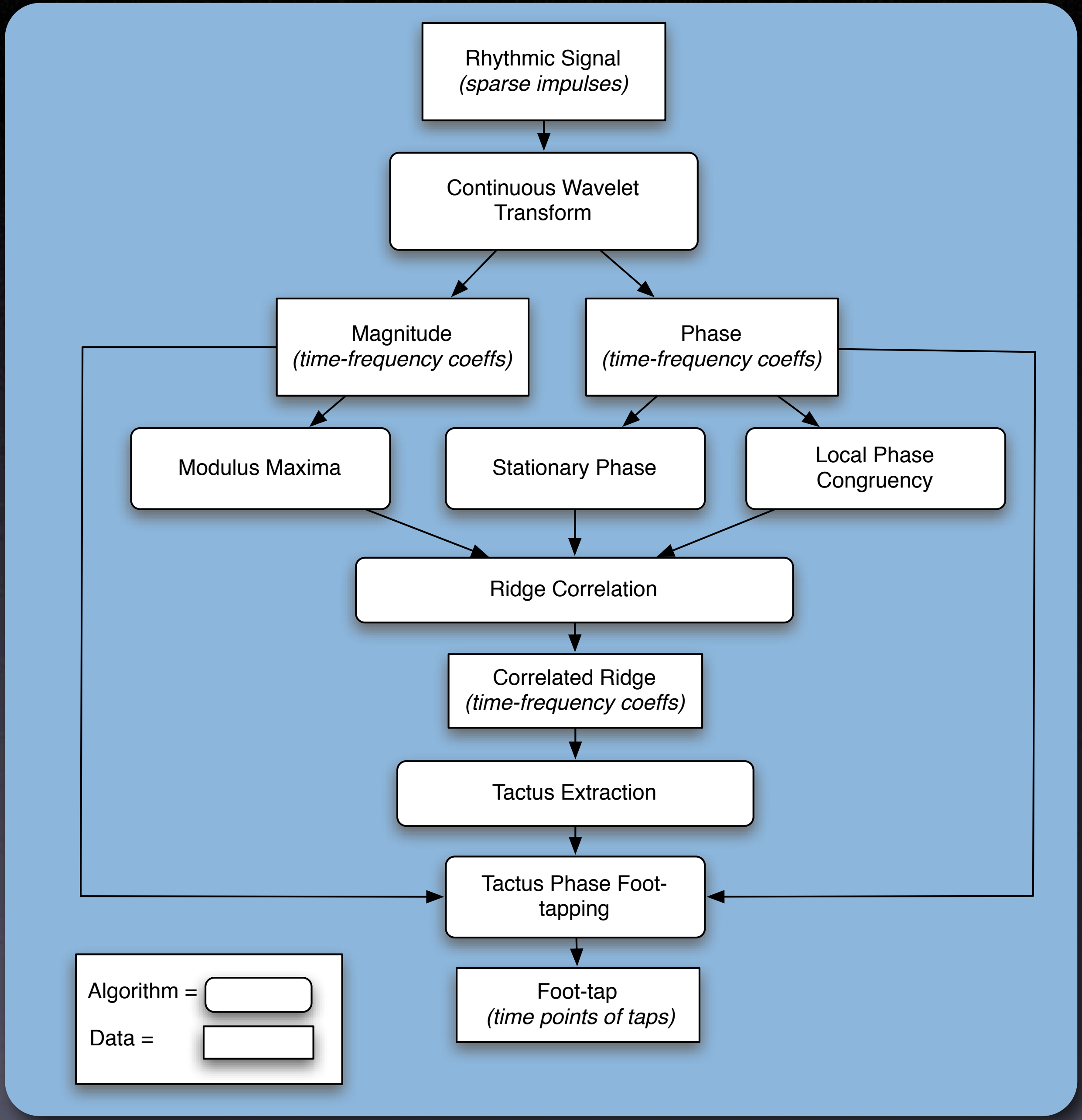
- The rhythm of "Greensleeves"...



Greensleeves



System Overview



Application:
Foot-tapping by reconstruction

Music
Cognition
Group

Application: Foot-tapping by reconstruction

- Suppress all but the magnitude coefficients of the extracted tactus ridge.

Music Cognition Group

Application: Foot-tapping by reconstruction

- Suppress all but the magnitude coefficients of the extracted tactus ridge.
- Invert the reduced magnitude and original phase planes back to the time domain.

Application: Foot-tapping by reconstruction

- Suppress all but the magnitude coefficients of the extracted tactus ridge.
- Invert the reduced magnitude and original phase planes back to the time domain.
- Produces a sinusoidal AM signal with an intact phase, and a period matching the foot-tap interval.

Application:
Foot-tapping by reconstruction

- Suppress all but the magnitude coefficients of the extracted tactus ridge.
- Invert the reduced magnitude and original phase planes back to the time domain.
- Produces a sinusoidal AM signal with an intact phase, and a period matching the foot-tap interval.
- Nominating a starting beat and noting its phase, all other foot-taps are generated for the same phase value.

Tapping to Greensleeves

- The rhythm of "Greensleeves" with computed foot-tap...

Tapping to Greensleeves

- The rhythm of "Greensleeves" with computed foot-tap...

Interpretation

Music
Cognition
Group

Interpretation

- CWT being an invertible transform, simply represents rhythm in the time-frequency domain.

Interpretation

- CWT being an invertible transform, simply represents rhythm in the time-frequency domain.
- Has no explicit model of rhythmic cognition $\Rightarrow$
  - Indicates how much structure is in the rhythmic signal.

Music Cognition Group

Interpretation

- CWT being an invertible transform, simply represents rhythm in the time-frequency domain.
- Has no explicit model of rhythmic cognition $\Rightarrow$
  - Indicates how much structure is in the rhythmic signal.
- Metrical durations from CWT suggests that rhythmic strata (ridges) may act as (bottom-up) cues to a metrical interpretation.

Interpretation

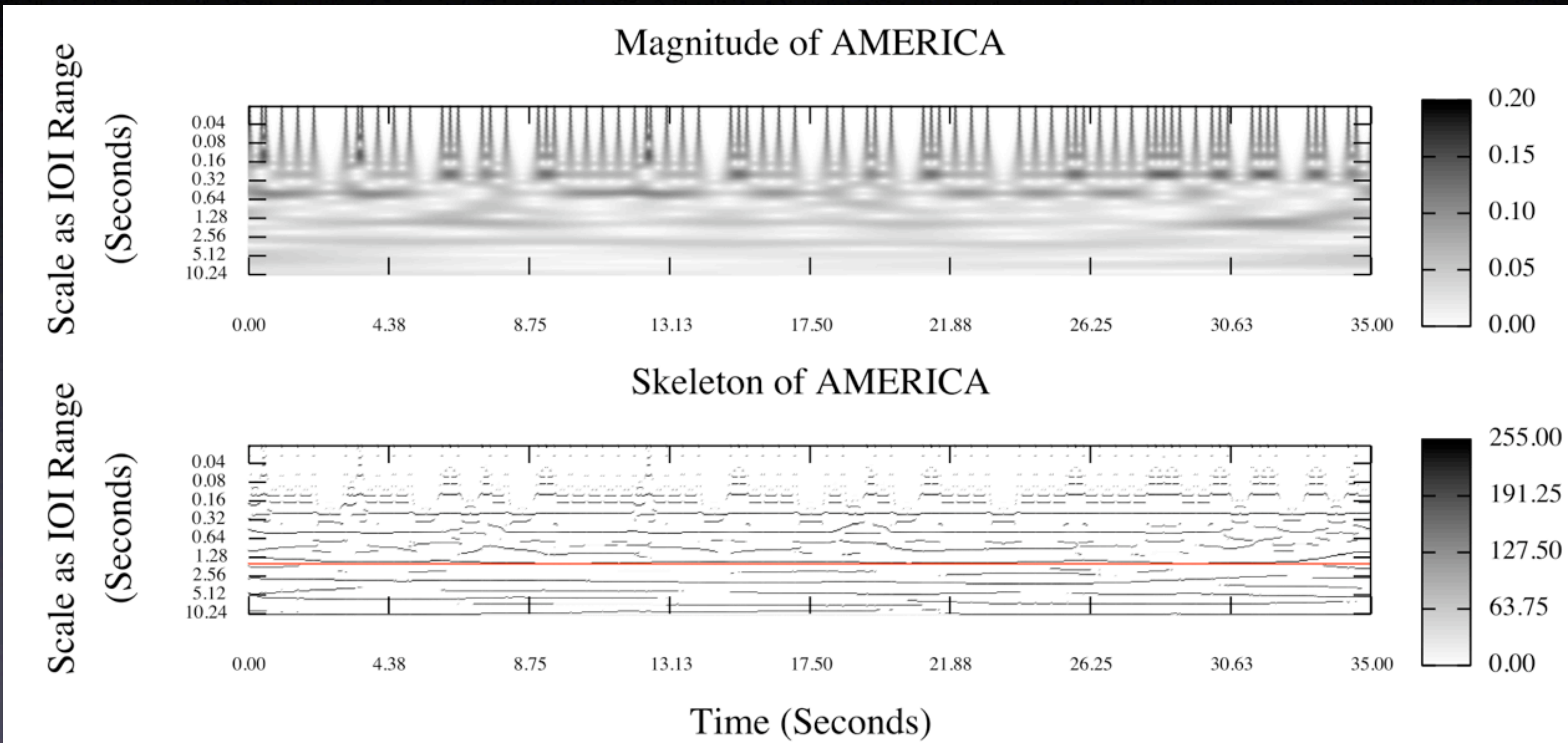
- CWT being an invertible transform, simply represents rhythm in the time-frequency domain.
- Has no explicit model of rhythmic cognition $\Rightarrow$
  - Indicates how much structure is in the rhythmic signal.
- Metrical durations from CWT suggests that rhythmic strata (ridges) may act as (bottom-up) cues to a metrical interpretation.
- Establishes a distinction between top-down expectation and bottom-up categorisation processes.

Evaluation Data: Anthems

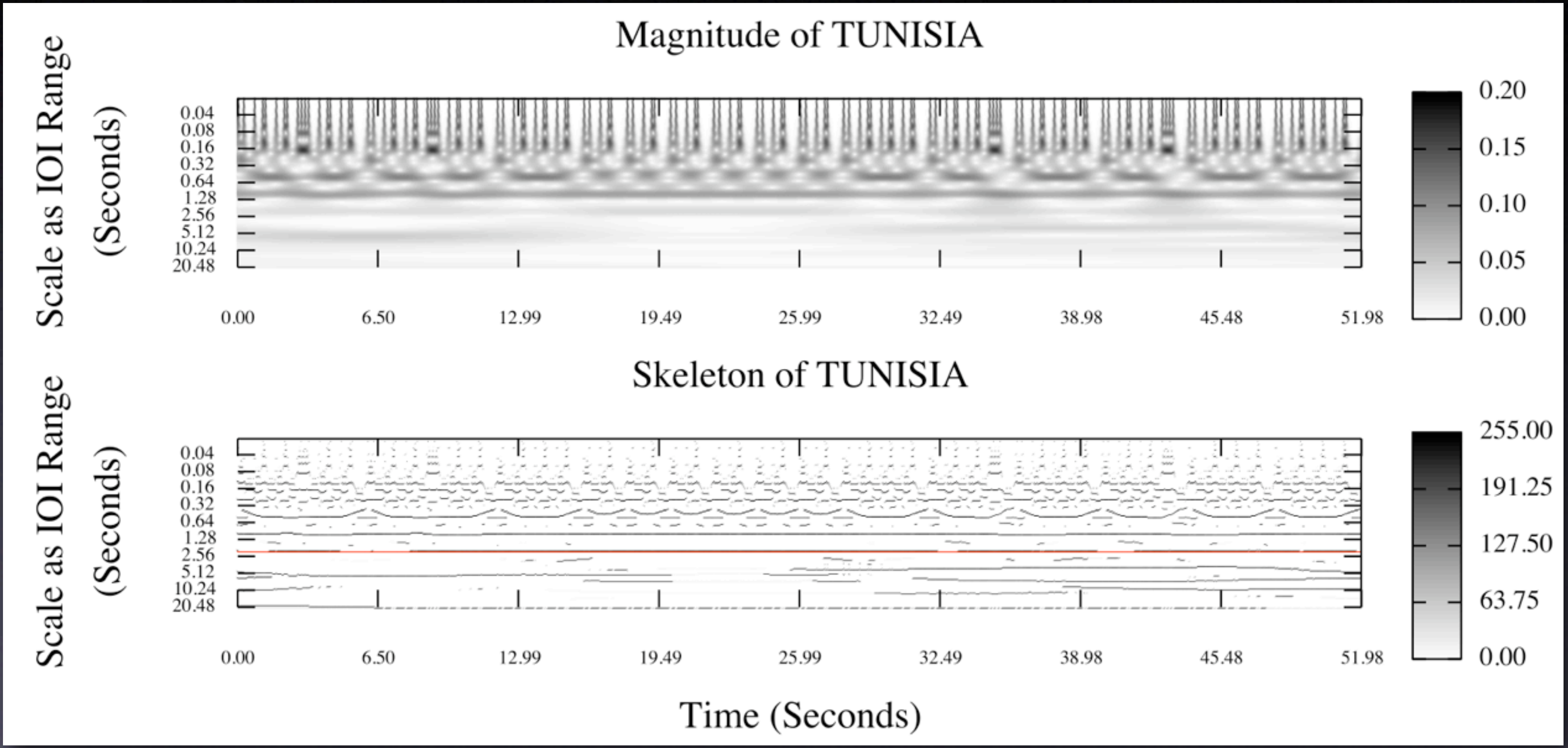
(Smith & Honing: ICoMCS 2007)

- 105 National Anthems (Shaw & Coleman 1960).
- Rhythms transcribed into interonset intervals (IOI), quarter-note & bar duration, anacrusis.
- No melodic, intensity or expression accents.
- Also used in analysis of rule-based systems of Longuet-Higgins & Lee (1982, 1985, 1991) (Desain & Honing 1999).
- Limited to a maximum length of 82 seconds each.

Example: America



Example: Tunisia



Music
Cognition
Group

Ridge Presence

- Ridge Presence: relative occurrence of a ridge (*r*) at each dilation scale (*a*), over the duration (*B*) of each rhythm.

Ridge Presence

- Ridge Presence: relative occurrence of a ridge (*r*) at each dilation scale (*a*), over the duration (*B*) of each rhythm.

$$P_a = \sum_{b=0}^{B-1} \frac{r(W_{b,a})}{B}$$

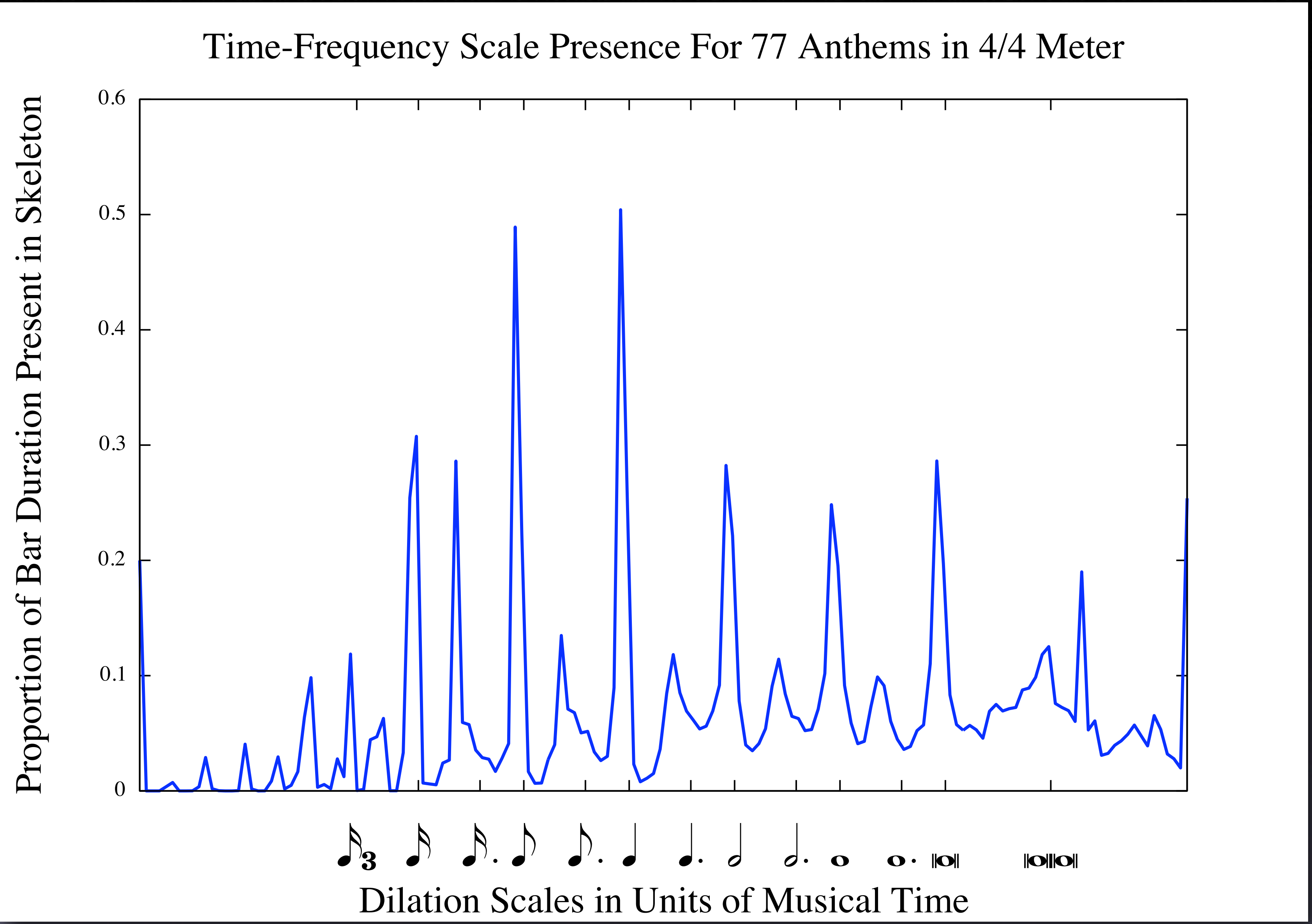
Ridge Presence

- Ridge Presence: relative occurrence of a ridge (*r*) at each dilation scale (*a*), over the duration (*B*) of each rhythm.

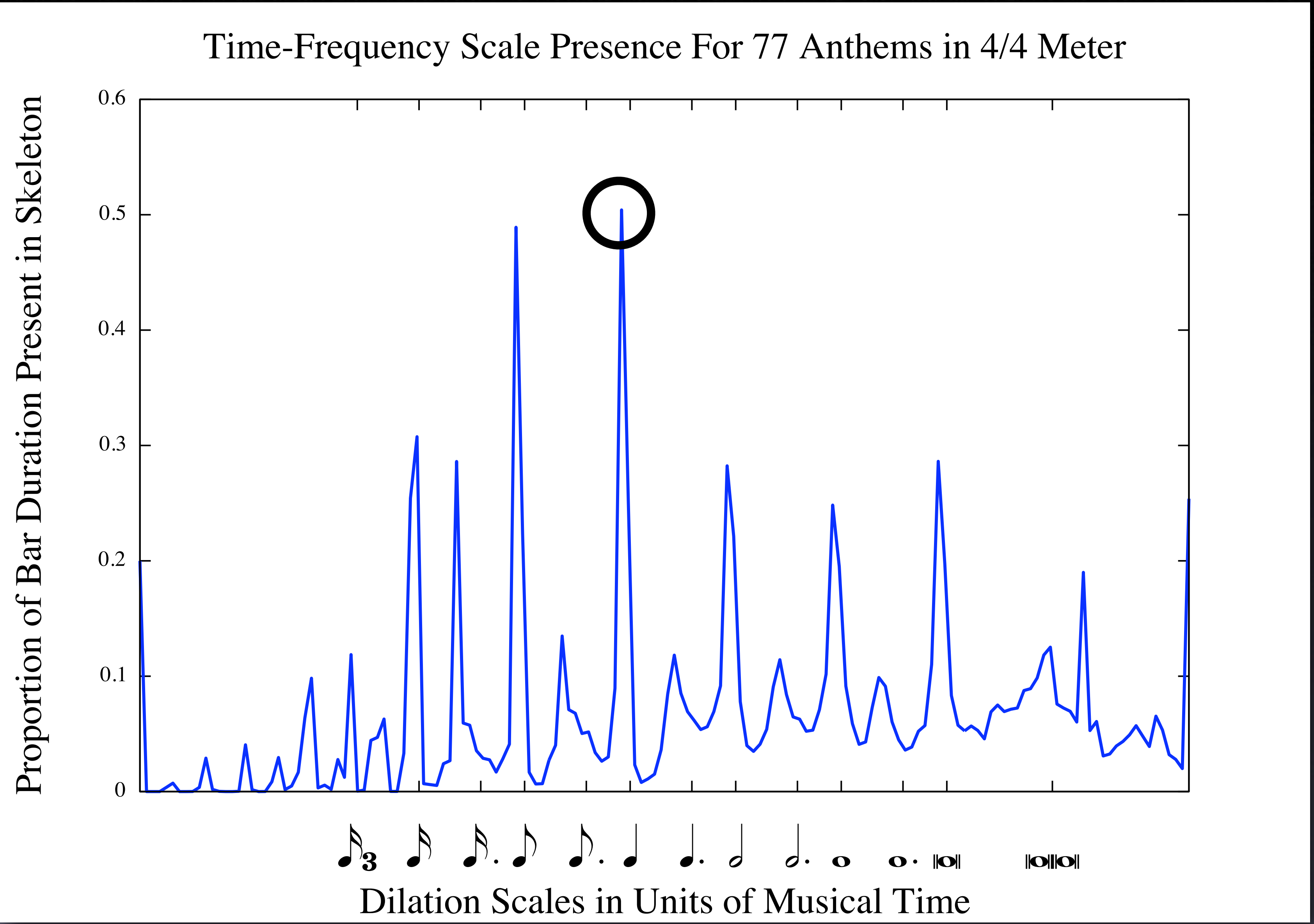
$$P_a = \sum_{b=0}^{B-1} \frac{r(W_{b,a})}{B}$$

- Average Ridge Presence: relative frequency of occurrence of each ridge averaged across *all* rhythms of a given meter.

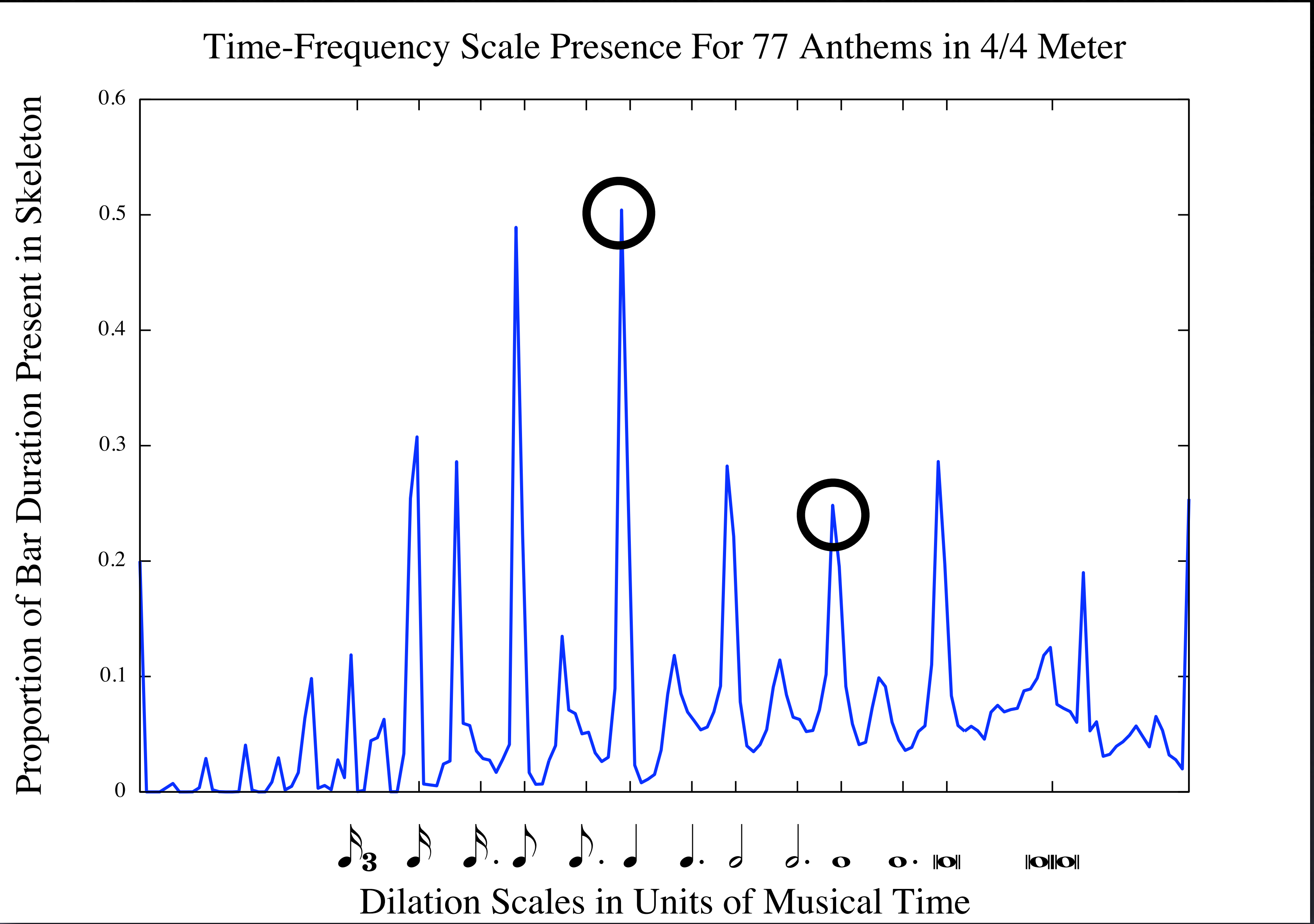
Average Ridge Presence



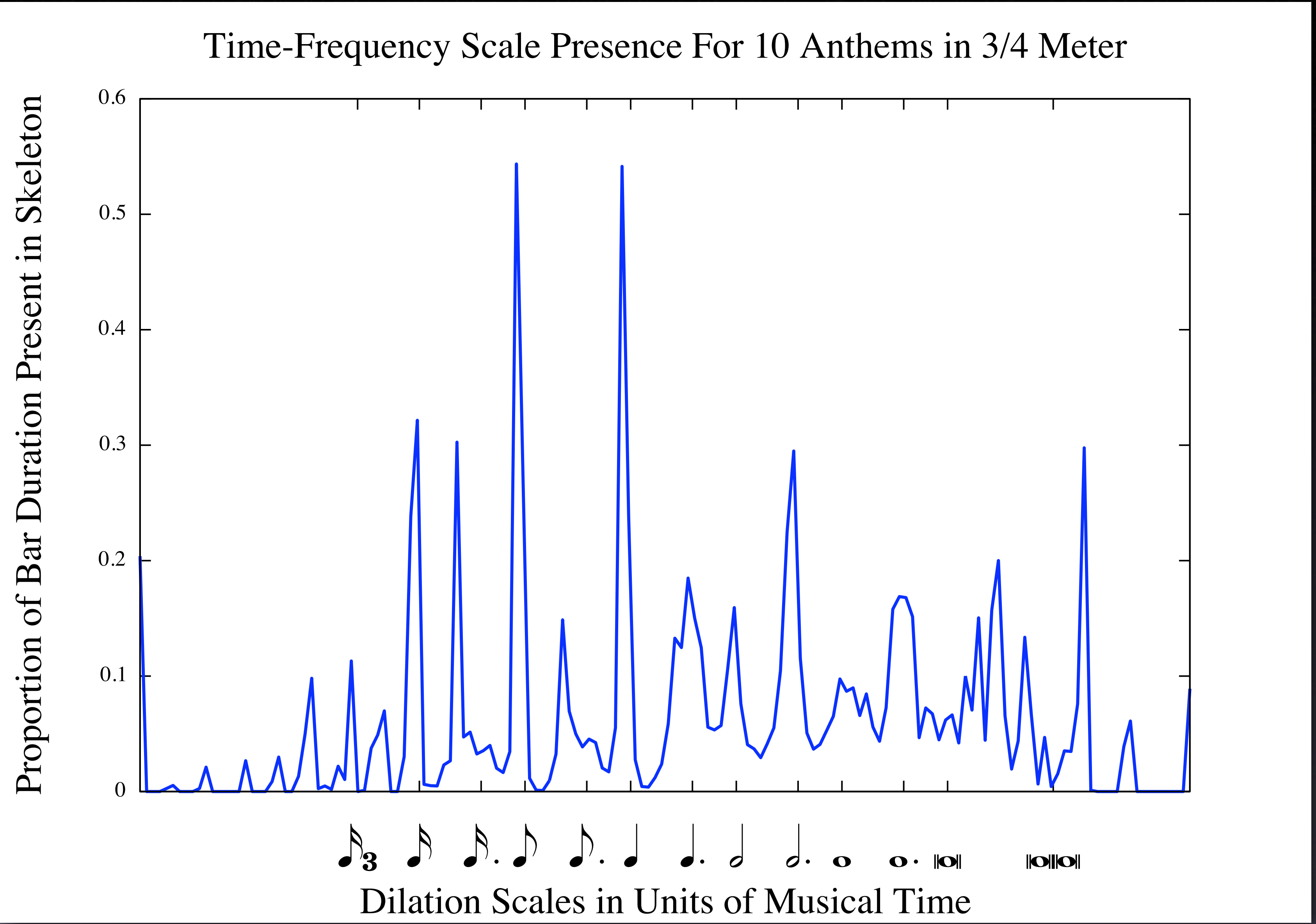
Average Ridge Presence



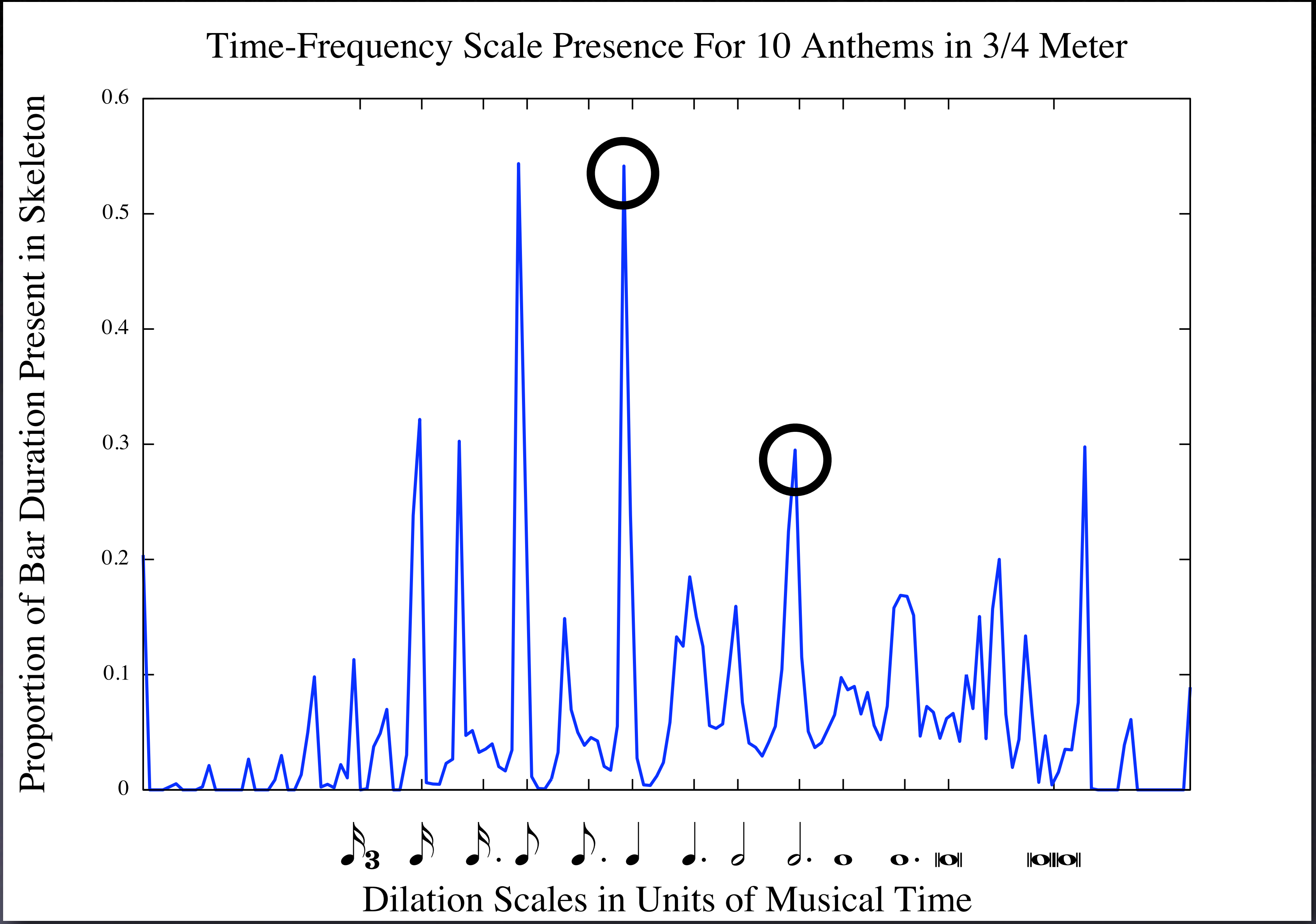
Average Ridge Presence



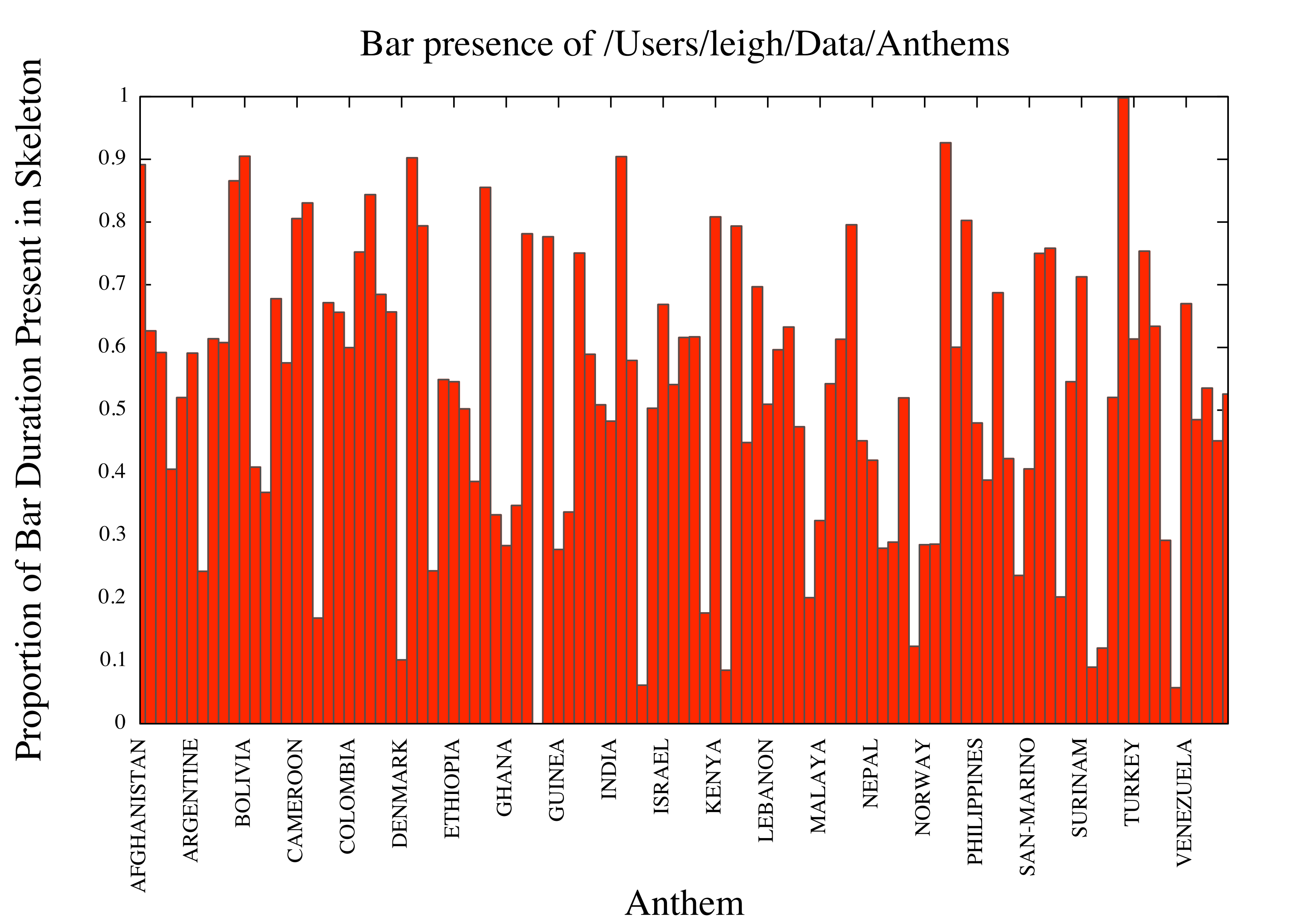
Triple Meter



Triple Meter



Presence of Bar Ridges



Music
Cognition
Group

Evaluation

Evaluation

- Decomposing the temporal structure of musical rhythms with CWT reveals durations of the notated beat and bar.

Music
Cognition
Group

Evaluation

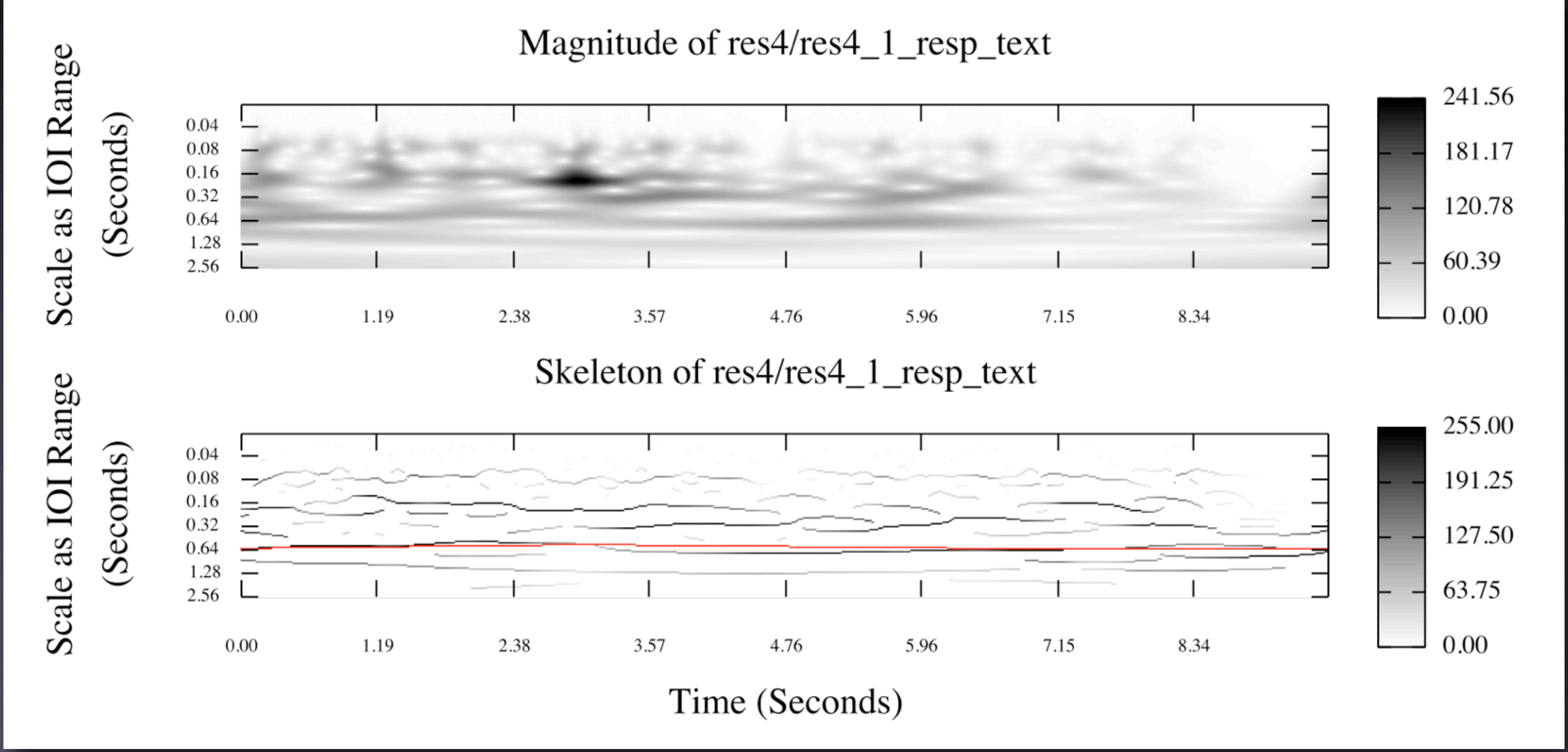
- Decomposing the temporal structure of musical rhythms with CWT reveals durations of the notated beat and bar.
- Stable over anthem database, exceptions probably due to lack of harmonic/melodic disambiguation.

Evaluation

- Decomposing the temporal structure of musical rhythms with CWT reveals durations of the notated beat and bar.
- Stable over anthem database, exceptions probably due to lack of harmonic/melodic disambiguation.
- Not simply statistical (only 33 anthems have any interonset-intervals of bar duration).

Memory Based Tactus

- Wavelet rhythm analysis is also applicable to continuous onset salience traces from auditory models (Coath et. al, to appear: Connection Science 2008).



Memory Based Tactus

Music
Cognition
Group

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.

Music
Cognition
Group

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.
- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute) using memory based derivation of tactus:

Music
Cognition
Group

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.
- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute) using memory based derivation of tactus:
- Example 1:

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.
- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute) using memory based derivation of tactus:
- Example 1: Original...

Memory Based Tactus

- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.
- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute) using memory based derivation of tactus:
- Example 1: Original... ...Original + Accompaniment

Music
Cognition
Group

Memory Based Tactus

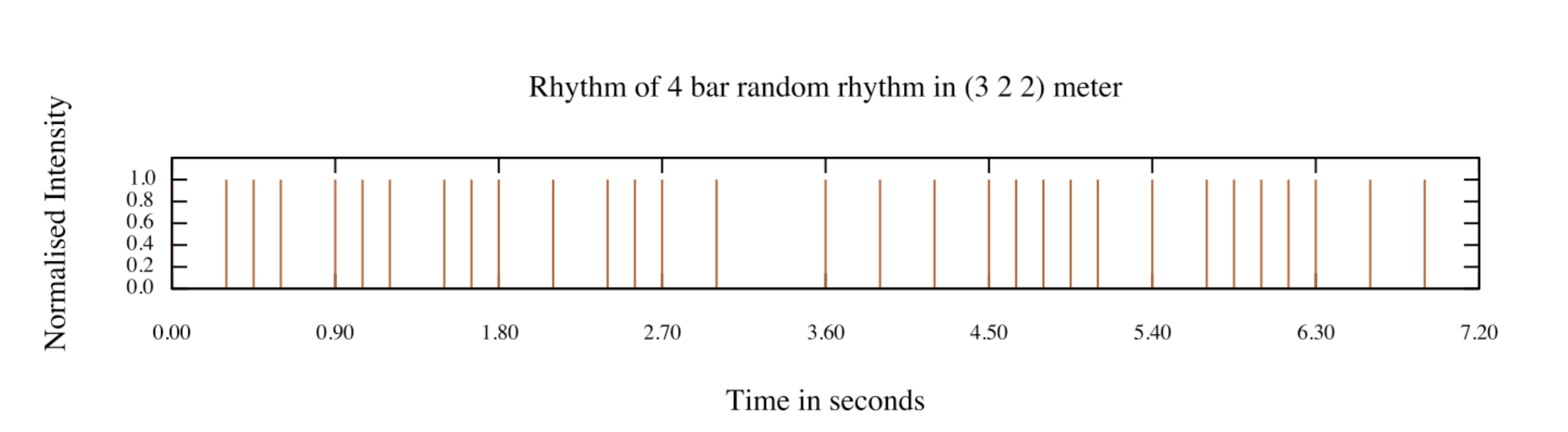
- Uses lossy windowed integrator to amass tactus likelihood.
- Invert the computed tactus and original phase plane back to the time domain. Creates single beat oscillation.
- Singing examples of Dutch folk songs from the "Onder de Groene Linde" collection (Meertens Institute) using memory based derivation of tactus:
- Example 1: Original... ...Original + Accompaniment
- Example 2: ...Original + Accompaniment

Expectation

- Generates future expectation times given a performed rhythm.
- Uses lossy windowed integrator to amass likelihood of projected time periods.
- Weighted by absolute tempo constraints.
- Uses CWT phase measures to correct the projected periods for phase at the edge of the time window.

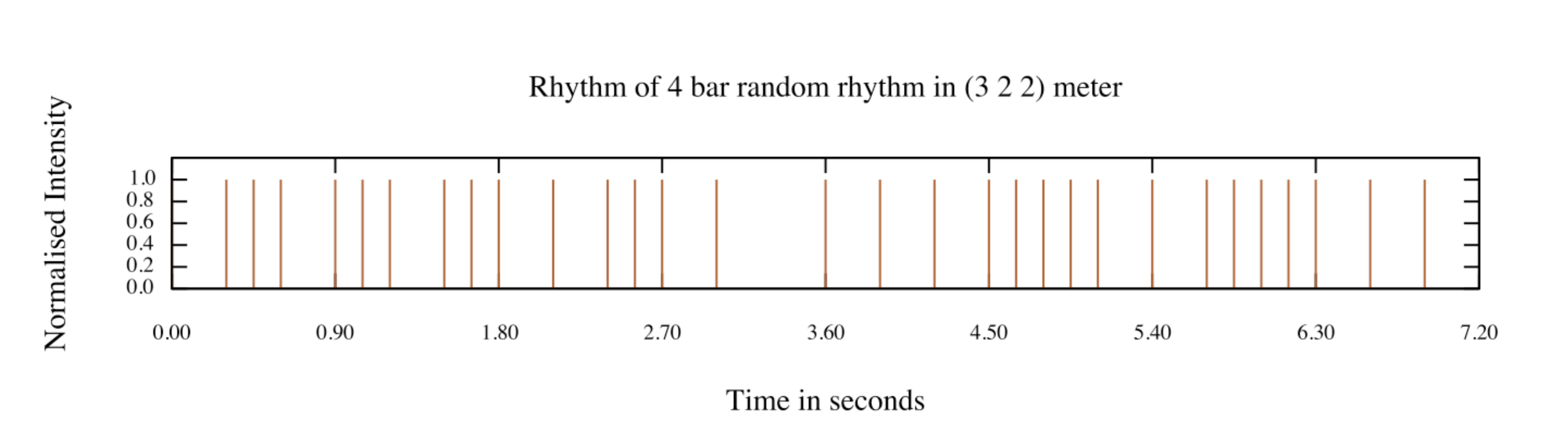
Emerging Metrical Context

- Example 3/4 rhythm (no accents)

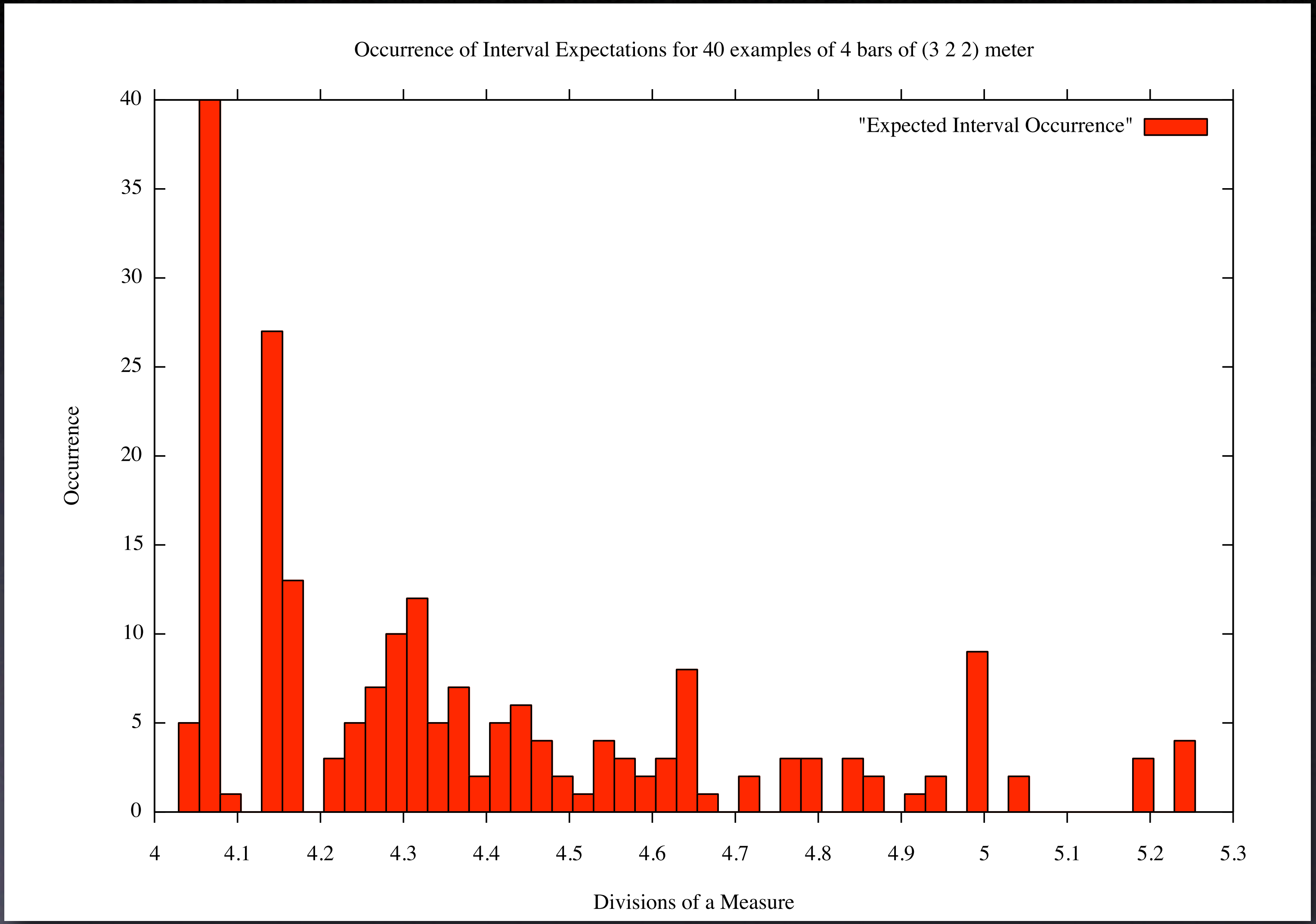


Emerging Metrical Context

- Example 3/4 rhythm (no accents)



Emerging Metrical Context



Further Work

Further Work

- Use of rhythmic phase that is available from the CWT to identify an anacrusis (upbeat).
- Compare performance against larger datasets (e.g MIREX).
- Derivation of causal multiresolution model combined with memory store for retrospection.

http://www.hum.uva.nl/mmm

http://www.science.uva.nl/~lsmith