CrossMark

Temporal Scale Selection in Time-Causal Scale Space

Tony Lindeberg[1]

Abstract When designing and developing scale selection mechanisms for generating hypotheses about characteristic scales in signals, it is essential that the selected scale levels reflect the extent of the underlying structures in the signal. This paper presents a theory and in-depth theoretical analysis about the scale selection properties of methods for automatically selecting local temporal scales in time-dependent signals based on local extrema over temporal scales of scale-normalized temporal derivative responses. Specifically, this paper develops a novel theoretical framework for performing such temporal scale selection over a time-causal and time-recursive temporal domain as is necessary when processing continuous video or audio streams in real time or when modelling biological perception. For a recently developed time-causal and time-recursive scale-space concept defined by convolution with a scale-invariant limit kernel, we show that it is possible to transfer a large number of the desirable scale selection properties that hold for the Gaussian scale-space concept over a non-causal temporal domain to this temporal scale-space concept over a truly time-causal domain. Specifically, we show that for this temporal scale-space concept, it is possible to achieve true temporal scale invariance although the temporal scale levels have to be discrete, which is a novel theoretical construction. The analysis starts from a detailed comparison of different temporal scale-space concepts and their relative advantages and disadvantages, leading the focus to a class of recently extended time-causal and time-recursive temporal scale-space concepts based on first-order integrators or equivalently truncated exponential kernels coupled in cascade. Specifically, by the discrete nature of the temporal scale levels in this class of time-causal scale-space concepts, we study two special cases of distributing the intermediate temporal scale levels, by using either a uniform distribution in terms of the variance of the composed temporal scale-space kernel or a logarithmic distribution. In the case of a uniform distribution of the temporal scale levels, we show that scale selection based on local extrema of scale-normalized derivatives over temporal scales makes it possible to estimate the temporal duration of sparse local features defined in terms of temporal extrema of first- or second-order temporal derivative responses. For dense features modelled as a sine wave, the lack of temporal scale invariance does, however, constitute a major limitation for handling dense temporal structures of different temporal duration in a uniform manner. In the case of a logarithmic distribution of the temporal scale levels, specifically taken to the limit of a time-causal limit kernel with an infinitely dense distribution of the temporal scale levels towards zero temporal scale, we show that it is possible to achieve true temporal scale invariance to handle dense features modelled as a sine wave in a uniform manner over different temporal durations of the temporal structures as well to achieve more general temporal scale invariance for any signal over any temporal scaling transformation with a scaling factor that is an integer power of the distribution parameter of the time-causal limit kernel. It is shown how these temporal scale selection properties developed for a pure temporal domain carry over to feature detectors defined over time-causal spatio-temporal and spectro-temporal domains.

✉ Tony Lindeberg
   tony@kth.se

1 Computational Brain Science Lab, Department of Computational Science and Technology, School of Computer Science and Communication, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

🖉 Springer

1 Introduction

When processing sensory data by automatic methods in
areas of signal processing such as computer vision or audio
processing or in computational modelling of biological per-
ception, the notion of receptive field constitutes an essential
concept [2,14,15,30,31,90].

For sensory data as obtained from vision or hearing, or
their counterparts in artificial perception, the measurement
from a single light sensor in a video camera or on the retina, or
the instantaneous sound pressure registered by a microphone
is hardly meaningful at all, since any such measurement is
strongly dependent on external factors such as the illumina-
tion of a visual scene regarding vision or the distance between
the sound source and the microphone regarding hearing.
Instead, the essential information is carried by the relative
relations between local measurements at different points and
temporal moments regarding vision or local measurements
over different frequencies and temporal moments regard-
ing hearing. Following this paradigm, sensory measurements
should be performed over local neighbourhoods over space–
time regarding vision and over local neighbourhoods in the
time–frequency domain regarding hearing, leading to the
notions of spatio-temporal and spectro-temporal receptive
fields.

Specifically, spatio-temporal receptive fields constitute a
main class of primitives for expressing methods for video
analysis [35,39,45,47,94,98,108,118–120,124], whereas
spectro-temporal receptive fields constitute a main class
of primitives for expressing methods for machine hearing
[4,18,29,40,87,96,97,103,123].

A general problem when applying the notion of receptive
fields in practice, however, is that the types of responses that
are obtained in a specific situation can be strongly dependent
on the scale levels at which they are computed. A spatio-
temporal receptive field is determined by at least a spatial
scale parameter and a temporal scale parameter, whereas a
spectro-temporal receptive field is determined by at least a
spectral and a temporal scale parameter. Beyond ensuring
that local sensory measurements at different spatial, tempo-
ral and spectral scales are treated in a consistent manner,
which by itself provides strong contraints on the shapes of
the receptive fields [68,75,78,79], it is necessary for com-
puter vision or machine hearing algorithms to decide what
responses within the families of receptive fields over different
spatial, temporal and spectral scales they should base their
analysis on.

Over the spatial domain, theoretically well-founded meth-
ods have been developed for choosing spatial scale levels

among receptive field responses over multiple spatial scales
[61,62,64,70,71] leading to, e.g. robust methods for image-
based matching and recognition [5,48,85,88,114,115,117]
that are able to handle large variations of the size of the
objects in the image domain and with numerous applications
regarding object recognition, object categorization, multi-
view geometry, construction of 3-D models from visual input,
human–computer interaction, biometrics and robotics.

Much less research has, however, been performed regard-
ing the topic of choosing local appropriate scales in temporal
data. While some methods for temporal scale selection have
been developed [46,60,121], these methods suffer from
either theoretical or practical limitations.

A main subject of this paper is present a theory for how
to compare filter responses in terms of temporal derivatives
that have been computed at different temporal scales, specif-
ically with a detailed theoretical analysis of the possibilities
of having temporal scale estimates as obtained from a tempo-
ral scale selection mechanism reflect the temporal duration of
the underlying temporal structures that gave rise to the feature
responses. Another main subject of this paper is to present a
theoretical framework for temporal scale selection that leads
to temporal scale invariance and enables the computation of
scale-covariant temporal scale estimates. While these topics
can for a non-causal temporal domain be addressed by the
non-causal Gaussian scale-space concept [21,32,41,43,57,
58,66,112,122], the development of such a theory has been
missing regarding a time-causal temporal domain.

1.1 Temporal Scale Selection

When processing time-dependent signals in video or audio
or more generally any temporal signal, special attention has
to be put to the facts that:

- the physical phenomena that generate the temporal sig-
  nals may occur at different speed—faster or slower, and
- the temporal signals may contain qualitatively different
  types of temporal structures at different temporal scales.

In certain controlled situations where the physical system
that generates the temporal signals that is to be processed is
sufficiently well known and if the variability of the tempo-
ral scales over time in the domain is sufficiently constrained,
suitable temporal scales for processing the signals may in
some situations be chosen manually and then be verified
experimentally. If the sources that generate the temporal sig-
nals are sufficiently complex and/or if the temporal structures
in the signals vary substantially in temporal duration by the
underlying physical processes occurring significantly faster
or slower, it is on the other hand natural to (i) include a
mechanism for processing the temporal data at multiple tem-

poral scales and (ii) try to detect in a bottom-up manner at what temporal scales the interesting temporal phenomena are likely to occur.

The subject of this article is to develop a theory for temporal scale selection in a time-causal temporal scale space as an extension of a previously developed theory for spatial scale selection in a spatial scale space [61,62,64,70,71], to generate bottom-up hypotheses about characteristic temporal scales in time-dependent signals, intended to serve as estimates of the temporal duration of local temporal structures in time-dependent signals. Special focus will be on developing mechanisms analogous to scale selection in non-causal Gaussian scale-space, based on local extrema over scales of scale-normalized derivatives, while expressed within the framework of a time-causal and time-recursive temporal scale space in which the future cannot be accessed and the signal processing operations are thereby only allowed to make use of information from the present moment and a compact buffer of what has occurred in the past.

When designing and developing such scale selection mechanisms, it is essential that the computed scale estimates reflect the temporal duration of the corresponding temporal structures that gave rise to the feature responses. To understand the prerequisites for developing such temporal scale selection methods, we will in this paper perform an in-depth theoretical analysis of the scale selection properties that such temporal scale selection mechanisms give rise to for different temporal scale-space concepts and for different ways of defining scale-normalized temporal derivatives.

Specifically, after an examination of the theoretical properties of different types of temporal scale-space concepts, we will focus on a class of recently extended time-causal temporal scale-space concepts obtained by convolution with truncated exponential kernels coupled in cascade [53,73,75,77]. For two natural ways of distributing the discrete temporal scale levels in such a representation, in terms of either a uniform distribution over the scale parameter τ corresponding to the variance of the composed scale-space kernel or a logarithmic distribution, we will study the scale selection properties that result from detecting local temporal scale levels from local extrema over scale of scale-normalized temporal derivatives. The motivation for studying a logarithmic distribution of the temporal scale levels, is that it corresponds to a uniform distribution in units of effective scale $\tau_{\mathrm{eff}} = A + B \log \tau$ for some constants A and B, which has been shown to constitute the natural metric for measuring the scale levels in a spatial scale space [41,55].

As we shall see from the detailed theoretical analysis that will follow, this will imply certain differences in scale selection properties of a temporally asymmetric time-causal scale space compared to scale selection in a spatially mirror symmetric Gaussian scale space. These differences in theoretical properties are in turn essential to take into explicit account

when formulating algorithms for temporal scale selection, in e.g. video analysis or audio analysis applications.

For the temporal scale-space concept based on a uniform distribution of the temporal scale levels in units of the variance of the composed scale-space kernel, it will be shown that temporal scale selection from local extrema over temporal scales will make it possible to estimate the temporal duration of local temporal structures modelled as local temporal peaks and local temporal ramps. For a dense temporal structure modelled as a temporal sine wave, the lack of true scale invariance for this concept will, however, imply that the temporal scale estimates will not be directly proportional to the wavelength of the temporal sine wave. Instead, the scale estimates are affected by a bias, which is not a desirable property.

For the temporal scale-space concept based on a logarithmic distribution of the temporal scale levels, and taken to the limit to scale-invariant time-causal limit kernel [75] corresponding to an infinite number of temporal scale levels that cluster infinitely close near the temporal scale level zero, it will on the other hand be shown that the temporal scale estimates of a dense temporal sine wave will be truly proportional to the wavelength of the signal. By a general proof, it will be shown this scale-invariant property of temporal scale estimates can also be extended to any sufficiently regular signal, which constitutes a general foundation for expressing scale invariant temporal scale selection mechanisms for time-dependent video and audio and more generally also other classes of time-dependent measurement signals.

As complement to this proposed overall framework for temporal scale selection, we will also present a set of general theoretical results regarding time-causal scale-space representations: (i) showing that previous application of the assumption of a semi-group property for time-causal scale-space concepts leads to undesirable temporal dynamics, which however can be remedied by replacing the assumption of a semi-group structure be a weaker assumption of a cascade property in turn based on a transitivity property, (ii) formulations of scale-normalized temporal derivatives for Koenderink's time-causal scale-time model [42], and (iii) ways of translating the temporal scale estimates from local extrema over temporal scales in the temporal scale-space representation based on the scale-invariant time-causal limit kernel into quantitative measures of the temporal duration of the corresponding underlying temporal structures and in turn based on a scale-time approximation of the limit kernel.

In these ways, this paper is intended to provide a theoretical foundation for expressing theoretically well-founded temporal scale selection methods for selecting local temporal scales over time-causal temporal domains, such as video and audio with specific focus on real-time image or sound streams. Applications of this scale selection methodology for

detecting both sparse and dense spatio-temporal features in video are presented in a companion paper [74].

1.2 Structure of this Article

As a conceptual background to the theoretical developments that will be performed, we will start in Sect. 2 with an overview of different approaches to handling temporal data within the scale-space framework including a comparison of relative advantages and disadvantages of different types of temporal scale-space concepts.

As a theoretical baseline for the later developments of methods for temporal scale selection in a time-causal scale space, we shall then in Sect. 3 give an overall description of basic temporal scale selection properties that will hold if the non-causal Gaussian scale-space concept with its corresponding selection methodology for a spatial image domain is applied to a one-dimensional non-causal temporal domain, e.g. for the purpose of handling the temporal domain when analysing pre-recorded video or audio in an offline setting.

In Sects. 4, 5 we will then continue with a theoretical analysis of the consequences of performing temporal scale selection in the time-causal scale space obtained by convolution with truncated exponential kernels coupled in cascade [53,73,75,77]. By selecting local temporal scales from the scales at which scale-normalized temporal derivatives assume local extrema over temporal scales, we will analyse the resulting temporal scale selection properties for two ways of defining scale-normalized temporal derivatives, by either variance-based normalization as determined by a scale normalization parameter γ or L_p-normalization for different values of the scale normalization power p.

With the temporal scale levels required to be discrete because of the very nature of this temporal scale-space concept, we will specifically study two ways of distributing the temporal scale levels over scale, using either a uniform distribution relative to the temporal scale parameter τ corresponding to the variance of the composed temporal scale-space kernel in Sect. 4 or a logarithmic distribution of the temporal scale levels in Sect. 5.

Because of the analytically simpler form for the time-causal scale-space kernels corresponding to a *uniform distribution of the temporal scale levels*, some theoretical scale-space properties will turn out to be easier to study in closed form for this temporal scale-space concept. We will specifically show that for a temporal peak modelled as the impulse response to a set of truncated exponential kernels coupled in cascade, the selected temporal scale level will serve as a good approximation of the temporal duration of the peak or be proportional to this measure depending on the value of the scale normalization parameter γ used for scale-normalized temporal derivatives based on variance-based normalization or the scale normalization power p for scale-

normalized temporal derivatives based on L_p-normalization. For a temporal onset ramp, the selected temporal scale level will on the other hand be either a good approximation of the time constant of the onset ramp or proportional to this measure of the temporal duration of the ramp. For a temporal sine wave, the selected temporal scale level will, however, not be directly proportional to the wavelength of the signal, but instead affected by a systematic bias. Furthermore, the corresponding scale-normalized magnitude measures will not be independent of the wavelength of the sine wave but instead show systematic wavelength-dependent deviations. A main reason for this is that this temporal scale-space concept does not guarantee temporal scale invariance if the temporal scale levels are distributed uniformly in terms of the temporal scale parameter τ corresponding to the temporal variance of the temporal scale-space kernel.

With a *logarithmic distribution of the temporal scale levels*, we will on the other hand show that for the temporal scale-space concept defined by convolution with *the time-causal limit kernel* [75] corresponding to an infinitely dense distribution of the temporal scale levels towards zero temporal scale, the temporal scale estimates will be perfectly proportional to the wavelength of a sine wave for this temporal scale-space concept. It will also be shown that this temporal scale-space concept leads to perfect scale invariance in the sense that (i) local extrema over temporal scales are preserved under temporal scaling factors corresponding to integer powers of the distribution parameter c of the time-causal limit kernel underlying this temporal scale-space concept and are transformed in a scale-covariant way for any temporal input signal and (ii) if the scale normalization parameter $\gamma = 1$ or equivalently if the scale normalization power $p = 1$, the magnitude values at the local extrema over scale will be equal under corresponding temporal scaling transformations. For this temporal scale-space concept, we can therefore fulfil basic requirements to *achieve temporal scale invariance also over a time-causal and time-recursive temporal domain*.

To simplify the theoretical analysis, we will in some cases temporarily extend the definitions of temporal scale-space representations over discrete temporal scale levels to a continuous scale variable, to make it possible to compute local extrema over temporal scales from differentiation with respect to the temporal scale parameter. Section 6 discusses the influence that this approximation has on the overall theoretical analysis.

Section 7 then illustrates how the proposed theory for temporal scale selection can be used for computing local scale estimates from 1-D signals with substantial variabilities in the characteristic temporal duration of the underlying structures in the temporal signal.

In Sect. 8, we analyse how the derived scale selection properties carry over to a set of spatio-temporal feature detec-

tors defined over both multiple spatial scales and multiple temporal scales in a time-causal spatio-temporal scale-space representation for video analysis. Section 9 then outlines how corresponding selection of local temporal and logspectral scales can be expressed for audio analysis operations over a time-causal spectro-temporal domain. Finally, Sect. 10 concludes with a summary and discussion.

To simplify the presentation, we have put some derivations and theoretical analysis in the appendix. "Appendix 1" presents a general theoretical argument of why a requirement about a semi-group property over temporal scales will lead to undesirable temporal dynamics for a time-causal scale space and argue that the essential structure of non-creation of new image structures from any finer to any coarser temporal scale can instead nevertheless be achieved with the less restrictive assumption about a cascade smoothing property over temporal scales, which then allows for better temporal dynamics in terms of shorter temporal delays.

In relation to Koenderink's scale-time model [42], "Appendix 2" shows how corresponding notions of scale-normalized temporal derivatives based on either variance-based normalization or L_p-normalization can be defined also for this time-causal temporal scale-space concept.

"Appendix 3" shows how the temporal duration of the time-causal limit kernel proposed in [75] can be estimated by a scale-time approximation of the limit kernel via Koenderink's scale-time model leading to estimates of how a selected temporal scale level $\hat{\tau}$ from local extrema over temporal scale can be translated into a estimates of the temporal duration of temporal structures in the temporal scale-space representation obtained by convolution with the time-causal limit kernel. Specifically, explicit expressions are given for such temporal duration estimates based on first- and second-order temporal derivatives.

2 Theoretical Background and Related Work

2.1 Temporal Scale-Space Concepts

For processing temporal signals at multiple temporal scales, different types of temporal scale-space concepts have been developed in the computer vision literature (see Fig. 1).

For offline processing of pre-recorded signals, a *non-causal Gaussian temporal scale-space* concept may in many situations be sufficient. A Gaussian temporal scale-space concept is constructed over the 1-D temporal domain in a similar manner as a Gaussian spatial scale-space concept is constructed over a D-dimensional spatial domain (Florack [21]; Iijima [32]; Koenderink [41]; Koenderink and van Doorn [43]; Lindeberg [57,58,66]; ter Haar Romeny [112]; Witkin [122]), with or without the difference that a model for temporal delays may or may not be additionally included [66].

When processing temporal signals in real time, or when modelling sensory processes in biological perception computationally, it is on the other hand necessary to base the temporal analysis on time-causal operations.

The first time-causal temporal scale-space concept was developed by Koenderink [42], who proposed to apply Gaussian smoothing on a logarithmically transformed time axis with the present moment mapped to the unreachable infinity. This temporal scale-space concept does, however, not have any known time-recursive formulation. Formally, it requires an infinite memory of the past and has therefore not been extensively applied in computational applications.

Lindeberg [53,73,75] and Lindeberg and Fagerström [77] proposed a time-causal temporal scale-space concept based on truncated exponential kernels or equivalently first-order integrators coupled in cascade, based on theoretical results by Schoenberg [104] (see also Karlin [38] and Schoenberg [105]) implying that such kernels are the only variation-diminishing kernels over a 1-D temporal domain that guarantee non-creation of new local extrema or equivalently zero-crossings with increasing temporal scale. This temporal scale-space concept is additionally time-recursive and can be implemented in terms of computationally highly efficient first-order integrators or recursive filters over time. This theory has been recently extended into a scale-invariant time-causal limit kernel [75], which allows for scale invariance over the temporal scaling transformations that correspond to exact mappings between the temporal scale levels in the temporal scale-space representation based on a discrete set of logarithmically distributed temporal scale levels.

Based on semi-groups that guarantee either self-similarity over temporal scales or non-enhancement of local extrema with increasing temporal scales, Fagerström [19] and Lindeberg [66] have derived time-causal semi-groups that allow for a continuous temporal scale parameter and studied theoretical properties of these kernels.

Concerning temporal processing over discrete time, Fleet and Langley [20] performed temporal filtering for optic flow computations based on recursive filters over time. Lindeberg [53,73,75] and Lindeberg and Fagerström [77] showed that first-order recursive filters coupled in cascade constitutes a natural time-causal scale-space concept over discrete time, based on the requirement that the temporal filtering over a 1-D temporal signal must not increase the number of local extrema or equivalently the number of zero-crossings in the signal. In the specific case when all the time constants in this model are equal and tend to zero while simultaneously increasing the number of temporal smoothing steps in such a way that the composed temporal variance is held constant, these kernels can be shown to approach the temporal Poisson kernel [77]. If on the other hand the time constants of the first-order integrators are chosen so that the temporal scale levels become logarithmically distributed, these tem-
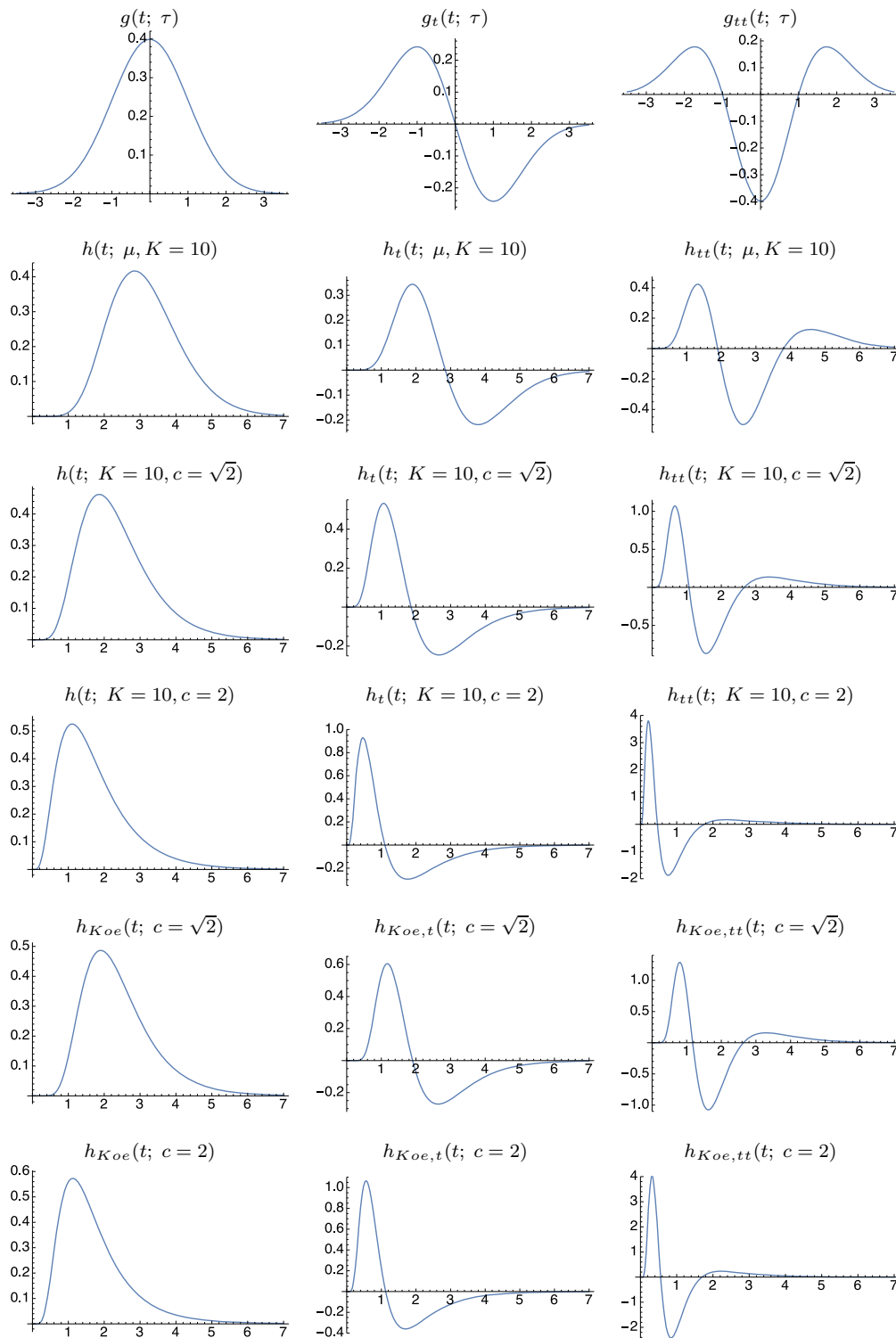
Fig. 1 Temporal scale-space kernels with composed temporal variance $\tau = 1$ for the main types of temporal scale-space concepts considered in this paper and with their first- and second-order temporal derivatives: (*top row*) the non-causal Gaussian kernel $g(t; \tau)$, (*second row*) the composition $h(t; \mu, K = 10)$ of $K = 10$ truncated exponential kernels with equal time constants, (*third row*) the composition $h(t; K = 10, c = \sqrt{2})$ of $K = 10$ truncated exponential kernels with logarithmic distribution of the temporal scale levels for $c = \sqrt{2}$, (*fourth row*) corresponding kernels $h(t; K = 10, c = 2)$ for $c = 2$, (*fifth row*) Koenderink's scale-time kernels $h_{\text{Koe}}(t; c = \sqrt{2})$ corresponding to Gaussian convolution over a logarithmically transformed temporal axis with the parameters determined to match the time-causal limit kernel corresponding to truncated exponential kernels with an infinite number of logarithmically distributed temporal scale levels according to (186) for $c = \sqrt{2}$, (*bottom row*) corresponding scale-time kernels $h_{\text{Koe}}(t; c = 2)$ for $c = 2$ (*horizontal axis* time t)

poral smoothing kernels approach a discrete approximation of the time-causal limit kernel [75].

Applications of using these linear temporal scale-space concepts for modelling the temporal smoothing step in visual and auditory receptive fields have been presented by Lindeberg [59,65,66,68,69,73,75], ter Haar Romeny et al. [113], Lindeberg and Friberg [78,79] and Mahmoudi [86]. Nonlinear spatio-temporal scale-space concepts have been proposed by Guichard [24]. Applications of the non-causal Gaussian temporal scale-space concept for computing spatio-temporal features have been presented by Laptev and Lindeberg [45–47], Klaser et al. [39], Willems et al. [121], Wang et al. [118], Shao and Mattivi [108] and others, see specifically Poppe [98] for a survey of early approaches to vision-based human action recognition, Jhuang et al. [35] and Niebles et al. [94] for conceptually related non-causal Gabor approaches, Adelson and Bergen [1] and Derpanis and Wildes [16] for closely related spatio-temporal orientation models and Han et al. [26] for a related mid-level temporal representation termed the video primal sketch.

Applications of the temporal scale-space model based on truncated exponential kernels with equal time constants coupled in cascade and corresponding to Laguerre functions (Laguerre polynomials multiplied by a truncated exponential kernel) for computing spatio-temporal features have presented by Rivero-Moreno and Bres [100], Shabani et al. [107] and Berg et al. [116] as well as for handling time scales in video surveillance (Jacob and Pless [33]), for performing edge preserving smoothing in video streams (Paris [95]) and is closely related to Tikhonov regularization as used for image restoration by Surya et al. [111]. A general framework for performing spatio-temporal feature detection based on the temporal scale-space model based on truncated exponential kernels coupled in cascade with specifically the both theoretical and practical advantages of using logarithmic distribution of the intermediated temporal scale levels in terms of temporal scale invariance and better temporal dynamics (shorter temporal delays) has been presented in Lindeberg [75].

2.2 Relative Advantages of Different Temporal Scale Spaces

When developing a temporal scale selection mechanism over a time-causal temporal domain, a first problem concerns what time-causal scale-space concept to base the multi-scale temporal analysis upon. The above reviewed temporal scale-space concepts have different relative advantages from a theoretical and computational viewpoint. In this section, we will perform an in-depth examination of the different temporal scale-space concepts that have been developed in the literature, which will lead us to a class of time-causal scale-space concepts that we argue is particularly suitable with respect to the set of desirable properties we aim at.

The non-causal Gaussian temporal scale space is in many cases the conceptually easiest temporal scale-space concept to handle and to study analytically [66]. The corresponding temporal kernels are scale invariant, have compact closed-form expressions over both the temporal and frequency domains and obey a semi-group property over temporal scales. When applied to pre-recorded signals, temporal delays can if desirable be disregarded, which eliminates any need for temporal delay compensation. This scale-space concept is, however, not time-causal and not time-recursive, which implies fundamental limitations with regard to real-time applications and realistic modelling of biological perception.

Koenderink's scale-time kernels [42] are truly time-causal, allow for a continuous temporal scale parameter, have good temporal dynamics and have a compact explicit expression over the temporal domain. These kernels are, however, not time-recursive, which implies that they in principle require an infinite memory of the past (or at least extended temporal buffers corresponding to the temporal extent to which the infinite support temporal kernels are truncated at the tail). Thereby, the application of Koenderink's scale-time model to video analysis implies that substantial temporal buffers are needed when implementing this non-recursive temporal scale-space in practice. Similar problems with substantial need for extended temporal buffers arise when applying the non-causal Gaussian temporal scale-space concept to offline analysis of extended video sequences. The algebraic expressions for the temporal kernels in the scale-time model are furthermore not always straightforward to handle and there is no known simple expression for the Fourier transform of these kernels or no known simple explicit cascade smoothing property over temporal scales with respect to the regular (untransformed) temporal domain. Thereby, certain algebraic calculations with the scale-time kernels may become quite complicated.

The temporal scale-space kernels obtained by coupling truncated exponential kernels or equivalently first-order integrators in cascade are both truly time-causal and truly time-recursive [53,73,75,77]. The temporal scale levels are on the other hand required to be discrete. If the goal is to construct a real-time signal processing system that analyses continuous streams of signal data in real time, one can however argue that a restriction of the theory to a discrete set of temporal scale levels is less of a constraint, since the signal processing system anyway has to be based on a finite amount of sensors and hardware/wetware for sampling and processing the continuous stream of signal data.

In the special case when all the time constants are equal, the corresponding temporal kernels in the temporal scale-space model based on truncated exponential kernels coupled in cascade have compact explicit expressions that are easy to handle both in the temporal domain and in the frequency

domain, which simplifies theoretical analysis. These kernels obey a semi-group property over temporal scales, but are not scale invariant and lead to slower temporal dynamics when a larger number of primitive temporal filters are coupled in cascade [73,75].

In the special case when the temporal scale levels in this scale-space model are logarithmically distributed, these kernels have a manageable explicit expression over the Fourier domain that enables some closed-form theoretical calculations. Deriving an explicit expression over the temporal domain is, however, harder, since the explicit expression then corresponds to a linear combination of truncated exponential filters for all the time constants, with the coefficients determined from a partial fraction expansion of the Fourier transform, which may lead to rather complex closed-form expressions. Thereby certain analytical calculations may become harder to handle. As shown in [75] and "Appendix 3", some such calculations can on the other hand be well approximated via a scale-time approximation of the time-causal temporal scale-space kernels. When using a logarithmic distribution of the temporal scales, the composed temporal kernels do however have very good temporal dynamics and much better temporal dynamics compared to corresponding kernels obtained by using truncated exponential kernels with equal time constants coupled in cascade. Moreover, these kernels lead to a computationally very efficient numerical implementation. Specifically, these kernels allow for the formulation of a time-causal limit kernel that obeys scale invariance under temporal scaling transformations, which cannot be achieved if using a uniform distribution of the temporal scale levels [73,75].

The temporal scale-space representations obtained from the self-similar time-causal semi-groups have a continuous scale parameter and obey temporal scale invariance [19,66]. These kernels do, however, have less desirable temporal dynamics (see "Appendix 1" for a general theoretical argument about undesirable consequences of imposing a temporal semi-group property on temporal kernels with temporal delays) and/or lead to pseudodifferential equations that are harder to handle both theoretically and in terms of computational implementation. For these reasons, we shall not consider those time-causal semi-groups further in this treatment.

2.3 Previous Work on Methods for Scale Selection

A general framework for performing scale selection for local differential operations was proposed in Lindeberg [56,57] based on the detection of local extrema over scale of scale-normalized derivative expressions and then refined in Lindeberg [61,62]—see Lindeberg [64,71] for tutorial overviews.

This scale selection approach has been applied to a large number of feature detection tasks over spatial image domains including detection of scale-invariant interest points (Linde-

berg [62,70]; Mikolajczyk and Schmid [88]; Tuytelaars and Mikolajczyk [115]), performing feature tracking (Bretzner and Lindeberg [9]), computing shape from texture and disparity gradients (Gårding and Lindeberg [23]; Lindeberg and Gårding [80]), detecting 2-D and 3-D ridges (Frangi et al. [22]; Krissian et al. [44]; Lindeberg [61]; Sato et al. [102]), computing receptive field responses for object recognition (Chomat et al. [12], Hall et al. [25]), performing hand tracking and hand gesture recognition (Bretzner et al. [8]) and computing time-to-collision (Negre et al. [93]).

Specifically, very successful applications have been achieved in the area of image-based matching and recognition (Bay et al. [5]; Lindeberg [67,72]; Lowe [85]). The combination of local scale selection from local extrema of scale-normalized derivatives over scales (Lindeberg [57,62]) with affine shape adaptation (Lindeberg and Gårding [81]) has made it possible to perform multi-view image matching over large variations in viewing distances and viewing directions (Lazebnik et al. [49]; Mikolajczyk and Schmid [88]; Mikolajczyk et al. [89]; Rothganger et al. [101]; Tuytelaars and van Gool [114]). The combination of interest point detection from scale-space extrema of scale-normalized differential invariants (Lindeberg [57,62]) with local image descriptors (Bay et al. [5]; Lowe [85]) has made it possible to design robust methods for performing object recognition of natural objects in natural environments with numerous applications to object recognition (Bay et al. [5]; Lowe [85]), object category classification (Bosch et al. [7]; Mutch and Lowe [92]), multi-view geometry (Hartley and Zisserman [27]), panorama stitching (Brown and Lowe [11]), automated construction of 3-D object and scene models from visual input (Agarwal et al. [3]; Brown and Lowe [10]), synthesis of novel views from previous views of the same object (Liu et al. [82]), visual search in image databases (Datta et al. [13]; Lew et al. [50]), human–computer interaction based on visual input (Jaimes and Sebe [34]; Porta [99]), biometrics (Bicego et al. [6]; Li [51]) and robotics (Se et al. [106]; Siciliano and Khatib [109]).

Alternative approaches for performing scale selection over spatial image domains have also been proposed in terms of (i) detecting peaks of weighted entropy measures (Kadir and Brady [36]) or Lyaponov functionals (Sporring et al. [110]) over scales, (ii) minimizing normalized error measures over scale (Lindeberg [63]), (iii) determining minimum reliable scales for edge detection based on a noise suppression model (Elder and Zucker [17]), (iv) determining at what scale levels to stop in nonlinear diffusion-based image restoration methods based on similarity measurements relative to the original image data (Mrazek and Navara [91]), (v) by comparing reliability measures from statistical classifiers for texture analysis at multiple scales (Kang et al. [37]), (vi) by computing image segmentations from the scales at which a supervised classifier delivers class labels with the highest reliability measure (Li et al. [52]; Loog et al. [84]), (vii)

selecting scales for edge detection by estimating the saliency of elongated edge segments (Liu et al. [83]) or (viii) considering subspaces generated by local image descriptors computed over multiple scales (Hassner et al. [28]).

More generally, spatial scale selection can be seen as a specific instance of computing invariant receptive field responses under natural image transformations, to (i) handle objects in the world of different physical size and to account for scaling transformations caused by the perspective mapping, and with extensions to (ii) affine image deformations to account for variations in the viewing direction and (iii) Galilean transformations to account for relative motions between objects in the world and the observer as well as to (iv) illumination variations [69].

Early theoretical work on temporal scale selection in a time-causal scale space was presented in Lindeberg [60] with primary focus on the temporal Poisson scale-space, which possesses a temporal semi-group structure over a discrete time-causal temporal domain while leading to long temporal delays (see "Appendix 1" for a general theoretical argument). Temporal scale selection in non-causal Gaussian spatio-temporal scale space has been used by Laptev and Lindeberg [46] and Willems et al. [121] for computing spatio-temporal interest points, however, with certain theoretical limitations that are explained in a companion paper [74].[1] The purpose of this article is to present a much further developed and more general theory for temporal scale selection in time-causal scale spaces over continuous temporal domains and to analyse the theoretical scale selection properties for different types of model signals.

3 Scale Selection Properties for the Non-causal Gaussian Temporal Scale-Space Concept

In this section, we will present an overview of theoretical properties that will hold if the Gaussian temporal scale-space concept is applied to a non-causal temporal domain, if additionally the scale selection mechanism that has been developed for a non-causal spatial domain is directly transferred to a non-causal temporal domain. The set of temporal scale-space properties that we will arrive at will then be used as a theoretical baseline for developing temporal scale-space properties over a time-causal temporal domain.

---

[1] The spatio-temporal scale selection method in (Laptev and Lindeberg [46]) is based on a spatio-temporal Laplacian operator that is not scale covariant under independent relative scaling transformations of the spatial versus the temporal domains [74], which implies that the spatial and temporal scale estimate will not be robust under independent variabilities of the spatial and temporal scales in video data. The spatio-temporal scale selection method applied to the determinant of the spatio-temporal Hessian in Willems et al. [121] does not make use of the full flexibility of the notion of γ-normalized derivative operators [74] and has not previously been developed over a time-causal spatio-temporal domain.

3.1 Non-causal Gaussian Temporal Scale-Space

Over a one-dimensional temporal domain, axiomatic derivations of a temporal scale-space representation based on the assumptions of (i) linearity, (ii) temporal shift invariance, (iii) semi-group property over temporal scale, (iv) sufficient regularity properties over time and temporal scale and (v) non-enhancement of local extrema imply that the temporal scale-space representation

$$L(\cdot; \ \tau, \delta) = g(\cdot; \ \tau, \delta) * f(\cdot) \tag{1}$$

should be generated by convolution with possibly time-delayed temporal kernels of the form [66]

$$g(t; \ \tau, \delta) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(t-\delta)^2}{2\tau}} \tag{2}$$

where τ is a temporal scale parameter corresponding to the variance of the Gaussian kernel and δ is a temporal delay. Differentiating the kernel with respect to time gives

$$g_t(t; \ \tau, \delta) = -\frac{(t-\delta)}{\tau} g(t; \ \tau, \delta) \tag{3}$$

$$g_{tt}(t; \ \tau, \delta) = \frac{((t-\delta)^2 - \tau)}{\tau^2} g(t; \ \tau, \delta) \tag{4}$$

see the top row in Fig. 1 for graphs. When analysing pre-recorded temporal signals, it can be preferable to set the temporal delay to zero, leading to temporal scale-space kernels having a similar form as spatial Gaussian kernels:

$$g(t; \ \tau) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{t^2}{2\tau}}. \tag{5}$$

3.2 Temporal Scale Selection from Scale-Normalized Derivatives

As a conceptual background to the treatments that we shall later develop regarding temporal scale selection in time-causal temporal scale spaces, we will in this section describe the theoretical structure that arises by transferring the theory for scale selection in a Gaussian scale space over a spatial domain to the non-causal Gaussian temporal scale space:

Given the temporal scale-space representation $L(t; \ \tau)$ of a temporal signal $f(t)$ obtained by convolution with the Gaussian kernel $g(t; \ \tau)$ according to (1), temporal scale selection can be performed by detecting *local extrema over temporal scales* of differential expressions expressed in terms of *scale-normalized temporal derivatives* at any scale τ according to [61,62,64,71]

$$\partial_{\zeta^n} = \tau^{n\gamma/2} \partial_{t^n}, \tag{6}$$

where $\zeta = t/\tau^{\gamma/2}$ is the scale-normalized temporal variable, n is the order of temporal differentiation and γ is a free parameter. It can be shown [62, Section 9.1] that this notion of γ-normalized derivatives corresponds to normalizing the nth-order Gaussian derivatives $g_{\zeta^n}(t;\ \tau)$ over a one-dimensional domain to constant L_p-norms over scale τ

$$\|g_{\zeta^n}(\cdot;\ \tau)\|_p = \left(\int_{t\in\mathbb{R}} |g_{\zeta^n}(t;\ \tau)|^p \, \mathrm{d}t\right)^{1/p} = G_{n,\gamma} \qquad (7)$$

with

$$p = \frac{1}{1 + n(1-\gamma)} \qquad (8)$$

where the perfectly scale-invariant case $\gamma = 1$ corresponds to L_1-normalization for all orders n of temporal differentiation.

Temporal scale invariance A general and very useful scale-invariant property that results from this construction of the notion of scale-normalized temporal derivatives can be stated as follows: consider two signals f and f' that are related by a temporal scaling transformation

$$f'(t') = f(t) \quad \text{with} \quad t' = S t, \qquad (9)$$

and assume that there is a local extremum over scales at $(t_0;\ \tau_0)$ in a differential expression $\mathcal{D}_{\gamma\text{-norm}}L$ defined as a homogeneous polynomial of Gaussian derivatives computed from the scale-space representation L of the original signal f. Then, there will be a corresponding local extremum over scales at $(t_0';\ \tau_0') = (S t_0;\ S^2 \tau_0)$ in the corresponding differential expression $\mathcal{D}_{\gamma\text{-norm}}L'$ computed from the scale-space representation L' of the rescaled signal f' [62, Section 4.1].

This scaling result holds for all homogeneous polynomial differential expression and implies that local extrema over scales of γ-normalized derivatives are preserved under scaling transformations. Specifically, this scale-invariant property implies that if a local scale temporal level in dimension of time $\sigma = \tau$ is selected to be proportional to the temporal scale estimate $\hat{\sigma} = \sqrt{\hat{\tau}}$ such that $\sigma = C\hat{\sigma}$, then if the temporal signal f is transformed by a temporal scale factor S, the temporal scale estimate and therefore also the selected temporal scale level will be transformed by a similar temporal factor $\hat{\sigma}' = S\hat{\sigma}$, implying that the selected temporal scale levels will automatically adapt to variations in the characteristic temporal scale of the signal. Thereby, such local extrema over temporal scale provide a theoretically well-founded way to automatically adapt the scale levels to local scale variations.

Specifically, scale-normalized scale-space derivatives of order n at corresponding temporal moments will be related according to

$$L'_{\zeta'^n}(t';\ \tau') = S^{n(\gamma-1)} L_{\zeta^n}(t;\ \tau) \qquad (10)$$

which means that $\gamma = 1$ implies perfect scale invariance in the sense that the γ-normalized derivatives at corresponding points will be equal. If $\gamma \neq 1$, the difference in magnitude can on the other hand be easily compensated for using the scale values of the corresponding scale-adaptive image features (see below).

3.3 Temporal Peak

For a temporal peak modelled as a Gaussian function with variance τ_0

$$g(t;\ \tau_0) = \frac{1}{\sqrt{2\pi\tau_0}}\mathrm{e}^{-\frac{t^2}{2\tau_0}}. \qquad (11)$$

it can be shown that scale selection from local extrema over scale of second-order scale-normalized temporal derivatives

$$L_{\zeta\zeta} = \tau^\gamma L_{tt} \qquad (12)$$

implies that the scale estimate at the position $t = 0$ of the peak will be given by Lindeberg [61, Equation (56)] [70, Equation (212)]

$$\hat{\tau} = \frac{2\gamma}{3 - 2\gamma}\tau_0. \qquad (13)$$

If we require the scale estimate to reflect the temporal duration of the peak such that

$$\hat{\tau} = q^2 \tau_0, \qquad (14)$$

then this implies

$$\gamma = \frac{3q^2}{2(q^2 + 1)} \qquad (15)$$

which in the specific case of $q = 1$ corresponds to [61, Section 5.6.1]

$$\gamma = \gamma_2 = \frac{3}{4} \qquad (16)$$

and in turn corresponding to L_p-normalization for $p = 2/3$ according to (8).

If we additionally renormalize the original Gaussian peak to having maximum value equal to one

$$p(t;\ t_0) = \sqrt{2\pi\tau_0}\, g(t;\ \tau_0) = \mathrm{e}^{-\frac{t^2}{2\tau_0}}, \qquad (17)$$

then if using the same value of γ for computing the magnitude response as for selecting the temporal scale, the maximum magnitude value over scales will be given by

$$L_{\zeta\zeta,\text{maxmagn}} = \frac{2^{\gamma}(2\gamma - 3)}{3\tau_0}\left(\frac{\gamma\,\tau_0}{3 - 2\gamma}\right)^{\gamma} \tag{18}$$

and will not be independent of the temporal scale τ_0 of the original peak unless $\gamma = 1$. If on the other hand using $\gamma = 3/4$ as motivated by requirements of scale calibration (14) for $q = 1$, the scale dependency will for a Gaussian peak be of the form

$$L_{\zeta\zeta,\text{maxmagn}}\big|_{\gamma=3/4} = \frac{1}{2\tau_0^{1/4}}. \tag{19}$$

To get a scale-invariant magnitude measure for comparing the responses of second-order temporal derivative responses at different temporal scales for the purpose of scale calibration, we should therefore consider a scale-invariant magnitude measure for peak detection of the form

$$L_{\zeta\zeta,\text{maxmagn,postnorm}}\big|_{\gamma=1} = \tau^{1/4}\,L_{\zeta\zeta,\text{maxmagn}}\big|_{\gamma=3/4} \tag{20}$$

which for a Gaussian temporal peak will assume the value

$$L_{\zeta\zeta,\text{maxmagn,postnorm}}\big|_{\gamma=1} = \frac{1}{2} \tag{21}$$

Specifically, this form of post-normalization corresponds to computing the scale-normalized derivatives for $\gamma = 1$ at the selected scale (14) of the temporal peak, which according to (8) corresponds to L_1-normalization of the second-order temporal derivative kernels.

3.4 Temporal Onset Ramp

If we model a temporal onset ramp with temporal duration τ_0 as the primitive function of the Gaussian kernel with variance τ_0

$$\Phi(t;\ \tau_0) = \int_{u=-\infty}^{t} g(u;\ \tau_0)\,\mathrm{d}u, \tag{22}$$

it can be shown that scale selection from local extrema over scale of first-order scale-normalized temporal derivatives

$$L_{\zeta} = \tau^{\gamma/2}L_t \tag{23}$$

implies that the scale estimate at the central position $t = 0$ will be given by Lindeberg [61, Equation (23)]

$$\hat{\tau} = \frac{\gamma}{1 - \gamma}\,\tau_0. \tag{24}$$

If we require this scale estimate to reflect the temporal duration of the ramp such that

$$\hat{\tau} = q^2\tau_0, \tag{25}$$

then this implies

$$\gamma = \frac{q^2}{q^2 + 1} \tag{26}$$

which in the specific case of $q = 1$ corresponds to [61, Section 4.5.1]

$$\gamma = \gamma_1 = \frac{1}{2} \tag{27}$$

and in turn corresponding to L_p-normalization for $p = 2/3$ according to (8).

If using the same value of γ for computing the magnitude response as for selecting the temporal scale, the maximum magnitude value over scales will be given by

$$L_{\zeta,\text{maxmagn}} = \frac{\gamma^{\gamma/2}}{\sqrt{2\pi}}\left(\frac{1 - \gamma}{\tau_0}\right)^{\frac{1}{2} - \frac{\gamma}{2}}, \tag{28}$$

which is not independent of the temporal scale τ_0 of the original onset ramp unless $\gamma = 1$. If using $\gamma = 1$ for temporal scale selection, the selected temporal scale according to (24) would, however, become infinite. If on the other hand using $\gamma = 1/2$ as motivated by requirements of scale calibration (25) for $q = 1$, the scale dependency will for a Gaussian onset ramp be of the form

$$L_{\zeta,\text{maxmagn}}\big|_{\gamma=1/2} = \frac{1}{2\sqrt{\pi}\,\sqrt[4]{\tau_0}}. \tag{29}$$

To get a scale-invariant magnitude measure for comparing the responses of first-order temporal derivative responses at different temporal scales, we should therefore consider a scale-invariant magnitude measure for ramp detection of the form

$$L_{\zeta,\text{maxmagn,postnorm}}\big|_{\gamma=1} = \tau^{1/4}\,L_{\zeta,\text{maxmagn}}\big|_{\gamma=1/2} \tag{30}$$

which for a Gaussian onset ramp will assume the value

$$L_{\zeta,\text{maxmagn,postnorm}}\big|_{\gamma=1} = \frac{1}{2\sqrt{\pi}} \approx 0.282 \tag{31}$$

Specifically, this form of post-normalization corresponds to computing the scale-normalized derivatives for $\gamma = 1$ at the selected scale (25) of the onset ramp and thus also to L_p-normalization of the first-order temporal derivative kernels for $p = 1$.

3.5 Temporal Sine Wave

For a signal defined as a temporal sine wave

$$f(t) = \sin(\omega_0 t), \tag{32}$$

it can be shown that there will be a peak over temporal scales in the magnitude of the nth-order temporal derivative $L_{\zeta^n} = \tau^{n\gamma/2}L_{t^n}$ at temporal scale [62, Section 3]

$$\tau_{\max} = \frac{n\gamma}{\omega_0^2}. \tag{33}$$

If we define a temporal scale parameter σ of dimension [time] according to $\sigma = \sqrt{\tau}$, then this implies that the scale estimate is proportional to the wavelength $\lambda_0 = 2\pi/\omega_0$ of the sine wave according to [62, Equation (9)]

$$\sigma_{\max} = \frac{\sqrt{\gamma n}}{2\pi} \lambda_0 \tag{34}$$

and does in this respect reflect a characteristic time constant over which the temporal phenomena occur. Specifically, the maximum magnitude measure over scale [62, Equation (10)]

$$L_{\zeta^n,\max} = \frac{(\gamma n)^{\gamma n/2}}{\mathrm{e}^{\gamma n/2}} \omega_0^{(1-\gamma)n} \tag{35}$$

is for $\gamma = 1$ independent of the angular frequency ω_0 of the sine wave and thereby scale invariant.

In the following, we shall investigate how these scale selection properties can be transferred to two types of time-causal temporal scale-space concepts.

4 Scale Selection Properties for the Time-Causal Temporal Scale Space-Concept Based on First-Order Integrators with Equal Time Constants

In this section, we will present a theoretical analysis of the scale selection properties that are obtained in the time-causal scale-space based on truncated exponential kernels coupled in cascade, for the specific case of a uniform distribution of the temporal scale levels in units of the composed variance of the composed temporal scale-space kernels, and corresponding to the time constants of all the primitive truncated exponential kernels being equal.

We will study three types of idealized model signals for which closed-form theoretical analysis is possible: (i) a temporal peak modelled as a set of K_0 truncated exponential kernels with equal time constants coupled in cascade, (ii) a temporal onset ramp modelled as the primitive function of the temporal peak model and (iii) a temporal sine wave. Specifically, we will analyse how the selected scale levels $\hat{K}$ obtained from local extrema of temporal derivatives over scale relate to the temporal duration of a temporal peak or a temporal onset ramp alternatively how the selected scale levels $\hat{K}$ depends on the wavelength of a sine wave.

We will also study how good approximation the scale-normalized magnitude measure at the maximum over temporal scales is compared to the corresponding fully scale-invariant magnitude measures that are obtained from the non-causal temporal scale concept as listed in Sect. 3.

4.1 Time-Causal Scale Space Based on Truncated Exponential Kernels with Equal Time Constants Coupled in Cascade

Given the requirements that the temporal smoothing operation in a temporal scale-space representation should obey (i) linearity, (ii) temporal shift invariance, (iii) temporal causality and (iv) guarantee non-creation of new local extrema or equivalently new zero-crossings with increasing temporal scale for any one-dimensional temporal signal, it can be shown [53,73,75,77] that the temporal scale-space kernels should be constructed as a cascade of truncated exponential kernels of the form

$$h_{\exp}(t; \ \mu_k) = \begin{cases} \frac{1}{\mu_k} \mathrm{e}^{-t/\mu_k} & t \ge 0, \\ 0 & t < 0. \end{cases} \tag{36}$$

If we additionally require the time constants of all such primitive kernels that are coupled in cascade to be equal, then this leads to a composed temporal scale-space kernel of the form

$$h_{\mathrm{composed}}(t; \ \mu, K) = \frac{t^{K-1} \mathrm{e}^{-t/\mu}}{\mu^K \Gamma(K)} = U(t; \ \mu, K) \tag{37}$$

corresponding to Laguerre functions (Laguerre polynomials multiplied by a truncated exponential kernel) and also equal to the probability density function of the Gamma distribution having a Laplace transform of the form

$$H_{\mathrm{composed}}(q; \ \mu) = \int_{t=-\infty}^{\infty} \left(*_{k=1}^{K} h_{\exp}(t; \ \mu_k) \right) \mathrm{e}^{-qt} \, \mathrm{d}t$$
$$= \frac{1}{(1+\mu q)^K} = \bar{U}(q; \ \mu, K). \tag{38}$$

Differentiating the temporal scale-space kernel with respect to time t gives

$$U_t(t; \ \mu, K) = -\frac{(t - (K-1)\mu)}{\mu t} U(t; \ \mu, K) \tag{39}$$

$$U_{tt}(t; \ \mu, K) = \frac{\left((K^2 - 3K + 2) \mu^2 - 2(K-1)\mu t + t^2 \right)}{\mu^2 t^2}$$
$$\times U(t; \ \mu, K), \tag{40}$$

see the second row in Fig. 1 for graphs. The L_1-norms of these kernels are given by

$$\|U_t(\cdot; \ \mu, K)\|_1 = \frac{2\mathrm{e}^{1-K}(K-1)^{K-1}}{\mu \, \Gamma(K)}, \tag{41}$$

$$\|U_{tt}(\cdot; \ \mu, K)\|_1 = 2\mathrm{e}^{-K-\sqrt{K-1}+1}$$
$$\times \left(\mathrm{e}^{2\sqrt{K-1}} \left(K + 2\sqrt{K-1} \right) \left(K - \sqrt{K-1} - 1 \right)^K \right.$$

$$+ \left(K - 2\sqrt{K-1} \right) \left(K + \sqrt{K-1} - 1 \right)^K \right) \Bigg/$$

$$\left((K-2)^2 \sqrt{K-1} \mu^2 \, \Gamma(K) \right). \tag{42}$$

The temporal scale level at level K corresponds to temporal variance $\tau = K\mu^2$ and temporal standard deviation $\sigma = \sqrt{\tau} = \mu\sqrt{K}$.

4.2 Temporal Peak

Consider an input signal defined as a time-causal temporal peak corresponding to filtering a delta function with K_0 first-order integrators with time constants μ coupled in cascade:

$$f(t) = \frac{t^{K_0-1} \, e^{-t/\mu}}{\mu^{K_0} \, \Gamma(K_0)} = U(t; \; \mu, K_0). \tag{43}$$

With regard to the application area of vision, this signal can be seen as an idealized model of an object with temporal duration $\tau_0 = K_0\mu^2$ that first appears and then disappears from the field of view, and modelled on a form to be algebraically compatible with the algebra of the temporal receptive fields. With respect to the application area of hearing, this signal can be seen as an idealized model of a beat sound over some frequency range of the spectrogram, also modelled on a form to be compatible with the algebra of the temporal receptive fields.

Define the temporal scale-space representation by convolving this signal with the temporal scale-space kernel (43) corresponding to K first-order integrators having the same time constants μ

$$L(t; \; \mu, K) = (U(\cdot; \; \mu, K) * f(\cdot))(t; \; \mu, K)$$
$$= \frac{e^{-\frac{t}{\mu}} \mu^{-K-K_0} t^{K+K_0-1}}{\Gamma(K+K_0)} = U(t; \; \mu, K_0 + K) \tag{44}$$

where we have applied the semi-group property that follows immediately from the corresponding Laplace transforms

$$\bar{L}(q; \; \mu, K) = \frac{1}{(1+\mu q)^K} \frac{1}{(1+\mu q)^{K_0}} = \frac{1}{(1+\mu q)^{K_0+K}}$$
$$= \bar{U}(q; \; \mu, K_0 + K). \tag{45}$$

By differentiating the temporal scale-space representation (44) with respect to time t, we obtain

$$L_t(t; \; \mu, K) = \frac{(\mu(K+K_0-1) - t)}{\mu t} L(t; \; \mu, K) \tag{46}$$

$$L_{tt}(t; \; \mu, K) = \left(\mu^2 \left(K^2 + K(2K_0-3) + K_0^2 - 3K_0 + 2 \right) \right.$$

$$\left. -2\mu t(K+K_0-1) + t^2 \right) \frac{L(t; \; \mu, K)}{\mu^2 t^2} \tag{47}$$

implying that the maximum point is assumed at

$$t_{\max} = \mu(K + K_0 - 1) \tag{48}$$

and the inflection points at

$$t_{\text{inflect1}} = \mu \left(K + K_0 - 1 - \sqrt{K + K_0 - 1} \right), \tag{49}$$

$$t_{\text{inflect2}} = \mu \left(K + K_0 - 1 + \sqrt{K + K_0 - 1} \right). \tag{50}$$

This form of the expression for the time of the temporal maximum implies that the temporal delay of the underlying peak $t_{\max,0} = \mu(K_0 - 1)$ and the temporal delay of the temporal scale-space kernel $t_{\max,U} = \mu(K - 1)$ are not fully additive, but instead composed according to

$$t_{\max} = t_{\max,0} + t_{\max,U} + \mu. \tag{51}$$

If we define the temporal duration d of the peak as the distance between the inflection points, if furthermore follows that this temporal duration is related to the temporal duration $d_0 = 2\mu\sqrt{K_0 - 1}$ of the original peak and the temporal duration $d_U = 2\mu\sqrt{K - 1}$ of the temporal scale-space kernel according to

$$d = t_{\text{inflect2}} - t_{\text{inflect1}} = 2\mu\sqrt{K + K_0 - 1}$$
$$= \sqrt{d_0^2 + d_U^2 + 4\mu^2}. \tag{52}$$

Notably these expressions are not scale invariant, but instead strongly dependent on a preferred temporal scale as defined by the time constant μ of the primitive first-order integrators that define the uniform distribution of the temporal scales.

Scale-normalized temporal derivatives When using temporal scale normalization by variance-based normalization, the first- and second-order scale-normalized derivatives are given by

$$L_\zeta(t; \; \mu, K) = \sigma^\gamma L_t(t; \; \mu, K) = (\mu\sqrt{K})^\gamma L_t(t; \; \mu, K) \tag{53}$$

$$L_{\zeta\zeta}(t; \; \mu, K) = \sigma^{2\gamma} L_{tt}(t; \; \mu, K) = (\mu^2 K)^\gamma L_{tt}(t; \; \mu, K) \tag{54}$$

where $\sigma = \sqrt{\tau}$, $\tau = K\mu^2$ and with $L_t(t; \; \mu, K)$ and $L_{tt}(t; \; \mu, K)$ according to (46) and (47).

When using temporal scale normalization by L_p-normalization, the first- and second-order scale-normalized derivatives are on the other hand given by Lindeberg [75, Equation (75)]

$$L_\zeta(t; \; \mu, K) = \alpha_{1,p}(\mu, K) L_t(t; \; \mu, K) \quad (55)$$

$$L_{\zeta\zeta}(t; \; \mu, K) = \alpha_{2,p}(\mu, K) L_{tt}(t; \; \mu, K) \quad (56)$$

with the scale normalization factors $\alpha_{n,p}(\mu, K)$ determined such that the L_p-norm of the scale-normalized temporal derivative computation kernel

$$L_{\zeta^n}(\cdot; \; \mu, K) = \alpha_{n,p}(\mu, K) h_{t^n}(\cdot; \; \mu, K) \quad (57)$$

equals the L_o-norm of some other reference kernel, where we here take the L_p-norm of the corresponding Gaussian derivative kernels [75, Equation (76)]

$$\|\alpha_{n,p}(\mu, K) h_{t^n}(\cdot; \; \mu, K)\|_p = \alpha_{n,p}(\mu, K) \|h_{t^n}(\cdot; \; \mu, K)\|_p$$
$$= \|g_{\xi^n}(\cdot; \; \tau)\|_p = G_{n,p} \quad (58)$$

for $\tau = \mu^2 K$, thus implying

$$L_\zeta(t; \; \mu, K) = \frac{G_{1,p}}{\|U_t(\cdot; \; \mu, K)\|_p} L_t(t; \; \mu, K) \quad (59)$$

$$L_{\zeta\zeta}(t; \; \mu, K) = \frac{G_{2,p}}{\|U_{tt}(\cdot; \; \mu, K)\|_p} L_{tt}(t; \; \mu, K) \quad (60)$$

where $G_{1,p}$ and $G_{2,p}$ denote the L_p-norms (7) of corresponding Gaussian derivative kernels for the value of γ at which they become constant over scales by L_p-normalization, and the L_p-norms $\|U_t(\cdot; \; \mu, K)\|_p$ and $\|U_{tt}(\cdot; \; \mu, K)\|_p$ of the temporal scale-space kernels U_t and U_{tt} for the specific case of $p = 1$ are given by (41) and (42).

Temporal scale selection Let us assume that we want to register that a new object has appeared by a scale-space extremum of the scale-normalized second-order derivative response.

To determine the temporal moment at which the temporal event occurs, we should formally determine the time where $\partial_\tau(L_{\zeta\zeta}(t; \; \mu, K)) = 0$, which by our model (54) would correspond to solving a third-order algebraic equation. To simplify the problem, let us instead approximate the temporal position of the peak in the second-order derivative by the temporal position of the peak t_{max} according to (48) in the signal and study the evolution properties over scale K of

$$L_{\zeta\zeta}(t_{max}; \; \mu, K) = L_{\zeta\zeta}(\mu(K + K_0 - 1); \; \mu, K). \quad (61)$$

In the case of variance-based normalization for a general value of γ, we have

$$L_{\zeta\zeta}(\mu(K + K_0 - 1); \; \mu, K)$$
$$= -\frac{K^\gamma \mu^{2\gamma-3} e^{-K-K_0+1}(K + K_0 - 1)^{K+K_0-2}}{\Gamma(K + K_0)} \quad (62)$$

and in the case of L_p-normalization for $p = 1$

$$L_{\zeta\zeta}(\mu(K + K_0 - 1); \; \mu, K)$$
$$= -C(K - 2)^2\sqrt{K - 1}e^{\sqrt{K-1}-K_0}\Gamma(K)\mu^{-K-K_0-1}$$
$$\times (\mu(K+K_0-1))^{K+K_0}/(K + K_0 - 1)^2 \Gamma(K + K_0)/$$
$$2\left(e^{2\sqrt{K-1}}\left(K + 2\sqrt{K - 1}\right)\left(K - \sqrt{K - 1} - 1\right)^K\right.$$
$$\left. + \left(K - 2\sqrt{K - 1}\right)\left(K + \sqrt{K - 1} - 1\right)^K\right). \quad (63)$$

To determine the scale $\hat{K}$ at which the local maximum is assumed, let us temporarily extend this definition to continuous values of K and differentiate the corresponding expressions with respect to K. Solving the equation

$$\partial_K(L_{\zeta\zeta}(\mu(K + K_0 - 1); \; \mu, K) = 0 \quad (64)$$

numerically for different values of K_0 then gives the dependency on the scale estimate $\hat{K}$ as function of K_0 shown in Table 1 for variance-based normalization with either $\gamma = 3/4$ or $\gamma = 1$ and L_p-normalization for $p = 1$.

As can be seen from the results in Table 1, when using variance-based scale normalization for $\gamma = 3/4$, the scale estimate $\hat{K}$ closely follows the scale K_0 of the temporal peak and does therefore imply a good approximate transfer of the scale selection property (14) to this temporal scale-space concept. If one would instead use variance-based normalization for $\gamma = 1$ or L_p-normalization for $p = 1$, then that would, however, lead to substantial overestimates of the temporal duration of the peak.

Furthermore, if we additionally normalize the input signal to having unit contrast, then the corresponding time-causal correspondence to the post-normalized magnitude measure (20)

$$L_{\zeta\zeta,\text{maxmagn,postnorm}}\big|_{\gamma=1}$$
$$= \tau^{1/4} L_{\zeta\zeta,\text{maxmagn}}\big|_{\gamma=3/4} = \frac{K}{K + K_0 - 1} \quad (65)$$

is for scale estimates proportional to the temporal duration of the underlying temporal peak $\hat{K} \sim K_0$ very close to constant under variations of the temporal duration of the underlying temporal peak as determined by the parameter K_0, thus implying a good approximate transfer of the scale selection property (21).

4.3 Temporal Onset Ramp

Consider an input signal defined as a time-causal onset ramp corresponding to the primitive function of K_0 first-order inte-

Table 1 Numerical estimates of the value of $\hat{K}$ at which the scale-normalized second-order temporal derivative assumes its maximum over temporal scale for a *temporal peak* (with the discrete expression over discrete temporal scales extended to a continuous variation) as function of K_0 and for either (i) variance-based normalization with $\gamma = 3/4$, (iii) variance-based normalization with $\gamma = 1$ and (iv) L_p-normalization with $p = 1$; For the case of variance-based normalization with $\gamma = 3/4$, (ii) the post-normalized magnitude measure $L_{\zeta\zeta,\mathrm{maxmagn,postnorm}}\big|_{\gamma=1}$ according to (20) and at the corresponding scale (i) is also shown

K_0	$\hat{K}$ (var, $\gamma = 3/4$)	$L_{\zeta\zeta,\mathrm{maxmagn,postnorm}}\big|_{\gamma=1}$ (var, $\gamma = 3/4$)	$\hat{K}$ (var, $\gamma = 1$)	$\hat{K}$ (L_p, $p = 1$)
Scale estimate $\hat{K}$ and maximum magnitude $L_{\zeta\zeta,\max}$ from temporal peak (uniform distr)				
4	3.1	0.504	6.1	10.3
8	7.1	0.502	14.1	18.3
16	15.1	0.501	30.1	34.3
32	31.1	0.500	62.1	66.3
64	63.1	0.500	126.1	130.3

Note that the temporal scale estimates $\hat{K}$ do for $\gamma = 3/4$ constitute a good approximation of the temporal scale $\hat{K}_0$ of the underlying structure and that the maximum magnitude estimates obtained at this temporal scale do for $\gamma = 1$ constitute a good approximation to a scale-invariant constant maximum magnitude measure over temporal scales

grators with time constants μ coupled in cascade:

$$f(t) = \int_{u=0}^{t} \frac{u^{K_0-1}\, e^{-u/\mu}}{\mu^{K_0}\, \Gamma(K_0)}\, du = \int_{u=0}^{t} U(u;\ \mu, K_0)\, du. \tag{66}$$

With respect to the application area of vision, this signal can be seen as an idealized model of a new object with temporal diffuseness $\tau_0 = K_0\, \mu^2$ that appears in the field of view and modelled on a form to be algebraically compatible with the algebra of the temporal receptive fields. With respect to the application area of hearing, this signal can be seen as an idealized model of the onset of a new sound in some frequency band of the spectrogram, also modelled on a form to be compatible with the algebra of the temporal receptive fields.

Define the temporal scale-space representation of the signal by convolution with the temporal scale-space kernel (43) corresponding to K first-order integrators having the same time constants μ

$$L(t;\ \mu, K) = (U(\cdot;\ \mu, K) * f(\cdot))(t;\ \mu, K)$$
$$= \int_{u=0}^{t} U(t;\ \mu, K_0 + K)\, du. \tag{67}$$

Then, the first-order temporal derivative is given by

$$L_t(t;\ \mu, K) = U(t;\ \mu, K_0 + K)$$
$$= \frac{t^{K_0+K-1}\, e^{-t/\mu}}{\mu^{K_0+K}\, \Gamma(K_0 + K)} \tag{68}$$

which assumes its temporal maximum at $t_{\mathrm{ramp}} = \mu(K_0 + K - 1)$.

Temporal scale selection Let us assume that we are going to detect a new appearing object from a local maximum in the first-order derivative over both time and temporal scales. When using variance-based normalization for a general value of γ, the scale-normalized response at the temporal maximum in the first-order derivative is given by

$$L_{\zeta,\max} = L_\zeta(\mu(K_0 + K - 1);\ \mu, K)$$
$$= \sigma^\gamma L_t(\mu(K_0 + K - 1);\ \mu, K)$$
$$= \frac{\left(\sqrt{K}\mu\right)^\gamma (K + K_0 - 1)^{K+K_0-1}e^{-K-K_0+1}}{\mu\, \Gamma(K + K_0)}. \tag{69}$$

When using L_p-normalization for a general value of p, the corresponding scale-normalized response is

$$L_{\zeta,\max} = L_\zeta(\mu(K_0 + K - 1);\ \mu, K)$$
$$= \frac{G_{1,p}}{\|U_t(\cdot;\ \mu, K)\|_p} L_t(\mu(K_0 + K - 1);\ \mu, K) \tag{70}$$

where the L_p-norm of the first-order scale-space derivative kernel can be expressed in terms of exponential functions, the Gamma function and hypergeometric functions, but is too complex to be written out here. Extending the definition of these expressions to continuous values of K and solving the equation

$$\partial_K(L_\zeta(\mu(K + K_0 - 1);\ \mu, K) = 0 \tag{71}$$

numerically for different values of K_0 then gives the dependency on the scale estimate $\hat{K}$ as function of K_0 shown in Table 2 for variance-based normalization with $\gamma = 1/2$ or L_p-normalization for $p = 2/3$.

Table 2 (Columns 2–3) Numerical estimates of the value of $\hat{K}$ at which the scale-normalized first-order temporal derivative assumes its maximum over temporal scale for a *temporal onset ramp* (with the discrete expression over discrete temporal scales extended to a continuous variation) as function of K_0 and for (i) variance-based normalization for $\gamma = 1/2$ and (ii) L_p-normalization for $p = 2/3$; (Columns 4–5) Maximum magnitude values $L_{\zeta,\max}$ at the corresponding temporal scales, with the magnitude values defined by (iii) variance-based normalization for $\gamma = 1$ and (iv) L_p-normalization for $p = 1$

K_0	$\hat{K}$ (var, $\gamma = 1/2$)	$\hat{K}$ (L_p, $p = 2/3$)	$L_{\zeta,\max}$ (var, $\gamma = 1$)	$L_{\zeta,\max}$ (L_p, $p = 1$)
Scale estimate $\hat{K}$ and maximum magnitude $L_{\zeta,\max}$ from temporal ramp (uniform distr)				
4	3.2	3.6	0.282	0.254
8	7.2	7.7	0.282	0.272
16	15.2	15.8	0.282	0.277
32	31.2	31.8	0.282	0.279
64	63.2	64.0	0.282	0.281

Note that for $\gamma = 1/2$ as well as for $p = 2/3$ the temporal scale estimates $\hat{K}$ constitute a good approximation of the temporal scale K_0 of the underlying onset ramp as well as that the scale-normalized maximum magnitude estimates $L_{\zeta,\max}$ computed for $\gamma = 1$ and $p = 1$ constitute a good approximation to a scale-invariant constant magnitude measure over temporal scales

As can be seen from the numerical results, for both variance-based normalization and L_p-normalization with corresponding values of γ and p, the numerical scale estimates in terms of $\hat{K}$ closely follow the diffuseness scale of the temporal ramp as parameterized by K_0. Thus, for both of these scale normalization models, the numerical results indicate an approximate transfer of the scale selection property (14) to this temporal scale-space model. Additionally, the maximum magnitude values according to (69) can according to Stirling's formula $\Gamma(n+1) \approx (n/e)^n \sqrt{2\pi n}$ be approximated by

$$L_{\zeta,\max} \approx \frac{\sqrt{K}}{\sqrt{2\pi}\sqrt{K + K_0 - 1}} \tag{72}$$

and are very stable under variations of the diffuseness scale K_0 of the ramp, and thus implying a good transfer of the scale selection property (31) to this temporal scale-space concept.

4.4 Temporal Sine Wave

Consider a signal defined as a sine wave

$$f(t) = \sin \omega_0 t. \tag{73}$$

This signal can be seen as a simplified model of a dense temporal texture with characteristic scale defined as the wavelength $\lambda_0 = 2\pi/\omega$ of the signal. In the application area of vision, this can be seen as an idealized model of watching some oscillating visual phenomena or watching a dense texture that moves relative the gaze direction. In the area of hearing, this could be seen as an idealized model of temporally varying frequencies around some fixed frequency in the spectrogram corresponding to vibrato.

Define the temporal scale-space representation of the signal by convolution with the temporal scale-space kernel (43) corresponding to K first-order integrators with equal time constants μ coupled in cascade

$$L(t; \ \mu, K) = (U(\cdot; \ \mu, K) * f(\cdot))(t; \ \mu, K)$$
$$= |\hat{U}(\omega_0; \ \mu, K)| \sin\left(\omega_0 t + \arg \hat{U}(\omega; \ \mu, K)\right) \tag{74}$$

where $|\hat{U}(\omega_0; \ \mu, K)|$ and $\arg \hat{U}(\omega; \ \mu, K)$ denote the magnitude and the argument of the Fourier transform $\hat{h}_{\text{composed}}(\omega; \ \mu, K)$ of the temporal scale-space kernel $U(\cdot; \ \mu, K)$ according to

$$\hat{U}(\omega; \ \mu, K) = \frac{1}{(1 + i\,\mu\,\omega)^K}, \tag{75}$$

$$|\hat{U}(\omega; \ \mu, K)| = \frac{1}{\left(1 + \mu^2\,\omega^2\right)^{K/2}}, \tag{76}$$

$$\arg \hat{U}(\omega; \ \mu, K) = -K \arctan(\mu\,\omega). \tag{77}$$

By differentiating (74) with respect to time t, it follows that the magnitude of the nth-order temporal derivative is given by

$$L_{t^n,\text{ampl}} = \frac{\omega_0^n}{\left(1 + \mu^2\,\omega_0^2\right)^{K/2}}. \tag{78}$$

Temporal scale selection Using variance-based temporal scale normalization, the magnitude of the corresponding scale-normalized temporal derivative is given by

$$L_{\zeta^n,\text{ampl}} = \sigma^n L_{t^n,\text{ampl}} = \frac{(K\mu^2)^{n\gamma/2}\omega_0^n}{\left(1 + \mu^2\,\omega_0^2\right)^{K/2}}. \tag{79}$$

Extending this expression to continuous values of K and differentiating with respect to K implies that the maximum over scale is assumed at scale

$$\hat{K} = \frac{\gamma n}{\log\left(1 + \mu^2\omega_0^2\right)} \tag{80}$$

with the following series expansion for small values of ω_0 corresponding to temporal structures of longer temporal duration

$$\hat{K} = \frac{\gamma n}{\mu^2\omega_0^2} + \frac{\gamma n}{2} - \frac{1}{12}\omega_0^2\left(\gamma\mu^2 n\right) + O\left(\omega_0^4\right). \tag{81}$$

Expressing the corresponding scale estimate $\hat{\sigma}$ in terms of dimension length and parameterized in terms of the wavelength $\lambda_0 = 2\pi/\omega_0$ of the sine wave

$$
\begin{aligned}
\hat{\sigma} = \sqrt{\hat{\tau}} = \mu\sqrt{\hat{K}} &= \mu\sqrt{\frac{\gamma n}{\log\left(1 + \frac{4\pi^2\mu^2}{\lambda_0^2}\right)}} \\
&= \frac{\sqrt{\gamma n}}{2\pi}\lambda_0\left(1 + \frac{\pi^2\mu^2}{\lambda_0^2} + O\left(\left(\frac{\mu}{\lambda_0}\right)^4\right)\right)
\end{aligned} \tag{82}
$$

we can see that the dominant term $\frac{\sqrt{\gamma n}\lambda_0}{2\pi}$ is proportional to the temporal duration of the underlying structures in the signal and in agreement with the corresponding scale selection property (34) of the scale-invariant non-causal Gaussian temporal scale-space concept, whereas the overall expression is not scale invariant.

If the wavelength λ_0 is much longer than the time constant μ of the primitive first-order integrators, then the scale selection properties in this temporal scale-space model will constitute a better approximation of the corresponding scale selection properties in the scale-invariant non-causal Gaussian temporal scale-space model.

The maximum value over scale is

$$L_{\zeta^n,\text{ampl,max}} = e^{-\frac{\gamma n}{2}}\omega_0^n\mu^{\gamma n}\left(\frac{\gamma n}{\log\left(1 + \mu^2\omega_0^2\right)}\right)^{\frac{\gamma n}{2}} \tag{83}$$

with the following series expansion for large $\lambda_0 = 2\pi/\omega_0$ and $\gamma = 1$:

$$
\begin{aligned}
L_{\zeta^n,\text{ampl,max}} = e^{-n/2}n^{n/2} \\
\times\left(1 + \frac{\pi^2\mu^2 n}{\lambda_0^2} + \frac{\pi^4\mu^4 n(3n - 10)}{6\lambda_0^4}\right. \\
\left. + O\left(\left(\frac{\mu}{\lambda_0}\right)^6\right)\right).
\end{aligned} \tag{84}
$$

Again we can note that the first term agrees with the corresponding scale selection property (35) for the scale-invariant non-causal Gaussian temporal scale space, whereas the higher-order terms are not scale invariant.

Figure 2 shows graphs of the scale estimate $\hat{\sigma}$ according to (82) for $n = 2$ and $\gamma = 3/4$ and the maximum response over scale $L_{\zeta^n,\text{ampl,max}}$ for $n = 2$ and $\gamma = 1$ as function of the wavelength λ_0 of the sine wave (marked in blue). For comparison, we also show the corresponding scale estimates (34) and magnitude values (35) that would be obtained using temporal scale selection in the scale-invariant non-causal Gaussian temporal scale space (marked in brown).

As can be seen from the graphs, both the temporal scale estimate $\hat{\sigma}(\lambda_0)$ and the maximum magnitude $L_{\zeta,\text{ampl,max}}(\lambda_0)$ obtained from a set of first-order integrators with equal time constants coupled in cascade approach the corresponding results obtained from the non-causal Gaussian scale space for larger values of λ_0 in relation to the time constant μ of the first-order integrators. The scale estimate obtained from a set of first-order integrators with equal time constants is, however, for lower values of λ_0 generally significantly higher than the scale estimates obtained from a non-causal Gaussian temporal scale space. The scale-normalized magnitude values, which should be constant over scale for a scale-invariant temporal scale space when $\gamma = 1$ according to the scale selection property (35), are for lower values of λ_0 much higher than the scale-invariant limit value when performing scale selection in the temporal scale-space concept obtained by coupling a set of first-order integrators with equal time constants in cascade. The scale selection properties (34) and (35) are consequently not transferred to this temporal scale-space concept for the sine wave model, which demonstrates the need for using a scale-invariant temporal scale-space concept when formulating mechanisms for temporal scale selection. The scale selection properties of such a scale-invariant time-causal temporal concepts will be analysed in Sect. 5, and showing that it is possible to obtain temporal scale estimates for a dense sine wave that are truly proportional to the wavelength of the signal, i.e. a characteristic estimate of the temporal duration of the temporal structures in the signal.

Concerning this theoretical analysis, it should be noted that we have here for the expressions (82), (83) and (84) disregarded the rounding of the continuous value $\hat{K}$ in (81) to the nearest integer upwards or downwards where it assumes its maximum value over temporal scales. Thereby, the graphs in Fig. 2 may appear somewhat different if such quantization effects because of discrete temporal scale levels are also included. The lack of true temporal scale invariance will, however, still prevail.

Concerning the motivation to the theoretical analysis in this section, while the purpose of this analysis has been to investigate how the temporal scale estimates depend on
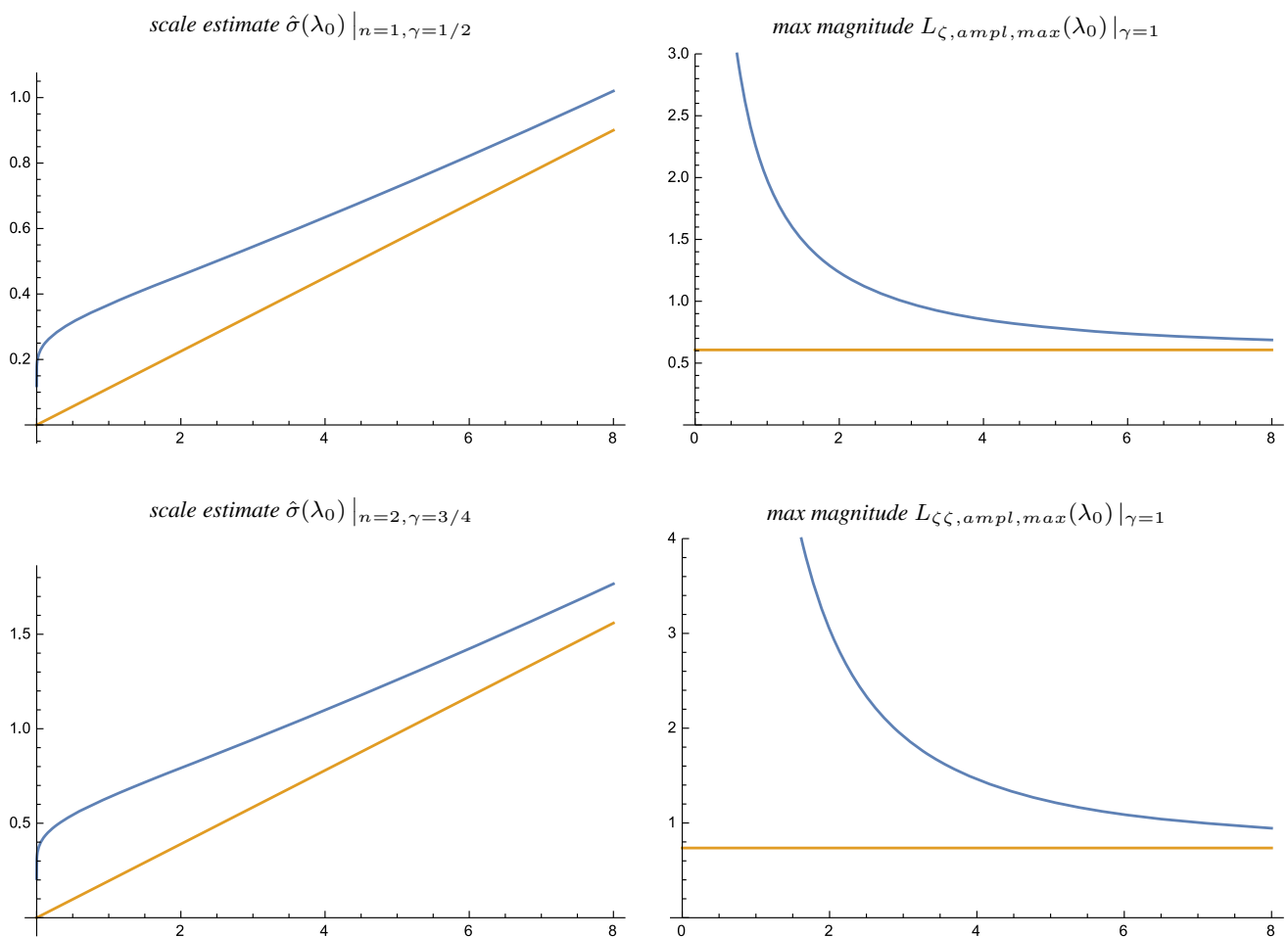
Fig. 2 (*Left column*) Temporal scale estimates $\hat{\sigma}$ (marked in *blue*) obtained from a *sine wave* using the maximum value over scale of the amplitude of either (*top row*) first-order scale-normalized temporal derivatives or (*bottom row*) second-order scale-normalized temporal derivatives and as function of the wavelength λ_0 for variance-based temporal scale normalization with $\gamma = 1/2$ for first-order derivatives and $\gamma = 3/4$ for second-order derivatives. Note that for this temporal scale-space concept based on truncated exponential kernels with equal time constants, the temporal scale estimates $\hat{\sigma}$ do not constitute a good approximation to the temporal scale estimates being proportional to the wavelength λ_0 of the underlying sine wave. Instead, the temporal scale estimates are affected by a systematic scale-dependent

bias. (*Right column*) The maximum scale-normalized magnitude for (*top row*) first-order scale-normalized temporal derivatives or (*bottom row*) second-order scale-normalized temporal derivatives as function of the wavelength λ_0 using variance-based temporal scale normalization with $\gamma = 1$ (marked in *blue*). The *brown curves* show corresponding scale estimates and magnitude values obtained from the scale-invariant but not time-causal Gaussian temporal scale space. Thus, when the temporal scale-space concept based on truncated exponential kernels with equal time constants coupled in cascade is applied to a sine wave, the maximum magnitude measures over temporal scales do not at all constitute a good approximation of temporal scale invariance (*horizontal axis* wavelength λ_0 in units of μ)

the frequency or the wavelength of the signal, it should be emphasized that the primary purpose has not been to develop a method for only estimating the frequency or the wavelength of a sine wave. Instead, the primary purpose has been to carry out a closed-form theoretical analysis of the properties of temporal scale selection when applied to a model signal for which such closed-form theoretical analysis can be carried out. Compared to using a Fourier transform for estimating the local frequency content in a signal, it should be noted that the computation of a Fourier transform requires a complementary parameter—a window

scale over which the Fourier transform is to be computed. The frequency estimate will then be an average of the frequency content over the entire interval as defined by the window scale parameter. Using the proposed temporal scale selection methodology it is on the other hand possible to estimate the temporal scale without using any complementary window scale parameter. Additionally, the temporal scale estimate will be instantaneous and not an average over multiple cycles of a periodic signal, see also the later experimental results that will be presented in Sect. 7 in particular Fig. 7.

5 Scale Selection Properties for the Time-Causal Temporal Scale-Space Concept Based on the Scale-Invariant Time-Causal Limit Kernel

In this section, we will analyse the scale selection properties for the time-causal scale-space concept based on convolution with the scale-invariant time-causal limit kernel.

The analysis starts with a detailed study of a sine wave, for which closed-form theoretical analysis is possible and showing that the selected temporal scale level $\hat{\sigma}$ measured in units of dimension time [time] according to $\hat{\sigma} = \sqrt{\hat{\tau}}$ will be proportional to the wavelength of the signal, in accordance with true scale invariance. We also show that despite the discrete nature of the temporal scale levels in this temporal scale-space concept, local extrema over scale will neverthe-less be preserved under scaling transformations of the form $\lambda_1 = c^j \lambda_0$, with c denoting the distribution parameter of the time-causal limit kernel.

Then, we present a general result about temporal scale invariance that holds for temporal derivatives of any order and for any input signal, showing that under a temporal scaling transformation of the form $t' = c^j t$, local extrema over scales are preserved under such temporal scaling transformations with the temporal scale estimates transforming according to $\tau' = c^j \tau$. We also show that if the scale normalization power $\gamma = 1$ corresponding to $p = 1$, the scale-normalized mag-nitude responses will be preserved in accordance with true temporal scale invariance.

5.1 Time-Causal Temporal Scale Space Based on the Scale-Invariant Time-Causal Limit Kernel

Given the temporal scale-space model based on truncated exponential kernels (36) coupled in cascade

$$h_{\text{composed}}(\cdot; \ \mu) = *_{k=1}^{K} h_{\exp}(\cdot; \ \mu_k), \tag{85}$$

having a composed Fourier transform of the form

$$H_{\text{composed}}(q; \ \mu) = \int_{t=-\infty}^{\infty} *_{k=1}^{K} h_{\exp}(\cdot; \ \mu_k)(t) \, e^{-qt} \, dt$$
$$= \prod_{k=1}^{K} \frac{1}{1 + \mu_k q}, \tag{86}$$

and as arises from the assumptions of (i) linearity, (ii) tem-poral shift invariance, (iii) temporal causality and (iv) non-creation of new local extrema or equivalently zero-crossings with increasing scale, it is more natural to distribute the tem-poral scale levels logarithmically over temporal scales

$$\tau_k = c^{2(k-K)} \tau_{\max} \qquad (1 \leq k \leq K) \tag{87}$$

so that the distribution in terms of effective temporal scale $\tau_{\text{eff}} = \log \tau$ [55] becomes uniform. This implies that time constants of the individual first-order integrators should for some $c > 1$ be given by [73,75]

$$\mu_1 = c^{1-K} \sqrt{\tau_{\max}}, \tag{88}$$
$$\mu_k = \sqrt{\tau_k - \tau_{k-1}} = c^{k-K-1} \sqrt{c^2 - 1} \sqrt{\tau_{\max}} \quad (2 \leq k \leq K). \tag{89}$$

Specifically, if one lets the number of temporal scale lev-els tend to infinity with the density of temporal scale levels becoming infinitely dense towards $\tau \to 0$, it can be shown that this leads to a *scale-invariant time-causal limit kernel* having a Fourier transform of the form [75, Section 5]

$$\hat{\Psi}(\omega; \ \tau, c) = \lim_{K \to \infty} \hat{h}_{\exp}(\omega; \ \tau, c, K)$$
$$= \prod_{k=1}^{\infty} \frac{1}{1 + i \, c^{-k} \sqrt{c^2 - 1} \sqrt{\tau} \, \omega}. \tag{90}$$

5.2 Temporal Sine Wave

Consider an input signal defined as a sine wave

$$f(t) = \sin \omega_0 t, \tag{91}$$

and taken as an idealized model of a oscillating signal with temporal structures having characteristic temporal duration $\lambda_0 = 2\pi/\omega_0$.

For the time-causal temporal scale-space defined by con-volution with the time-causal semi-group $\Psi(t; \ \tau, c)$, the temporal scale-space representation $L(t; \ \tau, c)$ is given by

$$L(t; \ \tau, c) = \left| \hat{\Psi}(\omega_0; \ \tau, c) \right| \sin \left(\omega_0 t + \arg \hat{\Psi}(\omega_0; \ \tau, c) \right) \tag{92}$$

where the magnitude $|\hat{\Psi}(\omega; \ \tau, c)|$ and the argument $\arg \hat{\Psi}(\omega; \ \tau, c)$ of the Fourier transform $\hat{\Psi}(\omega; \ \tau, c)$ of the time-causal limit kernel are given by

$$|\hat{\Psi}(\omega; \ \tau, c)| = \prod_{k=1}^{\infty} \frac{1}{\sqrt{1 + c^{-2k}(c^2 - 1)\tau \, \omega^2}}, \tag{93}$$

$$\arg \hat{\Psi}(\omega; \ \tau, c) = \sum_{k=1}^{\infty} \arctan \left(c^{-k} \sqrt{c^2 - 1} \sqrt{\tau} \, \omega \right). \tag{94}$$

Thus, the magnitude on the nth-order temporal derivative is given by

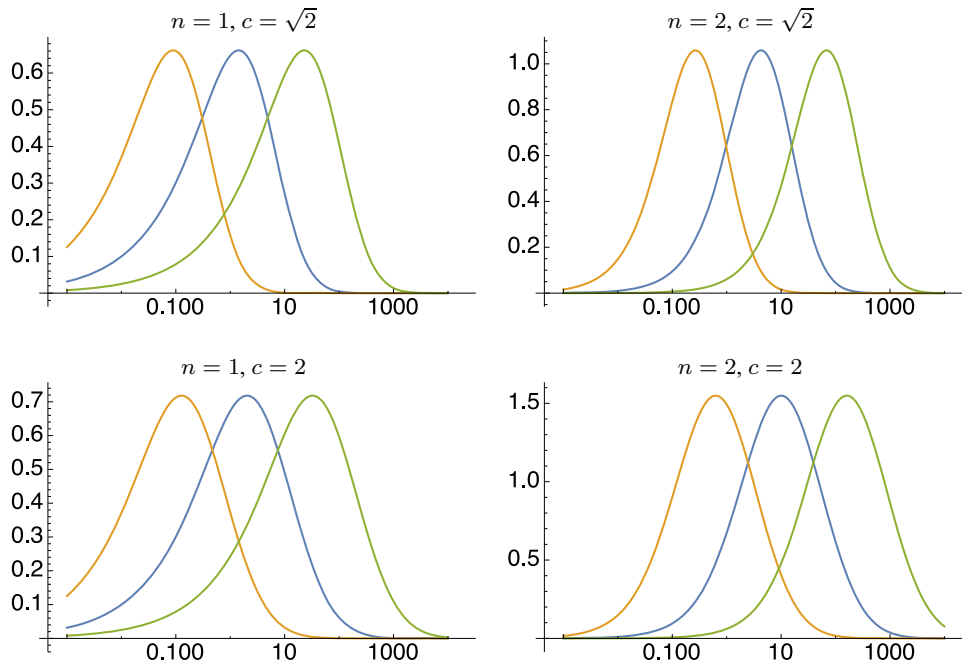$$L_{t^n, \text{ampl}} = \omega^n \, |\hat{\Psi}(\omega; \ \tau, c)| \tag{95}$$

Fig. 3 Scale-space signatures showing the variation over scales of the amplitude of scale-normalized temporal derivatives in the scale-space representation of *sine waves* with angular frequencies $\omega_0 = 1/4$ (*green curves*), $\omega_0 = 1$ (*blue curves*) and $\omega_0 = 4$ (*brown curves*) under convolution with the time-causal limit kernel approximated by the slowest $K = 32$ temporal smoothing steps and for $\gamma = 1$ corresponding to $p = 1$. (*Left column*) First-order temporal derivatives $n = 1$. (*Right column*) Second-order temporal derivatives $n = 2$. (*Top row*) Distribu-

tion parameter $c = \sqrt{2}$. (*Bottom row*) Distribution parameter $c = 2$. Note how these graphs reflect temporal scale invariance in the sense that (i) a variation in the angular frequency of the underlying signal corresponds to a mere shift of the scale-space signature to either finer or coarser temporal scales and (ii) the maximum magnitude response is independent of the frequency of the underlying signal (*horizontal axis* Temporal scale τ on a logarithmic axis)

Fig. 4 Corresponding results as in Fig. 3 above but with (*left column*) $\gamma = 1/2$ for first-order derivatives and (*right column*) $\gamma = 3/4$ for second-order derivatives. Note that the use of scale normalization powers $\gamma < 1$ implies that the maxima over temporal scales are moved to finer temporal scales and that the (uncompensated) maximum magnitude responses are no longer scale invariant
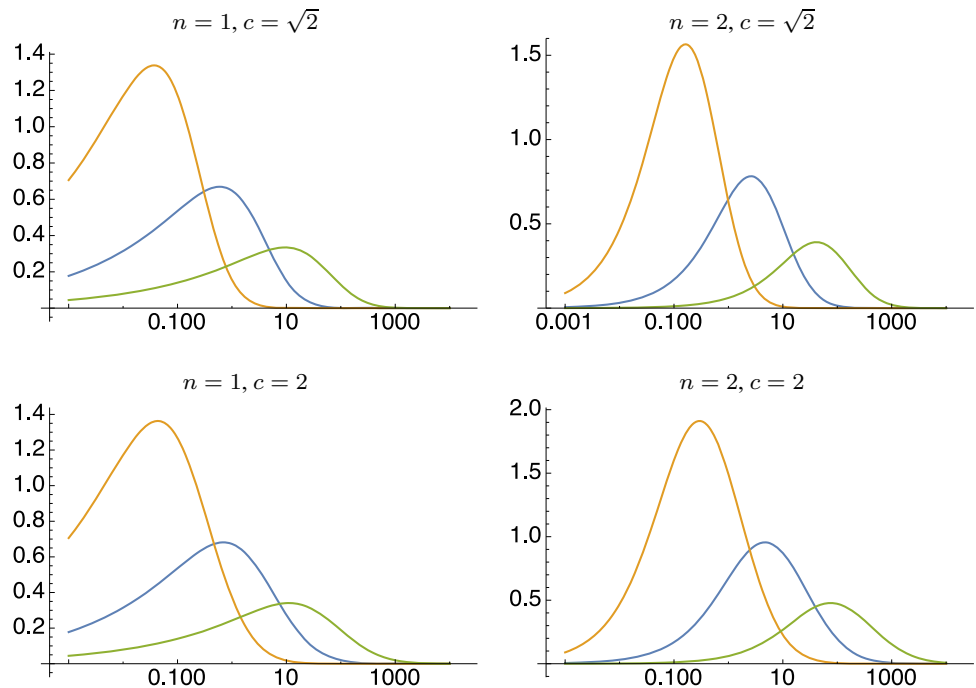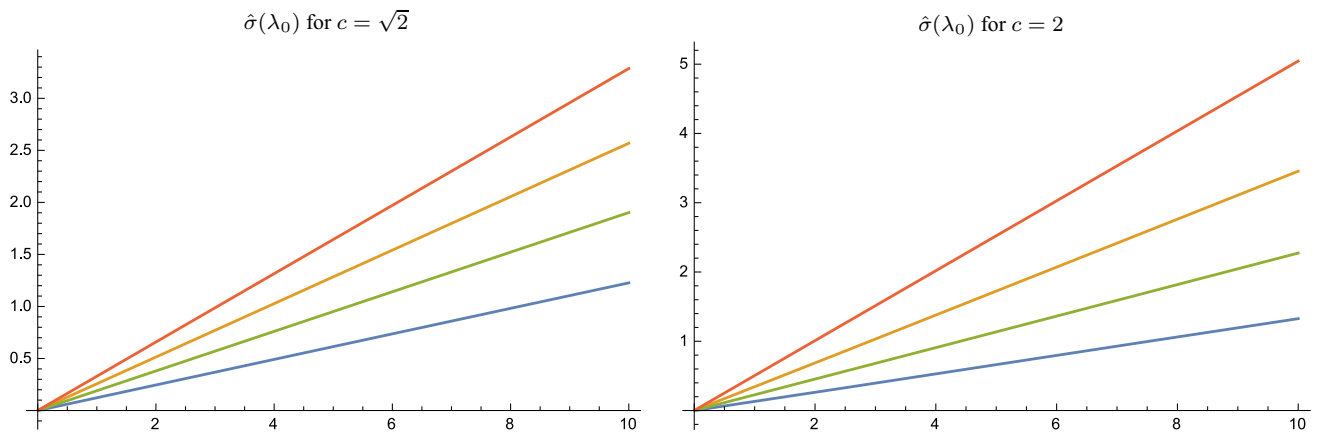
Fig. 5 Scale estimates $\hat{\sigma} = \sqrt{\hat{\tau}}$ as function of the wavelength λ_0 for *sine waves* of different angular frequencies $\omega_0 = 2\pi/\lambda_0$ and different orders n of temporal differentiation and different values of γ: (*blue curves*) $n = 1$ and $\gamma = 1/2$, (*brown curves*) $n = 2$ and $\gamma = 3/4$, (*green curves*) $n = 1$ and $\gamma = 1$, (*red curves*) $n = 2$ and $\gamma = 1$. Note how the scale estimates expressed in dimensions of time $\sigma = \sqrt{\tau}$

are directly proportional to the wavelength λ_0 of the sine wave and in agreement with temporal scale invariance (*horizontal axis* wavelength λ_0) (for $c = \sqrt{2}$ the time-causal limit kernel has been approximated by the slowest $K = 32$ temporal smoothing stages and for $c = 2$ by the slowest $K = 12$ temporal smoothing stages)

and for the nth-order scale-normalized derivate based on variance-based scale normalization the amplitude as function of scale is

$$
\begin{aligned}
L_{\zeta^n,\text{ampl}} &= \tau^{n\gamma/2} \omega^n |\hat{\Psi}(\omega;\ \tau, c)| \\
&= \tau^{n\gamma/2} \omega^n \prod_{k=1}^{\infty} \frac{1}{\sqrt{1 + c^{-2k}(c^2 - 1)\tau\,\omega^2}}.
\end{aligned} \tag{96}
$$

Figure 3 shows graphs of the variation of this entity as function of temporal scale for different angular frequencies ω_0, orders of temporal differentiation n and the distribution parameter c. As can be seen from the graphs, the maxima over scales are assumed at coarser scales with increasing wavelength of the sine wave. The maxima over temporal scale are also assumed at coarser scales for second-order derivatives than for first-order derivatives.

Specifically, when $\gamma = 1$ the magnitude values at the local extrema over scale are constant over scale, which implies a transfer of the scale selection property (10) to this temporal scale-space model. Notably, this situation is in clear contrast to the situation for the temporal scale-space generated by first-order integrators with equal time constants coupled in cascade. In Fig. 2 it was shown that because of the lack of true temporal scale invariance of that temoral scale-space concept, the maximum magnitude values are not constant over scales for $\gamma = 1$ as they should be according to the scale-invariant scale selection property (10).

When choosing lower values of γ as motivated from the determination of the parameter γ for scale selection in a Gaussian scale space to make the scale estimate reflect the width of a Gaussian peak for second-order derivatives or reflect the width of a diffuse ramp for first-order derivatives, which leads to $\gamma = 1/2$ for first-order derivatives (27) and

$\gamma = 3/4$ for second-order derivatives (16), the local extrema over scale are moved to finer scales (see Fig. 4). Then, however, the maximum magnitude values are no longer the same for sine waves of different frequencies, implying that a complementary magnitude normalization step is necessary (see Sect. 5.3 for additional details).

Local extrema over temporal scale Taking the logarithm of the expression (96) gives

$$
\begin{aligned}
\log L_{\zeta^n,\text{ampl}} = {}& \frac{n\gamma}{2} \log \tau + n \log \omega_0 \\
&- \frac{1}{2} \sum_{k=1}^{\infty} \log\left(1 + c^{-2k}(c^2 - 1)\tau\,\omega_0^2\right).
\end{aligned} \tag{97}
$$

If we treat τ as a continuous variable and differentiate with respect to τ we obtain

$$
\partial_\tau \left(\log L_{\zeta^n,\text{ampl}}\right) = \frac{n\gamma}{2\tau} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{c^{-2k}(c^2 - 1)\,\omega_0^2}{1 + c^{-2k}(c^2 - 1)\,\tau\,\omega_0^2}. \tag{98}
$$

Figure 5 shows graphs of how the scale estimate $\hat{\sigma}$ obtained by setting the derivative with respect to temporal scale to zero increases linearly with the wavelength λ_0 of the signal, with different slopes of the linear curve depending on the order of temporal differentiation and the value of the scale normalization parameter γ. This overall linear scaling behaviour can directly be proved by rewriting the expression $\partial_\tau \left(\log L_{\zeta^n,\text{ampl}}\right) = 0$ in (98) into

$$
\frac{n\gamma}{\tau\omega_0^2} - \sum_{k=1}^{\infty} \frac{c^{-2k}(c^2 - 1)}{1 + c^{-2k}(c^2 - 1)\,\tau\,\omega_0^2} = 0. \tag{99}
$$

Table 3 Scale estimates $\hat{\sigma}$ computed from local extrema over scale for a sine wave with angular frequency $\omega_0 = 1$ and different orders of temporal differentiation n and different scale normalization parameters γ with $\gamma_n = 1/2$ for $n = 1$ and $\gamma_n = 3/4$ for $n = 2$, and different values of the distribution parameter c (for $c = \sqrt{2}$ the time-causal limit kernel has been approximated by the slowest $K = 32$ temporal smoothing stages and for $c = 2$ by the slowest $K = 12$ temporal smoothing stages)

	$n = 1$	$n = 2$
Scale estimates $\hat{\sigma}$ for $c = \sqrt{2}$		
$\gamma = 1$	1.20	2.06
$\gamma = \gamma_n$	0.77	1.61
Scale estimates $\hat{\sigma}$ for $c = 2$		
$\gamma = 1$	1.43	3.17
$\gamma = \gamma_n$	0.83	2.17

Table 4 Ratios between the scale estimates $\hat{\sigma}$ in Table 3 and $\sqrt{n\gamma}$ computed from local extrema over scale for a sine wave with angular frequency $\omega_0 = 1$

	$n = 1$	$n = 2$
Ratios $\hat{\sigma}/\sqrt{n\gamma}$ for $c = \sqrt{2}$		
$\gamma = 1$	1.20	1.46
$\gamma = \gamma_n$	1.09	1.32
Ratios $\hat{\sigma}/\sqrt{n\gamma}$ for $c = 2$		
$\gamma = 1$	1.43	2.24
$\gamma = \gamma_n$	1.18	1.77

For the corresponding entities obtained from scale selection for a sine wave in a non-causal Gaussian scale space, this ratio is equal to one for all combinations of n and γ. These scale selection entities thus reveal a larger deviation from a Gaussian behaviour both for larger values of c and for temporal derivatives of higher order

Since this expression is a direct function of the dimensionless entity $\tau\omega_0^2$, it follows that the temporal scale estimates will be of the form

$$\hat{\tau} = \frac{\varphi(n\gamma, c)}{\omega_0^2} \tag{100}$$

for some function $\varphi(n\gamma, c)$, and thus obeying temporal scale invariance in the sense that the scale estimate in dimension length is proportional to the wavelength of the signal

$$\hat{\sigma} = \sqrt{\hat{\tau}} = \frac{\sqrt{\varphi(n\gamma, c)}}{2\pi} \lambda_0. \tag{101}$$

Notice how this situation is in contrast to the results of scale selection in the temporal scale-space concept obtained by coupling first-order integrators with equal time constants in cascade, where the scale estimate for a sine wave is not directly proportional to the wavelength of the temporal signal, but also affected by a wavelength-dependent temporal scale bias (see Eq. (82) and Fig. 2 in Sect. 4.4).

Table 3 shows numerical values of the differences between the results for different orders of differentiation n and different values of γ. These numerical entities become particularly illuminating by forming the ratio $\hat{\sigma}/\sqrt{n\gamma}$ as shown in Table 4. For a non-causal Gaussian scale space, this ratio should be equal to one for all combinations of n and γ [see Eq. (33)]. For this non-causal temporal scale space, we can, however, note that the deviation from one increases both with larger values of the distribution parameter[2] c and with increasing order of temporal differentiation n, which both lead to larger

degrees of temporal asymmetry due to the non-causal temporal dimension.

In the essential proportionality of the scale estimate $\hat{\sigma}$ to the wavelength λ_0 of the signal according to (101), the main component of the scale selection property (34) is thereby transferred to this temporal scale-space concept, although the proportionality constant has to be modified depending on the value of the temporal scale distribution parameter c, the order of temporal differentiation n and the scale normalization parameter γ.

Preservation of local extrema over temporal scale under temporal scaling transformations Let us assume that the continuous magnitude function (97) assumes a maximum over temporal scales for some pair (ω_0, τ_0) and that the derivative with respect to temporal scale is thereby zero

$$\partial_\tau \left(\log L_{\zeta^n, \text{ampl}} \right)\big|_{(\omega_0, \tau_0)}$$
$$= \frac{n\gamma}{2\tau_0} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{c^{-2k}(c^2 - 1)\,\omega_0^2}{1 + c^{-2k}(c^2 - 1)\,\tau_0\,\omega_0^2} = 0. \tag{102}$$

Let us next assume that we feed in a different sine wave with wavelength $\lambda_1 = c^j \lambda_0$ for some integer j (for the same value of c as used in the definition of the time-causal limit kernel) and corresponding to $\omega_1 = c^{-j} \omega_0$ with its matching scale $\tau_1 = c^{2j}\tau_0$. Then, it holds that

$$\partial_\tau \left(\log L_{\zeta^n, \text{ampl}} \right)\big|_{(\omega_1, \tau_1)}$$
$$= \frac{n\gamma}{2\tau_1} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{c^{-2k}(c^2 - 1)\,\omega_1^2}{1 + c^{-2k}(c^2 - 1)\,\tau_1\,\omega_1^2}$$
$$= c^{-2j} \left(\frac{n\gamma}{2\tau_0} - \frac{1}{2} \sum_{k=1}^{\infty} \frac{c^{-2k}(c^2 - 1)\,\omega_0^2}{1 + c^{-2k}(c^2 - 1)\,\tau_0\,\omega_0^2} \right)$$
$$= c^{-2j} \, \partial_\tau \left(\log L_{\zeta^n, \text{ampl}} \right)\big|_{(\omega_0, \tau_0)}. \tag{103}$$

---

[2] The reason why the these ratios depend on the distribution parameter c can specifically be explained by observing that the temporal duration of the temporal derivatives of the time-causal limit kernel $\Psi_{t^\alpha}(t; \ \tau, c)$ will depend on the distribution parameter c—see "Appendix 3" for a derivation and explicit estimates of the width of the temporal derivatives of the time-causal limit kernel.

This result implies that the sign of the derivate with respect to temporal scale is preserved between matching angular frequencies and scales (ω_0, τ_0) and (ω_1, τ_1). Specifically, local extrema over temporal scales are preserved under uniform scaling transformations of the temporal domain $t' = c^j t$, implying scale covariance of the temporal scales that are selected from local extrema over scales of scale-normalized temporal derivatives.

Note also that although the analysis in Eq. (103) is performed based on a temporary extension of τ into a continuous variable, the scale covariance still holds when the continuous function is sampled into a discrete set of temporal scale levels.[3]

5.3 General Scale Invariance Property Under Temporal Scaling Transformations

In Lindeberg [75, Appendix 3] it is shown that for two temporal signals f and f' that are related by a temporal scaling transform $f'(t') = f(t)$ for $t' = c^{j'-j} t$ with the corresponding transformation between corresponding temporal scale levels $\tau' = c^{2(j'-j)} \tau$, the scale-normalized temporal derivatives defined by either L_p-normalization or variance-based normalization in the scale-space representation defined by convolution with the time-causal limit kernel $\Psi(t;\ \tau, c)$ are *for any temporal input signal f* related according to

$$
\begin{aligned}
L'_{\zeta'^n}(t';\ \tau', c) &= c^{(j'-j)n(\gamma-1)} L_{\zeta^n}(t;\ \tau, c) \\
&= c^{(j'-j)(1-1/p)} L_{\zeta^n}(t;\ \tau, c).
\end{aligned}
\tag{104}
$$

This result specifically implies that the scale-space signatures, which are the graphs that show the variation in the strength of scale-normalized derivatives over scale, will be rescaled copies of each other for signals that are related by a uniform scaling transformation of the temporal domain.

Specifically, local temporal scale estimates $\hat{\tau}$ and $\hat{\tau}'$ as determined from local extrema over temporal scales in the two temporal domains will be assumed at corresponding temporal scale levels and will thus be transformed in a scale-covariant way for any temporal scaling transformation of the form $t' = c^{j'-j} t$. In units of the temporal variance, it holds that

$$
\hat{\tau}' = c^{2(j'-j)} \hat{\tau}
\tag{105}
$$

and in units of the temporal standard deviation

$$
\hat{\sigma}' = c^{j'-j} \hat{\sigma}.
\tag{106}
$$

If $\gamma = 1$ corresponding to $p = 1$, the magnitude values at corresponding temporal scale levels will be equal. If $\gamma \neq 1$ corresponding to $p \neq 1$, the magnitude values will be related according to (104). Thereby, this expression provides a way to normalize maximum strength measures between local extrema over scales assumed at different temporal scales as obtained e.g. in the scale-space signatures shown in Fig. 4.

Note that by this construction we have been able to transfer the temporal scale invariance property (9) and (10) that holds for a non-causal Gaussian temporal scale-space concept to also hold for a time-causal temporal scale-space concept, which is a novel type of theoretical construction. This property is, however, restricted to the temporal scale-space concept based on convolution with the time-causal limit kernel and does, for example, not hold for the time-causal temporal scale-space concept based on convolution with a cascade of truncated exponential kernels having equal time constants and corresponding to a uniform distribution of the temporal scale levels in units of the composed temporal variance.

6 Influence of Discrete Temporal Scale Levels on the Theoretical Analysis

In the theoretical analysis of scale selection properties of (i) a temporal peak in Sect. 4.2, (ii) a temporal onset ramp in Sect. 4.3 and (ii) a temporal sine wave in Sect. 4.4 and Sect. 5.2, we did first compute closed-form expressions for how the scale-normalized temporal magnitude measures depend upon the temporal scale levels $\tau_K = K \mu^2$ according to (62), (69) and (79) for the time-causal scale-space concept based on truncated exponential kernels with equal time constant coupled in cascade or how the scale-normalized magnitude measure depends on the temporal scale level $\tau_K = c^{2K} \tau_0$ according to (96) for the time-causal temporal scale-space concept based on the scale-invariant limit kernel:

$$
\begin{aligned}
M_{\text{peak,uni}}(K) &= -L_{\zeta\zeta}(\mu(K + K_0 - 1);\ \mu, K) \\
&= \frac{K^\gamma \mu^{2\gamma-3} e^{-K-K_0+1} (K + K_0 - 1)^{K+K_0-2}}{\Gamma(K + K_0)},
\end{aligned}
\tag{107}
$$

---

[3] A formal proof of the transfer of this preservation property of local extrema over temporal scales from a temporary extension of the temporal scale parameter τ into a continuous variable back into a restricted discrete set of temporal scale levels can be stated as follows: From Eq. (103) it follows that the continuous temporal scale-space signatures for the two sine waves of wavelengths λ_0 and $\lambda_1 = c^j \lambda_0$ will increase and decrease respectively at corresponding matching temporal scale levels τ_0 and $\tau_1 = c^{2j} \tau_0$. If we next sample these scale-space signatures at some discrete set of temporal scale levels $\tau_{0,k} = c^{2k}$ and $\tau_{1,k} = c^{2j} c^{2k}$, then it follows that the discrete maxima over temporal scales will also be related according to $\tau_{1,\text{max}} = c^{2j} \tau_{0,\text{max}}$. This preservation property of local extrema does, however, only hold for temporal scaling factors S that are integer powers of the distribution parameter c, i.e. only $S = c^j$. Alternatively, this preservation property can also be derived from the more general scale invariance property under temporal scaling transformations (104) that is stated in next section.

$$M_{\text{onset,uni}}(K) = L_\zeta(\mu(K_0 + K - 1); \; \mu, K)$$

$$= \frac{\left(\sqrt{K}\mu\right)^\gamma (K + K_0 - 1)^{K+K_0-1} e^{-K-K_0+1}}{\mu \, \Gamma(K + K_0)}, \qquad (108)$$

$$M_{\text{sine,uni}}(K) = L_{\zeta^n,\text{ampl}}(K)$$

$$= \frac{(K\mu^2)^{n\gamma/2} \omega_0^n}{\left(1 + \mu^2 \omega_0^2\right)^{K/2}}, \qquad (109)$$

$$M_{\text{sine,limit}}(K) = L_{\zeta^n,\text{ampl}}(c^{2K}\tau_0)$$

$$= (c^{2K}\tau_0)^{n\gamma/2} \omega^n \prod_{k=1}^{\infty} \frac{1}{\sqrt{1 + c^{-2k}(c^2-1)c^{2K}\tau_0\,\omega^2}}. \qquad (110)$$

Then, to compute the temporal scale levels at which the scale-normalized derivative responses assumed their maximum values over temporal scales, we temporarily extended these magnitude measures from being defined over discrete integer temporal scale levels K to a continuum over K, to be able to differentiate the closed-form expressions with respect to the temporal scale level.

A general question that could be raised in this context therefore concerns how good approximation the results from the continuous approximation of local extrema over scales are with respect to a setting where the temporal scale levels are required to be discrete. A common property of the four types of scale-space signatures according to equations (107)–(109) and shown in Fig. 6 is that they are unimodal, i.e., they assume a single maximum over temporal scales and do first increase and then decrease. Thereby, when the continuous variable K_c with its associated maximum over temporal scales $\hat{K}_c$ obtained from a continuous analysis is in a second stage restricted to be discrete, it follows that the discrete maximum over discrete temporal scales $\hat{K}_d$ is guaranteed to be assumed at either the nearest lower or the nearest higher integer. Thus, we obtain the discrete temporal scale estimate by rounding the continuous scale estimate $\hat{K}_c$ to either the nearest lower or the nearest higher integer.

Whether the value should be rounded upwards or downwards depends on how close the continuous estimate $\hat{K}_c$ is to the nearest downwards versus upwards integers and on the local degree of asymmetry of the scale-space signature around the maximum over temporal scales.

When implementing and executing a temporal scale selection algorithm in practice, the situation can on the other hand be reverse. Given a set of discrete temporal scale levels, we may detect a local maximum over temporal scales at some discrete temporal scale level $\hat{K}_d$. If we would like to use this temporal scale estimate for estimating the temporal duration of the underlying temporal structure that gave rise to the response, e.g. according to the methodology outlined in "Appendix 3", we may on the other hand would like to com-

pute a better continuous estimate $\hat{K}_c$ of the temporal scale level than as restricted by the discrete temporal scale levels.

A straightforward way of computing a more accurate temporal scale estimate in such a situation is by interpolating a parabola over the measurements over the temporal scale levels in an analogous way as subresolution spatial scale estimates can be obtained over a spatial scale-space representation [62,76,85]. Let (x_0, y_0) denote the scale level and the magnitude measure at the discrete maximum and let (x_0-1, y_{-1}) and (x_0+1, y_1) denote the corresponding scale level and magnitude measure at the nearest lower and upper temporal scales, respectively. Assuming the following form of the interpolating function

$$y(x) = a\,\frac{(x - x_0)^2}{2} + b\,(x - x_0) + c, \qquad (111)$$

the interpolation coefficients become

$$a = y_1 - 2y_0 + y_{-1} \qquad (112)$$

$$b = (y_1 - y_{-1})/2 \qquad (113)$$

$$c = y_0 \qquad (114)$$

with the corresponding subresolution estimate of the maximum over scales

$$\hat{x} = x_0 - \frac{b}{a} = x_0 - \frac{y_1 - y_{-1}}{2(y_1 - 2y_0 + y_{-1})}. \qquad (115)$$

Note that the correction offset $\Delta x = -b/a$ is restricted to the interval $\Delta x \in [-1/2, 1/2]$ implying that the location x_0 of the discrete maximum is guaranteed to be on the sampling grid point x_i nearest to the subresolution estimate $\hat{x}$.

7 Temporal Scale Selection for 1-D Temporal Signals

To illustrate the derived scale selection properties, we will in this section show the result of applying temporal scale selection to different types of purely temporal signals.

The bottom rows in Fig. 7 shows a one-dimensional model signal having a temporally varying frequency of the form

$$f(t) = \sin\left(\exp\left(\frac{b - t}{a}\right)\right) \qquad (116)$$

defined such that the local wavelength increases with time t. Figure 8 shows the temporal scale-space representation of the scale-normalized second-order temporal $-L_{\zeta\zeta}$ as function of scale for the maximally scale-invariant choice of $\gamma = 1$ corresponding to $p = 1$. Note how structures in the signal of longer temporal duration give rise responses at coarser temporal scales in agreement with the derived theoretical
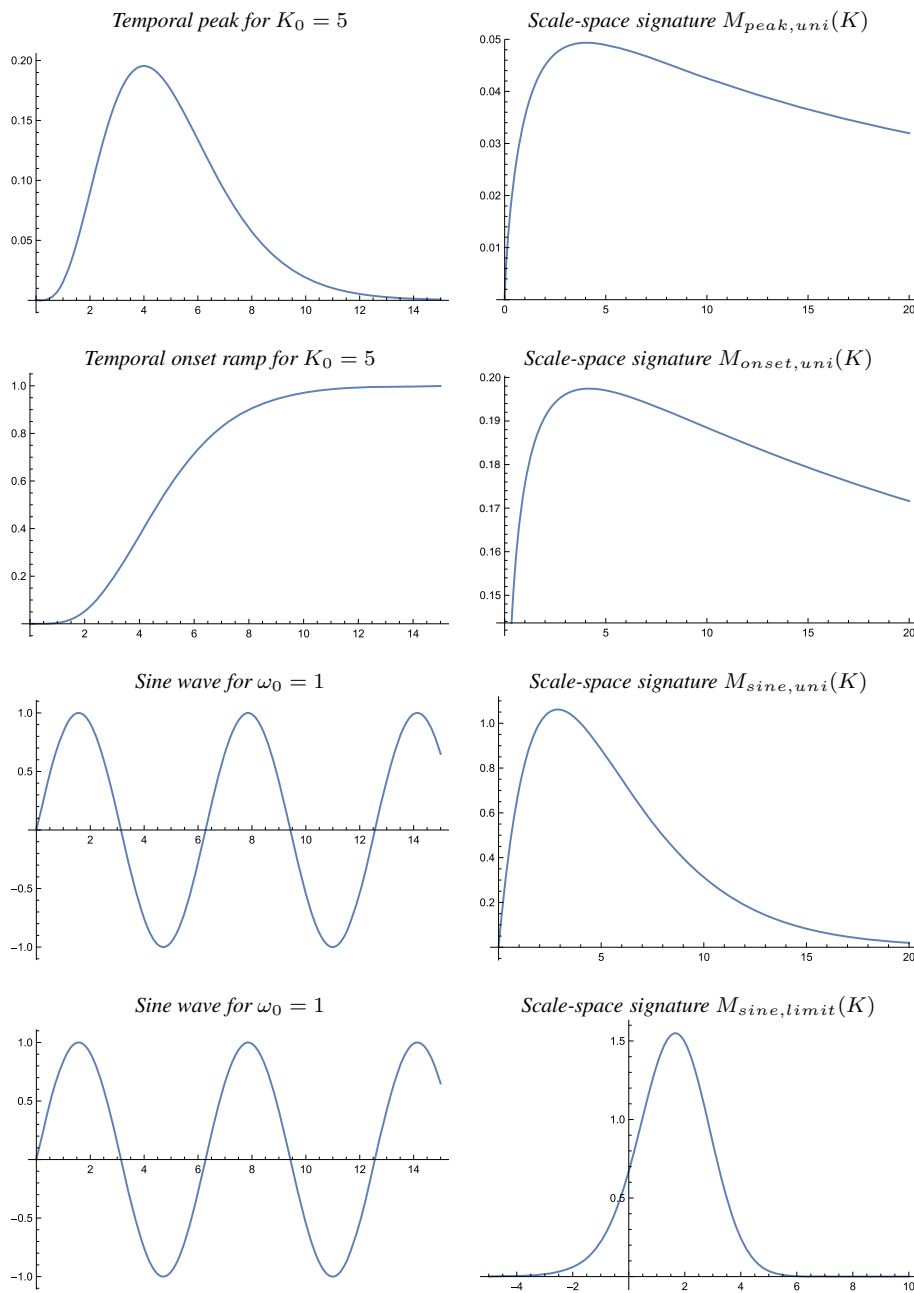
Fig. 6 Graphs of the three types of model signals for which closed-form expressions can be computed for how the scale-normalized derivative-based magnitude responses vary as function of the scale parameter: (*top row*) Time-causal temporal peak model according to (43) with temporal duration determined by $K_0 = 5$ for which the second-order scale-normalized temporal derivative response at the peak varies according to (107) here with $\gamma = 3/4$, (*second row*) Time-causal onset ramp model according to (66) with temporal duration determined by $K_0 = 5$ for which the first-order scale-normalized temporal derivative response at the ramp varies according to (108) here with $\gamma = 1/2$, (*third row*) Temporal sine wave according to (73) with angular frequency $\omega_0 = 1$ for which the amplitude of the second-order temporal derivative response varies according to (109) here with $\gamma = 1$ and (*bottom row*) Temporal sine wave according to (91) for which the ampli-

tude of the second-order temporal response measure varies according to (110) here with $\gamma = 1$. The scale-space signatures in the first three rows have been computed using the time-causal temporal scale-space concept based on truncated exponential kernels with equal time constants $\mu = 1$ coupled in cascade, whereas the scale-space signature in the *bottom row* has been computed using the time-causal temporal scale-space concept obtained by convolution with the scale-invariant limit kernel with distribution parameter $c = 2$. Note specifically that the shape of the scale-space signature in the third row is different from the shape in the *fourth row* because of the uniform distribution of temporal scale levels K in the *third row* and the logarithmic distribution in the *fourth row* (*horizontal axis* for the figures in the *left column*: time t) (*Horizontal axis* for the figures in the *right column*: temporal scale level K)

Scale-space extrema from the time-causal limit kernel

Scale-space extrema from the non-causal Gaussian kernel
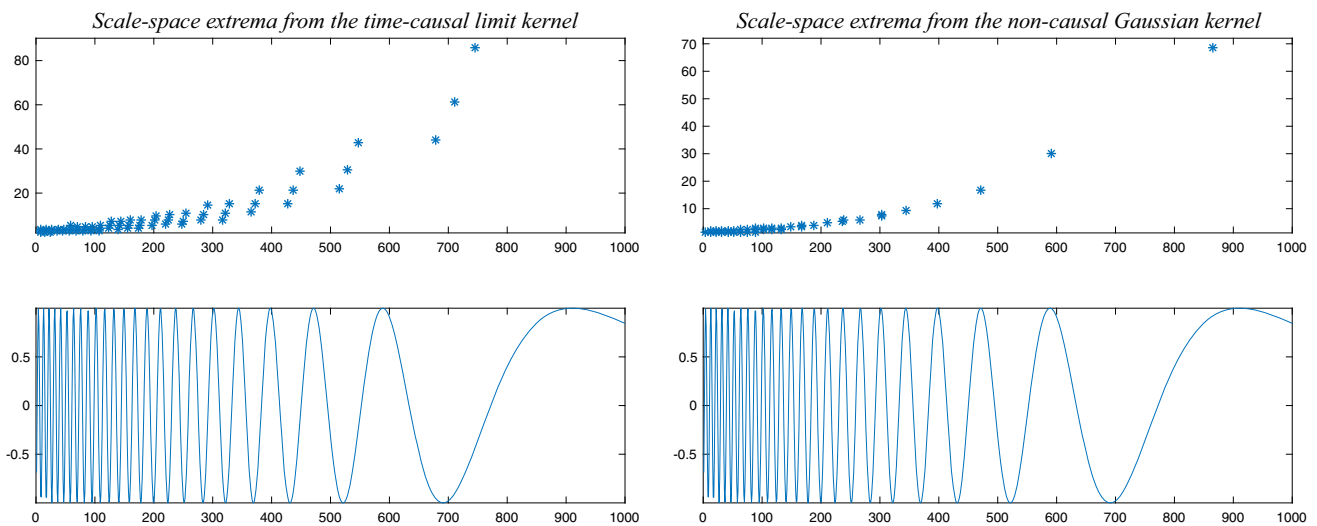

Fig. 7 Temporal scale selection by scale-space extrema detection applied to a synthetic sine wave signal $f(t) = \sin(\exp((b - t)/a))$ for $a = 200$ and $b = 1000$ with temporally varying frequency so that the wavelength increases with time t. (*Top left*) Temporal scale-space maxima of the scale-normalized second-order temporal derivative $-L_{\zeta\zeta}$ detected using the time-causal temporal scale-space concept corresponding to convolution with the time-causal limit kernel for $c = \sqrt{2}$ and with each scale-space maximum marked at the point $(\hat{t}, \hat{\sigma})$ with $\hat{\sigma} = \sqrt{\hat{\tau}}$ at which the scale-space maximum is assumed. (*Top right*)

Temporal scale-space maxima detected using the non-causal Gaussian temporal scale-space concept using five temporal scale levels per scale octave. For both temporal scale-space concepts, the scale-normalized temporal derivatives have been defined using l_p-normalization for $p = 2/3$ and corresponding to $\gamma = 3/4$ for second-order temporal derivatives (*horizontal axis* time t) (*Vertical axis in top row* Temporal scale estimates in units of $\sigma = \sqrt{\tau}$) (*vertical axis in bottom row* Signal strength $f(t)$)

properties of the scale-covariant scale-space concepts over a time-causal vs. a non-causal temporal domain.

In the top rows in Fig. 7 we show the results of applying local temporal scale selection by detecting local maxima over both time and scale of the scale-normalized second-order temporal derivative $-L_{\zeta\zeta}$ for the scale-calibrated choice of $\gamma = 3/4$ corresponding to $p = 2/3$. Each detected scale-space extremum has been marked by a star at the point in scale-space $(\hat{t}, \hat{\sigma}) = (\hat{t}, \sqrt{\hat{\tau}})$ at which the local maximum over temporal scale was assumed. Such temporal scale selection has been performed using two types of temporal scale-space concepts: (i) based on the time-causal scale-space representation corresponding to convolution with the scale-invariant limit kernel with distribution parameter $c = \sqrt{2}$ approximated by a finite number of the at least $K = 8$ slowest primitive smoothing steps at the finest level of scale or (ii) based on the non-causal Gaussian kernel using 5 temporal scale levels per scale octave.

The discrete implementation of the time-causal temporal scale-space representation corresponding to convolution with the scale-invariant limit kernel has been based on recursive filters over time according to the methodology described in Lindeberg [75, section 6.2] whereas the discrete implementation of the non-causal Gaussian temporal scale-space concept has been based on the discrete analogue of the

Gaussian kernel described in Lindeberg [53,54]. Discrete implementation of scale-normalized temporal derivatives has in turn been based on discrete l_p-normalization according to the methodology outlined in [75, section 7].

As can be seen from the results in the top rows in Fig. 7, the temporal scale estimates $\hat{\sigma}$ increase proportional to the local wavelength λ according to the derived scale-invariant properties of the temporal scale-space concepts based on convolution with the time-causal limit kernel or the non-causal Gaussian kernel. The scale estimates obtained using the time-causal versus the non-causal temporal scale-space concepts are, however, not equal in units of the standard deviation $\hat{\sigma}$ of the temporal scale-space kernel—see also the theoretical analysis in Sect. 5 with specifically the numerical comparisons in Table 4. Thus, different scale calibration factors are needed to transform the temporal scale estimates in units of the temporal standard deviation to units of the temporal duration of the signal for the time-causal versus the non-causal Gaussian temporal scale-space concepts—compare with the theoretical analysis in "Appendix 3".

Another qualitative difference that can be noted with regard to temporal scale selection in a time-causal temporal scale-space representation versus a non-causal Gaussian temporal scale-space representation is that any measurement performed in a time-causal temporal scale-space concept is associated with an inherent temporal delay δ, whereas the
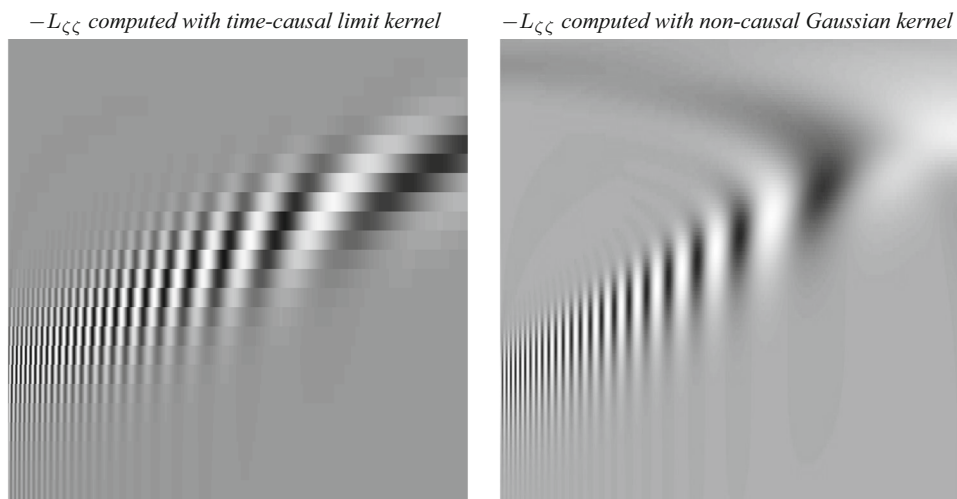
Fig. 8 Temporal scale-space representation of the scale-normalized second-order temporal derivative $-L_{\zeta\zeta}$ computed from a synthetic sine wave signal $f(t) = \sin(\exp((b-t)/a))$ for $a = 200$ and $b = 1000$ with temporally varying frequency so that the wavelength increases with time t. (*Left*) Using the time-causal temporal scale-space concept corresponding to convolution with the time-causal limit kernel for $c = \sqrt{2}$. (*Right*) Using the non-causal Gaussian temporal scale-space concept.

For both temporal scale-space concepts, the scale-normalized temporal derivatives have been defined using l_p-normalization for $p = 1$ and corresponding to $\gamma = 1$. Note how the notion of scale-normalized temporal derivatives implies that stronger responses are obtained at temporal scale levels proportional to the duration of the underlying structures in the temporal signal (*horizontal axis* time t) (*Vertical axis* effective temporal scale $\log \tau$)

temporal delay can be defined to be zero for a non-causal Gaussian temporal scale-space. Thus, the local extrema over scales will be assumed later with increasing temporal scale levels as induced by temporal structures of having longer temporal duration. Varying the parameter q in the scale calibration criteria (14) and (25) to values of $q < 1$ provides a straightforward way of enforcing responses to be obtained at finer temporal scales and thereby implying shorter temporal delays, at the potential cost of a larger likelihood of false positive responses by not detecting the underlying temporal structures at the same temporal scales as they occur.

Yet another side effect of the longer temporal delays at coarser temporal scales is that multiple local responses over scales may be obtained with respect to the same underlying temporal structure, with first responses obtained at finer temporal scales followed by later responses at coarser temporal scales. Because of the temporal shift caused by different temporal delays between adjacent temporal scale levels, scale-space extrema detection over a local 3×3 neighbourhood may not detect a single extremum over temporal scales as for the non-causal Gaussian temporal scale-space concept. Therefore, explicit handling of different temporal delays at different temporal scale levels is needed when performing temporal scale selection in a time-causal temporal scale-space representation. The presented theory of temporal scale selection properties is intended to be generally applicable with respect to different such strategies for handling the temporal delays in specific algorithms.
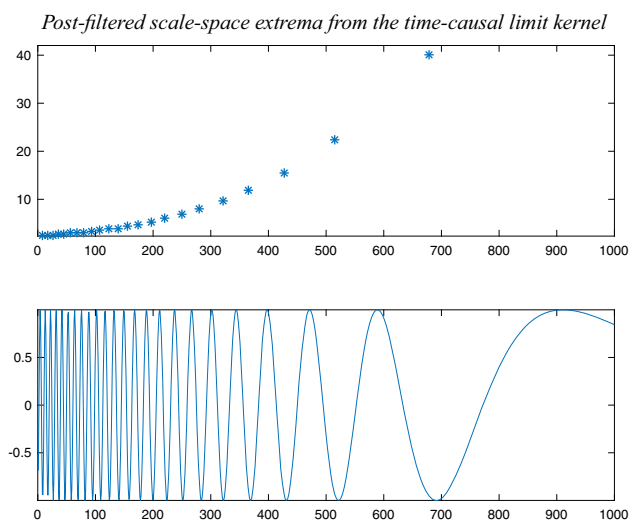

Fig. 9 The result of post-filtering the scale-space extrema shown in Fig. 7 (*left*) to result in a single extremum over temporal scales for each underlying temporal structure

7.1 Post-Filtering of Responses at Adjacent Temporal Scales

In view of the behaviour of image structures over temporal scales illustrated in Fig. 8, one way of suppressing multiple responses to the same underlying structure at different temporal scales is by performing an additional search around each scale-space extremum as follows: if a point $(\hat{t}, \hat{\tau})$ at temporal scale level with temporal scale index k is a scale-

space maximum, perform an additional search at the nearest finer temporal scale level $k-1$ to previous temporal moments $t - j$ as long as the scale-normalized values monotonically increase. When the monotone increase stops and a local temporal maximum has been found, then compare if the temporal maximum value at the nearest finer temporal scale is greater than the temporal maximum value at the current scale. If so, suppress the scale-space maximum at the current scale. In a corresponding manner, perform a search at the next coarser temporal scale $k + 1$ to the following temporal moments $t + j$ as long as the scale-normalized values monotonically increase. When the monotone increase stops and a local temporal maximum has been found, then compare if the temporal maximum value at the next coarser temporal scale is greater

than the temporal maximum value at the current scale. If so, suppress the scale-space maximum at the current scale. By this type of straightforward scale-space tracking over adjacent temporal scales, a single response will be obtained to the same underlying structure as illustrated in Fig. 9. Additionally, a more accurate temporal scale estimate can be computed by performing the parabolic interpolation according to (111) and (115) over the nearest backward and forward temporal maxima at the adjacent finer and coarser temporal scales as opposed to an interpolation over temporal scales at the same temporal moment as at which the temporal scale-space maximum was assumed.

Note that in a real-time situation, the necessary information needed to perform a search to the past can be stored
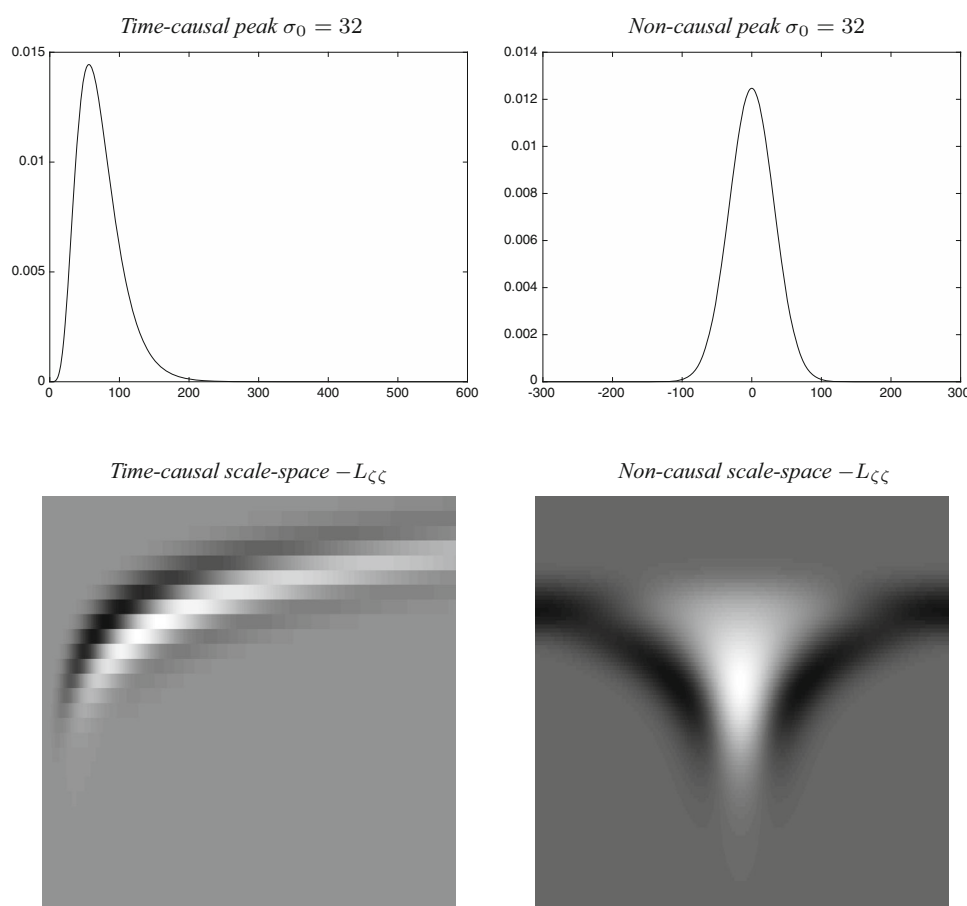

Fig. 10 Illustration of the temporal dynamics over multiple temporal scales that arises when a temporal scale-space representation responds to a temporal peak with temporal standard deviation $\sigma_0 = 32$ for (*left column*) the time-causal representation corresponding to convolution with the time-causal limit kernel and (*right column*) the non-causal Gaussian temporal scale space. (*Top row*) Graphs of the input kernels for the cases of a time-causal peak and a non-causal peak, respectively. (*Bottom row*) Scale-space maps of the scale-normalized second-order temporal derivative $-L_{\zeta\zeta}$. Note that if we slice the time-causal scale-space map vertically at a temporal moment before the full development

of the temporal scale-space maximum at the temporal scale corresponding to the temporal scale of the peak, we will get earlier temporal responses at finer temporal scales, whereas if we slice the time-causal scale-space map at a temporal moment after the full development of the temporal scale-space maximum, we will get later temporal responses at coarser temporal scales. When handling multiple temporal scale levels in a time-causal real-time situation, it is therefore natural to include explicit mechanisms for tracking and handling how the temporal structures evolve over temporal scales (*horizontal axis* time t) (*vertical axis in bottom row* effective temporal scale log τ)
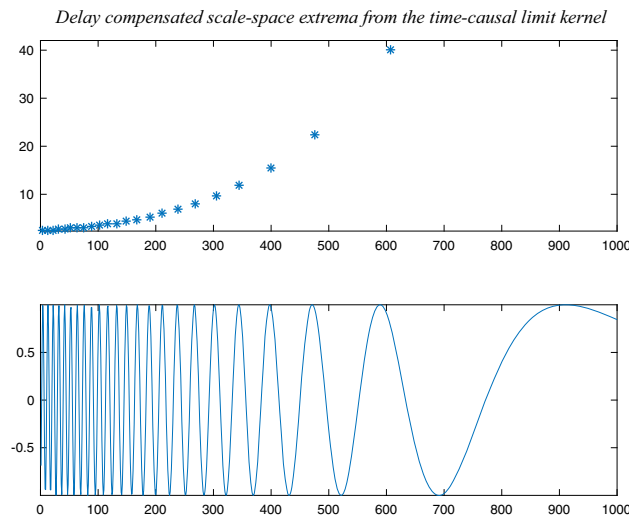
Fig. 11 The result of performing temporal delay compensation of the scale-space extrema shown in Fig. 9 by correcting the temporal moment $\hat{t}$ of the scale-space extremum by an estimate of the temporal delay computed from the position of the temporal maximum of the temporal scale-space kernel according to [75, Section 4]. Note that although the computations by necessity have to be associated with a temporal delay, we can nevertheless compute a good estimate of when the underlying event occurred that gave rise to the feature response

by a process that records the value of temporal maxima at each temporal scale. In a corresponding manner, later deletion or subresolution interpolation of the scale estimate of a scale-space maximum can only be performed when time has passed to the location of the next temporal maximum at the nearest coarser temporal scale. If a preliminary feature response has been obtained at any temporal scale while a potential response at the nearest coarser temporal scale is not available yet because of its longer temporal scale, a real-time system operating over multiple scales should preferably be designed with the ability to correct or adjust preliminary measurements when more information at coarser scales becomes available—see also Fig. 10 for an illustration of the underlying temporal dynamics that arises when processing signals at multiple scales using a time-causal temporal scale-space representation.

While these descriptions have been given regarding scale-space maxima and local maxima, the procedure for handling scale-space minima and local minima is analogous with the polarity of the signal reversed.

7.2 Temporal Delay Compensation

In Fig. 11 we have additionally adjusted the temporal location of every temporal scale-space maximum by an estimate of the temporal delay δ computed from the location of the temporal maximum of the underlying temporal scale-space kernel according to [75, Section 4]. Note that although any temporal event detected at a coarser temporal scale will by necessity be associated with nonzero temporal delay, we can nevertheless retrospectively compute a good estimate of when the underlying event occurred that gave rise to the registered feature response.

7.3 Temporal Scale Selection for a Real Measurement Signal

While a main purpose of this article is to develop a theory of temporal scale selection to be used in conjunction with a spatio-temporal scale-space concept for video analysis or a spectro-temporal scale-space concept for audio analysis, we argue that this theory is applicable to much larger classes of time-dependent measurement signals. For the purpose of isolating the effect to a purely one-dimensional measurement signal, we do in Fig. 12 show the result of applying corresponding temporal scale selection to a real measurement signal showing hourly measurements of the temperature at a weather station. Note how the temporal scale selection method based on scale-space extrema is able to extract the coarse scale temperature peaks although the signal contains substantial high-amplitude variations at finer scales.

For the purpose of having the local feature responses being less dependent on the local phase of the signal than for either first- or second-order temporal derivatives L_ζ or $L_{\zeta\zeta}$, we do in Fig. 13 show the result of computing a quasi quadrature measure

$$\mathcal{Q}_t L = \frac{L_\zeta^2 + C L_{\zeta\zeta}^2}{t^\Gamma} \tag{117}$$

derived in [74] to for

$$C = \frac{1}{\sqrt{(1-\Gamma)(2-\Gamma)}} \tag{118}$$

constitute an improved version of an earlier proposed quasi quadrature measure in Lindeberg [60].

Scale-space extrema from the time-causal limit kernel

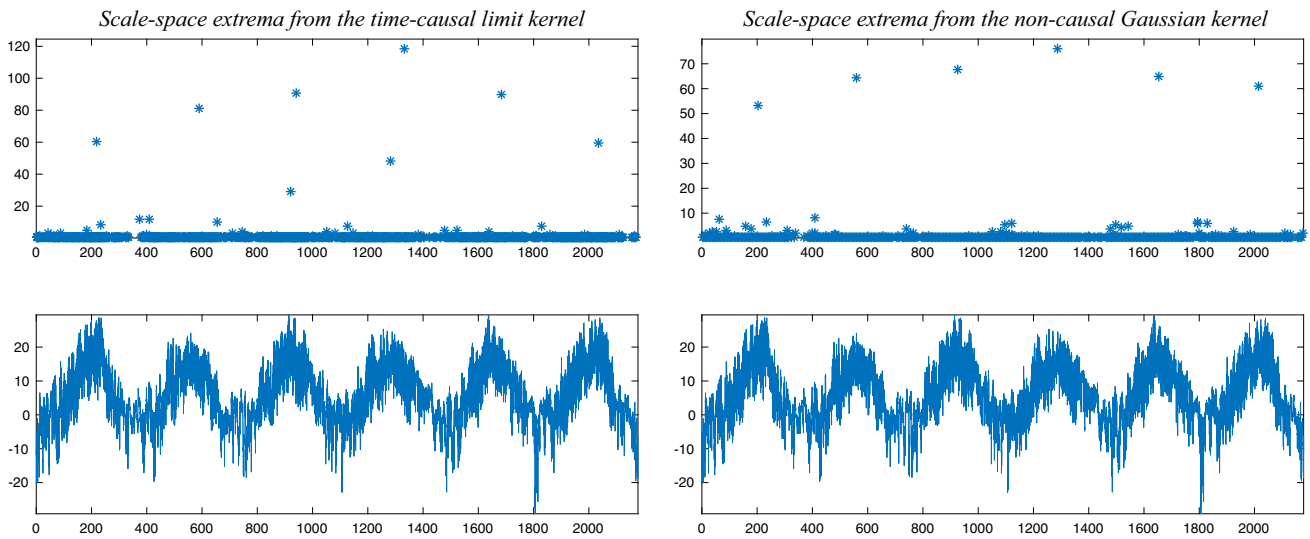*Scale-space extrema from the non-causal Gaussian kernel*

Fig. 12 Temporal scale selection by scale-space extrema detection applied to a measured temperature signal (showing hourly temperature measurements at Tullinge outside Stockholm during the years 1997–2002, open data courtesy of the Swedish Meteorological and Hydrological Institute, SMHI, in Sweden). (*Top left*) Temporal scale-space maxima of the scale-normalized second-order temporal derivative $-L_{\zeta\zeta}$ detected using the time-causal temporal scale-space concept corresponding to convolution with the time-causal limit kernel for $c = \sqrt{2}$ and with each scale-space maximum marked at the delay-compensated point $(\hat{t} - \delta, \hat{\sigma})$ with $\hat{\sigma} = \sqrt{\hat{\tau}}$ at which the scale-space maximum is assumed with the temporal delay estimate δ determined from the tem-

poral location of the peak of the corresponding temporal scale-space kernel at the given scale. (*Top right*) Temporal scale-space maxima detected using the non-causal Gaussian temporal scale-space concept using five temporal scale levels per scale octave. For both temporal scale-space concepts, the scale-normalized temporal derivatives have been defined using discrete l_p-normalization for $p = 2/3$ corresponding to $\gamma = 3/4$ for second-order temporal derivatives. Note how the temporal scale selection method is able to extract the coarse scale temporal phenomena although the signal contains very strong variations at finer temporal scales (*horizontal axis* time t in units of days) (*Vertical axis t in upper figure* temporal scale estimates in units of days)

Quasi quadrature $\sqrt{\mathcal{Q}_t L}$ computed with time-causal limit kernel

Quasi quadrature $\sqrt{\mathcal{Q}_t L}$ computed with non-causal Gaussian kernel
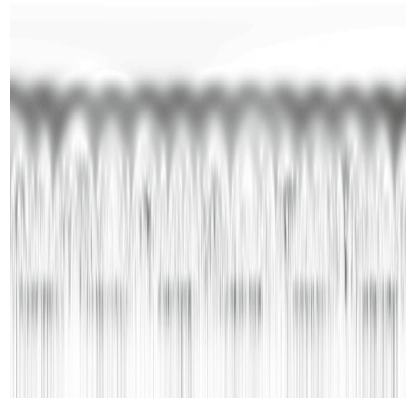

Fig. 13 Temporal scale-space representation of the scale-normalized quasi quadrature measure $\mathcal{Q}_t L = (L_{\zeta}^2 + C L_{\zeta\zeta}^2)/t^\Gamma$ for $C = 1/\sqrt{(1 - \Gamma)(2 - \Gamma)}$, $\gamma = 1$ and $\Gamma = 0$ computed from the temperature signal in Fig. 12. (*Left*) Using the time-causal temporal scale-space concept corresponding to convolution with the time-causal limit kernel for $c = \sqrt{2}$. (*Right*) Using the non-causal Gaussian temporal scale-space concept. For both temporal scale-space concepts, the

scale-normalized temporal derivatives have been defined using discrete l_p-normalization. Note how this operator responds to different types of temporal structures at different temporal scales with a particularly strong response due to temporal variations caused by the annual temperature cycle (*horizontal axis* time t) (*Vertical axis* effective temporal scale log τ)

Note how this operator responds to different types of temporal structures at different temporal scales with a particularly strong response due to the temporal variations caused

by the daily and annual temperature cycles. This effect becomes more immediately noticeable if we complement the above temporal scale selection method with a complemen-
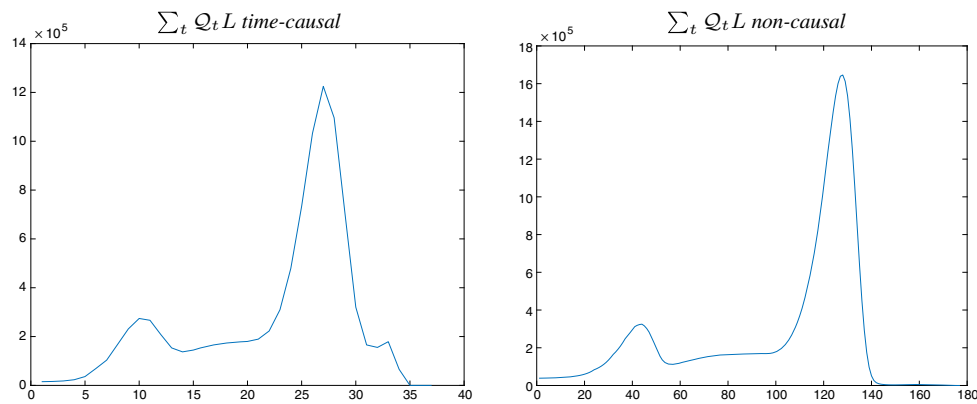
Fig. 14 Graphs that for each temporal scale level σ_k show the sum of the quasi quadrature responses in Fig. 13 over time t. With a complementary stationarity assumption, the peaks over temporal scales reflect the temporal scale levels σ_k at which there is the largest scale-normalized variability in terms of first- and second-order temporal derivative responses. For the time-causal model to the left, the local maxima over temporal scale are assumed at temporal scale levels $\hat{\sigma}_1 = 0.207$ days and $\hat{\sigma}_2 = 71.5$ days, respectively. For the non-causal model to the right, the local maxima over temporal scales are assumed at temporal scale levels $\hat{\sigma}_1 = 0.236$ days and $\hat{\sigma}_2 = 85.2$ days. The ratios between these temporal scale levels are 332 for the time-causal model and 345 for the non-causal model in good qualitative agreement with the finer scale peaks corresponding to the daily cycle and the coarser scale peaks corresponding to the yearly cycle of 365 days (*horizontal axis* temporal scale index k corresponding to a uniform sampling in terms of effective temporal scale $\log \tau$)

tary assumption about stationarity of the signal and sum up the scale-normalized feature responses of the quasi quadrature measure over time for every temporal scale level. In the graphs shown in Fig. 14 we do then obtain two major peaks over temporal scales, with the finer scale peak corresponding to the daily temperature cycle and the coarser scale peak to the annual temperature cycle.

Note that the purpose of this experiment is not primarily to develop an algorithm for detecting periodic variations in a signal but to illustrate that the temporal scale selection mechanism produces intuitively reasonable temporal scale estimates for a real-world 1-D signal with known properties.

7.4 Numerical Accuracy of the Temporal Scale Estimates

To investigate how well the temporal scale estimates generated by the resulting temporal scale selection mechanism reflect the temporal scale in the underlying temporal signal, we generated temporal model signals defined as either (i) a temporal peak modelled as time-causal limit kernel with temporal scale τ_0 and distribution parameter c for different values of τ_0 and c and (ii) a temporal onset ramp modelled as the primitive function of the temporal peak model. Then, we detected scale-space extrema of the second-order scale-normalized temporal derivative $-L_{\zeta\zeta}$ for peak detection or scale-space extrema of the first-order scale-normalized temporal derivative L_ζ for onset detection, using the time-causal temporal scale-space concept corresponding to convolution with the time-causal limit kernel for the same value of c and resulting in temporal scale estimates $\hat{\sigma} = \sqrt{\hat{\tau}}$.

Table 5 Experimental results of the accuracy of the temporal scale estimates when detecting temporal scale-space extrema of the scale-normalized derivative $-L_{\zeta\zeta}$ in a temporal peak defined as a time-causal limit kernel with temporal standard deviation $\sigma_0 = \sqrt{\tau_0}$ for distribution parameter c and then detecting the strongest temporal scale-space maximum with temporal scale estimate $\hat{\sigma} = \sqrt{\hat{\tau}}$ also using the same value of the distribution parameter c and for two ways of defining temporal scale-normalized derivatives by either (i) discrete l_p-normalization or (ii) variance-based normalization

$c = \sqrt{2}$			$c = 2$		
σ_0	$\hat{\sigma}_0\vert_{L_p}$	$\hat{\sigma}_0\vert_{\mathrm{var}}$	σ_0	$\hat{\sigma}_0\vert_{L_p}$	$\hat{\sigma}_0\vert_{\mathrm{var}}$
Temporal scale estimates from temporal peaks					
2	2.26	2.14	2	1.58	2.16
4	4.11	4.05	4	4.57	4.06
8	7.76	8.06	8	8.02	7.96
16	15.90	16.03	16	15.60	15.86
32	31.97	32.06	32	31.35	31.66
64	64.12	64.13	64	63.15	63.31

Tables 5 and 6 show results from this experiment for the time-causal temporal peak and the temporal onset ramp models, respectively. Note how well the theoretical temporal scale selection results derived for truncated exponential kernels with a uniform distribution of the temporal scale levels in Sects. 4.2 and 4.3 do also generalize to truncated exponential kernels with a logarithmic distribution of the temporal scale levels. Table 7 shows corresponding results for the non-causal Gaussian scale-space concept applied to non-causal temporal peak and temporal onset ramp models, respectively, and again with a very good agreement between the local scale

Table 6 Experimental results of the accuracy of the temporal scale estimates when detecting temporal scale-space extrema of the scale-normalized derivative L_ζ in a temporal onset ramp defined as the primitive function of a time-causal limit kernel with temporal standard deviation $\sigma_0 = \sqrt{\tau_0}$ for distribution parameter c and then detecting the strongest temporal scale-space maximum with temporal scale estimate $\hat\sigma = \sqrt{\hat\tau}$ also using the same value of the distribution parameter c and for two ways of defining temporal scale-normalized derivatives by either (i) discrete l_p-normalization or (ii) variance-based normalization

$c = \sqrt{2}$			$c = 2$		
σ_0	$\hat\sigma_0\vert_{L_p}$	$\hat\sigma_0\vert_{\text{var}}$	σ_0	$\hat\sigma_0\vert_{L_p}$	$\hat\sigma_0\vert_{\text{var}}$
Temporal scale estimates from temporal onset ramps					
2	2.01	1.96	2	2.04	1.97
4	3.66	3.90	4	3.68	3.98
8	7.71	8.02	8	7.60	7.94
16	15.68	16.02	16	15.42	15.90
32	32.00	32.04	32	31.48	31.77
64	64.08	64.08	64	63.44	63.54

Table 7 Experimental results of the accuracy of the temporal scale estimates for the non-causal Gaussian temporal scale-space concept when applied to the detection temporal scale-space maxima of the scale-normalized derivative $-L_{\zeta\zeta}$ for a temporal peak and to the scale-normalized derivative L_ζ for onset ramp detection for two ways of defining temporal scale-normalized derivatives by either (i) discrete l_p-normalization or (ii) variance-based normalization

Temporal peak			Temporal onset ramp		
σ_0	$\hat\sigma_0\vert_{L_p}$	$\hat\sigma_0\vert_{\text{var}}$	σ_0	$\hat\sigma_0\vert_{L_p}$	$\hat\sigma_0\vert_{\text{var}}$
Temporal scale estimates from non-causal Gaussian scale space					
2	2.95	1.92	2	1.59	2.07
4	4.04	3.97	4	3.86	4.03
8	8.10	7.98	8	7.92	8.02
16	15.89	15.99	16	15.94	16.01
32	32.06	32.00	32	31.97	32.00
64	64.02	64.00	64	63.97	64.00

estimates $\hat\sigma$ in relation to the inherent temporal scale σ_0 in the signal.

For this experiment, we can specifically note that variance-based temporal scale normalization does on average lead to slightly more accurate temporal scale estimates compared to discrete l_p-normalization, which can be contrasted to previous results regarding spatial scale selection in hybrid pyramids by Lindeberg and Bretzner [76], where l_p-normalization lead to much more accurate scale selection results compared to variance-based normalization.

8 Temporal Scale Selection in Spatio-Temporal Video Data

In this section, we will develop a basic proof of concept of applying the proposed theory and methodology for selecting local temporal scales in video data based on a small set of specific spatio-temporal feature detectors formulated in terms of spatio-temporal differential invariants. A more detailed treatment of this topic with examples for more general families of differential expressions for spatio-temporal scale selection with associated spatio-temporal image models for scale calibration is presented in a companion paper [74].

8.1 Spatio-Temporal Receptive Field Model

For applying the proposed framework for temporal scale selection to spatio-temporal video data, we follow the approach with idealized models of spatio-temporal receptive fields of the form

$$T(x_1, x_2, t; \; s, \tau; \; v, \Sigma) = g(x_1 - v_1 t, x_2 - v_2 t; \; s, \Sigma)\, h(t; \; \tau) \tag{119}$$

as previously derived, proposed and studied in Lindeberg [66,68,75] where

- $x = (x_1, x_2)^T$ denotes the image coordinates,
- t denotes time,
- s denotes the spatial scale,
- τ denotes the temporal scale,
- $v = (v_1, v_2)^T$ denotes a local image velocity,
- Σ denotes a spatial covariance matrix determining the spatial shape of an affine Gaussian kernel $g(x; \; s, \Sigma) = \frac{1}{2\pi s \sqrt{\det \Sigma}} e^{-x^T \Sigma^{-1} x / 2s}$,
- $g(x_1 - v_1 t, x_2 - v_2 t; \; s, \Sigma)$ denotes a spatial affine Gaussian kernel that moves with image velocity $v = (v_1, v_2)$ in space–time and
- $h(t; \; \tau)$ is a temporal smoothing kernel over time.

and we specifically here choose as temporal smoothing kernel over time either (i) the time-causal temporal scale-space kernel corresponding to a set of first-order integrators with equal time constants coupled in cascade (37)

$$h(t; \; \tau) = U(t; \; \mu, K) = \frac{t^{K-1}\, e^{-t/\mu}}{\mu^K\, \Gamma(K)} \tag{120}$$

with $\tau = K\mu^2$ or (ii) the time-causal limit kernel

$$h(t; \tau) = \Psi(t; \tau, c) \tag{121}$$

defined via its Fourier transform of the form (90).

For simplicity, we shall in this treatment restrict ourselves to space–time separable receptive fields obtained by setting the image velocity to zero $v = (v_1, v_2) = (0, 0)$ and to receptive fields that are rotationally symmetric over the spatial domain as obtained by setting the spatial covariance matrix to a unit matrix $\Sigma = I$. The resulting spatio-temporal receptive fields that we then obtain correspond to complementing our time-causal temporal scale-space concepts studied in Sects. 4, 5 with a rotationally symmetric spatial Gaussian scale-space concept over the spatial domain.

Specifically, the natural way of expressing spatio-temporal scale selection mechanisms within this space–time separable spatio-temporal scale-space concept

$$L(x_1, x_2, t; s, \tau) = (T(\cdot, \cdot, \cdot; s, \tau) * f(\cdot, \cdot, \cdot))(x_1, x_2, t; s, \tau) \tag{122}$$

is by studying scale-normalized partial derivates of the form [75, Section 8.5, Equation (108)]

$$L_{x_1^{m_1} x_2^{m_2} t^n, \text{norm}} = s^{(m_1+m_2)\gamma_s/2} \alpha_n(\tau) L_{x_1^{m_1} x_2^{m_2} t^n}. \tag{123}$$

where the factor $s^{(m_1+m_2)\gamma_s/2}$ transforms the regular partial spatial derivatives to corresponding scale-normalized spatial derivatives with γ_s denoting the spatial scale normalization parameter [62] and the factor $\alpha_n(\tau)$ is the scale normalization factor for scale-normalized temporal derivatives according to either variance-based normalization [75, Section 7.2, Equation (74)]

$$\alpha_n(\tau) = \tau^{n\gamma_\tau/2} \tag{124}$$

with γ_τ denoting the temporal scale normalization parameter for scale-normalized temporal derivatives according to L_p-normalization [75, Section 7.2, Equation (76)]

$$\alpha_n(\tau) = \frac{\|g_{\xi^n}(\cdot; \tau)\|_p}{\|h_{t^n}(\cdot; \tau)\|_p} = \frac{G_{n,\gamma_\tau}}{\|h_{t^n}(\cdot; \tau)\|_p}. \tag{125}$$

and G_{n,γ_τ} denotes the L_p-norm of the non-causal temporal Gaussian derivative kernel for the γ_τ-value for which this L_p-norm becomes constant over temporal scales.

8.2 Differential Entities for Spatio-Temporal Scale Selection

Inspired by the way neurons in the lateral geniculate nucleus (LGN) respond to visual input [14,15], which for many LGN cells can be modelled by idealized operations of the form [68, Equation (108)]

$$h_{LGN}(x, y, t; s, \tau) = \pm(\partial_{xx} + \partial_{yy}) g(x, y; s) \partial_{t^n} h(t; \tau), \tag{126}$$

let us consider the following differential entities [75, Section 8.4, Equations (95)–(96)]

$$\partial_t(\nabla^2_{(x,y)}L) = L_{xxt} + L_{yyt} \tag{127}$$

$$\partial_{tt}(\nabla^2_{(x,y)}L) = L_{xxtt} + L_{yytt} \tag{128}$$

which correspond to first- and second-order temporal derivatives of the spatial Laplacian operator and study the corresponding scale-normalized spatio-temporal derivative expressions for $\gamma_s = 1$:

$$\partial_{t,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L) = s\,\alpha_1(\tau)\,(L_{xxt} + L_{yyt}) = s\,\alpha_1(\tau)\,\partial_t(\nabla^2_{(x,y)}L), \tag{129}$$

$$\partial_{tt,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L) = s\,\alpha_2(\tau)\,(L_{xxtt} + L_{yytt}) = s\,\alpha_2(\tau)\,\partial_{tt}(\nabla^2_{(x,y)}L). \tag{130}$$

Notably, we do not focus on extending the previously established use of the spatial Laplacian operator for spatial scale selection to a spatio-temporal Laplacian operator for spatio-temporal scale selection, since the most straightforward way of defining such an operator $\nabla^2_{(x,y,t)}L = L_{xx} + L_{yy} + \kappa^2 L_{tt}$ for some κ is not covariant under independent rescaling of the spatial and temporal coordinates as occurs if observing the same scene with cameras having independently different spatial and temporal sampling rates. The differential entities $\partial_{t,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L)$ and $\partial_{tt,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L)$ are on the other hand truly covariant under independent rescalings of the spatial and temporal dimensions and therefore better candidates to be used as primitives in spatio-temporal scale selection algorithms.

8.3 Spatio-Temporal Scale Selection Properties

8.3.1 Response to a Localized Gaussian Blink

Consider a local idealized spatio-temporal image pattern defined as the combination of a rotationally symmetric Gaussian blob $g(x, y; s_0)$ over the spatial domain and a time-causal temporal peak $U(t; \mu, K_0)$ of the form (43) over the temporal domain

$$f(x, y, t) = g(x, y; s_0)\,U(t; \mu, K_0). \tag{131}$$

If we define the spatio-temporal scale-space representation of this spatio-temporal image pattern of the form (122) with the

temporal scale-space kernel chosen as the composed kernel $U(t; \mu, K)$ corresponding to a cascade of first-order integrators with equal time constants coupled in cascade (120), then it follows from the semi-group property of the spatial Gaussian kernel and the semi-group property (45) of the time-causal temporal kernel $U(t; \mu, K)$ that the spatio-temporal scale-space representation will be of the form

$$L(x, y, t; s, \tau) = g(x, y; s_0 + s) U(t; \mu, K_0 + K). \quad (132)$$

The second-order temporal derivative of the spatial Laplacian of the Gaussian Specifically, the scale-normalized differential entity $\partial_{tt,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L)$ constituting an idealized model of a "lagged" LGN cell [75, Figure 3(right)] will by a combination of the Laplacian response of Gaussian

$$\nabla^2 g(x, y; s) = (x^2 + y^2 - 2s)/s^2 g(x, y; s), \quad (133)$$

the second-order temporal derivative of $L_{tt}(t; \mu, K)$ of a time-causal peak in Eq. (47) and the temporal scale normalization operation

$$L_{\zeta\zeta}(t; \mu, K) = (\mu^2 K)^{\gamma_\tau} L_{tt}(t; \mu, K) \quad (134)$$

in Eq. (54) assume the form

$$
\begin{aligned}
&\partial_{tt,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L) \\
&= \frac{(\mu^2 K)^{\gamma_\tau}}{\mu^2 t^2} \left(\mu^2 \left(K^2 + K(2K_0 - 3) + K_0^2 - 3K_0 + 2 \right) \right. \\
&\quad \left. - 2\mu t (K + K_0 - 1) + t^2 \right) \\
&\quad \times \frac{(x^2 + y^2 - 2(s_0 + s))}{(s_0 + s)^2} g(x, y; s_0 + s) U(t; \mu, \\
&\quad K_0 + K).
\end{aligned}
\quad (135)
$$

Specifically, based on previously established scale selection properties of the spatial Laplacian of the Gaussian [62,70] and the second-order scale-normalized temporal derivatives of the time-causal scale space (Sect. 4.2), it follows that this spatio-temporal differential entity will for spatial and temporal scale normalization powers $\gamma_s = 1$ and $\gamma_\tau = 3/4$, respectively, assume its local extremum over both spatial and temporal scales at spatial scale

$$\hat{s} = s_0 \quad (136)$$

and at a temporal scale that is a good approximation of the temporal scale of the temporal peak $\hat{K} \approx K_0$ corresponding to (see Table 1)

$$\hat{\tau} \approx \tau_0. \quad (137)$$

Thus, simultaneous spatio-temporal scale selection using the differential entity $\partial_{tt,\text{norm}}(\nabla^2_{(x,y),\text{norm}}L)$ applied to the model signal (131) will estimate both the spatial extent and the temporal duration of the Gaussian blink.

The determinant of the spatio-temporal Hessian For general values of the spatial and temporal scale normalization parameters γ_s and γ_τ, the determinant of the spatio-temporal Hessian is given by

$$
\begin{aligned}
\det \mathcal{H}_{(x,y,t),\text{norm}} L = s^{2\gamma_s} \tau^{\gamma_\tau} \big(& L_{xx}L_{yy}L_{tt} + 2L_{xy}L_{xt}L_{yt} \\
& - L_{xx}L_{yt}^2 - L_{yy}L_{xt}^2 - L_{tt}L_{xy}^2 \big).
\end{aligned}
\quad (138)
$$

In the specific case when the spatio-temporal scale-space representation of a time-causal Gaussian blink is of the form (132), if we restrict the analysis to the spatial origin $(x, y) = (0, 0)$ where $g_x = g_y = 0$ and to the temporal maximum point $t_{\max}$ where $h_t = 0$, implying that $L_{xy} = L_{xt} = L_{yt} = 0$, it follows that the determinant of spatio-temporal Hessian at the spatio-temporal maximum reduces to the form

$$
\begin{aligned}
&\det \mathcal{H}_{(x,y,t),\text{norm}} L \big|_{(x,y)=(0,0), t=t_{\max}} \\
&= \left(s^{2\gamma_s} g\, g_{xx}\, g_{yy} \right) \Big|_{(x,y)=(0,0)} \left((\mu^2 K)^{\gamma_\tau} U^2 U_{tt} \right) \Big|_{t=t_{\max}},
\end{aligned}
\quad (139)
$$

we can observe that the spatial and temporal scale selection properties of the determinant of the spatio-temporal Hessian will be different from the scale selection properties of the second-order temporal derivative of the spatial Laplacian. In a companion paper [74], it is shown that for a corresponding non-causal Gaussian temporal scale-space concept, spatial and temporal scale normalization parameters equal to $\gamma_s = 5/4$ and $\gamma_\tau = 5/4$ lead to scale estimates $\hat{s}$ and $\hat{\tau}$ corresponding to the spatial and temporal extents $\hat{s} = s_0$ and $\hat{\tau} = \tau_0$ of a Gaussian blink.

By performing a corresponding study of the temporal scale selection properties of the purely temporal component of this expression

$$\theta_{\text{norm}} = \left((\mu^2 K)^{\gamma_\tau} U^2 U_{tt} \right) \Big|_{t=t_{\max}} \quad (140)$$

for the specific case of time-causal temporal scale-space representation based on a uniform distribution of the intermediate temporal scale levels as previously done for the temporal scale selection properties of the second-order temporal derivative of a temporal peak in Table 1, we obtain the results shown in Table 8. The column labelled $\hat{K}$ shows that the maximum over temporal scales is obtained at a temporal scale level near the temporal scale of the underlying temporal peak

$$\hat{\tau} \approx \tau_0, \qquad (141)$$

whereas the column labelled θ_{postnorm} shows that if we normalize the input signal $f(t) = U(t; \mu, K_0)$ to having unit contrast, then the corresponding post-normalized differential entity

$$\theta_{\text{postnorm}}\big|_{\gamma=1} = \frac{1}{\tau^{1/4}} \, \theta_{\text{norm}}\big|_{\gamma=5/4} = \frac{K}{K + K_0 - 1} \qquad (142)$$

is approximately constant for temporal peaks with different temporal duration as determined by the parameter K_0. In these respects, this time-causal scale selection method implies a good approximate transfer of the scale selection property of treating similar temporal structures of different temporal duration in a uniform manner.

8.3.2 Response to a Localized Gaussian Onset Blob

Consider a local idealized spatio-temporal image pattern defined as the combination of a rotationally symmetric Gaussian blob $g(x, y; s_0)$ over the spatial domain and a time-causal onset ramp $\int_{u=0}^{t} U(u; \mu, K_0) \, du$ of the form (66) over the temporal domain

$$f(x, y, t) = g(x, y; s_0) \int_{u=0}^{t} U(u; \mu, K_0) \, du. \qquad (143)$$

Again defining the temporal scale-space representation of this spatio-temporal image pattern of the form (122) with the temporal scale-space kernel chosen as the composed kernel $U(t; \mu, K)$ corresponding to a cascade of first-order integrators with equal time constants coupled in cascade (120), it follows from the semi-group property of the spatial Gaussian kernel and the semi-group property (45) of the time-causal temporal kernel $U(t; \mu, K)$ that the spatio-temporal scale-space representation will be of the form

$$L(x, y, t; s, \tau) = g(x, y; s_0 + s) \int_{u=0}^{t} U(u; \mu, K_0) \, du. \qquad (144)$$

The first-order temporal derivative of the spatial Laplacian of the Gaussian For the scale-normalized differential entity $\partial_{t,\text{norm}}(\nabla^2_{(x,y),\text{norm}} L)$ constituting an idealized model of a "non-lagged" LGN cell [75, Figure 3(left)] will by a combination of the Laplacian response of Gaussian $\nabla^2 g(x, y; s) = (x^2 + y^2 - 2s)/s^2 \, g(x, y; s)$, the first-order temporal derivative $L_t(t; \mu, K)$ of a time-causal onset ramp in Eq. (68) and the temporal scale normalization operation $L_\zeta(t; \mu, K) = (\mu\sqrt{K})^{\gamma_\tau} L_t(t; \mu, K)$ in Eq. (53) assume the form

Table 8 Numerical estimates of the value of $\hat{K}$ at which the temporal component of the determinant of the spatio-temporal Hessian assumes its maximum over temporal scale (with the discrete expression over discrete temporal scales extended to a continuous variation) as function of K_0 and for either (i) variance-based normalization with $\gamma = 5/4$, (iii) variance-based normalization with $\gamma = 1$ and (iv) L_p-normalization with $p = 1$; For the case of variance-based normalization with $\gamma = 5/4$, (ii) the post-normalized magnitude measure $L_{\zeta\zeta,\text{maxmagn,norm}}\big|_{\gamma=1}$ according to (20) and at the corresponding scale (i) is also shown

K_0	$\hat{K}$ (var, $\gamma = 5/4$)	$\theta_{\text{postnorm}}\big|_{\gamma=1}$ (var, $\gamma = 5/4$)
Scale estimate $\hat{K}$ and max magnitude θ_{postnorm} from temporal peak (uniform distr)		
4	3.1	0.508
8	7.1	0.503
16	15.1	0.501
32	31.1	0.502
64	63.1	0.501

Note that for $\gamma = 5/4$ the temporal scale estimate $\hat{K}$ constitutes a good approximation to the temporal scale estimate being proportional to the temporal scale of the underlying temporal peak and that the maximum magnitude estimate $\theta_{\text{postnorm}}\big|_{\gamma=1}$ constitutes a good approximation to the maximum magnitude measure being constant under variations of the temporal duration of the underlying spatio-temporal image structure

$$\partial_{t,\text{norm}}(\nabla^2_{(x,y),\text{norm}} L)$$
$$= \frac{(x^2 + y^2 - 2(s_0 + s))}{(s_0 + s)^2} \, g(x, y; s_0 + s)$$
$$\times (\mu\sqrt{K})^{\gamma_\tau} U(t; \mu, K_0 + K). \qquad (145)$$

Specifically, based on previously established scale selection properties of the spatial Laplacian of the Gaussian [62,70] and the first-order scale-normalized temporal derivatives of the time-causal scale space (Sect. 4.3), it follows that this spatio-temporal differential entity will for spatial and temporal scale normalization powers $\gamma_s = 1$ and $\gamma_\tau = 1/2$, respectively, assume its local extremum over both spatial and temporal scales at spatial scale

$$\hat{s} = s_0 \qquad (146)$$

and at a temporal scale that is a good approximation of the temporal scale of the temporal onset ramp $\hat{K} \approx K_0$ corresponding to (see Table 2)

$$\hat{\tau} \approx \tau_0. \qquad (147)$$

Thus, simultaneous spatio-temporal scale selection using the differential entity $\partial_{t,\text{norm}}(\nabla^2_{(x,y),\text{norm}} L)$ applied to the model signal (143) will estimate both the spatial extent and the temporal duration of a Gaussian onset blob.

8.4 General Scale Selection Property for Temporal Modelling and Time-Causal Scale Space Based on the Scale-Invariant Time-Causal Limit Kernel

If we instead model the Gaussian blink (131) and the onset Gaussian blob (143) by the scale-invariant time-causal limit kernel $\Psi(t; \tau, c)$ over the temporal domain

$$f(x, y, t) = g(x, y; s_0) \Psi(t; \tau, c), \tag{148}$$

$$f(x, y, t) = g(x, y; s_0) \int_{u=0}^{t} \Psi(u; \tau, c) \, du, \tag{149}$$

then by the general transformation property of the time-causal limit kernel under temporal scaling transformations by a temporal scaling factor $S = c^j$ that is an integer power of the distribution parameter c of the time-causal limit kernel [75, Equation (44)]

$$S \Psi(S t; S^2 \tau, c) = \Psi(t; \tau, c), \tag{150}$$

it holds that the corresponding spatio-temporal scale-space representations of two temporally scaled video sequences $f'(x', y', t) = f(x, y, t)$ for $(x', y', t') = (x, y, St)$ are related according to Lindeberg [75, Equation (46)]

$$L'(x', y'; , t'; s', \tau', c) = L(x, y, t; s, \tau, c) \tag{151}$$

for $(s', \tau') = (s, S^2 \tau)$ if $S = c^j$. The corresponding scale-normalized temporal derivatives are in turn related according to (104)

$$L'_{\zeta'^n}(x', y', t'; s', \tau', c) = S^{n(\gamma_\tau - 1)} L_{\zeta^n}(x, y, t; s, \tau, c). \tag{152}$$

If the scale-normalized temporal derivative $L_{\zeta^n}(x, y, t; s, \tau, c)$ computed from the original video sequence f assumes a local extremum over temporal scales at $(x, y, t; s, \tau) = (x_0, y_0, t_0; s_0, \tau_0)$, then by the general scale invariance property of temporal scale selection in the temporal scale-space concept based on the time-causal limit kernel, which is described in Sect. 5.3, it follows that the scale-normalized temporal derivative $L_{\zeta'^n}(x', y', t'; s', \tau', c)$ computed from the temporally scaled video sequence f' will assume a local extremum over temporal scales at

$$(x'_0, y'_0, t'_0; s'_0, \tau'_0) = (x_0, y_0, St_0; s_0, S^2 \tau_0). \tag{153}$$

This scale-invariant property can also be extended to spatio-temporal derivatives $L_{\xi^{m_1} \eta^{m_2} \zeta^n}$ and spatio-temporal differential invariants $\mathcal{D}_{\text{norm}}$ defined in terms of homogenous polynomials as well as homogeneous rational expressions of such scale-normalized spatio-temporal derivatives. In this

way, by performing both the temporal modelling of the underlying temporal signal in terms of the time-causal limit kernel and using a temporal scale-space concept based on the time-causal limit kernel, we can support fully scale-covariant temporal scale estimates for temporal scale selection in video data defined over a time-causal spatio-temporal domain.

The only component that remains is to determine how the original temporal scale estimate $\hat{\tau}$ depends on the distribution parameter c and the temporal scale normalization parameter γ_τ for some value of τ_0.

9 Temporal Scale Selection in Spectro-Temporal Audio Data

For audio signals, corresponding temporal scale selection methods can be applied to a time-causal spectro-temporal domain, with the 2-D spatial domain of video data over the spatial dimensions (x, y) conceptually replaced by a 1-D logspectral domain over the logspectral dimension ν in the spectrogram computed at any temporal moment using a time-causal receptive field model as proposed in Lindeberg and Friberg [78,79] or with the time-causal kernels in that model replaced by the time-causal limit kernel [75].

The analogous operations to the first- and second-order temporal derivatives of the spatial dimension would then be the first- and second-order temporal derivatives of the second-order derivative in the logspectral dimension

$$\partial_{t,\text{norm}}(L_{\nu\nu,\text{norm}}) = s^{\gamma_s} \tau^{\gamma_\tau/2} L_{\nu\nu t} \tag{154}$$

$$\partial_{tt,\text{norm}}(L_{\nu\nu,\text{norm}}) = s^{\gamma_s} \tau^{\gamma_\tau} L_{\nu\nu tt} \tag{155}$$

where s denotes the logspectral scale and τ the temporal scale.

By calibrating the logspectral scale normalization parameter γ_s such that the selected temporal scale should reflect the logspectral width of a spectral band, it follows that we should use $\gamma_s = 3/4$. By calibrating the temporal scale normalization parameter γ_τ such that the selected temporal scale of $\partial_{t,\text{norm}}(L_{\nu,\text{norm}})$ should reflect the temporal duration of an onset, if follows that we should use $\gamma_\tau = 1/2$ for this operator. By instead calibrating the temporal scale normalization parameter γ_τ such that the selected temporal scale of $\partial_{tt,\text{norm}}(L_{\nu\nu,\text{norm}})$ should reflect the temporal duration of a beat, if follows that we should use $\gamma_\tau = 3/4$ for that operator.

Note that these operations can be expressed both over frequency-time separable spectro-temporal receptive fields and over glissando-adapted spectro-temporal receptive fields if we for glissando-adapted receptive fields also replace the temporal derivative operator ∂_t by the corresponding glissando-adapted temporal derivative operator $\partial_{\bar{t}} = \partial_t + \nu \partial_\nu$, where ν denotes the glissando parameter.

10 Summary and Conclusions

In this treatment, we have proposed a new theoretical framework for temporal scale selection in a time-causal scale-space representation. Starting from a general survey of previously proposed temporal scale-space concepts and a detailed analysis of their relative advantages and disadvantages, we have focused our efforts on the time-causal scale-space concept based on first-order integrators coupled in cascade and analysed the extent to which scale-space properties that hold for the scale-invariant non-causal Gaussian temporal scale-space concept can be transferred to this time-causal scale-space concept. Specifically, we have analysed this time-causal scale-space concept for two specific ways of distributing the intermediate temporal scale levels using either (i) a uniform distribution over the temporal scales as parameterized by the variance of the temporal scale-space kernel and corresponding to convolution with temporal kernels that are Laguerre functions and in turn corresponding to temporal derivatives of the Gamma distribution or (ii) a logarithmic distribution taken to the limit of a recently proposed time-causal limit kernel with an infinitely dense distribution of temporal scale levels towards zero temporal scale.

For peak and ramp detection, we have shown that for the time-causal temporal scale-space concept based on first-order integrators with equal time constant coupled in cascade, we can reasonably well estimate the temporal scale of a localized temporal peak or a localized onset ramp with corresponding good approximation of constancy of appropriately post-normalized scale-normalized magnitude measures of the corresponding feature detectors under variations in the temporal duration of the underlying temporal peak or the underlying temporal ramp. For a non-localized sine wave signal, the lack of temporal scale invariance is, however, substantial both with regard to a systematic offset in temporal scale estimates and a lack of corresponding constancy of the magnitude measures over variations of the wavelength of the underlying sine wave.

For the time-causal temporal scale-space concept based on convolution with the time-causal limit kernel with an underlying logarithmic distribution of the temporal scale levels and taken to the limit of the time-causal limit kernel with an infinitely dense distribution of temporal scale levels near temporal scale zero, we have on the other hand shown that it is possible to achieve perfect temporal invariance in the respects that (i) the temporal scale estimates in dimension [time] are proportional to the wavelength of the underlying sine wave and (ii) the magnitude measures remain constant under variations of the wavelength of the sine wave.

Additionally, we have shown a general scale invariance result that holds for any temporal signal and which states that for temporal scaling transformation with a temporal scaling factor given as an integer power of the distribution parameter c of the time-causal limit kernel that is used for generating the temporal scale space, it holds that:

(i) local extrema over temporal scales of scale-normalized derivatives are preserved under this group of temporal scaling transformations with scaling factors of the form $t' = c^j t$ for integer j,

(ii) the corresponding scale estimates are transformed in a scale-covariant way corresponding to $\hat{\tau}' = c^{2j}\hat{\tau}$ in units of the variance of the temporal scale-space kernel,

(iii) if the scale normalization parameter γ of variance-based scale-normalized derivatives is chosen as $\gamma = 1$ and corresponding to $p = 1$ for L_p-normalization, then the magnitude values of the scale-normalized temporal derivatives are preserved under scaling transformations with any temporal scaling factor c^j that is an integer power of the distribution parameter c of the time-causal limit kernel and

(iv) for other values of the scale normalization parameters γ or p, the corresponding scale-normalized derivatives are transformed according to a scale-covariant power law (104), which is straightforward to compensate for by post-normalization.

In these respects, the proposed framework for temporal scale selection in the scale-space concept based on the time-causal limit kernel provides the necessary mechanisms to achieve temporal scale invariance while simultaneously being expressed over a time-causal and time-recursive temporal domain. From a theoretical perspective, this is a conceptually novel type of construction that has not previously been achieved based on any other type of time-causal temporal scale-space concept.

As experimental confirmation of the derived theoretical results regarding temporal scale selection properties, we have presented experimental results of applying two types of more specific temporal scale selection algorithms to one-dimensional temporal signals, based on either (i) sparse scale-space extrema detection by detecting local extrema of feature responses over both time and temporal scales or (ii) dense feature maps over temporal scales here specifically manifested in terms of a temporal quasi quadrature entity that constitutes an energy measure of the local strength of scale-normalized first- and second-order temporal derivates.

We have also described practical details to handle in time-causal scale selection algorithms in relation to the inherent temporal delays of time-causal image measurements and proposed specific mechanisms to handle the differences in temporal delays between time-causal scale-space representations at different temporal scales.

Experimental results presented for synthetic and real one-dimensional temporal signals show that it is possible to compute local estimates of temporal scale levels that in units

of the standard deviation of the underlying temporal scale-space kernel are proportional to the temporal duration of the underlying structures in the temporal signal that gave rise to the filter responses.

Beyond these two specific ways of expressing temporal scale selection mechanisms, we argue that the theoretical results presented in the paper should also more generally open up for extensions to other ways of comparing time-dependent filter responses at multiple temporal scales.

Experimental results obtained by applying this temporal scale selection theory to video analysis will be reported in a companion paper [74].

Appendix 1: Why a Semi-Group Property Over Temporal Scales Leads to Undesirable Temporal Dynamics in the Presence of Temporal Delays

One way of understanding why the assumption about a semi-group property over temporal scales may lead to undesirable temporal dynamics for a temporal scale space representation involving temporal delays can be obtained as follows:

Ideally, for a temporal scale-space concept involving a temporal delay one would like the temporal delay δ to be proportional to the temporal scale parameter σ in terms of dimension [time]

$$\delta = C\,\sigma \tag{156}$$

for some constant $C > 0$. For a temporal scale-space kernel with finite[4] temporal variance τ, this corresponds to letting

the temporal delay at any temporal scale be proportional to the square root of the temporal scale parameter τ according to

$$\delta = \phi(\tau) = C\sqrt{\tau}. \tag{157}$$

Let us next assume that we have to two temporal scale-space kernels $h(t;\ \tau_1, \delta_1)$ and $h(t;\ \tau_2, \delta_2)$ with finite temporal variances τ_1 and τ_2 and finite temporal means δ_1 and δ_2 from the same family of temporal kernels h. If the temporal kernels are to obey a semi-group property over temporal scales, then by the additive property of mean values and variances under convolution of positive functions, it follows that the composed temporal scale-space kernel should be given by

$$h(\cdot;\ \tau_1, \delta_1) * h(t;\ \tau_2, \delta_2) = h(t;\ \tau_1 + \tau_2, \delta_1 + \delta_2). \tag{158}$$

This property should for example hold for the non-causal Gaussian temporal scale-space kernels (2) if we require the kernels to obey a semi-group property over temporal scales. Combining this property with a fixed relationship between the temporal delay δ and the temporal scale τ according to $\delta = \phi(\tau)$ does, however, then lead to

$$\delta_1 + \delta_2 = \phi(\tau_1 + \tau_2) = \phi(\tau_1) + \phi(\tau_2). \tag{159}$$

This implies that the function ϕ must be additive in terms of its argument τ, implying increasingly longer temporal delays at coarser temporal scales and thus a violation of the desirable form of temporal dynamics $\delta = \phi(\tau) = C\sqrt{\tau}$.

We can, however, remedy the situation by replacing the temporal semi-group property with a weaker cascade smoothing property over temporal scales

$$L(\cdot;\ \tau_2, \delta_2) = h(\cdot;\ (\tau_1, \delta_1) \mapsto (\tau_2, \delta_2)) * L(\cdot;\ \tau_1, \delta_1), \tag{160}$$

where the temporal kernels should for any triplets of temporal scale values and temporal delays (τ_1, δ_1), (τ_2, δ_2) and (τ_3, δ_3)

---

[4] Regarding the assumption of a finite temporal variance, it is interesting to compare the situation with the time-causal semi-group kernel $\phi(t;\ \tau) = \frac{1}{\sqrt{2\pi}\, t^{3/2}}\, \tau\, e^{-\tau^2/2t}$ derived by Fagerström [19, Equation (27)] and Lindeberg [66, Equation (93)]. For this kernel, the first- and second-order temporal moments are not finite $\int_{t=0}^{\infty} t\,\phi(t;\ \tau)\,dt \rightarrow \infty$ and $\int_{t=0}^{\infty} t^2\,\phi(t;\ \tau)\,dt \rightarrow \infty$, implying that the analysis in this appendix breaks down if applied to the time-causal semi-group, since this analysis is based on the additive properties of mean values and variances for non-negative distributions. The fact that the first- and second-order temporal moments are infinite for the time-causal semi-group, does on the other hand also reflect undesirable temporal dynamics, since temporal smoothing with such a kernel leads to slow and smeared out temporal responses compared to temporal smoothing with a temporal kernel having finite first- and second-order temporal moments. If we measure the

temporal delay of the time-causal semi-group kernel by the position of the temporal maximum $\hat{t} = \tau^2/3$ [66, Equation (119)] and its temporal extent from the difference between the time instances at which the one-dimensional time-causal semi-group kernel assumes half its maximum value $\Delta t \approx 0.900\,\tau^2$ [66, Equation (122)], then the temporal delay and the temporal extent of the time-causal semi-group kernel are indeed proportional. Those measures of the temporal delay and the temporal extent of the temporal kernel are, however, not the same as used in the in the arguments in this appendix.

The example with the time-causal semi-group therefore demonstrates that at the cost of infinite first- and second-order temporal moments it is possible to find a temporal smoothing kernel that both obeys the semi-group property and a proportionality relation between measures of the temporal delay and the temporal extent in dimensions of [time]. Due to the infinite first- and second-order temporal moments, the temporal dynamics is, however, undesirable anyway.

obey the transitive property

$$h(\cdot; \ (\tau_1, \delta_1) \mapsto (\tau_2, \delta_2)) * h(\cdot; \ (\tau_2, \delta_2) \mapsto (\tau_3, \delta_3))$$
$$= h(\cdot; \ (\tau_1, \delta_1) \mapsto (\tau_3, \delta_3)) \tag{161}$$

and we can specifically for the non-causal Gaussian temporal scale-space concept with time-delayed Gaussian kernels of the form (2) choose

$$h(\cdot; \ (\tau_1, \delta_1) \mapsto (\tau_2, \delta_2)) = g(\cdot; \ \tau_2 - \tau_1, C(\sqrt{\tau_2} - \sqrt{\tau_1})). \tag{162}$$

Based on this form of cascade smoothing property over temporal scale, we can both (i) guarantee non-creation of new structures in the signal from finer to coarser temporal scales based on the scale-space properties of the temporal scale-space kernel g and (ii) achieve temporal delays that increase linearly with the temporal scale parameter in terms of dimension [time] such that $\delta_1 = C\sqrt{\tau_1}$ and $\delta_2 = C\sqrt{\tau_2}$ in (160).

In our temporal scale-space concept based on truncated exponential kernels coupled in cascade [53,73,75,77], we can specifically note that (i) the special case when all the time constants are equal implies a semi-group property over discrete temporal scales and longer temporal delays at coarser temporal scales (see the second row in Fig. 1) whereas the (ii) the special case with logarithmically distributed temporal scales implies that only a weaker cascade smoothing property holds and which enables much faster temporal response properties (see the third and fourth rows in Fig. 1).

Since any time-causal temporal scale-space representation will give rise to nonzero temporal delays, and we have shown in this section how the assumption of a semi-group structure over temporal scale leads to undesirable temporal dynamics in the presence of temporal delays, we argue that one should not require a semi-group structure over temporal scales for time-causal scale space and instead require a less restrictive cascade smoothing property over temporal scales.

Appendix 2: Scale Normalization of Temporal Derivatives in Koenderink's Scale-Time Model

In his scale-time model, Koenderink [42] proposed to perform a logarithmic mapping of the past via a time delay and then applied Gaussian smoothing in the transformed temporal domain. Following the slight modification of this model proposed in Lindeberg [75, Appendix 2] to have the temporal kernels normalized to unit L_1-norm such that a constant signal should remain unchanged under temporal smoothing, these kernels can be written on the form

$$h_{\text{Koe}}(t; \ \sigma, \delta) = \frac{1}{\sqrt{2\pi}\sigma\,\delta} e^{-\frac{\log^2\left(\frac{t}{\delta}\right)}{2\sigma^2} - \frac{\sigma^2}{2}}. \tag{172}$$

where δ represents the temporal delay and σ is a dimensionless temporal scale parameter relative to the logarithmically transformed temporal domain. The temporal mean of this kernel is [75, Appendix 2, Equation (152)]

$$\bar{t} = \int_{t=-\infty}^{\infty} t\,h_{\text{Koe}}(t; \ \sigma, \delta)\,\mathrm{d}t = \delta\,e^{\frac{3\sigma^2}{2}} \tag{173}$$

and the temporal variance [75, Appendix 2, Equation (153)]

$$\tau = \int_{t=-\infty}^{\infty} (t - \bar{t})^2\,h_{\text{Koe}}(t; \ \sigma, \delta)\,\mathrm{d}t = \delta^2 e^{3\sigma^2}\left(e^{\sigma^2} - 1\right). \tag{174}$$

In relation to the proportionality requirement (157) between the temporal delay and the temporal scale parameter in terms of dimension [time] used for the theoretical arguments in "Appendix 1"

$$\bar{t} = C\sqrt{\tau} \tag{175}$$

it follows that that this relation is satisfied if and only if

$$C = \frac{1}{\sqrt{e^{\sigma^2} - 1}} \tag{176}$$

is held constant between the temporal scale-time representations at different temporal scales, in other words only if a one-parameter family of scale-time representations is generated by keeping the dimensionless temporal scale parameter σ constant while varying only the temporal delay parameter δ. If proportionality between the temporal delay and the temporal scale parameter of dimension [time] is required, then the dimensionless scale parameter σ should therefore be determined from the proportionality constant C according to

$$\sigma = \sqrt{\log\left(1 + \frac{1}{C^2}\right)}. \tag{177}$$

Differentiating the kernel (172) with respect to time gives the expressions for the first-, second- and third-order temporal derivatives in equations (163), (164) and (165) in Table 9. The first-order temporal derivative has its zero-crossing at

$$t_{\max} = \delta, \tag{178}$$

the second-order temporal derivative has its zero-crossings at

$$t_{\text{inflect1}} = \delta\,e^{-\frac{1}{2}\sigma\left(\sqrt{\sigma^2+4}+\sigma\right)}, \tag{179}$$

$$t_{\text{inflect2}} = \delta\,e^{-\frac{1}{2}\sigma\left(\sigma-\sqrt{\sigma^2+4}\right)}. \tag{180}$$

and its peaks at the zero-crossings of the third-order derivative

$$t_{3,1} = e^{-\sigma\left(\sqrt{\sigma^2+3}+\sigma\right)}, \tag{181}$$

Table 9 Regular first-, second- and third-order temporal derivatives of Koenderink's scale-time kernel (renormalized according to (172))

Regular temporal derivatives of Koenderink's scale-time kernel

$$h_{t,Koe}(t;\ \sigma,\delta) = -\frac{\log(\frac{t}{\delta})\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})+\sigma^4}{2\sigma^2}}}{\sqrt{2\pi}\delta\sigma^3 t} \tag{163}$$

$$h_{tt,Koe}(t;\ \sigma,\delta) = \frac{\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})+\sigma^4}{2\sigma^2}}(\sigma^2\log(\frac{t}{\delta})+\log^2(\frac{t}{\delta})-\sigma^2)}{\sqrt{2\pi}\delta\sigma^5 t^2}, \tag{164}$$

$$h_{ttt,Koe}(t;\ \sigma,\delta) = \frac{\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})+\sigma^4}{2\sigma^2}}(-3\sigma^2\log^2(\frac{t}{\delta})+(3\sigma^2-2\sigma^4)\log(\frac{t}{\delta})-\log^3(\frac{t}{\delta})+3\sigma^4)}{\sqrt{2\pi}\delta\sigma^7 t^3}. \tag{165}$$

Table 10 Scale-normalized first- and second-order temporal derivatives of Koenderink's scale-time kernel using either variance-based normalization for a general value of γ, variance-based normalization for $\gamma = 1$ or L_P-normalization for $p = 1$

Scale-normalized temporal derivatives of Koenderink's scale-time kernel

Using variance-based normalization, the corresponding scale-normalized temporal derivatives are for a general value of γ given by

$$h_{\zeta,Koe}(t;\ \sigma,\delta) = \tau^{\gamma/2}h_{t,Koe}(t;\ \sigma,\delta) = \frac{\left(\delta^2\mathrm{e}^{3\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right)\right)^{\gamma/2}\log(\frac{\delta}{t})\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})+\sigma^4}{2\sigma^2}}}{\sqrt{2\pi}\delta\sigma^3 t}, \tag{166}$$

$$h_{\zeta\zeta,Koe}(t;\ \sigma,\delta) = \tau^{\gamma}h_{tt,Koe}(t;\ \sigma,\delta) = \frac{\left(\delta^2\mathrm{e}^{3\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right)\right)^{\gamma}\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})+\sigma^4}{2\sigma^2}}(\log(\frac{t}{\delta})(\log(\frac{t}{\delta})+\sigma^2)-\sigma^2)}{\sqrt{2\pi}\delta\sigma^5 t^2}, \tag{167}$$

which for the specific value of $\gamma = 1$ reduce to

$$h_{\zeta,Koe}(t;\ \sigma,\delta) = \sqrt{\tau}\,h_{t,Koe}(t;\ \sigma,\delta) = \frac{\sqrt{\mathrm{e}^{\sigma^2}-1}\log(\frac{\delta}{t})\mathrm{e}^{\sigma^2-\frac{\log^2(\frac{\delta}{t})}{2\sigma^2}}}{\sqrt{2\pi}\sigma^3 t}, \tag{168}$$

$$h_{\zeta\zeta,Koe}(t;\ \sigma,\delta) = \tau\,h_{tt,Koe}(t;\ \sigma,\delta) = \frac{\delta\left(\mathrm{e}^{\sigma^2}-1\right)\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})-5\sigma^4}{2\sigma^2}}(\log(\frac{t}{\delta})(\log(\frac{t}{\delta})+\sigma^2)-\sigma^2)}{\sqrt{2\pi}\sigma^5 t^2}, \tag{169}$$

or when using L_p-normalization for $p = 1$:

$$h_{\zeta,Koe}(t;\ \sigma,\delta) = \frac{G_{1,1}}{\|h_{t,Koe}(\cdot;\ \sigma,\delta)\|_1}\,h_{t,Koe}(t;\ \sigma,\delta) = -\frac{\log(\frac{t}{\delta})\mathrm{e}^{-\frac{\log^2(\frac{t}{\delta})}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2 t}, \tag{170}$$

$$h_{\zeta\zeta,Koe}(t;\ \sigma,\delta) = \frac{G_{2,1}}{\|h_{tt,Koe}(\cdot;\ \sigma,\delta)\|_1}\,h_{tt,Koe}(t;\ \sigma,\delta) = \frac{\sqrt{\frac{2}{\pi}}\delta\mathrm{e}^{-\frac{2\log^2(\frac{t}{\delta})+\sigma^4}{4\sigma^2}}(\log(\frac{t}{\delta})(\log(\frac{t}{\delta})+\sigma^2)-\sigma^2)}{\sigma^3 t^2\left(\sigma\sinh\left(\frac{1}{4}\sigma\sqrt{\sigma^2+4}\right)+\sqrt{\sigma^2+4}\cosh\left(\frac{1}{4}\sigma\sqrt{\sigma^2+4}\right)\right)}. \tag{171}$$

$$t_{3,2} = \delta\mathrm{e}^{-\sigma^2}, \tag{182}$$

$$t_{3,3} = \delta\mathrm{e}^{\sigma\left(\sqrt{\sigma^2+3}-\sigma\right)}. \tag{183}$$

The L_1-norms of the first- and second-order temporal scale-space kernels are thereby given by

$$\|h_{t,Koe}(\cdot;\ \sigma,\delta)\|_1 = 2h_{Koe}(t_{max};\ \sigma,\delta) = \frac{\sqrt{\frac{2}{\pi}}\mathrm{e}^{-\frac{\sigma^2}{2}}}{\delta\sigma}, \tag{184}$$

$$\begin{aligned}
&\|h_{tt,Koe}(\cdot;\ \sigma,\delta)\|_1 \\
&= 2\left(h_{t,Koe}(t_{inflect1};\ \sigma,\delta) - h_{t,Koe}(t_{inflect2};\ \sigma,\delta)\right) \\
&= \frac{\sqrt{\frac{2}{\pi}}\mathrm{e}^{-\frac{\sigma^2}{4}-\frac{1}{2}}}{\delta^2\sigma^2}\left(\sigma\sinh\left(\frac{1}{4}\sigma\sqrt{\sigma^2+4}\right)\right. \\
&\quad\left. + \sqrt{\sigma^2+4}\cosh\left(\frac{1}{4}\sigma\sqrt{\sigma^2+4}\right)\right).
\end{aligned} \tag{185}$$

Based on these characteristics, we can define scale-normalized temporal derivatives of Koenderink's scale-time kernel according to Equations (166)–(171) in Table 10.

Appendix 3: Estimating the Temporal Duration of Underlying Temporal Structures from Scale-Time Approximations of Temporal Derivatives of the Time-Causal Limit Kernel

In [75, Appendix 2] the following transformation between the parameters in Koenderink's scale-time kernels and the time-causal limit kernel (90) is derived

$$\begin{cases}
\tau = \delta^2\,\mathrm{e}^{3\sigma^2}\left(\mathrm{e}^{\sigma^2}-1\right) \\
c = \frac{\mathrm{e}^{\sigma^2}}{2-\mathrm{e}^{\sigma^2}}
\end{cases}
\qquad
\begin{cases}
\sigma = \sqrt{\log\left(\frac{2c}{c+1}\right)} \\
\delta = \frac{(c+1)^2\sqrt{\tau}}{2\sqrt{2}\sqrt{(c-1)c^3}}
\end{cases} \tag{186}$$

Fig. 15 Scale-time approximations of the temporal durations $d_1(\tau, c)$ and $d_2(\tau, c)$ of the first- and second-order temporal derivatives of the time-causal limit kernel according to (193) and (194) as function of the distribution parameter c for $\tau = 1$
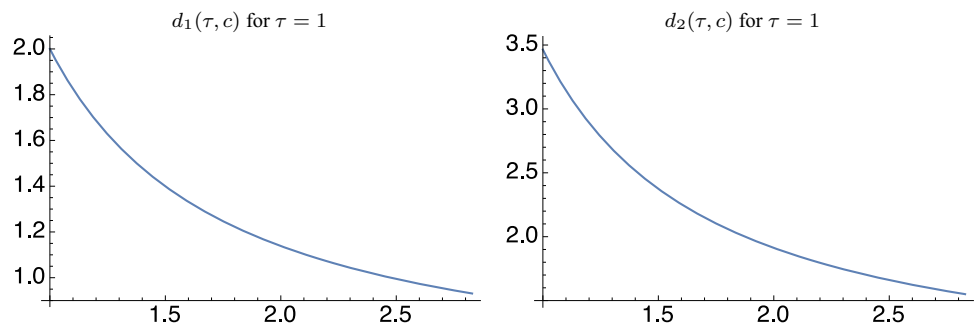
under the conditions $c > 1$ and $\sigma < \sqrt{\log 2} \approx 0.832$ by requiring the first- and second-order temporal moments of the kernels in the two families to be equal.

Given this approximate mapping between the time-causal limit kernel and the temporal kernels in Koenderink's scale-time model, we can approximate the positions of the temporal peak, the peaks in the first- and second-order temporal derivatives of the time-causal limit kernel based on our previously derived expressions for the maximum point t_{max}, the inflection points $t_{inflect1}$ and $t_{inflect2}$ as well as the zero-crossings of the third-order derivative $t_{3,1}$, $t_{3,2}$ and $t_{3,2}$ according to (178)–(183) in "Appendix 2":

$$t_{max} \approx \delta = \frac{(c+1)^2}{2\sqrt{2}\sqrt{\frac{(c-1)c^3}{\tau}}}, \tag{187}$$

$$t_{inflect1} \approx \delta\, e^{-\frac{1}{2}\sigma\left(\sqrt{\sigma^2+4}+\sigma\right)}$$
$$= \frac{(c+1)^{5/2}\sqrt{\frac{\tau}{c-1}}e^{-\frac{1}{2}\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+4\right)}}}{4c^2}, \tag{188}$$

$$t_{inflect2} \approx \delta\, e^{-\frac{1}{2}\sigma\left(\sigma-\sqrt{\sigma^2+4}\right)}$$
$$= \frac{(c+1)^{5/2}\sqrt{\frac{\tau}{c-1}}e^{\frac{1}{2}\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+4\right)}}}{4c^2}, \tag{189}$$

$$t_{3,1} \approx e^{-\sigma\left(\sqrt{\sigma^2+3}+\sigma\right)}$$
$$= \frac{(c+1)^3\sqrt{\frac{\tau}{c-1}}e^{-\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+3\right)}}}{4\sqrt{2}c^{5/2}}, \tag{190}$$

$$t_{3,2} \approx \delta e^{-\sigma^2}$$
$$= \frac{(c+1)^3}{4\sqrt{2}\sqrt{\frac{(c-1)c^5}{\tau}}}, \tag{191}$$

$$t_{3,3} \approx \delta e^{\sigma\left(\sqrt{\sigma^2+3}-\sigma\right)}$$
$$= \frac{(c+1)^3 e^{\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+3\right)}}}{4\sqrt{2}\sqrt{\frac{(c-1)c^5}{\tau}}} \tag{192}$$

Specifically, this leads to the following estimates of how the temporal width of the first- and second-order temporal derivatives depend on the distribution parameter c

$$d_1 = t_{inflect2} - t_{inflect1} \approx 2\delta e^{-\sigma^2/2} \sinh\left(\frac{\sigma}{2}\sqrt{\sigma^2+4}\right)$$
$$= \frac{(c+1)^{5/2}}{2c^2\sqrt{c-1}}$$
$$\times \sinh\left(\frac{1}{2}\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+4\right)}\right)\sqrt{\tau} \tag{193}$$

$$d_2 = t_{3,3} - t_{3,2} \approx 2\delta e^{-\sigma^2} \sinh\left(\sigma\sqrt{\sigma^2+3}\right)$$
$$= \frac{(c+1)^3}{2\sqrt{2}\,c^2\sqrt{c(c-1)}}$$
$$\times \sinh\left(\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+3\right)}\right)\sqrt{\tau} \tag{194}$$

which can be compared to the corresponding width measures for the non-causal Gaussian kernel

$$d_1 = t_{inflect2} - t_{inflect1} = (\delta+\sqrt{\tau}) - (\delta-\sqrt{\tau}) = 2\sqrt{\tau}, \tag{195}$$

$$d_2 = t_{3,3} - t_{3,1} = (\delta+\sqrt{3\tau}) - (\delta-\sqrt{3\tau}) = 2\sqrt{3}\sqrt{\tau}. \tag{196}$$

Figure 15 shows graphs of the width measures (193) and (194) of the first- and second-order temporal derivatives time-causal limit kernel obtained as obtained from a scale-time approximation. As can be seen from the graphs, the width measures vary by about 30 % when the distribution parameter is varied between $c = \sqrt{2}$ and $c = 2$. Thus, the value of the distribution parameter c must be taken into explicit account when transferring the temporal scale parameter τ to a characteristic length estimate on the temporal axis. Notably when the distribution parameter tends to $c \to 1$, the temporal width estimates approach the corresponding width estimates (195) and (196) of the Gaussian kernel.

Given that a temporal feature has been detected from a local maximum over temporal scales in either the first- or second-order temporal derivative of the time-causal limit kernel, if we use the behaviour of the Gaussian temporal scale-space model (195) and (196) for additional calibration of the proportionality constant of the scale estimate, we do then obtain the following estimates $\hat{d}$ of the temporal duration of the corresponding temporal feature as function of the temporal scale estimate $\hat{\tau}$ and the distribution parameter c:

$$\hat{d}_1 = \frac{d_1(\hat{\tau}, c)}{2}$$
$$= \frac{(c+1)^{5/2}}{4c^2\sqrt{c-1}}$$
$$\times \sinh\left(\frac{1}{2}\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+4\right)}\right)\sqrt{\hat{\tau}},$$
(197)

$$\hat{d}_2 = \frac{d_2(\hat{\tau}, c)}{2\sqrt{3}}$$
$$= \frac{(c+1)^3}{4\sqrt{6}\,c^2\sqrt{c\,(c-1)}}$$
$$\times \sinh\left(\sqrt{\log\left(\frac{2c}{c+1}\right)\left(\log\left(\frac{2c}{c+1}\right)+3\right)}\right)\sqrt{\hat{\tau}}.$$
(198)

References

1. Adelson, E., Bergen, J.: Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A **2**, 284–299 (1985)
2. Aertsen, A.M.H.J., Johannesma, P.I.M.: The spectro-temporal receptive field: a functional characterization of auditory neurons. Biol. Cybern. **42**(2), 133–143 (1981)
3. Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R.: Building Rome in a day. In: Proceedings of International Conference on Computer Vision (ICCV 2009), pp. 72–79 (2009)
4. Alías, F., Socoró, J.C., Sevillano, X.: A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. Appl. Sci. **6**(5), 143 (2016)
5. Bay, H., Ess, A., Tuytelaars, T., van Gool, L.: Speeded up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
6. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proceedings of Computer Vision and Pattern Recognition Workshop (CVPRW 2006), p. 35 (2006)
7. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: Proceedings of International Conference on Computer Vision (ICCV 2007), pp. 1–8. Rio de Janeiro, Brazil (2007)
8. Bretzner, L., Laptev, I., Lindeberg, T.: Hand-gesture recognition using multi-scale colour features, hierarchical features and particle filtering. In: Proceedings of Face and Gesture, pp. 63–74. Washington D.C., USA (2002)
9. Bretzner, L., Lindeberg, T.: Feature tracking with automatic selection of spatial scales. Comput. Vis. Image Underst. **71**(3), 385–392 (1998)
10. Brown, M., Lowe, D.G.: Unsupervised 3d object recognition and reconstruction in unordered datasets. In: Proceedings of 3-D Digital Imaging and Modeling (3DIM 2005), pp. 56–63 (2005)
11. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vis. **74**(1), 59–73 (2007)
12. Chomat, O., de Verdiere, V., Hall, D., Crowley, J.: Local scale selection for Gaussian based description techniques. In: Proceedings of European Conference on Computer Vision (ECCV 2000), Lecture Notes in Computer Science, vol. **1842**, pp. I:117–133. Springer, Dublin, Ireland (2000)
13. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: ideas, influences, and trends of the new age. ACM Comput. Surv. **40**(2), 5 (2008)
14. DeAngelis, G.C., Anzai, A.: A modern view of the classical receptive field: Linear and non-linear spatio-temporal processing by V1 neurons. In: Chalupa, L.M., Werner, J.S. (eds.) The Visual Neurosciences, vol. 1, pp. 704–719. MIT Press, Cambridge (2004)
15. DeAngelis, G.C., Ohzawa, I., Freeman, R.D.: Receptive field dynamics in the central visual pathways. Trends Neurosci. **18**(10), 451–457 (1995)
16. Derpanis, K.G., Wildes, R.P.: Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. IEEE Trans. Pattern Anal. Mach. Intell. **34**(6), 1193–1205 (2012)
17. Elder, J., Zucker, S.: Local scale control for edge detection and blur estimation. IEEE Trans. Pattern Anal. Mach. Intell. **20**(7), 699–716 (1998)
18. Ezzat, T., Bouvrie, J.V., Poggio, T.: Spectro-temporal analysis of speech using 2-D Gabor filters. In: INTERSPEECH, pp. 506–509 (2007)
19. Fagerström, D.: Temporal scale-spaces. Int. J. Comput. Vis. **2–3**, 97–106 (2005)
20. Fleet, D.J., Langley, K.: Recursive filters for optical flow. IEEE Trans. Pattern Anal. Mach. Intell. **17**(1), 61–67 (1995)
21. Florack, L.M.J.: Image Structure. Series in Mathematical Imaging and Vision. Springer, Berlin (1997)
22. Frangi, A.F., Hoogeveen, N.W., van Walsum, T., Viergever, M.A.: Model-based quantitation of 3D magnetic resonance angiographic images. IEEE Trans. Med. Imaging **18**(10), 946–956 (2000)
23. Gårding, J., Lindeberg, T.: Direct computation of shape cues using scale-adapted spatial derivative operators. Int. J. Comput. Vis. **17**(2), 163–191 (1996)
24. Guichard, F.: A morphological, affine, and Galilean invariant scale-space for movies. IEEE Trans. Image Process. **7**(3), 444–456 (1998)
25. Hall, D., de Verdiere, V., Crowley, J.: Object recognition using coloured receptive fields. In: Proceedings of European Conference on Computer Vision (ECCV 2000), Lecture Notes in Computer Science, vol. 1842, pp. I:164–177. Springer, Dublin, Ireland (2000)
26. Han, Z., Xu, Z., Zhu, S.C.: Video primal sketch: a unified middle-level representation for video. J. Math. Imaging Vis. **53**(2), 151–170 (2015)
27. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
28. Hassner, T., Mayzels, V., Zelnik-Manor, L.: On SIFTs and their scales. In: Proceedings of Computer Vision and Pattern Recognition (CVPR 2012), pp. 1522–1528. Providence, Rhode Island (2012)
29. Heckmann, M., Domont, X., Joublin, F., Goerick, C.: A hierarchical framework for spectro-temporal feature extraction. Speech Commun. **53**(5), 736–752 (2011)
30. Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurones in the cat's striate cortex. J. Physiol. **147**, 226–238 (1959)

31. Hubel, D.H., Wiesel, T.N.: Brain and Visual Perception: The Story of a 25-Year Collaboration. Oxford University Press, Oxford (2005)
32. Iijima, T.: Observation theory of two-dimensional visual patterns. Technical Report, Papers of Technical Group on Automata and Automatic Control, IECE, Japan (1962)
33. Jacobs, N., Pless, R.: Time scales in video surveillance. IEEE Trans. Circuits Syst. Video Technol. **18**(8), 1106–1113 (2008)
34. Jaimes, A., Sebe, N.: Multimodal human-computer interaction: a survey. Comput. Vis. Image Underst. **108**(1), 116–134 (2007)
35. Jhuang, H., Serre, T., Wolf, L., Poggio, T.: A biologically inspired system for action recognition. In: International Conference on Computer Vision (ICCV'07), pp. 1–8 (2007)
36. Kadir, T., Brady, M.: Saliency, scale and image description. Int. J. Comput. Vis. **45**(2), 83–105 (2001)
37. Kang, Y., Morooka, K., Nagahashi, H.: Scale invariant texture analysis using multi-scale local autocorrelation features. In: Proceedings of Scale Space and PDE Methods in Computer Vision (Scale-Space'05), Lecture Notes in Computer Science, vol. 3459, pp. 363–373. Springer (2005)
38. Karlin, S.: Total Positivity. Stanford University Press, Palo Alto (1968)
39. Kläser, A., Marszalek, M., Schmid, C.: A spatio-temporal descriptor based on 3D-gradients. In: Proceedings of British Machine Vision Conference, Leeds, U.K. (2008)
40. Kleinschmidt, M.: Methods for capturing spectro-temporal modulations in automatic speech recognition. Acta Acust. United Acust. **88**(3), 416–422 (2002)
41. Koenderink, J.J.: The structure of images. Biol. Cybern. **50**, 363–370 (1984)
42. Koenderink, J.J.: Scale-time. Biol. Cybern. **58**, 159–162 (1988)
43. Koenderink, J.J., van Doorn, A.J.: Generic neighborhood operators. IEEE Trans. Pattern Anal. Mach. Intell. **14**(6), 597–605 (1992)
44. Krissian, K., Malandain, G., Ayache, N., Vaillant, R., Trousset, Y.: Model-based detection of tubular structures in 3D images. Comput. Vis. Image Underst. **80**(2), 130–171 (2000)
45. Laptev, I., Caputo, B., Schuldt, C., Lindeberg, T.: Local velocity-adapted motion events for spatio-temporal recognition. Comput. Vis. Image Underst. **108**, 207–229 (2007)
46. Laptev, I., Lindeberg, T.: Space-time interest points. In: Proceedings of International Confernce on Computer Vision (ICCV 2003), pp. 432–439. Nice, France (2003)
47. Laptev, I., Lindeberg, T.: Local descriptors for spatio-temporal recognition. Proceedings of ECCV'04 Workshop on Spatial Coherence for Visual Motion Analysis. Lecture Notes in Computer Science, vol. 3667, pp. 91–103. Springer, Prague, Czech Republic (2004)
48. Larsen, A.B.L., Darkner, S., Dahl, A.L., Pedersen, K.S.: Jet-based local image descriptors. In: Proceedings of European Conference on Computer Vision (ECCV 2012), Lecture Notes in Computer Science, vol. 7574, pp. III:638–650. Springer (2012)
49. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1265–1278 (2005)
50. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimed. Comput. Commun. Appl. **2**(1), 1–19 (2006)
51. Li, S.Z. (ed.): Encyclopedia of Biometrics. Springer, Berlin (2009)
52. Li, Y., Tax, D.M.J., Loog, M.: Supervised scale-invariant segmentation (and detection). Proceedings of Scale Space and Variational Methods in Computer Vision (SSVM 2011). Lecture Notes in Computer Science, vol. 6667, pp. 350–361. Springer, Ein Gedi, Israel (2012)
53. Lindeberg, T.: Scale-space for discrete signals. IEEE Trans. Pattern Anal. Mach. Intell. **12**(3), 234–254 (1990)
54. Lindeberg, T.: Discrete derivative approximations with scale-space properties: a basis for low-level feature extraction. J. Math. Imaging Vis. **3**(4), 349–376 (1993)
55. Lindeberg, T.: Effective scale: a natural unit for measuring scale-space lifetime. IEEE Trans. Pattern Anal. Mach. Intell. **15**(10), 1068–1074 (1993)
56. Lindeberg, T.: On scale selection for differential operators. Proceedings of 8th Scandinavian Conf. on Image Analysis (SCIA'93), pp. 857–866. Norwegian Society for Image Processing and Pattern Recognition, Tromsø Norway (1993)
57. Lindeberg, T.: Scale-Space Theory in Computer Vision. Springer, Berlin (1993)
58. Lindeberg, T.: Scale-space theory: a basic tool for analysing structures at different scales. J. Appl. Stat. **21**(2), 225–270 (1994). http://www.csc.kth.se/~tony/abstracts/Lin94-SI-abstract.html
59. Lindeberg, T.: Linear spatio-temporal scale-space. In: ter Haar Romeny, B.M., Florack, L.M.J., Koenderink, J.J., Viergever, M.A. (eds.) Scale-Space Theory in Computer Vision: Proceedings of First International Conference on Scale-Space'97, Lecture Notes in Computer Science, vol. 1252, pp. 113–127. Springer, Utrecht, The Netherlands (1997)
60. Lindeberg, T.: On automatic selection of temporal scales in time-casual scale-space. In: Sommer, G., Koenderink, J.J. (eds.) Proceedings of AFPAC'97: Algebraic Frames for the Perception-Action Cycle. Lecture Notes in Computer Science, vol. 1315, pp. 94–113. Springer, Kiel, Germany (1997)
61. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. Int. J. Comput. Vis. **30**(2), 117–154 (1998)
62. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vis. **30**(2), 77–116 (1998)
63. Lindeberg, T.: A scale selection principle for estimating image deformations. Image Vis. Comput. **16**(14), 961–977 (1998)
64. Lindeberg, T.: Principles for automatic scale selection. In: Handbook on Computer Vision and Applications, pp. 239–274. Academic Press, Boston, USA (1999). http://www.csc.kth.se/cvap/abstracts/cvap222.html
65. Lindeberg, T.: Linear spatio-temporal scale-space. Technical Report ISRN KTH/NA/P–01/22–SE, Department of Numerical Analysis and Computer Science, KTH (2001). http://www.csc.kth.se/cvap/abstracts/cvap257.html
66. Lindeberg, T.: Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. J. Math. Imaging Vis. **40**(1), 36–81 (2011)
67. Lindeberg, T.: Scale invariant feature transform. Scholarpedia **7**(5), 10,491 (2012)
68. Lindeberg, T.: A computational theory of visual receptive fields. Biol. Cybern. **107**(6), 589–635 (2013)
69. Lindeberg, T.: Invariance of visual operations at the level of receptive fields. PLoS One **8**(7), e66,990 (2013)
70. Lindeberg, T.: Scale selection properties of generalized scale-space interest point detectors. J. Math. Imaging Vis. **46**(2), 177–210 (2013)
71. Lindeberg, T.: Scale selection. In: Ikeuchi, K. (ed.) Computer Vision: A Reference Guide, pp. 701–713. Springer, Berlin (2014)
72. Lindeberg, T.: Image matching using generalized scale-space interest points. J. Math. Imaging Vis. **52**(1), 3–36 (2015)
73. Lindeberg, T.: Separable time-causal and time-recursive spatio-temporal receptive fields. In: Proc. Scale-Space and Variational Methods for Computer Vision (SSVM 2015), Lecture Notes in Computer Science, vol. 9087, pp. 90–102. Springer (2015)
74. Lindeberg, T.: Spatio-temporal scale selection in video data (in preparation) (2016)
75. Lindeberg, T.: Time-causal and time-recursive spatio-temporal receptive fields. J. Math. Imaging Vis. **55**(1), 50–88 (2016)
76. Lindeberg, T., Bretzner, L.: Real-time scale selection in hybrid multi-scale representations. In: Griffin, L., Lillholm, M.

(eds.) Proc. Scale-Space Methods in Computer Vision (Scale-Space'03), Lecture Notes in Computer Science, vol. 2695, pp. 148–163. Springer, Isle of Skye, Scotland (2003)

77. Lindeberg, T., Fagerström, D.: Scale-space with causal time direction. Proceedings of European Conference on Computer Vision (ECCV'96). Lecture Notes in Computer Science, vol. 1064, pp. 229–240. Springer, Cambridge, UK (1996)

78. Lindeberg, T., Friberg, A.: Idealized computational models of auditory receptive fields. PLoS One **10**(3), e0119,032:1–e0119,032:58 (2015)

79. Lindeberg, T., Friberg, A.: Scale-space theory for auditory signals. In: Proceedings of Scale-Space and Variational Methods for Computer Vision (SSVM 2015), Lecture Notes in Computer Science, vol. 9087, pp. 3–15. Springer (2015)

80. Lindeberg, T., Gårding, J.: Shape from texture from a multi-scale perspective. In: Nagel, T.S.H.H.-H., Shirai, Y. (eds.) Proceedings of International Conference on Computer Vision (ICCV'93), pp. 683–691. IEEE Computer Society Press, Berlin, Germany (1993)

81. Lindeberg, T., Gårding, J.: Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure. Image Vis. Comput. **15**, 415–434 (1997)

82. Liu, C., Yuen, J., Torralba, A.: SIFT flow: dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 978–994 (2011)

83. Liu, X.M., Wang, C., Yao, H., Zhang, L.: The scale of edges. In: Proceedings of Computer Vision and Pattern Recognition (CVPR 2012), pp. 462–469 (2012)

84. Loog, M., Li, Y., Tax, D.: Maximum membership scale selection. Multiple Classifier Systems. Lecture Notes in Computer Science, vol. 5519, pp. 468–477. Springer, Berlin (2009)

85. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)

86. Mahmoodi, S.: Linear neural circuitry model for visual receptive fields. J. Math. Imaging Vis. **54**(2), 1–24 (2016)

87. Meyer, B.T., Kollmeier, B.: Optimization and evaluation of Gabor feature sets for ASR. In: INTERSPEECH, pp. 906–909 (2008)

88. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comput. Vis. **60**(1), 63–86 (2004)

89. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., van Gool, L.: A comparison of affine region detectors. Int. J. Comput. Vis. **65**(1–2), 43–72 (2005)

90. Miller, L.M., Escabi, N.A., Read, H.L., Schreiner, C.: Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. **87**(1), 516–527 (2001)

91. Mrázek, P., Navara, M.: Selection of optimal stopping time for nonlinear diffusion filtering. Int. J. Comput. Vis. **52**(2–3), 189–203 (2003)

92. Mutch, J., Lowe, D.G.: Object class recognition and localization using sparse features with limited receptive fields. Int. J. Comput. Vis. **80**(1), 45–57 (2008)

93. Negre, A., Braillon, C., Crowley, J.L., Laugier, C.: Real-time time-to-collision from variation of intrinsic scale. Exp. Robot. **39**, 75–84 (2008)

94. Niebles, J.C., Wang, H., Fei-Fei, L.: Unsupervised learning of human action categories using spatial-temporal words. Int. J. Comput. Vis. **79**(3), 299–318 (2008)

95. Paris, S.: Edge-preserving smoothing and mean-shift segmentation of video streams. Proceedings of European Conference on Computer Vision (ECCV 2008). Lecture Notes in Computer Science, pp. 460–473. Springer, Marseille, France (2008)

96. Patterson, R.D., Allerhand, M.H., Giguere, C.: Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. **98**(4), 1890–1894 (1995)

97. Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M.: Complex sounds and auditory images. Audit. Physiol. Percept. **83**, 429–446 (1992)

98. Poppe, R.: A survey on vision-based human action recognition. Image Vis. Comput. **28**(6), 976–990 (2010)

99. Porta, M.: Vision-based user interfaces: methods and applications. int. j. hum. comput. stud. **57**, 27–73 (2002)

100. Rivero-Moreno, C.J., Bres, S.: Spatio-temporal primitive extraction using Hermite and Laguerre filters for early vision video indexing. Image Anal. Recogn. **3211**, 825–832 (2004)

101. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. Int. J. Comput. Vis. **66**(3), 231–259 (2006)

102. Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., Kikinis, R.: 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Med. Image Anal. **2**(2), 143–168 (1998)

103. Schlute, R., Bezrukov, L., Wagner, H., Ney, H.: Gammatone features and feature combination for large vocabulary speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'07), vol. IV, pp. 649–652 (2007)

104. Schoenberg, I.J.: On Pòlya frequency functions. ii. Variation-diminishing integral operators of the convolution type. Acta Sci. Math. (Szeged) **12**, 97–106 (1950)

105. Schoenberg, I.J.: I. J. Schoenberg Selected Papers, vol. 2. Springer, Berlin (1988). Edited by C. de Boor

106. Se, S., Lowe, D.G., Little, J.J.: Vision-based global localization and mapping for mobile robots. IEEE Trans. Robot. **21**(3), 364–375 (2005)

107. Shabani, A.H., Clausi, D.A., Zelek, J.S.: Improved spatio-temporal salient feature detection for action recognition. In: British Machine Vision Conference (BMVC'11), pp. 1–12. Dundee, U.K. (2011)

108. Shao, L., Mattivi, R.: Feature detector and descriptor evaluation in human action recognition. In: Proceedings of ACM International Conference on Image and Video Retrieval CIVR'10, pp. 477–484. Xian, China (2010)

109. Siciliano, B., Khatib, O. (eds.): Springer Handbook of Robotics. Springer, Berlin (2008)

110. Sporring, J., Colios, C.J., Trahanias, P.E.: Generalized scale selection. In: Proceedings of International Conference on Image Processing (ICIP'00), pp. 920–923. Vancouver, Canada (2000)

111. Surya, P.V.B., Vorotnikov, D., Pelapur, R., Jose, S., Seetharaman, G., Palaniappan, K.: Multiscale Tikhonov-total variation image restoration using spatially varying edge coherence exponent. IEEE Trans. Image Process. **24**(12), 5220–5235 (2015)

112. ter Haar Romeny, B.: Front-End Vision and Multi-Scale Image Analysis. Series in mathematical imaging and vision. Springer, Berlin (2003)

113. ter Haar Romeny, B., Florack, L., Nielsen, M.: Scale-time kernels and models. In: Proceedings of International Conference on Scale-Space and Morphology in Computer Vision (Scale-Space'01), Lecture Notes in Computer Science. Springer, Vancouver, Canada (2001)

114. Tuytelaars, T., van Gool, L.: Matching widely separated views based on affine invariant regions. Int. J. Comput. Vis. **59**(1), 61–85 (2004)

115. Tuytelaars, T., Mikolajczyk, K.: A Survey on Local Invariant Features, Foundations and Trends in Computer Graphics and Vision, vol. **3**(3). Now Publishers (2008)

116. van der Berg, E.S., Reyneke, P.V., de Ridder, C.: Rotational image correlation in the Gauss–Laguerre domain. In: Third SPIE Conference on Sensors, MEMS and Electro-Optic Systems: Proceedings of SPIE, vol. 9257, pp. 92,570F-1–92,570F-17 (2014)

117. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010)

118. Wang, H., Ullah, M.M., Kläser, A., Laptev, I., Schmid, C.: Evaluation of local spatio-temporal features for action recognition. In: Proceedings of British Machine Vision Conference, London, U.K. (2009)

119. Wang, L., Qiao, Y., Tang, X.: Action recognition with trajectory-pooled deep-convolutional descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), pp. 4305–4314 (2015)

120. Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. Comput. Vis. Image Underst. **115**(2), 224–241 (2011)

121. Willems, G., Tuytelaars, T., van Gool, L.: An efficient dense and scale-invariant spatio-temporal interest point detector. Proceedings of European Conference on Computer Vision (ECCV 2008). Lecture Notes in Computer Science, vol. 5303, pp. 650–663. Springer, Marseille, France (2008)

122. Witkin, A.P.: Scale-space filtering. In: Proceedings of 8th International Joint Conference Artificial Intelligence, pp. 1019–1022. Karlsruhe, Germany (1983)

123. Wu, Q., Zhang, L., Shi, G.: Robust multifactor speech feature extraction based on Gabor analysis. IEEE Trans. Audio Speech Lang. Process. **19**(4), 927–936 (2011)

124. Zelnik-Manor, L., Irani, M.: Event-based analysis of video. In: Proceedings of Computer Vision and Pattern Recognition, pp. II:123–130. Kauai Marriott, Hawaii (2001)

Tony Lindeberg is a Professor of Computer Science at KTH Royal Institute of Technology in Stockholm, Sweden. He was born in Stockholm in 1964, received his M.Sc. degree in 1987, his Ph.D. degree in 1991, became docent in 1996, and was appointed professor in 2000. He was a Research Fellow at the Royal Swedish Academy of Sciences between 2000 and 2010. His research interests in computer vision relate to scale-space representation, image features, object recognition, spatio-temporal recognition, focus-of-attention and computational modelling of biological vision. He has developed theories and methodologies for continuous and discrete scale-space representation, visual and auditory receptive fields, detection of salient image structures, automatic scale selection, scale-invariant image features, affine invariant features, affine and Galilean normalization, temporal, spatio-temporal and spectro-temporal scale-space concepts as well as spatial and spatio-temporal image descriptors for image-based recognition. He has also worked on topics in medical image analysis and gesture recognition. He is author of the book Scale-Space Theory in Computer Vision.