DISCRETE IMPLEMENTATION OF THE FIRST ORDER SYSTEM CASCADE AS THE BASIS FOR A MELODIC SEGMENTATION MODEL

Maja Šerman

Computer Science and
Information Systems Department
University of Limerick
maja.serman@ul.ie

ABSTRACT

The basis for a low-level melodic segmentation model and its discrete implementation is presented. The model is based on the discrete approximation of the one-dimensional convective transport mechanism. In this way, a physically plausible mechanism for achieving multi-scale representation is obtained. Some aspects of edge detection theory thought to be relevant for solving similar problems in auditory perception are briefly introduced. Two examples presenting the dynamic behaviour of the model are shown.

1. INTRODUCTION

The problem of the detection of intensity changes in early visual processing has been computationally investigated through various methods ever since Marr's influential work in that area [1, 2, 3]. While both theoretical and experimental research in melodic perception recognise that the processing of the auditory signal takes place over multiple time scales (which corresponds to different spatial resolution in the low-level vision theory), a similar approach has not been widely adopted in computational modelling of melodic perception. In particular, the current computational models of melodic segmentation rely on an exclusive set of discrete rules applied to notated melodies. In this paper we first discuss some aspects of edge detection theory which are believed to be important for auditory processing. Secondly, the basis for a dynamic model of melodic segmentation and its discrete implementation is presented.

2. THE DETECTION OF THE INTENSITY CHANGES

Edge detection theory[1] proposes three stages for detecting intensity changes in images. The first stage incorporates techniques for smoothing the image, the second addresses methods for differentiating such smoothed image intensities. In the third stage, extraction of features essential for describing image structure (for instance peaks, or zero-crossings) from the smoothed and differentiated images is addressed. This comes as a consequence of two fundamental properties of images; a) the detection of changes in an image can be obscured by the effect of noise and b) significant changes in an image frequently occur at different resolutions. In order to reduce the noise and to enable changes at different resolutions to come through the following steps are taken. The image is first smoothed, for instance by convoluting image intensities with a smoothing operator of different scales. The intensities of such smoothed image are differentiated through the application of the 1st and 2nd derivation operators, to extract the significant changes. One possible choice for the smoothing operator is the Gaussian filter (2-dimensional Gaussian function is described in equation 1), shown in figure 1, together with its first and second derivatives.

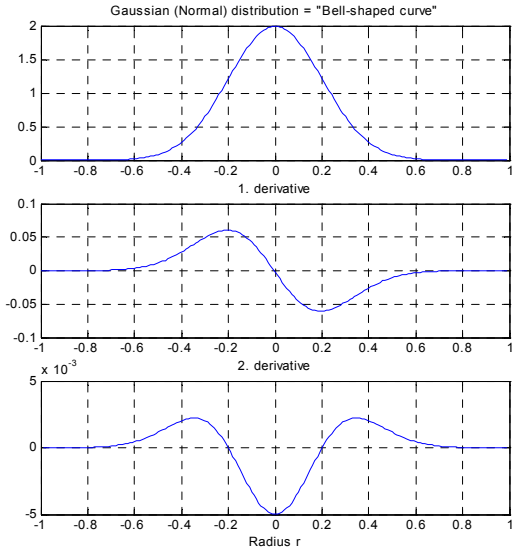

Figure 1. *Gaussian function (σ=0.2) and its derivatives.*

$$g(r,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1)$$

where:
r is the radius (polar coordinates).
σ is the space-constant.

The changes in the function give rise to peaks in the first derivative and zero-crossings in the second at the points r =-0.2 and r = 0.2.

The detection of the peaks or zero-crossings in the smoothed and differentiated images together with peak's position, sharpness and height, result in the *primal sketch* representation. This is believed to approximate the first stage in the human vision processing and accepted as such in machine vision [4].

3. MODELLING THE AUDITORY SIGNAL PROPAGATION DYNAMICS

The combination of the Gaussian filter and its second derivative (*Laplacian* operator) has several properties that make it a 'favourite' linear filter choice in edge detection theory and applications. It is considered to be an optimal edge-detection operator [3], as well as the only operator that exhibits 'nice' behaviour in the scale-space representation [5]. Also, both the Gaussian filter behaviour and that of its derivatives have been found to be remarkably similar to the response of some cells in human visual pathways [6]. Here, we are interested in the temporal domain filtering, i.e. in the smoothing that emerges when propagating the signal through an echoic memory neural map.

One possible mechanism for performing the multi-scale analysis of temporal signals was introduced in [7], along with a number of issues important for the development of auditory signal representations similar to the *primal sketch*.[2] Todd proposed a temporal approximation of the Gaussian, and implemented this in the form of a low-pass filter bank model of the echoic store. According to Todd, the resulting multi-scale mechanism (*rhythmogram*) can be understood as a model of auditory sensory memory as well as of a number of other important auditory phenomena.

However, Todd does not discuss the underlying dynamics of the signal propagation through the echoic store modelled by the proposed Gaussian filter temporal approximation. The close relationship between the Gaussian filter concept and diffusion mechanism [8, 9], indicates a possible analogy between the multi-scale mechanism proposed in [7] and the diffusion - e.g. one-dimensional heat conduction according to equation 2.

$$\frac{\partial \vartheta(x,t)}{\partial t} = k \frac{\partial \vartheta^2(x,t)}{\partial x^2} \tag{2}$$

where:
$\vartheta(x,t)$ is the temperature at point x, moment t.
k is the constant related to the physical properties of the material.

Under specific initial and boundary conditions the function $\vartheta(x,t_k)$, i.e. the solution to equation 1, at the time instant t_k, takes a form of the Gaussian curve shown in figure 1.

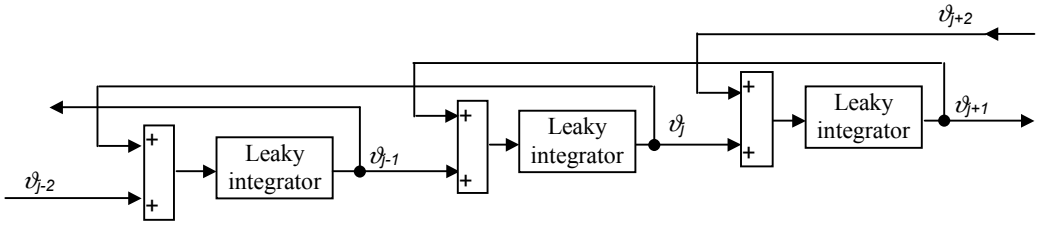

Figure 2. *Simplified signal propagation model based on the Gaussian filter approximation.*

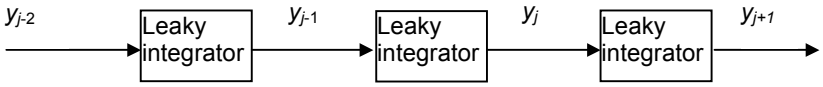

Figure 3. *The leaky integrators cascade.*

By use of the *control volume* approach the approximate model of one-dimensional heat conduction can be expressed by the set of *n* ordinary differential equations. Apart from the *boundary volumes* equations, the equations are of the following form:

$$\frac{d\vartheta_j}{dt} = C\left(\vartheta_{j-1} + \vartheta_{j+1} - 2\vartheta_j\right) \quad \text{for } j = 2, ..., n\text{-}1 \tag{3}$$

The structure approximating the equations 3 is shown in figure 2. The basic element of the structure is the first order system with a time constant, often referred to as *leaky integrator*. Since the *leaky integrator* is known as one possible model of neural signal transmission, the fact that the signal passes in both directions in the above propagation model implies backward connections in the neural map i.e. its bi-directional conductivity.

Considering the physical and neuro-anatomical constraints, we have chosen a pure cascade of *leaky integrators* (figure 3) for modelling signal propagation through the neural map instead of the diffusion model.

Similarly to the interpretation of the Gaussian-based model through the heat conduction dynamics, the *leaky integrators* cascade can be interpreted as a spatially discrete approximation of the one-dimensional convective transport mechanism (see equation 4).

$$\frac{\partial y}{\partial t} + w\frac{\partial y}{\partial x} = 0 \tag{4}$$

where:
$y(x,t)$ is the signal at point x, time t.
w is the signal carrier velocity.

Apart from exhibiting the characteristic smoothing behaviour of the sensory transduction, the *leaky integrators* cascade model (figure 3) enables a relatively straightforward implementation of the appropriate derivative estimator that is important for a fully developed melodic segmentation model.

4. DISCRETE REALISATION OF THE LEAKY INTEGRATORS CASCADE

Finally, the question of the computational implementation of the above proposed model is addressed, taking into account the fact that the *leaky integrator* inherently belongs to the class of continuous systems, and that the model inputs are temporally discrete. A linear first order system with a time constant (i.e. *linear leaky integrator*) is described as:

$$T \cdot \frac{dy}{dt} + y = x \tag{5}$$

where:
T is the time constant.

Assuming the linearity of the signal propagation dynamics, equation 5 can be rewritten for the j[th] leaky integrator in an analogue cascade:

$$T\frac{dy_j}{dt} + y_j = y_{j-1} \tag{6}$$

where:
j is the *leaky integrator* index.

The *leaky integrator* was digitally implemented in the form of the first order ARMA (autoregressive moving average) filter, defined here in index notation:

$$y_{j,k} = a \cdot y_{j,k-1} + (1-a) \cdot y_{j-1,k} \tag{7}$$

where:
k is the current moment index.
$y_{j,k}$ is the current output of the j^{th} integrator.
$y_{j,k\text{-}1}$ is the previous output of the j^{th} integrator.
$y_{j\text{-}1,k}$ is the current input to the j^{th} integrator.
a is the ARMA constant.

The connection between the integrator described by the equation 6 and its discrete realisation (equation 7) can be elaborated as follows. For the discrete form of the derivation as a function of time:

$$\frac{dy_j}{dx} \approx \frac{y_{j,k} - y_{j,k-1}}{h} \tag{8}$$

where:
h is the time step (t_k-$t_{k\text{-}1}$).
k is the current moment index.
k-1 is the previous moment index.

By incorporating equation 8 into equation 6, we obtain the discrete form of the *leaky integrator* equation:

$$T \cdot \frac{y_{j,k} - y_{j,k-1}}{h} + y_{j,k} = y_{j-1,k} \tag{9}$$

From equation 9, $y_{j,k}$ can then be expressed explicitly:

$$y_{j,k} = \frac{T}{T+h} \cdot y_{j,k-1} + \frac{h}{T+h} y_{j-1,k} \tag{10}$$

Equation 10 is the discrete approximation of equation 6. The approximation improves as the time step h decreases. The condition for identity of equation 10 and equation 7 is easily obtained from the two equations:

$$a = \frac{T}{T+h} \tag{11}$$

Equation 11 represents the relation between the time constant of the leaky integrator (T), and the coefficient of the ARMA filter (a).

5. EXAMPLES

The input to the described segmentation model consists of the melodic descriptors extracted directly from a recorded melody [10]. The descriptors are pitch, loudness and one aspect of timbre - centroid frequency descriptor. In figure 4, we show the dynamic behaviour of the proposed cascade, through the model's response to a synthesized pitch input consisting of the notes of the major scale (roof change) and a tritone in the same scale (2 abrupt changes). The cascade has 15 filters, and all 15 responses are shown.

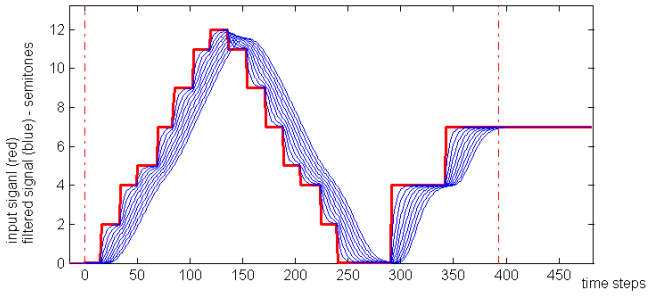

Figure 4. *Smoothing properties of the proposed model for the scale and tritone input.*

Observe that by the time the signal reaches the last of the filters, higher 'change frequencies' (scale steps) have been filtered out leaving the ramp shape, while the stronger abrupt changes (tritone) are correctly reproduced.

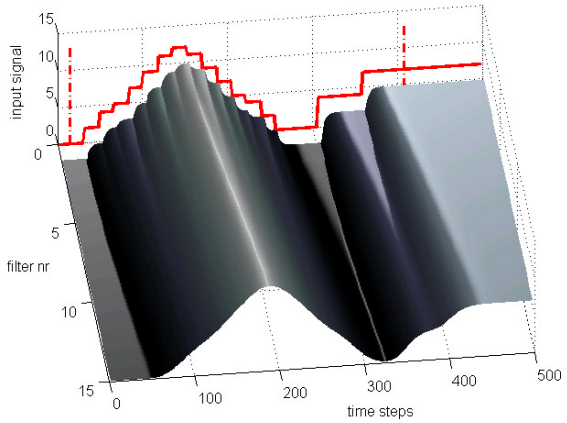

Figure 5. *Smoothing properties of the proposed model, scale and tritone input.*

The dynamic behaviour of the model representing a form of memory, can best be observed in 3D view (figure 5). The model shows the same properties (smoothing and remembering) in the case of the loudness indicator. For instance, the "Tennessee air" phrase input propagation is shown in figure 6.

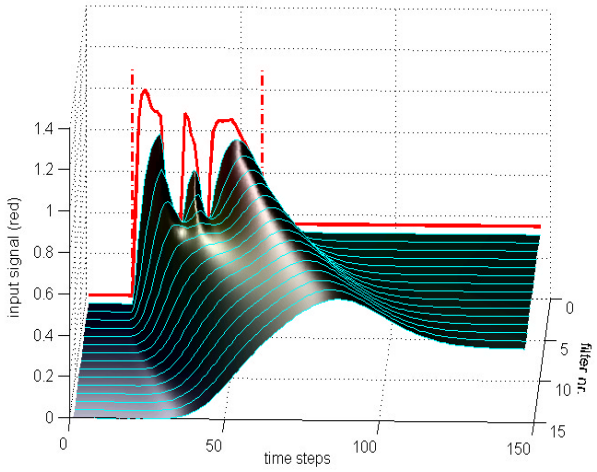

Figure 6. *The propagation of the "Tennessee air" phrase input.*

6. FUTURE WORK

Based on the extensive classification of the Western Tonal Music - melodic shapes, provided in Narmour's work [11] we can speculate that the two most common changes in any melodic descriptor are an abrupt (step) change and a roof (ramp) change. Detection of these two change types requires a mechanism sensitive to the changes in a melodic descriptor (first derivation) and to the changes in the descriptor's gradient (second derivation).

Such a mechanism has indeed been designed and integrated into the proposed model structure and is currently being evaluated. The first and the second differences obtained for the same input as in figure 4, are shown in figure 7. As is the case with the spatial structure, the information obtained from different channels (filters) must be combined in such a way to obtain meaningful entities across different scales. This is the next phase in this research.

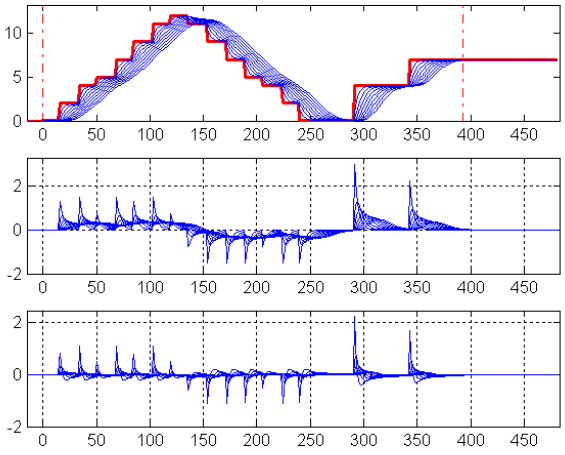

Figure 7. *Pitch Input from figure 4 with its first and second derivations.*

7. ACKNOWLEDGMENTS

My thanks go to Niall Griffith and Nikola Šerman, for sharing their knowledge, thoughts, and insights with me. Without such generosity and support on their part this work would not have been possible. Thanks to Ciara Finnegan, for providing her lovely Northern Irish accent in the 'Tennessee air' phrase.

8. REFERENCES

[1] Marr, D., "Early processing of visual information", *Philosophical Transactions of the Royal Society (London) Series B*, 275:483-524, 1976.

[2] Marr, D., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco, 1982.

[3] Hildreth, E., "Edge detection", Artificial Intelligence Laboratory, Massachusetts Institute of Technology, A.I. Memo No.858, 1985.

[4] Blake, A. and Troscianko, T. (Eds.), *AI and the eye*, John Wiley & Sons, Chichester, England, 1990.

[5] Yuille, A.L. and Poggio, T., "Scaling Theorems for Zero-Crossings", Artificial Intelligence Laboratory, Massachusetts Institute of Technology, A.I. Memo No. 722, 1983.

[6] Hendee, W.R. and Wells, P.N.T., (Eds.), *The Perception of Visual Information*, Springer-Verlag, New York, 1997.

[7] Todd, N.P.M., "The Auditory "Primal Sketch": A Multiscale Model of Rhythmic Grouping", *Journal of New Music Research*, 23:25-70, 1994.

[8] Aurich, V. and Weule, J., "Non-Linear Gaussian Filters Performing Edge Preserving Diffusion", Proc. 17. DAGM-Symposium, Bielefeld, pp. 538-545, 1995.

[9] Lindeberg, T., "On Scale Selection for Differential Operators", Proc. 8th Scandinavian Conference on Image Analysis, Tromso, Norway, 1993, pp.857-866.

[10] Narmour, E., *The Analysis and Cognition of Basic Melodic Structures*, University of Chicago Press, Chicago, 1990.

[11] Serman, M., Griffith, N. & Serman, N., (2000) "MusicTracker: a system for modelling melodic dynamics in music performance", ICMC2000, Berlin, pp. 217-220, 2000.

[1] Edge detection is a commonly used if somewhat misleading term. What is actually detected are changes in intensity and other properties of images that are thought to enable recognition of edges in the real world.

[2] The multi-scale mechanism proposed by Todd is more or less specific to the auditory rhythm perception.