2017 American Control Conference
Sheraton Seattle Hotel
May 24–26, 2017, Seattle, USA

On fast retrieval of relational experiences

Weibo Gong*
Dept. of Electrical and Computer Engineering and College of Information and Computer Sciences
University of Massachusetts
Amherst, MA 01002, USA
gong@ecs.umass.edu

Yu-Chi Ho
John A. Paulson School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02139, USA
ho@deas.harvard.edu

***Abstract*— Various intelligent systems are needed for cyber-physical systems. Such intelligent systems need to learn from the human intelligence about concept abstraction and analogical thinking in order to resolve complex issues using past experiences. The algorithms for abstraction and analogies are based on quick memory recall with clever information coding and processing. The quick and accurate memory recall is based on the fact that the memory mostly records the relations among the constituents of the stimulating signals, rather than the constituents themselves. Relational memories can be stored in the form of networks of neuron clusters capable of resonating to particular signal sequences. However, similarity testing for such network representations is difficult. We suggest that linear dynamic systems that relate the system matrix and the output time function can be used as a conversion mechanism between the network matrix and the temporal representations of the signals. This leads to algorithms for relational similarity testing and concept abstraction. Transient behavior based selection rules in ordinal optimization is important in achieving quickness in our development.**

I. INTRODUCTION

Intelligence in cyber-physical systems often needs to resolve problems in understanding highly complex information and make decision choices based on analogically similar experiences. One of the key challenges is the complexity in robust and accurate object recognition and concept abstraction. In this paper we propose a set of algorithmic schemes in response to these challenges.

Most machine learning algorithms are based on convergence of iterative procedures, which is in sharp contrast with the quickness of human and animal recognition. On the other hand, it has been the foundational knowledge in control theory that the steady-state frequency domain behavior of a linear dynamic system and its transient responses are just the two sides of the same coin. Nature is likely to take advantage of such duality. In other words neuronal networks could act in time domain to produce some seemingly stationary capacities such as memory and recognition that are based on frequency domain characterizations. In this paper we develop algorithms along this line. Our algorithms are aimed at computer implementations with biological plausibility as a guiding principle.

We take a quite different approach to machine intelligence than the current main stream developments centered in deep learning networks and statistical methods. The achievements of these modern neural networks with supervising objective functions and back propagating gradient learning are undoubtedly impressive and surpass all the past records in multiple areas of machine learning. However it is hard to see a clear path toward cognitive functions such as metaphorical analogies that is considered a hallmark of the human intelligence [13][11]. Our view toward the development of intelligent systems is that one needs to aim at the capability of abstract concept similarity testing and work downward to specify the requirements for the lower level algorithms. In future intelligent systems one will encounter problems requiring human level abstraction to handle unforeseeable events. In these scenarios the past experiences that are similar in abstract ways are very useful. Interestingly our development indicates that the biological intelligence based on abstract similarity testing is an evolutionary inevitability, since the recognition algorithms for abstract relational similarity were already developed for concrete geometric shape similarity. Furthermore as we argue later the seemingly advanced intelligence capacities are also used in concrete object recognition tasks because relations among parts of the object is crucial in identifying the object. In this sense the division of being abstract vs concrete is artificial.

To deliver human mind-like capabilities an intelligent system has to let the sensory signals excite stored experiences accurately, robustly and analogically. When we see a tree leave, we see many details accurately. This allows us to pick a subset of the details and their relations to form a robust recognition of the kind of the trees. It also allows us to tell the difference from another leaf of the same tree. We can also be reminded of another tree half earth away and many years ago, and this may poetically lead to the analogy of a leaf and a person's life.

All these can be accomplished if the intelligent system has a coding scheme that allows a huge memory capacity with quick and accurate recall. Superficially these goals conflict with each other. Nature must address such conflicts with all plausible resources, including resonance to oscillatory signal-

* This work is supported in part by the United States National Science Foundation CNS-1239102 and Army Research Office Contract W911NF-12-1-0385.

s, time ordered sequences for memory, and spatial-temporal signal conversions. These are used in our developments.

II. Resonance transients for signal component recognition

We consider a signal as superposition of many exponentially decayed sine functions (sometimes referred as Boils functions). The expressing power of such a superposition has been illustrated in literature such as [1][2]. For the current discussion we limit ourselves to the case where all components start at time zero. The generalizations to signals with delayed components is straightforward.

Motivated by biological arguments we study the transient behavior of a second order dynamic system stimulated by a decayed sine function input. Specifically we check

$$\begin{aligned} h(t) &= e^{-\lambda t}\sin(nt) * e^{-\mu t}\sin(mt) \\ &= \int_0^t e^{-\lambda t}\sin(n\tau)e^{-\mu t}\sin(m(t-\tau))d\tau. \end{aligned}$$

We assume the neural circuits are efficient high Q filters with a very small μ. We also assume the input signal does not decay very fast and thus is with a small μ. The latter can be relaxed. Under these assumptions we carry out a transient behavior analysis.

We have (treating both λ and μ as zero):

$$\begin{aligned} & h(t) \\ = & \frac{m\sin(nt) - n\sin(mt)}{m^2-n^2} \\ = & \frac{m}{m^2-n^2}\left[\sin(nt)-\sin(mt)\right] + \frac{m-n}{m^2-n^2}\sin(mt) \\ = & \frac{m}{m^2-n^2}\left[\sin(nt)-\sin(mt)\right] + \frac{1}{m+n}\sin(mt) \\ = & \frac{2m}{m^2-n^2}\left[\sin(\frac{n-m}{2}t)\cos(\frac{n+m}{2}t)\right] \\ + & \frac{1}{m+n}\sin(mt). \end{aligned}$$

When m and n are large the impact of the last term is minimal. The first term is a sine wave at frequency $|n+m)|/2$ modulated by a low frequency sine wave at frequency $|n-m|/2$. When $n-m \to 0$ it can be shown that this term goes to t, a basic phenomena in resonance. When n and m are close this term results in an envelop starting going down at around $0.25/|n-m|$. When the difference $|n-m|$ increases this point moves closer to the time origin. We can check the location of this point in the filter response to decide whether the input signal contains the component or not. Note that this is independent of the amplitude of the components in the incoming signal. Check Figure 1 for a small scale experiment.

functions and legends of Figure 1.

```
t=0.0:0.01:20.0;
a=exp(-0.1*t).*sin(500*t);
b=exp(-0.1*t).*sin(505*t);
c=exp(-0.1*t).*sin(510*t);
d=exp(-0.1*t).*sin(515*t);
e=exp(-0.1*t).*sin(495*t);
```

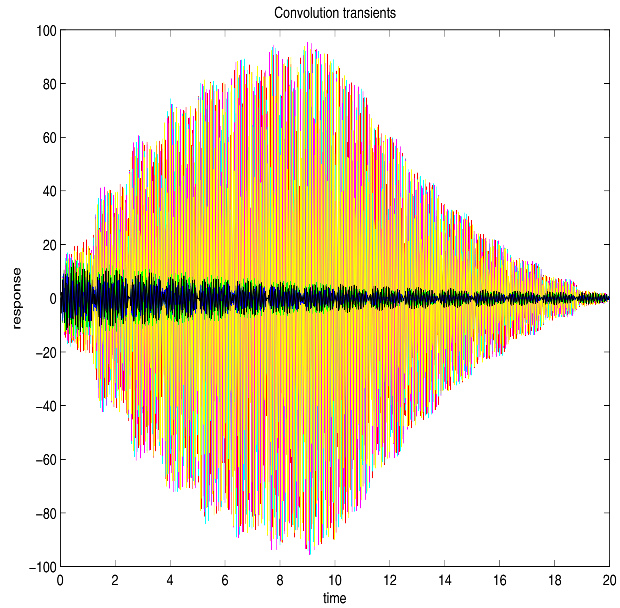

Fig. 1. Using Resonance Transients to Recognize Input Components

```
f=exp(-0.1*t).*sin(490*t);
g=exp(-0.1*t).*sin(520*t);
x=conv(a+b+c+d,a, red;
y=conv(a+b+c+d,b), cyan;
z=conv(a+b+c+d,c), magenta;
w=conv(a+b+c+d,d), yellow;
u=conv(a+b+c+d,e), green;
v=conv(a+b+c+d,f), blue;
s=conv(a+b+c+d,g), black
```

The usefulness of recognizing an exponentially decayed sinusoid component has been demonstrated in many works in speech signal processing. More importantly the spatial geometric shapes can be represented by superpositions of exponentially decayed sinusoids based on spectral analysis. A simple example in this regard is to consider two "white noise" images generated by independent random numbers. They look very similar to our eyes, suggesting that our perception is based on the frequency domain signals since the two white noise images has the same spectrum of sinusoid decompositions despite that they are "mathematically" independent.

III. Memory unit

To take advantage of the resonance transient behavior exhibited above we assume that in an intelligent system the concepts are represented by exponentially decayed sinusoid (EDS) signal and their combinations. The question now is how to code the EDS signal combinations in memory. To this end we consider a set of resonators sequentially arranged so that each would get excited in the time order in which the matching frequency occurs in the input signal. We call this arrangement a Resonant Chain Unit, or RCU. The importance of time ordering in memory coding has been emphasized in the immensely successful Long Short-Term

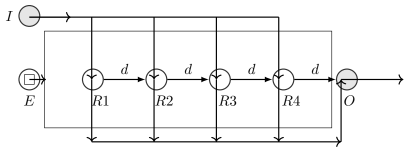

Fig. 2. Basic structure of a Resonator Chain Unit (RCU)

Memory (LSTM) in machine learning [16]. Our unit is much closer to physical layer of intelligence and is not designed or even suitable for the back propagation training. RCU depends on the resonance mechanism to code information in a unsupervised manner.

A RCU is composed of several resonators with mostly different resonating frequencies. There are fixed time delays between the successive resonators. For an illustrative example consider the case of a RCU consisted of 4 resonators $R1 \sim R4$ with frequencies $f_1 \sim f_4$, respectively. The RCU is build as $f_1 \to d \to f_2 \to d \to f_3 \to d \to f_4$. When f_1 is excited by a matching frequency in the input its output go through the delay d and get the f_2 resonator ready. But f_2 has to also receive a signal with a matching frequency to get excited. This process goes on until all resonators are excited, or the unit excitation is aborted. The RCUs in a region can get all excited by receiving a signal with many frequency components. Figure 2 is a schematic basic structure of a RCU. RCU could be forced to excite by a signal at the circled box E. Such signals could be due to group connection via simultaneity, attention, and other learning-based connections.

When a RCU is excited it generates output as the sum of all its resonators with the delays between them. Note that in Figure 2 the sum signal can go out only when R4 is excited. The sum signal goes on to other parts looking for similar RCUs to excite. In the areas immediately behind the signal sensors the RCUs reset to rest condition shortly after the excitation input signal vanishes. However the RCUs in the memory region would sustain the excitation states longer in order to form connections with other excited RCUs to form a network recording the simultaneously excited RCU signals. For computer implementation such period could be very small.

Take visual signal processing as an example. The photo sensory signals are converted into time functions and go through a random fabric that channels the time function to many RCUs. The spatial-temporal signal conversation algorithm is discussed later, which is generally applicable to convert the signal of a small cluster of spatially connected RCUs to a time function.

Now consider two clusters of RCUs, A and B. When the two clusters send RCU output signals to each other the RCUs that resonant would form strong connection links and those who do not would remain connected only in their original clusters. The resonated ones form a new cluster $A \star B$. The others, suppressed by the resonant ones form clusters $C = A - A \star B$ and $D = B - A \star B$, respectively. Now we have clusters $A, B, A \star B, C, D$. This way A, B and their generated clusters together form the relations between A and B. These clusters can be modeled as a complex valued time varying matrix. The matrix can be converted into temporal representations via the spatial-temporal conversion scheme discussed in the next section. The temporal representations would be coded into a new cluster of RCUs in memory as the relations between A and B.

A RCU could also serve as a node in a super RCU to code more complicated sequences. There are other uses of RCUs to code spatial-temporal information and to construct hierarchies of memory trees to facilitate fast retrieval of information coded by long RCU sequences.

The RCU interactions are quick for the following reasons. When an oscillatory time function hits a group of second order filters its components will selectively excite the resonators with natural frequencies very close to the frequency of one of the components, a phenomena similar to the famous Barton's pendulums. While most resonance phenomena takes a while to signify and most textbooks analyze only the steady state behavior, resonance effect in fact accumulates from the moment when external signal arrives. In other words, when a sinusoid function hits a group of resonators the resonating response curve deviates from others immediately. While the initial differences are small, they are enough for inhibiting other resonators. This is similar to the selection rule in the ordinal optimization method [8] referred as horse race rule. As the name implies, simulation based optimization decision can often be decided long before reaching steady state. Often times the transient behavior of a system reveals its potential when ordinal comparison is the decision base. In our current situation it is known in neuroscience that most neurons when excited tend to inhibit other neighboring neurons and render them silent. Such feature is known to be crucial for edge detection in human and animal vision systems, and is observed in other brain neurons. If we restrict to a mathematically simplified scenario where a causal sinusoid function convolve with the impulse response function of a seconder order linear system, it can be seen that with the above convolutional horse race selection the input sinusoid function would very quickly select a receiver system that resonances with it. We believe this is the basis for the quickness of image recognition exhibited in human and many animals. In fact we believe this is also the base for the quickness of abstract thoughts where relevant concepts in the memory are recalled very quickly.

Resonance transients based quick recall assumes linear signal processing and enables signal aggregation and decomposition in large scale, which are necessary for all kinds of intelligence activities. In previous works [5][6][7] we connect the spatial-temporal conversion in linear control theory to the researches relating the geometric shapes to their Laplacian spectrums [10], [14]. In contrast to the biological memory theories based on nonlinear circuits, we suggest that signal processing in linear dynamic systems has advantages in making a large scale content addressable memory. Although nonlinear mechanisms are ubiquitous in biological systems including the brain, linear signal processing in these systems could nevertheless be essential in certain signal range. The usage of nonlinear elements in neural networks is to facilitate function approximation so the focus is different than ours.

IV. Spatial-temporal signal conversion

As we mentioned the most obvious case for spatial-temporal signal conversion is in image processing, although the idea is generally applicable for converting a cluster of RCUs to a summarizing time function. Various attempts of using diffusion for obtaining multiscale images, generally referred to as scale space method, are based on the property of Gaussian-like diffusion kernels that they do not introduce extrinsic features into the diffused image. This is fundamentally due to the fact that Gaussian distribution is a model for the sum of many nearly i.i.d. random components. As such if one views the diffusion on image as to spread the intensity of every pixel around with random jolts for nearly i.i.d directions and strengths then the above property is not surprising. One does not want to use deterministic averaging since the boundary and other parameter choices bring extrinsic features. This has led researchers to conclude that the best blurring is via Gaussian kernel.

Our visual system has to code the image information and transmit it elsewhere for memory recall. In human-made engineering systems such coding and transmission are often based on resonance of oscillatory signals. Since an image is represented as a static matrix of pixel intensity we need to code the information into oscillatory signals. Recent study [15] indicates that the microsaccades during fixation of the fovea would turn a stationary image patch into a dynamical one in order to maintain the sensitivity of the photo sensors in the retina. It is reasonable to assume that the visual signals generated from images are composed of oscillatory time signals. We will represent visual images as spatially connected RCUs. For each time point we consider a complex $n \times n$ matrix $V(x)$ where x is the RCU position. The adjacency matrix of $V(x)$ is a $n^2 \times n^2$ matrix denoted as A with the entries representing the link weights between the row node and the column node as

$$a_{xy} = \frac{G|V(x) - V(y)|}{\|x - y\|^2}$$

which is a complex number. G is a real constant. Now we consider the Laplacian matrix for $V(x)$ as

$$L = D - A$$

where D is the diagonal matrix with the row sum of A as the diagonal entries. We use L to form a linear differential equation for the "state" vector $\psi(t)$ as

$$\dot{\psi}(t) = -L\psi(t), \quad \psi(0) = \text{uniform}$$

and the complex valued $\psi(t)$ is turned to a real time function by

$$y(t) = C\psi(t)\psi^*(t)C^*$$

with vector C representing a summation over the given image region. This time function $y(t)$ is the temporal representation of the image patch defined by C that we are pursuing. We note that only a small segment of C is non-zero to reflect the physical range of a summarizing neuron. It also enables the coding of the spectral component phase as the sum of the eigenvector entries at the non-zero C segment. This deviates from the usual approaches using Dirichlet or other well-known boundary conditions for Laplacian spectral analysis of domain shapes.

The above procedure assumes that a spatial signal is a complex number matrix. Then the diffusion like equation is to use peer pressure to smooth out the value changes in the neighborhood. The diffusion currents resulted from such smoothing action are collected as the temporal representation of the spatial signal. When the system matrix is real the realization theory of linear time invariant systems guarantees that one can recover the original spatial signal from such temporal representations under general conditions. This can be generalized to time varying complex spatial signals and indicates that the information is preserved in such conversions. One can see that the above resonance transients based "realization" is much less demanding on the numerical accuracy than the Markov parameter based realization algorithms.

V. Abstraction and analogical thinking

One of the magic the biological intelligence perform is to abstract, which is the basis for analogy. In [13] Steven Pinker argued forcefully that metaphorical analogy is the niche for human intelligence evolution. To perform metaphorical analogy it is necessary that signals in brain travel in aggregated form and find content in other places with many similar components. This demands signal superposition and decomposition which are the hallmarks of linear signal processing. We outline an algorithmic model for concept abstraction here.

Abstraction is the process of sifting the similarities from the instances or the differences among the instances. One of the impressive examples of abstract similarity testing is the psychology phenomena referred as analogical reminding, where a sequence of current events remind a possibly remote experience that is only similar in an abstract manner [9][12]. Analogical reminding is featured by the quickness, the abstractness, and the involuntariness of a memory recall carried out by the relatively slow neurons. The quickness calls for large scale content addressable memory. The abstractness needs the ability to sift commons from instances. And the involuntariness demands automatical recall. The spatial/temporal signal conversion described above could help achieving these.

To explain the idea of abstraction we use an example of shape relation similarity testing. The shapes in the visual field are represented as clusters of connected RCUs. Using the above conversion scheme one has the temporal representation of a RCU cluster, which codes the spectral information of the shape images. A temporal function is just a vector when discretized and the relation between two vectors can be coded in a matrix/network/RCU cluster. Now we consider simple analogies in which one tests the similarity between the relations that relate more concrete concepts. Consider two pairs of geometrical shapes (A, B) and (X, Y). Suppose it makes sense to say that "A is to B as X is to Y", for example "a square A is to a rectangle B as a circle X is to an ellipse Y". The following diagram shows the relational matrices R_{AB} and R_{XY} (which would be coded into RCU clusters) between the shape pairs.

Now the relational similarity statement that "A is to B as X is to Y" can be understood as the similarity of matrices R_{AB} and R_{XY}, which is algorithmically the same as the similarity testing of the two concrete shapes such as two squares.

$$\begin{tikzcd} A \arrow[r] \arrow[d, "R_{AB}"'] & X \arrow[d, "R_{XY}"] \\ B \arrow[r] & Y \end{tikzcd}$$

More concretely let A be a square and X be a circle, and B and Y are their stretched versions namely a horizontal ellipse and a horizontal rectangle, respectively. In terms of the RCU cluster representations one checks the similarity of the similar components between A and B and those between X and Y, as well as the similarities of the differences between A and B and those between X and Y. The information that X is a stretched version of A and Y is a stretched version of B would emerge from these comparisons. In fact the concept of "stretching a shape" would be represented by the commonalities emerged from such comparisons, forming the RCU cluster representation of the abstract concept of "stretching a shape". Specifically the RCU cluster representations for the circle and the ellipse would differ since the spatial-temporal conversion would make different temporal representations for them. The differences would be coded in the RCU cluster representations and such differences would have similar components with the differences between the square and the rectangle. This commonality is a prototype of the concept "shape stretching".

The similarity testing for relations discussed above does not need conscious instructional effort and leads to an involuntary experience, not unlike recognizing facial expressions. Recognizing facial expressions such as a smile is crucial for human interactions. Consider two faces A and X, and their smile versions B and Y. The relational matrices R_{AB} and R_{XY} are similar and the similarity defines the concept of "smile". Human babies perhaps sift this out from smile faces early on. The relational network of smile could make connections to relevant concepts and it is possible that the silly smiles and sounds made by adults when holding a baby form the neurological base for the sense of humor, which would be triggered by concrete or abstract silliness depends on the storage of abstractions in the mind.

The relational networks R_{AB} and R_{XY} are generated automatically (due to resonance among components of the temporal representations A, B, X, Y) and stored in the memory as part of the experiences. Such relational networks, and the relational networks for these relational networks, are all generated in subconscious and sitting there ready to be excited. This may help explaining the analogical reminding phenomena mentioned before.

Furthermore, since the above scheme in testing the relational similarity is the same as the "concrete" shape similarity, the process can be repeated to extract multiple levels of relations that exists between different lower level relations. The spatial-temporal signal conversion plays a central role in this recursion. Since a RCU cluster can be represented by a matrix, the algorithm capable of converting the information carried in a matrix or by a time function back and forth enables such recursion. This process generates many RCU clusters representing all sorts of relations which serve as cues and constraints for memory recall.

Historically mathematician Mark Kac asked "Can one hear the shape of a drum" in [10] where the mathematical question is can one recover the geometrical shape of a continuous domain by checking only the heat kernel time function generated from a linear dynamic system acting on the domain. His topic and approach has a long and broad impact in mathematical research. Our discussion here can be expressed as a similar question of "Can one hear the shape of a concept?"[5]. Basically if a concept, represented by a network of RCUs, can also be represented by a (short) time function, then the above scheme will be able to generate many relational networks of RCUs to be stored in memory and to be used for intelligence functions such as metaphorical analogies.

We emphasize again the following important point in the above discussions. As far as representations in the brain is concerned, there is no difference between abstract concepts and concrete objects. All contents in memory are represented by the same type of RCU clusters with their connecting networks. An object or concept is the collection of relations among its constituent components as well as the relations with other objects or concepts. Memory recall is achieved by matching the relation constraints through the similarity testing algorithm described above. In this process oftentimes a portion of relation matching is enough to single out the wanted object or content, similar to reading emails with misspelled words. This "80-20" phenomena has been studied in the topic of ordinal optimization in discrete event systems [8][3].

VI. Stationary vs Transient - Some Connections to Systems and Control Research

The topic of recovering the shape of a domain from the Laplacian eigenvalues belongs to the spectral inversion problem and has a long history. It is interesting to systems and control people to note that originally Hendrik Lorentz formulated the problem as a wave problem in 1910. He said that in an enclosure with a perfectly reflecting surface there can form standing electromagnetic waves analogous to tones of an organ pipe. He then conjectured that for a 2D domain (a membrane) the number of Laplacian eigenvalues less than λ would approach $\|D\|\lambda/2\pi$ when $\lambda \to \infty$ which was proved by Herman Weyl soon after. The original problem and the approaches were all about waves and wave equations (including the Schrodinger equation) and thus the phrase "hearing the drum". It was Mark Kac who started treating the problem using diffusion theory, making use of the fact that both waves and diffusions are the acting of the same Laplacian operator. While in the wave approach the steady state behavior was the focus of analysis, in the diffusion approach of Kac the subject was the transient behavior of the diffused "stuff". We all know in control theory there are two alternative approaches to system identification for a linear time invariant system, namely either to use the sinusoidal input to obtain steady state frequency domain data, or to use an impulse or step input for transient time domain data. The latter is less practical in control engineering. However as we discussed in this paper it turns out that the time domain transient method becomes quite useful in the identification of a sinusoid components early on, and this can serve for tasks such as abstract relational similarity testing.

In the study of using heat diffusion quantities to recover the domain shapes along the lines of Mark Kac researchers have considered a uniform initial distribution of temperature and formulate a problem like the following [17].

Consider a compact Riemannian manifold (M, g) with a smoothly closed domain $D \in M$. Consider a Laplace operator acting on functions with a Dirichlet boundary condition. Let $p_D(x, y, t)$ be the kernel associated to D, and dg be the volume form associated to the metric, let

$$u(x,t) = \int_D p_D(x,y,t)dg(y)$$

be the solution to the initial value problem

$$\frac{1}{2}\Delta u = \frac{\partial u}{\partial t} \quad \text{on} \quad D \times (0, \infty),$$

$$u(x,0) = 1 \quad \text{if} \quad x \in D \quad \text{and} \quad 0 \quad \text{if} \quad x \in \partial D,$$

$$u(x,t) = 0 \quad \text{if} \quad x \in \partial D.$$

Letting $q(t)$ be the heat content of D at time t:

$$q(t) = \int_D u(x,t)dg,$$

[17] has shown that $q(t)$ admits a small time asymptotic expansion

$$q(t) \sim \sum_{n=0}^{\infty} q_n t^{n/2}$$

where the coefficients q_n are geometric invariants of D.

The line of mathematical works pioneered by Mark Kac [10] provide a motivation for the development of the random walk algorithms for concept abstraction [4][6]. Random walks on graphs serve as a high level of analogy to neuronal spikes moving around a neural network. In computer algorithms one can use spectral computation as we discussed. In fact even the computation of the resonance transients can be significantly simplified in computer implementations.

VII. Concluding remarks

Linear dynamic systems are powerful for signal aggregation, decomposition and spatial-temporal conversion. The superposition property enables the signal aggregation and decompostion required for searching the similar components in other different regions via a fabric of random connections. Linear systems with oscillatory responses are particularly useful for signal transmission and similarity testing. We suggest that linear systems with complex valued system matrices can be used to perform conversions between the temporal and spatial representations of signals. We also propose using clusters of Resonator Chain Unit (RCU) to code information and to form relations between time signals. While time functions are good for transmission, RCU cluster are plausible for memory and reliable retrieval . Combined, they offer possibilities to explain certain functions in biological intelligence such as the quick recall of similar contents and analogical thinking, as well as for the development of computer algorithms having similar functions. The transient behavior based selection rules in ordinal optimization is very useful in this development and may lead to efficient implementations for intelligent systems

References

[1] G. Beylkin and L. Monzon. On approximation of functions by exponential sums. *Applied and Computational Harmonic Analysis*, 19(1):17 – 48, 2005.

[2] O. Derrien, R. Badeau, and G. Richard. Parametric audio coding with exponentially damped sinusoids. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7):1489–1501, July 2013.

[3] P. Glynn and S. Juneja. Ordinal optimization: A nonparametric framework. *Proceedings of the 2011 Winter Simulation Conference*, 2011.

[4] W. Gong. Are stochastic dynamics the foundation of intelligence? In *http://www.ieeecss-oll.org/lecture/are-stochastic-dynamics-foundation-intelligence*. 50th IEEE Conference on Decision and Control, 2011.

[5] W. Gong. Can one hear the shape of a concept? In *Proceedings of 31st Chinese Control Conference*, pages 22–26, 2012.

[6] W. Gong. Transient response functions for graph structure addressable memory. In *Proceedings of 52nd IEEE Conference on Decision and Control*, pages 6978–6985, 2013.

[7] W. Gong. Spatial/temporal coding for involuntary concept abstraction using linear dynamic systems. In *Proceedings of 2015 European Control Conference*, 2015.

[8] Y.-C. Ho, Q. Zhao, and Q. Jia. *Ordinal Optimization: Soft Optimization for Hard Problems*. Springer, 2007.

[9] D. Hofstadter and E. Sander. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books, 2013.

[10] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73(4, part2):1–23, 1966.

[11] J. Markoff. Taking baby steps toward software that reasons like humans. *New York Times*, March 2016.

[12] S. Pinker. *The Stuff of Thought*. Viking Penguin, 2007.

[13] S. Pinker. The cognitive niche: Coevolution of intelligence,sociality, and language. *PNAS*, 107(May):8993–8999, 2010.

[14] M. Reuter, F. Wolter, and N. Peinecke. Laplace-beltrami spectra as "shape-dna" of surfaces and solids. *Computer-Aided Design*, 38(4):342–366, 2006.

[15] M. Rucci and J.D. Victor. The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences*, 38 4:195–206, 2015.

[16] J. Schmidhuber S. Hochreiter. Long short-term memory. *Neural Computation*, 9(8):1735–1780, November 1997. Published By MIT Press.

[17] M. van den Berg and P. Gilkey. Heat content asymptotics of a riemannian manifold with boundary. *J. Funct. Anal.*, 120:48–71, 1994.