Conclusions: We have proposed a reduced-complexity TTCM-GA-MUD scheme that is capable of maintaining a performance similar to that of the optimum-MUD. For $K = 10$ users and a BER of 10^{-4}, the TTCM-GA-MUD aided QPSK and 16QAM systems achieved an SNR reduction of 3.8 and 2.9 dBs at the effective throughputs of 1 and 3 bit/symbol in comparison to the single user bounds of uncoded BPSK and uncoded 8-level PSK (8PSK), respectively, without extending the required bandwidth. The complexity reduction factors of $F = 1.31 \times 10^3$ and $F = 8.59 \times 10^6$ were achieved for $M = 4$ and $M = 16$, respectively, when supporting $K = 10$ users.

19 April 2002
Electronics Letters Online No: 20020489
DOI: 10.1049/el:20020489

S.X. Ng, K. Yen and L. Hanzo (*Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom*)

E-mail: lh@ecs.soton.ac.uk

References

1 VERDÚ, S.: 'Multiuser detection' (Cambridge University Press, New York, 1998)
2 GOLDBERG, D.E.: 'Genetic algorithms in search, optimization, and machine learning' (Addison-Wesley, Reading, MA, USA, 1989)
3 YEN, K., and HANZO, L.: 'Genetic algorithm assisted joint multiuser symbol detection and fading channel estimation for synchronous CDMA systems', *IEEE J. Sel. Areas Commun.*, 2001, **19**, pp. 985–998
4 ABEDI, S., and TAFAZOLLI, R.: 'Genetically modified multiuser detection for code division multiple access systems', *IEEE J. Sel. Areas Commun.*, 2002, **20**, pp. 463–473
5 ROBERTSON, P., and WÖRZ, T.: 'Bandwidth-efficient turbo trellis-coded modulation using punctured component codes', *IEEE J. Sel. Areas Commun.*, 1998, **16**, pp. 206–218

Learning of physical-like sound synthesis models by adaptive spline recurrent neural networks

F. Iannelli and A. Uncini

A recently introduced neural networks architecture, 'adaptive spline neural networks' with FIR/IIR synapse, is used to define a general class of physical-like sound synthesis model. To reduce computational cost, use is made of power-of-two synapses followed by a CR-spline-based flexible activation function the shape of which can be modified through its control points. The learning phase is performed by an efficient combinatorial optimisation algorithm, Tabu Search, for both power-of-two weights and CR-spline control points.

Introduction: The physical model paradigms are in general based on the subdivision of the synthesiser in a nonlinear excitation part in connection with other linear parts as delay lines and/or filters. The most famous model-based technique is the so-called digital waveguide filter [1]. One of the main problems with model-based synthesis techniques is the determination of the model parameters. Usually, several analyses of the original signal are necessary to design the filters correctly, and many simplifications are made to describe the nonlinear excitation mechanism (NLEM). The NLEM, in fact, is very important in so far as it characterises the timbre of the instrument.

In this Letter we propose a new recurrent-network-based synthesis model for single reed NLEM. Recently, a neural networks architecture, based on a flexible Catmul-Rom spline (CR-spline) activation function, the so-called 'adaptive spline neural networks' (ASNNs), has been proposed [2]. ASNNs have universal approximation property and are very suitable for many nonlinear signal processing applications. In general, the NLEM, as in the vibrating reed, is a nonlinear system with memory, so a static network cannot adequately model this system. To take this nonlinear dynamic into account, neural networks with FIR/IIR synapse and flexible activation function are used [2, 3]. Moreover, to obtain an efficient hardware/software implementation, the synaptic weights are constrained to be power-of-two terms.

The learning phase has been carried out by an efficient combinatorial optimisation algorithm, the so-called Tabu Search, recently used for a power-of-two adaptive filter [4].

To demonstrate the effectiveness of the proposed model, experiments on single-reed woodwind instruments have been carried out.

Physical-like synthesis model: A model for a woodwind instrument is shown in Fig. 1. It is composed of a delay line, a linear element (filter) and an NLEM [1]. The single-reed and mouthpiece arrangement act as a pressure controlled valve, which transmits energy into the instrument for the initialisation and maintenance of oscillations in the acoustic resonator. The reed can be modelled as a damped nonlinear oscillator so that the motion of a second-order mass-spring system is given by:

$$m_r\left[\frac{d^2x}{dt^2} + \mu\omega_r\frac{dx}{dt} + \omega_r^2(x - x_0)\right] = g(p_\Delta(t), U(t)) \qquad (1)$$

where m_r is the equivalent reed mass, μ is the damping factor, ω_r is fundamental reed frequency, $p_\Delta(t)$ is the difference between the player's oral cavity and the pressure in the reed channel $p_\Delta = (p_{oc} - p_r)$ and $U(t)$ is the steady volume flow through the reed and $g(\cdot)$ is a hard nonlinear function [5, 6]. The reed excitation mechanism (1) is a nonlinear system with memory and estimation of its parameters can be very difficult. In our approach, to define a more general nonlinear model we use neural networks with FIR-IIR synapses and activation functions implemented through an adaptive CR-spline interpolated curve. The parameters that we optimise [2] are the weights of the filter and the control point of the CR-spline activation curve.

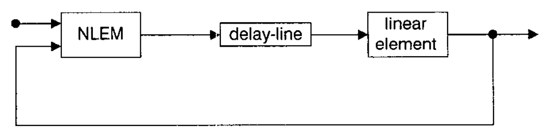

Fig. 1 *Simplified physical model*

The flexible activation function greatly reduces the number of neurons and connections to approximate the NLEM, as it gives an increased expressive power to each neuron. The synthesis network (here called 'adaptive spline recurrent network' (ASRN)) constructed on the basis of the previously described physical model of a single reed instrument is shown in Fig. 2. Once the parameters are fixed with the learning algorithm, the instrument works as a typical physical model instrument.

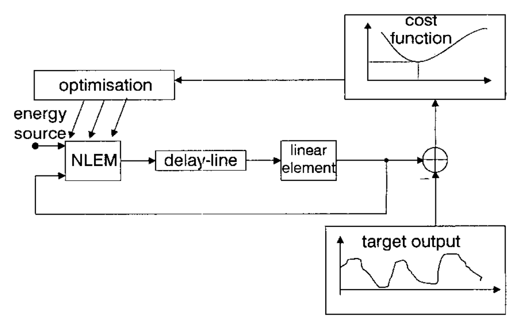

Fig. 2 *Proposed learning based pseudo-physical model*

To obtain a computationally efficient synthesiser, the ASRN makes use of IIR-FIR synapse with power-of-two (or a sum of power-of-two) coefficients. This represents a great advantage in the case of hardware realisation: multipliers can be built using a few simple and fast shift registers instead of slower floating-point arithmetic, and such a strategy can reduce both VLSI silicon area and computational time. Moreover, as specified in [2], the activation function can be easily efficiently implemented both in hardware and software or after learning, simply realised through a lookup table.

ASRN learning by Tabu Search: Several batch or on-line algorithms to train locally recurrent neural networks have been developed [3].

However, when the net topology is sophisticated, as in some synthesis schemes, it is not easy to calculate the gradient. Moreover, if other constraints are added, as in the proposed model where synaptic weights are defined over the power-of-two domain, the correct gradient computation is impossible.

Let $\mathbf{w}$ be the N dimensional vector containing all the synaptic weights defined in the discrete domain ($\mathbf{w} \in D^N$), such that

$$D = \left\{ \begin{array}{l} \alpha\!:\ \alpha = \sum_{i=1}^{\lambda} c_i 2^{-g_i}, \\ c_i \in \{-1, 0, 1\},\ g_i \in \{0, 1, \ldots, B\} \end{array} \right\} \tag{2}$$

Let M be the number of control points collected in a vector $\mathbf{q}$ defined, instead, in a continuous real domain ($\mathbf{q} \in R^M$). Let $\mathbf{x} = \mathbf{w} \cup \mathbf{q}$, the vector containing all the free parameters of the network, learning consists of a search of an optimum $\mathbf{x}^*$ over the space defined by the Cartesian product ($S = R^M \times D^N$). Thus, learning represents a difficult mixed continuous-discrete optimisation problem.

In this Letter we make use of the Tabu Search (TS) algorithm recently used for power-of-two adaptive filters [4]. Unlike other combinatorial optimisation algorithms, such as genetic algorithms or simulated annealing, where each step is performed independently from the previous moves, TS keeps trace of the visited region in the search space in order to avoid entrapment in local minima and to prevent cycling. The search mechanism can be viewed as a dynamic system that makes it possible to reduce the amount of inspected points and, consequently, the computational burden. Moreover, depending on the setting of parameter, TS does not present a difficult problem.

Basic TS algorithm: Let us define our notation to describe a basic version of the TS algorithm:

- X the discrete domain of the optimisation problem; it is also named 'feasible region';
- $x \in X$ the variables array;
- $f(\cdot)$ the objective function: its value, evaluated on a point x, is indicated as $z = f(x)$;
- $x^* \in X$ abscissa of minimum value $z^* = f(x^*)$;
- $\mu(\cdot)$ a function, usually called *move*, which generates a new point during the search phase;
- $CL(x)$ the candidate list represents a set of different moves applicable to the point x;
- T is the Tabu List that collects the forbidden moves.

The following pseudocode illustrates a basic version of TS ($\leftarrow$ denotes an assignment operation).

Step 1. Initialisation phase:

$$x \leftarrow x_0 \in X, x^* \leftarrow x, z^* \leftarrow f(x^*), T \leftarrow 0$$
$$k \leftarrow 0(counter)$$

Step 2. Compute candidate list, $CL(x) = \{\mu / \mu(x \in X), \forall \mu\}$
Step 3. If $CL(x) - T = 0$ go to step 6.
Step 4. Choose the best move in $CL(x) - T$
Step 5. If $f(x) < z^*$ then $x^* \leftarrow x$, $z^* \leftarrow f(x^*)$
Step 6. Update T, increment the iteration counter k
Step 7. If $k < k_{\max}$ go to Step 2.

END: the solution is x^* and $z^* = f(x^*)$.

Experimental results: We have tested this model with clarinet and saxophone sounds. The learning phase consisted of 1000 TS cycles for clarinet, having taken into account a cost function over a window of 1024 samples. This cost function depends on the difference between the real instrument sound and the generated one. We have tested the model with several FIR-IIR delay line lengths. For clarinet we have obtained good results even without FIR synapses, just with an adaptive CR-spline neuron, while for the saxophone model it was better to use IIR synapses but real time was compromised. Using the model in Fig. 3, we obtained good results for clarinet, and acceptable results for saxophone-like sound: we are now testing a new solution to improve the saxophone, such as using adaptive all-pass filters as termination of the delay-line. Fig. 4 shows the nonlinear activation function obtained by learning from clarinet sound.

Our important goal is not that we reproduce the identical target sound, which is impossible without complicating the model, but that our instrument can learn the parameters able to reproduce different types of sound, so that the same model can be used to reproduce a class of different musical instruments.

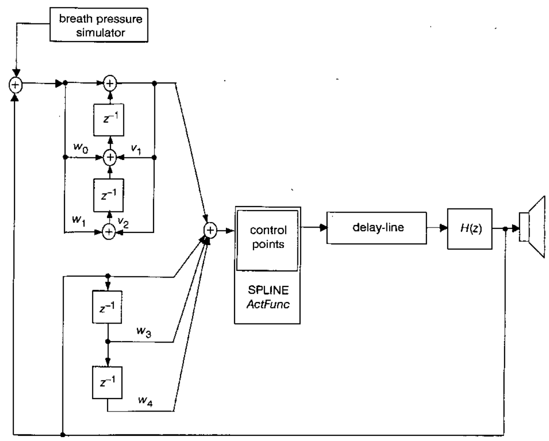

Fig. 3 *Example of mass-reed model*

Once all the parameters are fixed for a single instrument, by minimising the cost function (learning stage) the instrument can be played (forward stage) as a normal physical-model instrument.

Conclusions: We have proposed the use of ASRN as a physical-like model for sound synthesis of reed instruments. The ASRN represents a general computational model and we believe it possible to make it able to reproduce a very large class of instruments such as plucked, bowed and woodwind instruments. Our base model has been made as simple as possible in order to be played in real time (we tested the model with an AMD K7 Athlon 600 MHz).

The use of IIR-FIR synapse with power-of-two (or a sum of power-of-two) coefficients represents a great advantage in the case of hardware realisation. Multipliers can be built using simple and fast shift registers instead of slower floating-point arithmetic, and such a strategy can reduce both VLSI silicon area and computational time.

Real time implementations of the models discussed in this Letter were carried out using a platform independent, floating-point, C++ environment created by P.R. Cook and G. Scavone and called 'Synthesis ToolKit' (STK) [5].

3 October 2001
Electronics Letters Online No: 20020486
DOI: 10.1049/el:20020486

F. Iannelli and A. Uncini (*Dipartimento INFOCOM, Università di Roma 'La Sapienza', Via Eudossiana, 18, 00184 Roma, Italy*)

References

1 SMITH, J.O.: 'Physical modeling using digital wave-guides', *Comput. Music J.*, 1992, **16**, (4), pp. 74–91
2 GUARNIERI, S., PIAZZA, F., and UNCINI, A.: 'Multilayer feedforward networks with adaptive spline activation function', *IEEE Trans. Neural Netw.*, 1999, **10**, (3), pp. 672–683
3 CAMPOLUCCI, P., UNCINI, A., PIAZZA, F., and RAO, B.D.: 'On-line learning algorithms for locally recurrent neural networks', *IEEE Trans. Neural Netw.*, 1999, **10**, (2), pp. 253–271
4 TRAFERRO, S., and UNCINI, A.: 'Power-of-two adaptive filters using Tabu Search', *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, 2000, **47**, (6)
5 SCAVONE, G.P., and COOK, P.: 'Real time computer modeling of woodwind instruments'. Int. Symp. on Musical Acoustics, Leavenworth, WA, USA, 1998
6 FLETCHER, N.H., and ROSSING, T.D.: 'The physics of musical instruments' (Springer-Verlag, New York, 1981)