MODULAR INTERACTIONS AND HYBRID MODELS: A CONCEPTUAL MAP FOR MODEL-BASED SOUND SYNTHESIS

Cumhur Erkut

Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology
P.O. Box 3000, FI-02015 TKK, Espoo, Finland
phone: + 358 9 451 2496, fax: + 358 9 460 224, email: Cumhur.Erkut@tkk.fi
web: http://www.acoustics.hut.fi

ABSTRACT

This paper presents a conceptual map for model-based sound synthesis (MBSS). This map emphasizes the role of modular interaction between sub-models and it provides a natural framework for hybrid models. The MBSS techniques are grouped according to their variables, and their possible interactions are investigated.

1. INTRODUCTION

Model-based sound synthesis (MBSS; more widely known as physics-based sound synthesis or physical modeling) focuses on developing efficient digital audio processing algorithms built upon the essential physical behavior of various sound production mechanisms. The model-based representation of audio can be used in many digital audio applications, including digital sound synthesis, structural analysis of sounds, automatic transcription of musical signals, and parametric audio coding. MBSS is currently one of the most active research areas in audio signal processing [1, 2, 3]. Many refinements to existing DSP algorithms, as well as several novel techniques are emerging, albeit somewhat scattered. They may be easier to follow within a conceptual framework.

A classification for physics-based methods is proposed in [2]. Although very useful for understanding and implementing each technique, focusing on a general framework for MBSS has additional requirements. In particular, modular interactions deserve special attention, since they are favored by the *object-oriented* (OO) approach. Moreover, explicit modular interactions support important OO concepts such as *aggregation* and *encapsulation*.

Some current research trends were also enlisted in [2]. The list of hot topics include combining different physical modeling approaches, among others. Using hybrid approaches in sound synthesis to maximize strengths and minimize weaknesses of each technique, has been previously addressed in [4]. It has been pointed out that hybridization typically shows up after a technique has been around for some time and its characteristics have been extensively explored. A corollary of this remark is that six commonly used techniques of MBSS in [2] result in 21 possible hybrid combinations[1]. These combinations soon become unmanageable without a conceptual model.

This paper aims at placing the current research in MBSS within a conceptual map and emphasizes the role of modular interaction between sub-models, and thus provides a natural framework for hybrid models. It is an overview and synthesis of various related lines of effort, aimed toward unification of several paradigms in a modular modeling framework.

[1]An upper-triangular 6×6 matrix. Combining discrete models of the same technique is a non-trivial task due to computability problems and deserves in most cases a special attention.

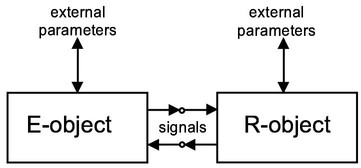

Figure 1: Excitation plus resonator paradigm of MBSS.

2. BACKGROUND

A computational object is an entity that has a unique *identity*, a set of operations that can be applied to it (*behavior*), and a *state* that stores the effect of the operations. A system is informally defined as a collection of objects united by some form of interaction or interdependence. A physical model is a discrete spatio-temporal structure that is based on the internal physical states and behavior of the original system. According to the *domain* of the system, the model variables correspond to different physical quantities. Common physical domains are mechanical, acoustical, or electrical domains.

Quantitative description of a physical system is obtained through measurable quantities that typically come in pairs of variables. For example in the acoustical domain, the deviation from the steady-state pressure $p(\mathbf{x},t)$ and the volume velocity $u(\mathbf{x},t)$ is such a pair, where $\mathbf{x}$ is the spatial vector variable, and t is the time. Force and velocity in the mechanical domain and voltage and current in the electrical domain form similar pairs. An MBSS structure that uses such a pair of variables in computation is referred to as a *Kirchhoff model*, or a K-model for short. The ratio of a variable pair yields the impedance Z in LTI systems. The admittance is the inverse of Z, i.e., $Y = 1/Z$.

A generic MBSS system can be decomposed to *exciter* and *resonator* [5] (abbreviated in Fig. 1 as E-object and R-object, respectively). In this generic scheme, exciter is usually nonlinear, whereas resonator is usually linear, and can be decomposed into sub-models. The interaction between the objects is usually handled implicitly within the system.

Discretization of K-variables leads to difficulties for modular construction of MBSS systems by interconnection of elements and consistent ordering of operations. Whereas in physical systems an infinitesimal delay between spatial points is a reasonable assumption and the causal order of events is well-defined, in an MBSS structure a unit delay is the shortest possible interval for an explicit two-way interaction between sub-structures, such as the one between exciter and resonators in Fig. 1. This leads easily to *delay-free loops* [6], i.e., implicit equations with instantaneous input/output dependency.

The delay-free loops are easier to avoid if computations are formulated by wave (shortly W) variables instead of K-variables. A pair of W-variables are obtained by splitting the corresponding K-variable into directional wave components. The wave components of the other K-variable may be computed implicitly via the impedance. This yields in the acoustical domain

$$P_i = P_i^+ + P_i^- \text{ and } U_i^+ = Y_i P_i^+, \tag{1}$$

where i is the discrete spatial index, the plus and minus superscripts denote (so far) arbitrary but opposite directions, and capital letters denote a transform variable. For instance, P_i is the z-transform of the signal $p_i(n)$. The MBSS models are grouped below according to the type of their variables.

3. MODULAR INTERACTIONS

When interconnecting objects, energy transfer between the modules should be computed according to Kirchhoff type of *continuity laws*, ensuring that the total energy is preserved. The W-models and K-models differ in their provision of modular interactions in terms of continuity laws. In this section, a brief summary of each technique with an emphasis on the modularity and interaction is provided. A more detailed description of each technique is reported in [2].

3.1 K-models

When an exciter is represented by a signal generator, a resonator by a time-varying filter, and the bi-directional signal exchange between them is reduced to unidirectional signal flow from the exciter towards the resonator, we obtain a *source-filter model*. Since the signal flow is strictly unidirectional, this technique does not provide good means for interactions. However, the resonators may be decomposed to arbitrary number of subblocks, and outputs of several exciters may be added. Thus, to a certain degree, the modularity is provided.

As basic physical systems can be described by partial differential equations (PDEs), they can be numerically solved by *finite differences*. An early example of using finite differences in MBSS can be found in [7]. Currently, Chaigne systematically extends this line of research [8]. Although finite-difference models provide means for bidirectional interaction, they are not modular. Specifically, when coupled ER systems are discretized, the resulting equations are in general implicit, hence result in delay-free loops.

In *modal synthesis*, the linear resonators are described in terms of their modal characteristics (frequency, damping factor, and mode shape for each mode), whereas connections (representing all non-linear aspects) describe the mode of interaction between objects (e.g. strike, pluck, or bow) [9]. The descriptors of resonators are typically obtained by experimental modal analysis, and implemented as a parallel resonator bank. Modal analysis is modular and supports the bi-directional interaction (usually by iteration).

A new technique related to the modal synthesis is the *Functional Transformation Method* (FTM) [10]. In FTM, the modal description of a resonator is obtained directly from the governing PDEs by applying two consecutive integral transforms (Laplace and Sturm-Liouville) to remove the temporal and spatial partial derivatives, respectively[2].

Mass-spring networks decompose the original physical system in its structural atoms [11]. These structural atoms are masses, springs, and dash-pots in the mechanical domain, although domain analogies may also be used. The interactions between the atoms are managed via explicit interconnection elements. By imposing a constraint on the causality of action and reaction, and using finite-difference formalism, modularity is also achieved [11]. Thus, it is possible to construct complex mass-spring networks by this technique.

3.2 W-models

The *wave digital filter* (WDF) theory is originally formulated for conversion of analog filters into digital filters [6]. In MBSS, a physical system is first converted to an equivalent electrical circuit using the domain analogies, then each circuit element is discretized (usually by the bilinear transform). Each object is assigned a port impedance and the energy transfer between objects is carried out by explicit interconnection objects (*adaptors*), which implement Kirchhoff laws and eliminate the delay-free loops. WDF models are mostly used as exciters, but are also applicable to resonators.

The *digital waveguide* (DWG) theory is specifically formulated for MBSS [12]. A DWG is a bi-directional delay line pair with an assigned port admittance Y and it accommodates the W-variables of any physical domain. The change in Y across a junction of the waveguide sections causes *scattering*, and the scattering junctions of interconnected ports have to be formulated. For instance, in a parallel junction of N waveguides in the acoustical domain, the Kirchhoff constraints are

$$P_1 = P_2 = \cdots = P_N = P_J \tag{2}$$
$$U_1 + U_2 + \cdots + U_N + U_{\text{ext}} = 0, \tag{3}$$

where P_i and U_i are the total pressure and volume velocity of the ith branch, respectively, P_J is the common pressure of coupled branches and U_{ext} is an external volume velocity to the junction. When port pressures are represented by incoming wave components P_i^+, outgoing wave components by P_i^-, admittances attached to each port by Y_i, and

$$P_i = P_i^+ + P_i^- \text{ and } U_i^+ = Y_i P_i^+, \tag{4}$$

the junction pressure P_J can be obtained as:

$$P_J = \frac{1}{Y_{\text{tot}}} \left(U_{\text{ext}} + 2 \sum_{i=1}^{N} Y_i P_i^+ \right), \tag{5}$$

where $Y_{\text{tot}} = \sum_{i=1}^{N} Y_i$ is the sum of all admittances to the junction. Outgoing pressure waves are obtained from Eq. (4) to yield

$$P_i^- = P_J - P_i^+. \tag{6}$$

[2] Unlike Laplace transform, Sturm-Liouville transform utilizes a non-unique kernel that depends on the boundary conditions.

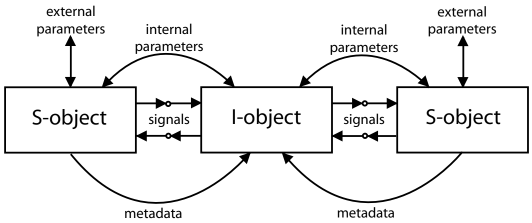

Figure 2: Modular interaction diagram.

3.3 A general modular interaction scheme

With the exception of source-filter and finite difference models, all other MBSS techniques use explicit interaction elements and provide some means for modularity. Therefore, it is advantageous to consider a generic modular interaction scheme to describe most of the MBSS techniques regardless of their domain or variables. A modular system with explicit local interactions is illustrated in Fig. 2. This scheme was first proposed in [5], but only recently it is being used for implementing MBSS systems.

In Fig. 2, an *S-object* represents a synthesis module that can correspond to both the exciter and the resonator of Fig. 1. An *I-object* is an explicit interconnection object (connector), such as a DWG junction or a WDF adaptor. Each synthesis module has *internal* and *external* parameters, with a reference of their accessibility from the connector. Internal parameters of a synthesis module (such as port admittances) are used by a connector for distributing the outgoing signals; they are only meaningful if the objects are linked. The external parameters are specific attributes of a synthesis module. Finally, metadata contains descriptors such as the domain or the type of the synthesis module. Note that locality implies that only neighboring synthesis modules are connected to a connector.

The modular interaction diagram in Fig. 2 has some interesting computational properties. First of all, it requires a complete authoring system, as the existing audio processing software systems fall short for managing the scheduling of bidirectional interactions. Moreover, most of the computational power in run-time is consumed by connectors. Therefore, for efficient implementation of detailed models, the synthesis modules must be very simple. This promotes the idea of atomic synthesis modules, and therefore the modularity.

Once the importance and the corollaries of the modular interactions are understood, the next step is to formulate the two remaining K-models that do not use explicit interaction elements (source-filter and finite-difference models) in a modular manner. As stated above, the source-filter model has a unidirectional signal flow, and therefore it is not difficult to connect it to a connector. On the other hand, finite-difference models deserve special attention.

A modular structure based on finite-differences in the time-domain (FDTD) has been reported in [13], where the essential FDTD operations are reformulated in accordance with Fig. 2. The resulting structure is illustrated in Fig. 3, where synthesis module s are abstracted as signal pipes (K-pipes in the figure) with internal parameters (port admittances), and all the computations are done in connectors (K-nodes). The essential difference between this structure and other modular K-models is in the connector computations. In this model, the distribution of K-variables is instantaneous, whereas in others they are either obtained by iteration, or contain an arbitrary unit delay.

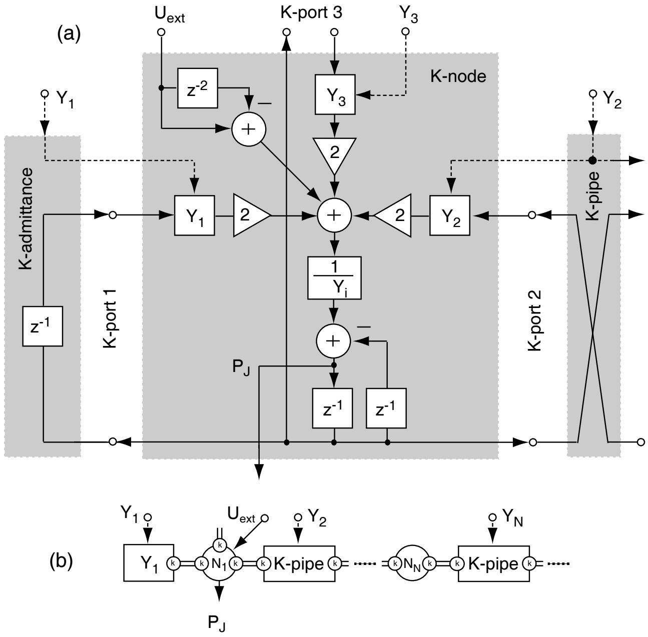

Figure 3: (a) Digital filter structure for finite difference approximation of a three-port scattering node with port admittances Y_i. Only total velocity P_J (K-variable) is explicitly available. (b) Abstract representation of the K-node in (a). After [13].

4. HYBRID MODELS

Once the MBSS structures are grouped as K-models and W-models, and their modularity is emphasized, only three cases remain to be considered for hybrid modeling. This section discusses them separately.

4.1 K-hybrids

Although the modular K-models could be theoretically interconnected, these connections would be suboptimal. For instance, the connectors of the mass-spring networks in [11] contain an arbitrary unit delay, and compute two additional intermediate mechanical K-variables (displacement and acceleration) besides the necessary K-pair (force and velocity). Therefore a good practice is to reformulate the K-models for instantaneous interaction using the guidelines set in [13] (see also Sec. 4.3).

4.2 W-hybrids

The W-models are compatible and many W-hybrids exist. See [3, 12] for examples.

4.3 KW-hybrids

One way of constructing KW-hybrids is to formulate a particular modular K-model with explicit instantaneous interaction elements, and then use a special KW-converter. The advantage of this approach is that the full dynamics of the K-model is preserved and its scheduling is made similar to that of the W-model. The disadvantage of this approach is that it is not

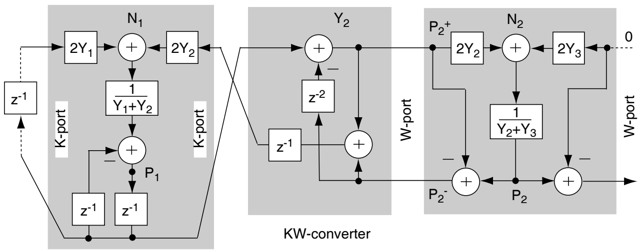

Figure 4: FDTD node (left) and a DWG node (right) forming a part of a hybrid waveguide. There is a KW-converter between K- and W-models. Y_i are wave admittances of W-lines and K-pipes. P_1 and P_2 are the junction pressures of the K-node and W-node, respectively.

general, as each K-model should be formulated separately for instantaneous modular interactions. Such a formulation is carried out in [13] for finite-difference structures. Here, the operation of a KW-hybrid model (shown in Fig. 4) in the acoustical domain is outlined.

In Fig. 4, a K-node N_1 (left) and a W-node N_2 (right), aligned with the spatial grids $i = 1$ and 2, respectively. A K-node was previously shown in Fig. 3, whereas a W-node corresponds to an connector that calculates the junction pressure in Eq. (5) and distributes the outgoing waves according to Eq. (6). Note that the junction pressures are available in both types of nodes, but in the DWG case not at the W-ports. However, the similarity of operations may be used to obtain the following transfer matrix of the 2-port KW-converter element

$$\begin{bmatrix} P_2^+ \\ z^{-1}P_2 \end{bmatrix} = \begin{bmatrix} 1 & -z^{-2} \\ 1 & (1-z^{-2}) \end{bmatrix} \begin{bmatrix} z^{-1}P_1 \\ P_2^- \end{bmatrix} \tag{7}$$

The KW-converter in Fig. 4 essentially performs the calculations given in Eq. (7) and interconnects the K-type port of an FDTD node and the W-type port of a DWG node.

Another way of constructing KW-hybrids is to formulate the K-models within the state-space formalism (as a black-box with added ports), and choose the port resistance to break instantaneous input-output path to avoid delay-free loops. The advantage of this approach is its generality, as any LTI K-model can be formulated as a state-space structure [12]. The disadvantage of this approach is that the dynamics of the K-model is hidden and its scheduling has to be separately authorized. A KW-hybrid modeling formulation based on the state-space formalism is presented in [14].

5. CONCLUSIONS AND FUTURE PLANS

In this paper, a conceptual map for MBSS is proposed. Implementing a general MBSS framework based on this map remains a challenging future task. Reformulation of other modular K-models for instantaneous interactions is an important requirement in this respect.

We may also expect that the atomic synthesis modules would be further simplified and connectors would be made more sophisticated in future generation of MBSS systems. This requires a modification in the local modular interaction scheme in Fig. 2, so that it becomes a special case of globally interacting synthesis modules. This may be a key element for authoring nonlinear, time-varying, or *multi-agent* systems that model complex group behavior of a large population of simple objects, such as an applause.

6. ACKNOWLEDGMENTS

This research has been funded by the Academy of Finland within the CAPSAS (104934) and MAPS (105651) projects. The author wishes to thank Matti Karjalainen, Vesa Välimäki, and Jyri Pakarinen for fruitful discussions.

REFERENCES

[1] V. Välimäki, A. Sarti, M. Karjalainen, R. Rabenstein, and L. Savioja, “Editorial,” *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 923–925, July 2004. Special issue on model-based sound synthesis.

[2] V. Välimäki, “Physics-based modeling of musical instruments,” *Acta Acustica united with Acustica*, vol. 90, pp. 611–617, July/Aug. 2004.

[3] J. O. Smith, “Virtual acoustic musical instruments: Review and update,” *J. New Music Research*, vol. 33, 2005. To appear.

[4] D. A. Jaffe, “Ten criteria for evaluating synthesis techniques,” *Computer Music J.*, vol. 19, no. 1, pp. 76–87, 1995.

[5] G. Borin, G. D. Poli, and A. Sarti, “Algorithms and structures for synthesis usign physical models,” *Computer Music J.*, vol. 16, no. 4, pp. 30–42, 1992.

[6] A. Fettweis, “Wave digital filters: Theory and practice,” *Proc. IEEE*, vol. 74, pp. 270–327, February 1986.

[7] L. Hiller and P. Ruiz, “Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 1,” *J. Audio Eng. Soc.*, vol. 19, pp. 462–470, June 1971.

[8] A. Chaigne, “Numerical simulations of stringed instruments – today’s situation and trends for the future,” *Catgut Acoustical Society Journal*, vol. 4, pp. 12–20, May 2002.

[9] J.-M. Adrien, “The missing link: modal synthesis,” in *Representations of Musical Signals* (G. D. Poli, A. Piccialli, and C. Roads, eds.), pp. 269–297, Cambridge, Massachusetts, USA: The MIT Press, 1991.

[10] L. Trautmann and R. Rabenstein, *Digital Sound Synthesis by Physical Modeling Using the Functional Transformation Method*. New York, USA: Kluwer Academic/Plenum Publishers, 2003.

[11] C. Cadoz, A. Luciani, and J.-L. Florens, “CORDIS-ANIMA. a modeling and simulation system for sound and image synthesis. the general formalism,” *Computer Music J.*, vol. 17, pp. 19–29, Spring 1993.

[12] J. O. Smith III, *Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects, Draft*. http://ccrma.stanford.edu/˜jos/pasp/, August 2004.

[13] M. Karjalainen and C. Erkut, “Digital waveguides versus finite difference structures: Equivalence and mixed modeling,” *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 978–989, July 2004.

[14] S. Petrausch and R. Rabenstein, “Interconnection of state space structures and wave digital filters,” *IEEE Transactions on Circuits and Systems, Part II (Express Briefs)*, vol. 52, pp. 90–93, Feb. 2005.