

Computer Music – Âm Nhạc Máy Tính

Nguyễn Quân Bá Hồng*

Ngày 15 tháng 3 năm 2025

Tóm tắt nội dung

This text is a part of the series *Some Topics in Advanced STEM & Beyond*:

URL: https://nqbh.github.io/advanced_STEM/.

Latest version:

- *Computer Music – Âm Nhạc Máy Tính*.

PDF: URL: [.pdf](#).

T_EX: URL: [.tex](#).

- .

PDF: URL: [.pdf](#).

T_EX: URL: [.tex](#).

Mục lục

| | | |
|----------|--|-----------|
| 1 | Basic Computer Music | 1 |
| 1.1 | RENATO FABBRI, VILSON VIEIRA DE SILVA JUNIOR, ANTÔNIO CARLOS SILVANO PESSOTTI, DÉBORA CRISTINA CORRÊA, OSVALDO N. OLIVEIRA JR. Musical Elements in Discrete-Time Representation of Sound | 1 |
| 1.2 | [Väl+06]. VESA VÄLIMÄKI, JYRI PAKARINEN, CUMHUR ERKUT, MATTI KARJALAINEN. Discrete-Time Modeling of Musical Instruments | 5 |
| 2 | Wikipedia | 9 |
| 2.1 | Wikipedia/computer music | 9 |
| 2.1.1 | History | 9 |
| 2.1.2 | Advances | 9 |
| 2.1.3 | Research | 9 |
| 2.1.4 | Machine improvisation | 9 |
| 2.1.5 | Live coding | 9 |
| 2.2 | Wikipedia/transcription (music) | 9 |
| 2.2.1 | Adaptation | 9 |
| 2.2.2 | Transcription aids | 10 |
| 2.2.3 | Automatic music transcription (AMT) | 10 |
| 3 | Miscellaneous | 11 |
| | Tài liệu | 11 |

1 Basic Computer Music

- 1.1 RENATO FABBRI, VILSON VIEIRA DE SILVA JUNIOR, ANTÔNIO CARLOS SILVANO PESSOTTI, DÉBORA CRISTINA CORRÊA, OSVALDO N. OLIVEIRA JR. **Musical Elements in Discrete-Time Representation of Sound**

[2 citations]

- **Abstract.** Representation of basic elements of music in terms of discrete audio signals is often used in software for musical creation & design. Nevertheless, there is no unified approach that relates these elements to discrete samples of digitized sound. In this article, each musical element is related by equations & algorithms to discrete-time samples of sounds, & each of these relations are implemented in scripts within a software toolbox, referred to as MASS (Music & Audio in Sample Sequences). Fundamental element, musical note with duration, volume, pitch, & timbre, is related quantitatively to characteristics of digital signal. Internal variations of a note, e.g. tremolos, vibratos, & spectral fluctuations, are also considered, which enables

*A Scientist & Creative Artist Wannabe. E-mail: nguyenquanbahong@gmail.com. Bến Tre City, Việt Nam.

synthesis of notes inspired by real instruments & new sonorities. With this representation of notes, resources are provided for generation of higher scale musical structures, e.g. rhythmic meter, pitch intervals & cycles. This framework enables precise & trustful scientific experiments, data sonification & is useful for education & art. Efficacy of MASS is confirmed by synthesis of small musical pieces using basic notes, elaborated notes & notes in music, which reflects organization of toolbox & thus of this article. Possible to synthesize whole albums through collage of scripts & settings specified by user. With open source paradigm, toolbox can promptly scrutinized, expanded in co-authorship processes & used with freedom by musicians, engineers, & other interested parties. In fact, MASS has already been employed for diverse purposes which include music production, artistic presentations, psychoacoustic experiments & computer language diffusion where appeal of audiovisual artifacts is exploited for education.

- CCS Concepts: Applied computing → Sound & music computing. Computing methodologies → Modeling methodologies. General & reference → Surveys & overviews, Reference works.
- Additional Key Words & Phrases: music, acoustics, psychophysics, digital audio, signal processing.
- 1. Introduction. Music is usually defined as art whose medium is sound. Definition might also state: medium includes silences & temporal organization of structures, or music is also a cultural activity or product. In physics & in this document, sounds are longitudinal waves of mechanical pressure. Human auditory system perceives sounds in frequency bandwidth between 20Hz & 20kHz, with actual boundaries depending on person, climate conditions & sonic characteristics themselves. Since speed of sound ≈ 343.2 m/s, such frequency limits corresponds to wavelengths of $\frac{343.2}{20} \approx 17.6$ m & $\frac{343.2}{20000} \approx 17.16$ mm. Hearing involves stimuli in bones, stomach, ears, transfer functions of head & torso (thân mình), & processing by nervous system. Ear is a dedicated organ or appreciation of these waves, which decomposes them into their sinusoidal spectra & delivers to nervous system. Sinusoidal components are crucial to musical phenomena, as one can recognize in constitution of sounds of musical interest (e.g. harmonic sounds & noises, discussed in Sects. 2–3), & higher level musical structures (e.g. tunings, scales, & chords, Sect. 4) [55]

Representation of sound can take many forms, from musical scores & texts in a phonetic language to electric analog signals & binary data. It includes sets of features e.g. wavelet or sinusoidal components. Although terms ‘audio’ & ‘sound’ are often used without distinction & ‘audio’ has many definitions which depend on context & author, audio most often means a representation of amplitude through time. In this sense, audio expresses sonic waves yield by synthesis or input by microphones, although these sources are not always neatly distinguishable e.g. as captured sounds are processed to generate new sonorities (âm thanh). Digital audio protocols often imply in quality loss (to achieve smaller files, ease storage & transfer) & are called *lossy* [47]. This is case e.g. of MP3 & Ogg Vorbis. Non-lossy representations of digital audio, called *lossless* protocols or formats, on other hand, assures perfect reconstruction of analog wave within any convenient precision. Standard paradigm of lossless audio consists of representing sound with samples equally spaced by a duration δ_s , & specifying amplitude of each sample by a fixed number of bits. This is linear Pulse Code Modulation (LPCM) representation of sound, herein referred to as PCM. A PCM audio format has 2 essential attributes: a sampling frequency $f_s = \frac{1}{\delta_s}$ (also called e.g. sampling rate or sample rate), which is number of samples used for representing a second of sound; & a bit depth, which is number of bits used for specifying amplitude of each sample. Fig. 1. Example of PCM audio: a sound wave is represented by 25 samples equally spaced in time where each sample has an amplitude specified with 4 bits. shows 25 samples of a PCM audio with a bit depth of 4, which yields $2^4 = 16$ possible values for amplitude of each sample & a total of $4 \cdot 25 = 100$ bits for representing whole sound.

Fixed sampling frequency & bit depth yield quantization error or quantization noise. This noise diminishes as bit depth increases while greater sampling frequency allows higher frequencies to be represented. Nyquist theorem asserts: sampling frequency is twice maximum frequency: represented signal can contain [49]. Thus, for general musical purposes, suitable to use a sample rate of at least twice highest frequency heard by humans, i.e., $f_s \geq 2 \cdot 20 \text{ kHz} = 40 \text{ kHz}$. This is basic reason for adoption of sampling frequencies e.g. 44.1 kHz & 48 kHz, which are standards in Compact Disks (CD) & broadcast systems (radio & television), resp.

Within this framework for representing sounds, musical notes can be characterized. Note often stands as ‘fundamental unit’ of musical structures (e.g. atoms in matter or cells in macroscopic organisms) &, in practice, it can unfold into sounds that uphold other approaches to music. This is of capital importance because science & scholastic artists widened traditional comprehension of music in 20th century to encompass discourse without explicit rhythm, melody or harmony. This is evident, e.g., in concrete, electronic, electroacoustic, & spectral musical styles. In 1990s, it became evident: popular (commercial) music had also incorporated sound amalgams & abstract discursive arcs. [There are well known incidences of such characteristics in ethnic music, e.g. in Pygmy music, but western theory assimilated them only in last century [74].] Notes are also convenient for another reason: average listener – & a considerable part of specialists – presupposes rhythmic & pitch organization (made explicit in Sect. 4) as fundamental musical properties, & these are developed in traditional musical theory in terms of notes. Thereafter, in this article describe musical notes in PCM audio through equations & then indicate mechanisms for deriving higher level musical structures. Understand: this is not unique approach to mathematically express music in digital audio, but musical theory & practice suggest: this is a proper framework for understanding & making computer music, as should become patent in reminder of this text & is verifiable by usage of MASS toolbox. Hopefully, interested reader or programmer will be able to use this framework to synthesize music beyond traditional conceptualizations when intended.

This document provides a fundamental description of musical structures in discrete-time audio. Results include mathematical relations, usually in terms of musical characteristics & PCM samples, concise musical theory considerations, & their implementation as software routines both as very raw & straightforward algorithms & in context of rendering musical pieces. Despite

general interests involved, there are only a few books & computer implementations that tackle subject directly. These mainly focus on computer implementations & way to mimic traditional instruments, with scattered mathematical formalisms for basic notations. Articles on topic appear to be lacking, to best of our knowledge, although advanced & specialized developments are often reported. A compilation of such works & their contributions is in Appendix G of [21]. Although current music software uses analytical descriptions presented here, there is no concise mathematical description of them, & far from trivial to achieve equations by analyzing available software implementations.

Accordingly, objectives of this paper:

1. Present a concise set of mathematical & algorithmic relations between basic musical elements & sequences of PCM audio samples.
2. Introduce a framework for sound & musical synthesis with control at sample level which entails potential uses in psychoacoustic experiments, data sonification & synthesis with extreme precision (recap in Sect. 5).
3. Provide a powerful theoretical framework which can be used to synthesize musical pieces & albums.
4. Provide approachability to developed framework [All analytic relations presented in this article are implemented as small scripts in public domain. They constitute MASS toolbox, available in an open source Git repository [9]. These routines are written in Python & make use of Numpy, which performs numerical routines efficiently (e.g. through LAPACK), but language & packages are by no means mandatory. Part of scripts has been ported to JavaScript (which favors their use in Web browsers e.g. Firefox & Chromium) & native Python [48, 56, 70]. These are all open technologies, published using licenses that grant permission for copying, distributing, modifying & usage in research, development, art & education. Hence, work presented here aims at being compliant with recommended practices for availability & validation & should ease co-authorship processes [43, 52].]
5. Provide a didactic (mang tính giáo huấn) presentation of content, which is highly multidisciplinary, involving signal processing, music, psychoacoustics & programming.

Reminder of this article is organized as follows: Sect. 2 characterizes basic musical note; Sect. 3 develops internal dynamics of musical notes; Sect. 4 tackles organization of musical notes into higher level musical structures [14, 41, 42, 54, 62, 72, 74, 76]. As these descriptions require knowledge on topics e.g. psychoacoustics, cultural traditions, & mathematical formalisms, text points to external complements as needed & presents methods, results, & discussions altogether. Sect. 5 is dedicated to final considerations & further work.

- 1.1. **Additional material.** 1 Supporting Information document [27] holds commented listings of all equations, figures, tables, & sects in this document & scripts in MASS toolbox. Another Supporting Information document [28] is a PDF version of code that implements equations & concepts in each sect [Toolbox contains a collection of Python scripts which

- * implements each of equations
- * render music & illustrate concepts
- * render each of figures used in this article.

Documentation of toolbox consists of this article, Supporting Information documents & scripts themselves.]. Git repository [26] holds all PDF documents & Python scripts. Rendered musical pieces are referenced when convenient & linked directly through URLs, & constitute another component of framework. They are not very traditional, which facilitates understanding of specific techniques & extrapolation of note concept. There are MASS-based software packages [23, 25] & further musical pieces that are linked in Git repository.

- 1.2. **Synonymy, polysemy & theoretical frames (disclaimer).** Given: main topic of this article (expression of musical elements in PCM audio) is multidisciplinary & involves art, reader should be aware: much of vocabulary admits different choices of terms & defs. More specifically, often case where many words can express same concept & where 1 word can carry different meanings. This is a very deep issue which might receive a dedicated manuscript. Reader might need to read rest of this document to understand this small selection of synonymy & polysemy (đa nghĩa) in literature, but important to illustrate point before more dense sects:

- * a “note” can mean a pitch or an abstract construct with pitch & duration or a sound emitted from a musical instrument or a specific note in a score or a music.
- * Sampling rate is also called *sampling frequency* or *sample rate*.
- * A harmonic in a sound is most often a sinusoidal component which is in harmonic series of fundamental frequency. Many times, however, terms harmonic & component are not distinguished. A harmonic can also be a note performed in an instrument by preventing certain overtones (components).
- * Harmony can refer to chords or to note sets related to chords or even to “harmony” in a more general sense, as a kind of balance & consistency.
- * A “tremolo” can mean different things: e.g. in a piano score, a tremolo is a fast alternation of 2 notes (pitches) while in computer music theory it is (most often) an oscillation of loudness.

Strived to avoid nomenclature clashes & use of more terms than needed. Also, there are many theoretical standpoints for understanding musical phenomena, which is an evidence: most often there is not a single way to express or characterize musical structures. Therefore, in this article, adjectives e.g. “often”, “commonly”, & “frequently” are abundant & they would

probably be even more numerous if wanted to be pedantically precise. Some of these issues are exposed when content is convenient, e.g. in 1st considerations of timbre.

– Cố gắng tránh xung đột danh pháp & sử dụng nhiều thuật ngữ hơn mức cần thiết. Ngoài ra, có nhiều quan điểm lý thuyết để hiểu các hiện tượng âm nhạc, đây là bằng chứng: thường không có một cách duy nhất để diễn đạt hoặc mô tả các cấu trúc âm nhạc. Do đó, trong bài viết này, các tính từ như “thường xuyên”, “thường xuyên”, & “thường xuyên” rất nhiều & chúng có thể còn nhiều hơn nữa nếu muốn chính xác về mặt học thuật. Một số vấn đề này được nêu ra khi nội dung thuận tiện, ví dụ như trong những cân nhắc đầu tiên về âm sắc.

- 2. Characterization of musical note in discrete-time audio. In diverse artistic & theoretical contexts, music is conceived as constituted by fundamental units referred to as notes, “atoms” that constitute music itself [44, 72, 74]. In a cognitive perspective, notes are understood as discernible elements that facilitate & enrich transmission of information through music [41, 55]. Canonically, basic characteristics of a musical note are duration, loudness, pitch, & timbre (âm sắc) [41]. All relations described in this sect are implemented in file `src/sections/eqs2.1.py`. Musical pieces *5 sonic portraits* & *reduced-fi* are also available online to corroborate & illustrate concepts.

- 2.1. Duration. Sample frequency f_s is defined as number of samples in each sec of discrete-time signal. Let $T = \{t_i\}$ be an ordered set of real samples separated by $\delta_s = \frac{1}{f_s}$ secs ($f_s = 44.1$ kHz $\Rightarrow \delta_s = \frac{1}{44100} \approx 0.023$ ms). A musical note of duration Δ secs can be expressed as a sequence T^Δ with $\Lambda = \lfloor \Delta \cdot f_s \rfloor$ samples. I.e., integer part of multiplication is considered, & an error of $\leq \delta_s$ missing secs is admitted, which is usually fine for musical purposes. Thus

$$T^\Delta = \{t_i\}_{i=0}^{\lfloor \Delta f_s \rfloor - 1} = \{t_i\}_0^{\Lambda - 1}.$$

- 2.2. Loudness. Loudness [Loudness & “volume” are often used indistinctly. In technical contexts, loudness is used for subjective perception of sound intensity while volume might be used for some measurement of loudness or to a change in intensity of signal by equipment. Accordingly, one can perceive a sound as loud or soft & change volume by turning a knob. Will use term loudness & avoid more ambiguous term volume.] is a perception of sonic intensity that depends on reverberation, spectrum, & other characteristics described in Sect. 3 [11]. One can achieve loudness variations through power of wave [11]:

$$\text{pow}(T) = \frac{\sum_{i=0}^{\Lambda-1} t_i^2}{\Lambda}.$$

Final loudness is dependent on amplification of signal by speakers. Thus, what matters: relative power of a note in relation to the others around it, or power of a musical sect in relation to the rest. Differences in loudness are result of complex psychophysical phenomena but can often be reasoned about in terms of decibels, calculated directly from amplitudes through energy or power:

$$V_{\text{dB}} = 10 \log_{10} \frac{\text{pow}(T')}{\text{pow}(T)}.$$

- 2.3. Pitch.
- 2.4. Timbre.
- 2.5. Spectra of sampled sounds.
- 2.6. Basic note.
- 2.7. Spatialization: localization & reverberation.
- 2.8. Musical usages.
- 3. Variation in Basic Note.
 - 3.1. Lookup table.
 - 3.2. Incremental variations of frequency & intensity.
 - 3.3. Application of digital filters.
 - 3.4. Noise.
 - 3.5. Tremolo & vibrato, AM & FM.
 - 3.6. Musical usages.
- 4. Organization of notes in music.
 - 4.1. Tuning, intervals, scales, & chords.
 - 4.2. Atonal & tonal harmonies, harmonic expansion & modulation.
 - 4.3. Counterpoint.
 - 4.4. Rhythm.
 - 4.5. Repetition & variation: motifs & larger units.
 - 4.6. Directional structures.

- 4.7. Cyclic structures.
- 4.8. Serialism & post-serial techniques.
- 4.9. Musical idiom?
- 4.10. Musical usages.
- 5. Conclusions & Further Developments.

1.2 [Väl+06]. VESA VÄLIMÄKI, JYRI PAKARINEN, CUMHUR ERKUT, MATTI KARJALAINEN. **Discrete-Time Modeling of Musical Instruments**

[283 citations]

Question 1 (Cf. Continuous-time modeling vs. Discrete-time modeling). *How about continuous-time modeling of musical instruments?*

Question 2 (Cf. Mathematical modeling technique vs. Physical modeling technique). *Compare Mathematical modeling technique vs. Physical modeling technique.*

- **Abstract.** This article describes physical modeling techniques that can be used for simulating musical instruments. Methods are closely related to digital signal processing. They discretize system w.r.t. time, because aim: run simulation using a computer. Physics-based modeling methods can be classified as mass-spring, modal, wave digital, finite difference, digital waveguide & source-filter models. Present basic theory & a discussion on possible extensions for each modeling technique. For some methods, a simple model example is chosen from existing literature demonstrating a typical use of method. E.g., in case of digital waveguide modeling technique a vibrating string model is discussed, & in case of wave digital filter technique, present a classical piano hammer model. Tackle some nonlinear & time-varying models & include new results on digital waveguide modeling of a nonlinear string. Discuss current trends & future directions in physical modeling of musical instruments.
- **1. Introduction.** Musical instruments have historically been among most complicated mechanical systems made by humans. They have been a topic of interest for physicists & acousticians for over a century. Modeling of musical instruments using computers is newest approach to understanding how these instruments work.

This paper presents an overview of physics-based modeling of musical instruments. Specifically, this paper focuses on sound synthesis methods derived using physical modeling approach. Several previously published tutorial & review papers discussed physical modeling synthesis techniques for musical instruments sounds [73, 129, 251, 255, 256, 274, 284, 294]. Purpose of this paper: give a unified introduction to 6 main classes of discrete-time physical modeling methods, namely mass-spring, modal, wave digital, finite difference, digital waveguide & source-filter models. This review also tackles mixed & hybrid models in which usually 2 different modeling techniques are combined.

Physical models of musical instruments have been developed for 2 main purposes: research of acoustical properties & sound synthesis. Methods discussed in this paper can be applied to both purpose, but here main focus is sound synthesis. Basic idea of physics-based sound synthesis: build a simulation model of sound production mechanism of a musical instrument & to generate sound with a computer program or signal processing hardware that implements that model. Motto of physical modeling synthesis: when a model has been designed properly, so that it behaves much like actual acoustic instrument, synthetic sound will automatically be natural in response to performance. In practice, various simplifications of model cause sound output to be similar to, but still clearly different from, original sound. Simplifications may be caused by intentional approximations that reduce computational cost or by inadequate knowledge of what is actually happening in acoustic instrument. A typical & desirable simplification: linearization of slightly nonlinear phenomena, which may avert unnecessary complexities, & hence may improve computational efficiency.

In speech technology, idea of accounting for physics of sound source, human voice production organs, is an old tradition, which has led to useful results in speech coding & synthesis. While 1st experiments on physics-based musical sound synthesis were documented several decades ago, 1st commercial products based on physical modeling synthesis were introduced in 1990s. Thus, topic is still relatively young. Research in field has been very active in recent years.

1 of motivations for developing a physically based sound synthesis: musicians, composers, & other users of electronic musical instruments have a constant hunger for better digital instruments & for new tools for organizing sonic events. A major problem in digital musical instruments has always been how to control them. For some time, researchers of physical models have hoped: these models would offer more intuitive, & in some ways better, controllability than previous sound synthesis methods. In addition to its practical applications, physical modeling of musical instruments is an interesting research topic for other reasons. It helps to solve old open questions, e.g. which specific features in a musical instrument's sound make it recognizable to human listeners or why some musical instruments sound sophisticated while others sound cheap. Yet another fascinating aspect of this field: when physical principles are converted into computational methods, possible to discover new algorithms. This way, possible to learn new signal processing methods from nature.

- **2. Brief history.** Modeling of musical instruments is fundamentally based on understanding of their sound production principles. 1st person attempting to understand how musical instruments work might have been Pythagoras, who lived in ancient Greece around 500 BC. At that time, understanding of musical acoustics was very limited & investigations focused on tuning of string

instruments. Only after late 18th century, when rigorous mathematical methods e.g. PDEs were developed, was it possible to build formal models of vibrating strings & plates.

Earliest work on physics-based discrete-time sound synthesis was probably conducted by KELLY & LOCHBAUM in context of vocal-tract modeling [145]. A famous early musical example is ‘Bicycle Built for 2’ (1961), where singing voice was produced using a discrete-time model of human vocal tract. This was result of collaboration between MATHEWS, KELLY & LOCHBAUM [43]. 1st vibrating string simulations were conducted in early 1970s by HILLER & RUIZ [113, 114], who discretized wave equation to calculate waveform of a single point of a vibrating string. Computing 1 s of sampled waveform took minutes. A few years later, CADOZ & his colleagues developed discrete-time mass-spring models & built dedicated computing hardware to run real-time simulations [38].

In late 1970s & early 1980s, MCINTYRE, WOODHOUSE, & SCHUMACHER made important contributions by introducing simplified discrete-time models of bowed strings, clarinet & flute [173, 174, 235], & KARPLUS & STRONG [144] invented a simple algorithm that produces string-instrument-like sounds with few arithmetic operations. Based on these ideas & their generalizations, SMITH & JAFFE introduced a signal-processing oriented simulation technique for vibrating strings [120, 244]. Soon thereafter, SMITH proposed term ‘digital waveguide’ & developed general theory [247, 249, 253].

1st commercial product based on physical modeling synthesis, an electronic keyboard instrument by Yamaha, was introduced in 1994 [168]; it used digital waveguide techniques. More recently, digital waveguide techniques have been also employed in MIDI synthesizers on personal computer soundcards. Currently, much of practical sound synthesis is based on software, & there are many commercial & freely available pieces of synthesis software that apply 1 or more physical modeling methods.

- 3. General concepts of physics-based modeling. In this sect, discuss a number of physical & signal processing concepts & terminology that are important in understanding modeling paradigms discussed in subsequent sects. Each paradigm is also characterized briefly in end of this sect. A reader familiar with basic concepts in context of physical modeling & sound synthesis may go directly to Sect. 4.

- 3.1. Physical domains, variables, & parameters. Physical phenomena can be categorized as belonging to different ‘physical domains’. Most important ones for sound sources e.g. musical instruments are acoustical & mechanical domains. In addition, electrical domain is needed for electroacoustic instruments & as a domain to which phenomena from other domains are often mapped. Domains may interact with one another, or they can be used as analogies (equivalent models) of each other. Electrical circuits & networks are often applied as analogies to describe phenomena of other physical domains.

Quantitative description of a physical system is obtained through measurable quantities that typically come in pairs of variables, e.g. force & velocity in mechanical domain, pressure & volume velocity in acoustical domain or voltage & current in electrical domain. Members of such dual variable pairs are categorized generically as ‘across variable’ or ‘potential variable’, e.g. voltage, force or pressure, & ‘through variable’ or ‘kinetic variable’, e.g. current, velocity or volume velocity. If there is a linear relationship between dual variables, this relation can be expressed as a parameter, e.g. impedance $Z = \frac{U}{I}$ being ratio of voltage U & current I , or by its inverse, admittance $Y = \frac{I}{U}$. An example from mechanical domain is mobility (mechanical admittance) defined as ratio of velocity & force. When using such parameters, only 1 of dual variables is needed explicitly, because the other one is achieved through constraint rule.

Modeling methods discussed in this paper use 2 types of variables for computation, ‘K-variables’ & ‘wave variables’ (also denoted as ‘W-variables’). ‘K’ comes from Kirchhoff & refers to Kirchhoff continuity rules of quantities in electric circuits & networks [185]. ‘W’ is shortform for wave, referring to wave components of physical variables. Instead of pairs of across & through as with K-variables, wave variables come in pairs of incident & reflected wave components. Details of wave modeling are discussed in Sects. 7–8, while K-modeling is discussed particularly in Sects. 4 & 10. It will become obvious: these are different formulations of same phenomenon, & possibility to combine both approaches in hybrid modeling will be discussed in Sect. 10.

Decomposition into wave components is prominent in such wave propagation phenomena where opposite-traveling waves add up to actual observable K-quantities. A wave quantity is directly observable only when there is no other counterpart. It is, however, a highly useful abstraction to apply wave components to any physical case, since this helps in solving computability (causality) problems in discrete-time modeling.

- 3.2. Modeling of physical structure & interaction. Physical phenomena are observed as structures & processes in space & time. In sound source modeling, interested in dynamic behavior that is modeled by variables, while slowly varying or constant properties are parameters. Physical interaction between entities in space always propagates with a finite velocity, which may differ by orders of magnitude in different physical domains, speed of light being upper limit.

‘Causality’ is a fundamental physical property that follows from finite velocity of interaction from a cause to corresponding effect. In many mathematical relations used in physical models causality is not directly observable. E.g., relation of voltage across & current through an impedance is only a constraint, & variables can be solved only within context of whole circuit. Requirement of causality (more precisely temporal order of cause preceding effect) introduces special computability problems in discrete-time simulation, because 2-way interaction with a delay shorter than a unit delay (sampling period) leads to ‘delay-free loop problem’. Use of wave variables is advantageous, since incident & reflected waves have a causal relationship. In particular, wave digital filter (WDF) theory, discussed in Sect. 8, carefully treats this problem through use of wave variables & specific scheduling of computation operations.

Taking finite propagation speed into account requires using a spatially distributed model. Depending on case at hand, this can be a full 3D model e.g. used for room acoustics, a 2D model e.g. for a drum membrane (discarding air loading) or a 1D

model e.g. for a vibrating string. If object to be modeled behaves homogeneously enough as a whole, e.g. due to its small size compared with wavelength of wave propagation, it can be considered a lumped entity that does not need a description of spatial dimensions.

– Việc tính đến tốc độ lan truyền hữu hạn đòi hỏi phải sử dụng một mô hình phân bố không gian. Tùy thuộc vào trường hợp cụ thể, đây có thể là mô hình 3D đầy đủ, ví dụ như được sử dụng cho âm học phòng, mô hình 2D, ví dụ như cho màng trống (loại bỏ tải trọng không khí) hoặc mô hình 1D, ví dụ như cho một cú chích rung. Nếu vật thể được mô hình hóa hoạt động đủ đồng nhất như một tổng thể, ví dụ như do kích thước nhỏ so với bước sóng truyền sóng, thì nó có thể được coi là một thực thể tập trung không cần mô tả về kích thước không gian.

- **3.3. Signals, signal processing, & discrete-time modeling.** In signal processing, signal relationships are typically represented as 1-directional cause-effect chains. Contrary to this, bi-directional interaction is common in (passive) physical systems, e.g. in systems where reciprocity principle is valid. In true physics-based modeling, 2-way interaction must be taken into account. I.e., from signal processing viewpoint, such models are full of feedback loops, which further implicates: concepts of computability (causality) & stability become crucial.

In this paper, apply digital signal processing (DSP) approach to physics-based modeling whenever possible. Motivation for this: DSP is an advanced theory & tool that emphasizes computational issues, particularly maximal efficiency. This efficiency is crucial for real-time simulation & sound synthesis. Signal flow diagrams are also a good graphical means to illustrate algorithms underlying simulations. Assume: reader is familiar with fundamentals of DSP, e.g. sampling theorem [242] to avoid aliasing (also spatial aliasing) due to sampling in time & space as well as quantization effects due to finite numerical precision.

An important class of systems is those that are linear & time invariant (LTI). They can be modeled & simulated efficiently by digital filters. They can be analyzed & processed in frequency domain through linear transforms, particularly by Z-transform & discrete Fourier transform (DFT) in discrete-time case. While DFT processing through fast Fourier transform (FFT) is a powerful tool, it introduces a block delay & does not easily fit to sample-by-sample simulation, particularly when bi-directional physical interaction is modeled.

Nonlinear & time-varying systems bring several complications to modeling. Nonlinearities create new signal frequencies that easily spread beyond Nyquist limit, thus causing aliasing, which is perceived as very disturbing distortion. In addition to aliasing, delay-free loop problem & stability problems can become worse than they are in linear systems. If nonlinearities in a system to be modeled are spatially distributed, modeling task is even more difficult than with a localized nonlinearity. Nonlinearities will be discussed in several sects of this paper, most completely in Sect. 11.

- **3.4. Energetic behavior & stability.** Product of dual variables e.g. voltage & current gives power, which, when integrated in time, yields energy. Conservation of energy in a closed system is a fundamental law of physics that should also be obeyed in true physics-based modeling. In musical instruments, resonators are typically passive, i.e. they do not produce energy, while excitation (plucking, bowing, blowing, etc.) is an active process that injects energy to passive resonators.

Stability of a physical system is closely related to its energetic behavior. Stability can be defined so that energy of system remains finite for finite energy excitations. From a signal processing viewpoint, stability may also be defined so that variables, e.g. voltages, remain within a linear operating range for possible inputs in order to avoid signal clipping & distortion.

In signal processing systems with 1-directional input–output connections between stable subblocks, an instability can appear only if there are feedback loops. In general, impossible to analyze such a system’s stability without knowing its whole feedback structure. Contrary to this, in models with physical 2-way interaction, if each element is passive, then any arbitrary network of such elements remains stable.

- **3.5. Modularity & locality of computation.** For a computational realization, desirable to decompose a model systematically into blocks & their interconnections. Such an object-based approach helps manage complex models through use of modularity principle. Abstractions to macro blocks on basis of more elementary ones helps hiding details when building excessively complex models.

For 1-directional interactions used in signal processing, enough to provide input & output terminals for connecting blocks. For physical interaction, connections need to be done through ports, with each port having a pair of K- or wave variables depending on modeling method used. This allows mathematical principles used for electrical networks [185]. Details on block-wise construction of models will be discussed in following sects for each modeling paradigm.

Locality of interaction is a desirable modeling feature, which is also related to concept of causality. For a physical system with a finite propagation speed of waves, enough: a block interacts only with its nearest neighbors; it does not need global connections to compute its task & effect automatically propagates throughout system.

In a discrete-time simulation with bi-directional interactions, delays shorter than a unit delay (including 0 delay) introduce delay-free loop problem that we face several times in this paper. While possible to realize fractional delays [154], delays shorter than unit delay contain a delay-free component. There are ways to make such ‘implicit’ system computable, but cost in time (or accuracy) may become prohibitive for real-time processing.

- **3.6. Physics-based discrete-time modeling paradigms.** This paper presents an overview of physics-based methods & techniques for modeling & synthesizing musical instruments. Have excluded some methods often used in acoustics, because they do not easily solve task of efficient discrete-time modeling & synthesis. E.g., finite element & boundary element methods (FEM & BEM) are generic & powerful for solving system behavior numerically, particularly for linear systems, but focus on inherently time-domain methods for sample-by-sample computation.

Main paradigms in discrete-time modeling of musical instruments can be briefly characterized as follows.

- * 3.6.1. Finite difference models.
- * 3.6.2. Mass-spring networks.
- * 3.6.3. Modal decomposition methods.
- * 3.6.4. Digital waveguides.
- * 3.6.5. Wave digital filters.
- * 3.6.6. Source-filter models.
- 4. Finite difference models.
 - 4.1. Finite difference models for an ideal vibrating string.
 - 4.2. Boundary conditions & string excitation.
 - 4.3. Finite difference approximation of a lossy string.
 - 4.4. Stiffness in finite difference strings.
- 5. Mass-spring networks.
 - 5.1. Basic theory.
 - 5.2. CORDIS-ANIMA.
 - 5.3. Other mass-spring systems.
- 6. Modal decomposition methods.
 - 6.1. Modal synthesis.
 - 6.2. Filter-based modal methods.
 - 6.3. Functional transform method.
- 7. Digital waveguides.
 - 7.1. From wave propagation to digital waveguides.
 - 7.2. Modeling of losses & dispersion.
 - 7.3. Modeling of waveguide termination & scattering.
 - 7.4. Digital waveguide meshes & networks.
 - 7.5. Reduction of a DWG model to a single delay loop structure.
 - 7.6. Commuted DWG synthesis.
 - 7.7. Case study: modeling & synthesis of acoustic guitar.
 - 7.8. DWG modeling of various musical instruments.
- 8. Wave digital filters.
 - 8.1. What are wave digital filters?
 - 8.2. Analog circuit theory.
 - 8.3. Wave digital building blocks.
 - 8.4. Interconnection & adaptors.
 - 8.5. Physical modeling using WDFs.
 - 8.6. Current research.
- 9. Source-filter models.
 - 9.1. Subtractive synthesis in computer music.
 - 9.2. Source-filter models in speech synthesis.
 - 9.3. Instrument body modeling by digital filters.
 - 9.4. Karplus-Strong algorithm.
 - 9.5. Virtual analog synthesis.
- 10. Hybrid models.
 - 10.1. KW-hybrids.
 - 10.2. KW-hybrids modeling examples.
- 11. Modeling of nonlinear & time-varying phenomena.
 - 11.1. Modeling of nonlinearities in musical instruments.
 - 11.2. Case study: nonlinear string model using generalized time-varying allpass filters.
 - 11.3. Modeling of time-varying phenomena.
- 12. Current trends & further research.
- 13. Conclusions.

2 Wikipedia

2.1 Wikipedia/computer music

“*Computer music* is application of **computing technology** in **music composition**, to help human composers create new music or to have computers independently create music, e.g. with **algorithmic composition** programs. it includes theory & application of new & existing computer software technologies & basic aspects of music, e.g. **sound synthesis**, **digital signal processing**, **sound design**, sonic diffusion, **acoustics**, **electrical engineering**, & **psychoacoustics**. Field of computer music can trace its roots back to origins of **electric music**, & 1st experiments & innovations with electronic instruments at turn of 20th century.

2.1.1 History

2.1.2 Advances

2.1.3 Research

2.1.4 Machine improvisation

2.1.5 Live coding

” – **Wikipedia/computer music**

2.2 Wikipedia/transcription (music)

“A J. S. BACH keyboard piece transcribed for guitar. In music, *transcription* is practice of **notating** a piece or a sound which was previously unnotated &/or unpopular as a written music, e.g., a **jazz improvisation** or a **video game soundtrack**. When a musician is tasked with creating **sheet music** from a recording & they write down notes that make up piece in **music notation**, it is said: they created a *musical transcription* of that recording. Transcription may also mean rewriting a piece of music, either solo or **ensemble**, for another instrument or other instruments than which it was originally intended. **Beethoven Symphonies** transcribed for solo piano by **Franz Liszt** are an example. Transcription in this sense is sometimes called **arrangement**, although strictly speaking transcriptions are faithful adaptations, whereas arrangements change significant aspects of original piece.

Further examples of music transcription include **ethnomusicological** notation of **oral traditions** of folk music, e.g. Béla Bartók’s & Ralph Vaughan Williams’ collections of national folk music of Hungary & England resp. French composer Olivier Messiaen transcribed **birdsong** in wild, & incorporated it into many of his compositions, e.g. his **Catalogue d’oiseaux** for solo piano. Transcription of this nature involves scale degree recognition & harmonic analysis, both of which transcriber will need **relative** or **perfect pitch** to perform.

In popular music & rock, there are 2 forms of transcription. Individual performers copy a note-for-note guitar solo or other melodic line. As well, music publishers transcribe entire recordings of guitar solos & bass lines & sell sheet music in bound books. Music publishers also publish PVG (piano/vocal/guitar) transcriptions of popular music, where melody line is transcribed, & then accompaniment on recording is arranged as a piano part. Guitar aspect of PVG label is achieved through guitar chords written above melody. Lyrics are also included below melody.

2.2.1 Adaptation

Some composers have rendered homage to other composers by creating “identical” versions of earlier composers’ pieces while adding their own creativity through use of completely new sounds arising from difference in instrumentation. Most widely known example of this is RAVEL’s arrangement for orchestra of MUSSORGSKY’s piano piece ***Pictures at an Exhibition***. WEBERN used his transcription for orchestra of 6-part **ricercar** from BACH’s ***The Musical Offering*** to analyze structure of Bach piece, by using different instruments to play different subordinate **motifs** of Bach’s themes & melodies.

In transcription of this form, new piece can simultaneously imitate original sounds while recomposing them with all technical skills of an expert composer in such a way that it seems: piece was originally written for new medium. But some transcriptions & arrangements have been done for purely pragmatic or contextual reasons. E.g., in Mozart’s time, overtures & songs from this popular operas were transcribed for small **wind ensemble** simply because such ensembles were common ways of providing popular entertainment in public places. MOZART himself did this in his own opera ***The Marriage of Figaro***. A more contemporary example is STRAVINSKY’s transcription for 4 hands piano of ***The Rite of Spring***, to be used on ballet’s rehearsals. Today musicians who play in cafes or restaurants will sometimes play transcriptions or arrangements of pieces written for a larger group of instruments.

Other examples of this type of transcription include BACH’s arrangement of VIVALDI’s 4-violin concerti for 4 keyboard instruments & orchestra; MOZART’s arrangement of some Bach **fugues** from ***The Well-Tempered Clavier*** for string **trio**; BEETHOVEN’s arrangement of his ***Große Fuge***, originally written for **string quartet**, for **piano duet**, & his arrangement of his **Violin Concerto** as a **piano concerto**; Franz Liszt’s piano arrangements of works of many composers, including **symphonies of Beethoven**; TCHAIKOVSKY’s arrangement of 4 Mozart piano pieces into an **orchestral suite** called “**Mozartiana**”; MAHLER’s re-orchestration of SCHUMANN symphonies; & SCHOENBERG’s arrangement for orchestra of BRAHMS’s piano quintet & BACH’s “St. Anne” Prelude & Fugue for organ.

Since piano became a popular instrument, a large literature has sprung up of transcriptions & arrangements for piano of works for orchestra or chamber music ensemble. These are sometimes called “**piano reductions**”, because multiplicity of orchestral parts – in an orchestral piece there may be as many as 2 dozen separate instrumental parts being played simultaneously – has

to be reduced to what a single pianist (or occasionally 2 pianists, or 1 or 2 pianos, e.g. different arrangements for GEORGE GERSHWIN's *Rhapsody in Blue*) can manage to play.

Piano reductions are frequently made of orchestral accompaniments to choral works, for purposes of rehearsal or of performance with keyboard alone.

Many orchestral pieces have been transcribed for **concert band**.

2.2.2 Transcription aids

- **Notation software.** Since advent of desktop publishing, musicians can acquire **music notation software**, which can receive user's mental analysis of notes & then store & format those notes into standard music notation for personal printing or professional publishing of sheet music. Some notation software can accept a Standard **MIDI File** (SMF) or MIDI performance as input instead of manual note entry. These notation applications can export their scores in a variety of formats like **EPS**, **PNG**, & **SVG**. Often software contains a sound library that allows user's score to be played aloud by application for verification.
- **Slow-down software.** Prior to invention of digital transcription aids, musicians would slow down a record or a tape recording to be able to hear melodic lines & chords at a slower, more digestible pace. Problem with this approach was: it also changed pitches, so once a piece was transcribed, it would then have to be transposed into correct key. Software designed to slow down tempo of music without changing pitch of music can be very helpful for recognizing pitches, melodies, chords, rhythms, & lyrics when transcribing music. However, unlike slow-down effect of a record player, pitch & original octave of notes will stay same, & not descend in pitch. This technology is simple enough that it is available in many free software applications.

Software generally goes through a 2-step process to accomplish this. 1st, audio file is played back at a lower sample rate than that of original file. This has same effect as playing a tape or vinyl record at slower speed – pitch is lowered meaning music can sound like it is in a different key. 2nd step: use **Digital Signal Processing** (or DSP) to shift pitch back up to original pitch level or musical key.

- **Pitch tracking software.** Main article: **Wikipedia/pitch tracker**. As mentioned in the Automatic music transcription sect, some commercial software can roughly track pitch of dominant melodies in polyphonic musical recordings. Note scans are not exact, & often need to be manually edited by user before saving to file in either a proprietary file format or in Standard MIDI File Format. Some pitch tracking software also allows scanned note lists to be animated during audio playback.

2.2.3 Automatic music transcription (AMT)

Term “automatic music transcription” was 1st used by audio researchers JAMES A. MOORER, MARTIN PISZCZALSKI, & BERNARD GALLER in 1977. With their knowledge of digital audio engineering, these researchers believed: a computer could be programmed to analyze a **digital recording** of music s.t. pitches of melody lines & chord patterns could be detected, along with rhythmic accents of percussion instruments. Task of AMT concerns 2 separate activities: making an analysis of a musical piece, & printing out a score from that analysis.

This was not a simple goal, but one that would encourage academic research for at least another 3 decades. Because of close scientific relationship of speech to music, much academic & commercial research that was directed toward more financially resourced **speech recognitions** technology would be recycled into research about music recognition technology. While many musicians & educators insist that manually doing transcriptions is a valuable exercise for developing musicians, motivation for AMT remains same as motivation for sheet music: musicians who do not have intuitive transcription skills will search for sheet music or a chord chart, so that they may quickly learn how to play a song. A collection of tools created by this ongoing research could be of great aid to musicians. Since much recorded music does not have available sheet music, an automatic transcription device could also offer transcriptions that are otherwise unavailable in sheet music. To date, no software application can yet completely fulfill JAMES MOORER's definition of AMT. However, pursuit of AMT has spawned creation of many software applications that can aid in manual transcription. Some can slow down music while maintaining original pitch & octave, some can track pitch of melodies, some can track chord changes, & others can track beat of music.

Automatic transcription most fundamentally involves identifying pitch & duration of performed notes. This entails tracking pitch & identifying note onsets. After capturing those physical measurements, this information is mapped into traditional music notation, i.e., sheet music.

Digital Signal Processing is branch of engineering that provides software engineers with tools & algorithms needed to analyze a digital recording in terms of pitch (note detection of melodic instruments), & energy content of un-pitched sounds (detection of percussion instruments). Musical recordings are sampled at a given recording rate & its frequency data is stored in any digital wave format in computer. Such format represents sound by **digital sampling**.

- **Pitch detection.** **Pitch detection** is often detection of individual **notes** that might make up a **melody** in music, or notes in a **chord**. When a single key is pressed upon a piano, what we hear is not just **1 frequency** of sound vibration, but a *composite* of multiple sound vibrations occurring at different mathematically related frequencies. Elements of this composite of vibrations at differing frequencies are referred to as **harmonics** or partials.

E.g., if note A_3 (220 Hz) is played, individual **frequencies** of composite's **harmonic series** will start at 220 Hz as **fundamental frequency**: 440 Hz would be 2nd harmonic, 660 Hz would be 3rd harmonic, 880 Hz would be 4th harmonic, etc. These are integer multiples of fundamental frequency (e.g., $2 \cdot 220 = 440$, 2nd harmonic). While only about 8 harmonics are really needed to audibly recreate note, total number of harmonics in this mathematical series can be large, although higher harmonic's numerical weaker magnitude & contribution of that harmonic. Contrary to intuition, a musical recording at its lowest physical

level is not a collection of individual **notes**, but is really a collection of individual harmonics. That is why very similar-sounding recordings can be created with differing collections of instruments & their assigned notes. As long as total harmonics of recording are recreated to some degree, it does not really matter which instruments or which notes were used.

- **Beat detection.**
- **How ATM works.**
- **Detailed computer steps behind AMT.**

” – [Wikipedia/transcription \(music\)](#)

3 Miscellaneous

Tài liệu

- [Väl+06] Vesa Välimäki, Jyri Pakarinen, Cumhur Erkut, and Matti Karjalainen. “Discrete-time modelling of musical instruments”. In: *Rep. Prog. Phys.* 69.1 (2006), pp. 1–78. DOI: [10.1088/0034-4885/69/1/R01](#). URL: [https://iopscience.iop.org/article/10.1088/0034-4885/69/1/R01](#).