Научная статья
УДК 51-78
DOI:10.31854/1813-324X-2022-8-2-76-81

Some New Mathematical Models of Synthesized Sound Signals

Gleb Rogozinsky[1]✉, gleb.rogozinsky@gmail.com
Michael Chesnokov[2], chesnokov.inc@gmai.com
Alexandra Kutlyiarova[3], mr.echelon17@gmail.com

[1]The Bonch-Bruevich Saint Petersburg State University of Telecommunications, St. Petersburg, 193232, Russian Federation
[2]Technological Center «YAFI», St. Petersburg, 194223, Russian Federation
[3]Saint Petersburg State Institute of Film and Television, St. Petersburg, 191119, Russian Federation

Abstract: *Modern sound synthesis systems make it possible to implement various signal generation algorithms of higher complexity. The theory of sound synthesis actively uses the mathematical apparatus of analog and digital radio engineering and signal processing, however, it should be noted that the classical signal models used in acoustics are not adequate to real-world synthesized signals, mainly due to the significant complexity of the latter. This article presents some models of synthesized signals typical for practical use.*

Keywords: *sound synthesis systems, mathematical models, additive synthesis, subtractive synthesis, frequency modulation, granular synthesis, sound objects, Csound language, computer music*

For citation: Rogozinsky G., Chesnokov M., Kutlyiarova A. Some New Mathematical Models of Synthesized Sound Signals. *Proc. of Telecom. Universities.* 2022;8(2):76–81. (in Russ.) DOI:10.31854/1813-324X-2022-8-2-76-81

Несколько новых математических моделей синтезированных звуковых сигналов

Глеб Гендрихович Рогозинский[1]✉, gleb.rogozinsky@gmail.com
Михаил Александрович Чесноков[2], chesnokov.inc@gmai.com
Александра Александровна Кутлыярова[3], mr.echelon17@gmail.com

[1]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
[2]НТЦ «ЯФИ», Санкт-Петербург, 194223, Российская Федерация
[3]Санкт-Петербургский государственный институт кино и телевидения, Санкт-Петербург, 19119, Российская Федерация

Аннотация: *Современные системы синтеза звука позволяют реализовать в значительной мере сложные алгоритмы генерации сигналов. Теория синтеза звука активно использует математический аппарат аналоговой и цифровой радиотехники и обработки сигналов, однако следует отметить, что классические модели сигналов, применяемые в акустике, не являются адекватными реальным синтезированным сигналам, главным образом, в силу значительной сложности последних. В данной статье представлены модели реальных синтезированных сигналов, характерных для практического использования.*

Ключевые слова: *системы синтеза звука, математические модели, аддитивный синтез, субтрактивный синтез, частотная модуляция, гранулярный синтез, звуковые объекты, язык Csound, компьютерная музыка*

Ссылка для цитирования: Рогозинский Г.Г., Чесноков М.А., Кутлыярова А.А. Несколько новых математических моделей синтезированных звуковых сигналов // Труды учебных заведений связи. 2022. Т. 8. № 2. С. 76–81. DOI:10.31854/1813-324X-2022-8-2-76-81

Introduction

The very first sound synthesis systems began to appear in the second half of the 20th century. To date, there is a significant amount of methodological information about the programming of synthesizers and the implementation of various classes of sound objects on their basis. The specificity of using synthesizers determines, first of all, practical methods. At the same time, the need to improve sound synthesis algorithms and the development of new approaches require a theoretical approach to modeling. The theory of sound synthesis actively uses the mathematical apparatus of analog and digital radio engineering and signal processing. However, one should note that classical signal models used in the radio engineering theory are not adequate to real synthesized signals, mainly due to the significant complexity of the latter. For example, consider frequency modulation synthesis (also known as FM synthesis). The classical radio-engineering model of a frequency-modulated signal assumes the presence of only one carrier frequency with only one modulating frequency. At the same time, a typical FM synthesis model operates with up to eight oscillators (or operators), between which one can have a number of interconnections; thus, it requires a much more complex description. Therefore, an urgent task for the development of the theory of sound synthesis is the application of a systematic approach, including to existing methods of synthesis.

The paper presents some new models of "classical" sound synthesis methods. These models possess a detailed mathematical view of synthesis algorithms.

A Brief Overview of Sound Synthesis Methods

Today there are several theoretical works devoted to sound synthesis, among which it is necessary to name the works by Manning [1], Chowning [2], Roads [3], Lazzarini *et al*. [4], Cook [5], and others. Certain questions of the wavelet theory can be attributed to the sound synthesis, for example, the work by Kudumakis and Sandler [6] and others. A rather complex mathematical apparatus can be found in the works by Ishutkin and Uvarov [7] devoted to the Hilbert-based modulation theory of sound. The most significant works on physical modeling are those by Smith III [8]. The additive and subtractive methods of sound synthesis are the basis for the classical theory; many publications on computer music describe their principles pretty well. The works by Chowning [2] and other authors are devoted to the synthesis based on frequency modulation. Roads [9] discloses numerous techniques of granular synthesis in detail in his monograph.

It should be noted that with the exception of physical modeling [8], other types of sound synthesis are poorly described in terms of mathematical models. Their modeling and implementation are often based on special programming languages, for example, Csound [4], and the established approach is the synthesizer operation algorithm in the form of a program code, which sometimes complicates the systematic approach.

The following section highlights four "classical" methods of sound synthesis and presents the corresponding models.

Additive Synthesis

In the simplest form, sound objects based on additive synthesis are the linear combination of harmonic signals of different amplitudes, frequencies, and initial phases:

$$S_{A_1}(t) = \sum_{i=1}^{N} \gamma_i \cdot \sin(\omega_i t + \phi_i), \qquad (1)$$

where γ_i – partial amplitude; ω_i – partial frequency; φ_i – the initial phase of the *i*-th partial.

In a practical case, e. g. Morphine (https://www.image-line.com/fl-studio-learning/fl-studio-online-manual/html/plugins/Morphine.htm), each partial $\sin(\omega_i t + \phi_i)$ possesses its own amplitude envelope as a function of time $\gamma_i(t)$, also it can be a common envelope $A(t)$ for the whole array of signals:

$$S_{A_2}(t) = A(t) \cdot \sum_{i=1}^{N} \gamma_i(t) \cdot \sin(\omega_i t + \phi_i). \qquad (2)$$

In a more general case, harmonic signals (1) and (2) can be substituted with other signals, localized around frequency ω_0 with $\Delta\omega$ spread in the frequency domain, e. g. narrow-banded noise $\Gamma(t) * h_i(\omega_0, \Delta\omega)$, $h_i(\omega_0, \Delta\omega)$ – impulse response of a band-pass filter, which forms the corresponding *i*-th band; $\Gamma(t)$ – arbitrary signal from L^2: $\operatorname{supp}\Gamma(\omega) > \Delta\omega, \omega_0 \in \operatorname{supp}\Gamma(\omega)$.

Thus, one can re-write (2) as follows:

$$S_{A_3}(t) = A(t) \cdot \sum_{i=1}^{N} \gamma_i(t) \cdot [\Gamma(t) * h_i(\omega_0, \Delta\omega)]. \qquad (3)$$

One should also take into account the important role of modulation (in the sense of parameter changing over time) of various parameters of the sound synthesis algorithm:

$$S_{A_4}(t) = A(t) \cdot \sum_{i=1}^{N} \gamma_i(t) \times \times \left[\Gamma(t) * h_i\left(\omega_{0,i}(t), \Delta\omega_i(t)\right)\right]. \quad (4)$$

The hardware implementation of additive synthesis is complicated by the complexity of the interface of the corresponding device or virtual plug-in (for 100 partials with individual ADSR-type envelopes, at least 400 controllers are required). The software models are easy to implement, but still difficult to manage. The solutions are either a macro parametric approach, e.g. management of groups of parameters through one control element, or a graphical method (as in the famous Russian ANS synthesizer). Figure 1 gives a waveform and a spectrogram of 60 secs of rendered audio (normalized to −1.0 dB).

The realization of (4), coded in Csound, is given below. For the sake of space, we restrict the synthesizer code to only three additive components.

```
opcode Voice, a, aik ; UDO definition for one of Additive Synth's voice
      aIn,iF0,kM xin   ; inputs – audio signal, central frequency, mod level
      kEnv[] init 2
      kEnv[0] jspline 0.5,0.05,0.3  ; random envelope for amplitude
      kEnv[1] jspline 0.5,0.02,0.5  ; random envelope for filter frequency
      kEnv += 0.5  ; DC shift for both envelopes
      kF = iF0 + kEnv[1]*kM  ; filter frequency modulation
      a1 butterbp aIn, kF ,iF0*0.1 ; two stage Butterworth band-pass filtering
      a1 butterbp a1, kF,iF0*0.1
      xout a1*kEnv[0] ; applying amplitude envelope & route to UDO out
endop

instr Additive Synth
      iF0[] fillarray 164.814, 195.998, 261.626 ; (E,G,C) pitch set in Hz
      kEnvA linenr 1, 2, 2,.01  ; overall synth envelope A(t)
      aOut[] init 3
      aNoise rand 1,2,1  ; white noise generator
      aOut[0] Voice aNoise,iF0[0],iF0[0]*.5  ; obtain three voices using UDO
      aOut[1] Voice aNoise,iF0[1],iF0[1]*.5
      aOut[2] Voice aNoise,iF0[2],iF0[2]*.5
      out (aOut[0]+aOut[1]+aOut[2])*kEnvA ; mixing and applying A(t)
endin
```

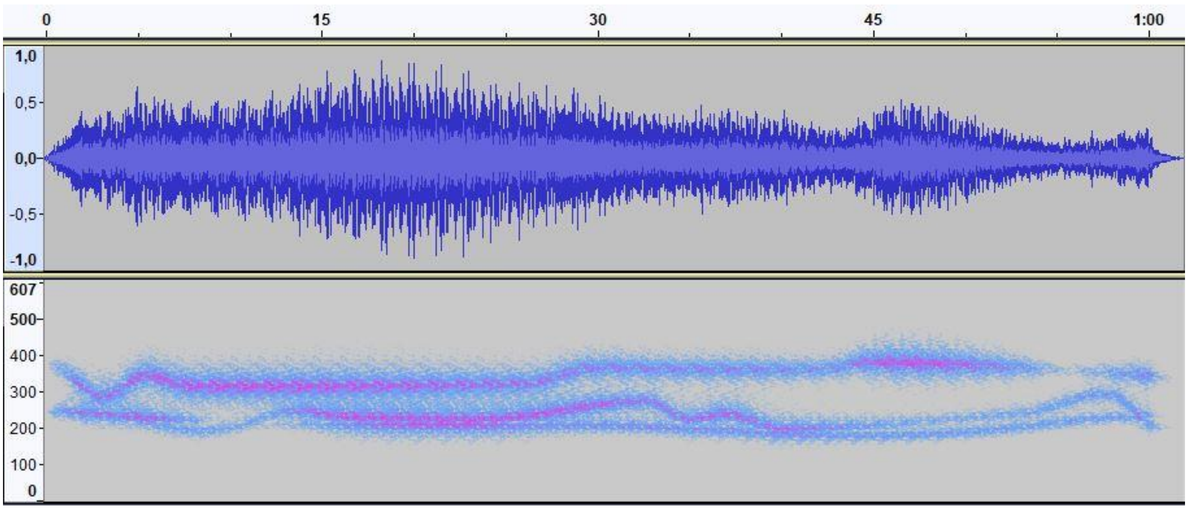

Fig. 1. The Waveform (Top) and Spectrogram (Bottom) of the Sound Object Obtained Using Csound Realization of Additive Model (4)

Subtractive Synthesis

Obtained through the subtractive synthesis, sound objects can be modeled as a convolution of initial polyharmonic signal $S_n(t)$, or noise with a given probability function, and the impulse response of the filter $h(t)$, also typically featuring the common envelope $A(t)$. In most cases, the filter is the object of modulation $M(t)$, especially its cut-off frequency ω_0:

$$S_{A_5}(t) = A(t) \cdot S_n(t) * h(M(t) \cdot \omega_0). \quad (5)$$

Among the original polyharmonic signals, the most commonly used are (6–9):

Sawtooth signal with a limited number of harmonics up to the N-th harmonic (alias-free):

$$S_{\text{saw}}(t) = \sum_{k=1}^{N} \frac{\sin(k \cdot \omega_0 t)}{k}, \quad N \cdot \omega_0 < \frac{\omega_S}{2}, \quad (6)$$

where $\frac{\omega_S}{2}$ – Nyquist frequency (the half of the sampling frequency); ω_0 – fundamental frequency.

The array of detuned saws:

$$S_{\text{nsaw}}(t) = \sum_{m=1}^{L}\sum_{k=1}^{N} A_m \frac{\sin(k\cdot(\omega_0+\Delta\omega_m)t)}{k}, \quad N\cdot\omega_0 < \frac{\omega_S}{2}. \tag{7}$$

Square signal (pulse with 50% duty), band-limited:

$$S_{\text{sq}}(t) = \sum_{k=0}^{N} \frac{\sin\big((2k+1)\cdot\omega_0 t\big)}{2k+1}, \quad (2N+1)\cdot\omega_0 < \frac{\omega_S}{2}. \tag{8}$$

Triangle signal, band-limited:

$$S_{\text{tri}}(t) = \sum_{k=0}^{N} \frac{\sin\big((2k+1)\cdot\omega_0 t\big)}{(2k+1)^2}, \quad (2N+1)\cdot\omega_0 < \frac{\omega_S}{2} \tag{9}$$

The subtractive method of sound synthesis is the most common. This is due to its rather simplistic approach to control having just a small number of basic parameters and an intuitive representation of the signal changing results in the frequency domain. At the same time, subtractive synthesizers lack the flexibility to control individual components. Their timbres are often very recognizable and monotonous, or overused in music. The difference is achieved, in many ways, using different processing effects.

Frequency/Phase Modulation Synthesis (FM/PM Synthesis)

In contrast to the radio engineering understanding of frequency-modulated signals, sound synthesis systems based on frequency or phase modulation are characterized by cascades of several modulations and feedbacks (self-modulation). Though the original synthesis method is known as FM, most of its implementations are associated with phase modulation. Thus, w PM equations will be used below.

The unitary element of an FM sound synthesis system is the operator. Its mathematical model can be represented as:

$$S_{OP}(t) = A(t)\cdot\sin\big(\omega_0(t)t + S_M(t)\big), \tag{10}$$

where $A(t)$ – the amplitude envelope of the operator; ω_0 – the carrier of the operator; $S_M(t)$ – the frequency modulation function, which in practice can be arbitrary.

All the operators S_{OPi} are organized according to the connection algorithm, modeled with a square matrix X ($n\times n$), in which a_{ij} – the modulation caused by operator i on operator j:

$$X = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}. \tag{11}$$

The output may include various numbers of operators (from 1 to n).

$$S_{FM}(t) = \sum_{j=1}^{n} A_j(t)\sin\left(\omega_j(t)t + \sum_{i=1}^{n} a_{ij}S_i(t)\right). \tag{12}$$

In the existing FM-based sound synthesis systems [11], the operator frequency is defined by the frequency ratio $R_j = \frac{\omega_j}{\omega_0}$, where ω_j – the carrier frequency of the j-th operator's, ω_0 – the active note frequency.

$$S_{FM}(t) = \sum_{j=1}^{n} A_j(t)\sin\left(R_j\cdot\omega_0(t)t + \sum_{i=1}^{n} a_{ij}S_i(t)\right) \tag{13}$$

FM-based synthesis produces complex timbres that usually combine both harmonic and inharmonic components. In this case, the spectral composition can vary significantly over time, depending on the envelopes of each operator. A disadvantage of FM synthesis is the complexity of programming timbres in view of the difficulty of representing the resulting spectra.

Granular Synthesis

Granular-based sound objects can be modeled as a composition of signals S, taken with a window function ω and probability Q:

$$S_G(t) = \sum_{k=0}^{N} Q\big(p(k)\big)\cdot\omega\left(\frac{t-kT}{\alpha}\right)\cdot S\left(\frac{t-kT-\tau}{\beta}\right), \tag{14}$$

where $Q\big(p(k)\big) = \begin{cases}1, p(k)\ge p_0\\ 0, p(k) < p_0\end{cases}$; α – window ω scale factor; β – signal S scale factor; τ – signal S transition factor.

Generally, the factors α, β, and τ can be random values for each time k, thus defined by the probability functions p_α, p_β and p_τ. In addition, one can introduce the grain amplitude A, also randomly variating with the probability function p_A:

$$S_G(t) = \sum_{k=0}^{N} Q\big(p_Q(k)\big)\cdot A\big(p_A(k)\big)\cdot\omega\left(\frac{t-kT}{\alpha\big(p_\alpha(k)\big)}\right)\times \times S\left(\frac{t-kT-\tau\big(p_\tau(k)\big)}{\beta\big(p_\beta(k)\big)}\right). \tag{15}$$

The grain size rarely exceeds 200 ms [9], and it is not practical to manage a single granule. Therefore, a whole cloud of grains is controlled through a tuple of macro parameters:

$$G = \langle D(p_Q), A(p_A), \alpha(p_\alpha), \beta(p_\beta), \tau(p_\tau)\rangle. \tag{16}$$

Granular sound synthesis, on the one hand, is aimed at creating specific timbres formed by a combination of a large number of very short sounds, and on the other hand, it is a composition method that determines the position of various sound objects in time according to probabilistic laws. Granular synthesis is not designed to produce sound objects similar to the sounds of acoustic musical instruments (or electronic instruments that play a similar role) and is mainly used in various avant-garde composition techniques.

Table 1 presents a brief summary of the proposed models. The table uses the following abbreviations: E – the amplitude envelope set of parameters, e. g. $<A, D, S, R>$ in most of the classical cases, A – attack time; D – decay time; S – sustain level, R – release time; M – modulation set, e. g. $<L, R, F, T>$, L – modulation level; R – modulation rate; F – modulation frequency; T – modulation waveform type; F – filter parameter set, e. g. $<C, R, T>$, C – filter cut-off frequency; R – filter resonance; T – filter type.

TABLE 1. Sound Synthesis Model Survey

Synthesis Method	N_1	N_{mod}	AN	NT
Additive	$E + 3$	$2M$	E	$K_A \cdot (E + 3) \times$ $\times 2M + AE$
Subtractive	6	M	$2E + F$	$K_S \cdot (6 \cdot M) +$ $+ E + F$
FM	$E + 2$	$2M$	E	$K_F \cdot (E + 2) \times$ $\times 2M + E$
Granular (single)	5	$3M$	E	$K_{G1} \cdot 15M + E$
Granular (cloud)	–	–	$5M$	$K_{G2} \cdot 5M$

In Table 1, N_1 – the estimated max number of the model parameters per each element, i.e. oscillator, operator, etc.; N_{mod} – the estimated number of possible modulation parameters per each element; AN – estimated parameter increasing after mixing elements; NT – the estimated overall number of parameters.

Using the typical numeric values for given E, M and F, e.g. $E = 4$, $M = 4$, $F = 3$, one can estimate the complexity of control. In addition, each synthesis method operates with various numbers of elements K_x, so it can be assumed that $K_A \subset [6, 15]$, $K_S \subset [3, 5]$, $K_F \subset [3, 8]$, $K_{G1} \subset [200, 2000]$, $K_{G2} \subset [1, 5]$. Surely, the single grain model should be eliminated from the further comparison due to the overwhelming number of parameters. Thus, for the given values, some results are obtained, presented in Fig. 2 below. Blue columns correspond to the lower edges of elements number, and orange do the same for the higher ones.

It can be seen that the additive synthesis and FM synthesis are the most complicated in terms of the parameter number used to control it. Granular synthesis and subtractive synthesis are much easier to control. Though these results may seem obvious, meanwhile the attempt of numeric estimation can lead to novel approaches in the sound synthesis study.

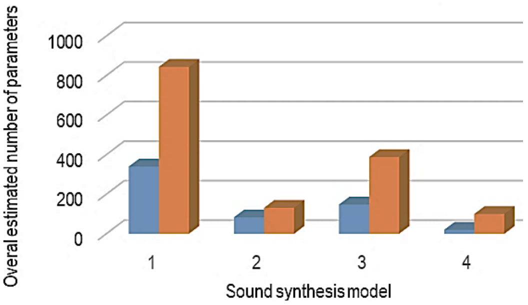

Fig. 2. Sound Synthesis Models Parameter Estimation: 1 – Additive Synthesis; 2 – Subtractive Synthesis; 3 – FM-Synthesis; 4 – Granular Synthesis

Comparison with Existing Models

It is not common to use mathematical models in the world of sound synthesis, due to its practical aspects. Typically, the sound design starts rather from the algorithm than from the theoretical description. Meanwhile, some known models can be mentioned. Smith III [12] gives an additive synthesis model combined with noise, which is close to (2) and (3), but he does not generalize his model somewhat close to (3). Schottstaedt [13] gives a mathematical description for the 3-operator FM signals, also without generalization to the n-operator case. Regarding granular synthesis, it is common to give a set of parameters (on micro and macro levels) (see Roads [9]), though without putting them together into a mathematical model. The authors were not able to find any mathematical model for subtractive synthesis, except for the trivial models of sawtooth, square, and triangle signals.

Conclusions

Several new models for the classical sound synthesis methods were presented. On the one side, the mathematical models may seem excessive and of no practical use when having an algorithmic representation that is much closer to exact sound design. Nevertheless, these mathematical models can be used for system analysis purposes, making a convenient connection between the rather specific (at least in terms of terminology) world of computer music and the more formalistic domain of system analysis. Such connection is highly needed for sonification, as the perspective intersection of computer music technologies, sound design, human interfaces, and telecommunications. Also, the availability of adequate models of synthesized signals will improve the design of various systems using artificially created sounds.

References

1. Manning P. *Electronic and Computer Music*. Oxford: Oxford University Press; 2013. 576 p.
2. Chowning J.M. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation. *Journal of the Audio Engineering Society*. 1973;21(7):526–534.
3. Roads C. *The Computer Music Tutorial*. Boston: MIT Press, 1996. 1234 p.
4. Lazzarini V., Yi S., Ffitch J., Heintz J., Brandtsegg Ø., McCurdy I. *Csound: A Sound and Music Computing System*. Cham: Springer; 2016. 516 p.
5. Cook P.R. *Real Sound Synthesis for Interactive Applications*. Boca Raton: CRC Press; 2002. 263 p.
6. Kudumakis P.E., Sandler, M. Synthesis and coding of audio signals using wavelet transforms for multimedia applications: An overview // Proceedings of the DSP'96 Technical Conference (London, UK, 3–4 December 1996). 1996. PP. 88–92
7. Ishutkin Yu.M., Uvarov V.K. *The Foundations of Modulation Transformations of Sound Signals*. St. Petersburg: Saint Petersburg State University of Film and Television Publ.; 2004. 102 p. (in Russ.)
8. Smith III J.O. Efficient Synthesis of Stringed Musical Instruments. *Proceedings of the Conference on International Computer Music, ICMC, Tokyo, Japanese*. San Francisco: Computer Music Association; 1993. p.64–71.
9. Roads C. *Microsound*. Boston: MIT Press; 2004. 424 p.
10. Rogozinsky G.G., Podolsky D.A., Goryachev N.V. *Wintermute v.0.2.8*. Patent RF, no 2019613755, 19.03.2019. (in Russ.)
11. Rogozinsky G., Goryachev N. Method of Modeling of Sound Synthesizers. *Proceedings of Telecommunication Universities*. 2019;5(1):25–30. (in Russ.) DOI:10.31854/1813-324X-2019-5-1-25-30
12. Smith III J.O. *Spectral Audio Signal Processing*. Stanford: W3K Publishing; 2011. 674 p.
13. Schottstaedt B. *An Introduction to FM*. URL: https://ccrma.stanford.edu/software/snd/snd/fm.html [Accessed 06th February 2022]

Список источников

1. Manning P. Electronic and Computer Music. Oxford: Oxford University Press, 2013. 576 p.
2. Chowning J.M. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation // Journal of the Audio Engineering Society. 1973. Vol. 21. Iss. 7. PP. 526–534.
3. Roads C. The Computer Music Tutorial. Boston: MIT Press, 1996. 1234 p.
4. Lazzarini V., Yi S., Ffitch J., Heintz J., Brandtsegg Ø., McCurdy I. Csound: A Sound and Music Computing System. Cham: Springer, 2016. 516 p.
5. Cook P.R. Real Sound Synthesis for Interactive Applications. Boca Raton: CRC Press, 2002. 263 p.
6. Kudumakis P.E., Sandler, M. Synthesis and coding of audio signals using wavelet transforms for multimedia applications: An overview. Proceedings of the DSP'96 Technical Conference, 3–4 December 1996, London, UK. 1996. p.88–92
7. Ишуткин Ю.М., Уваров В.К. Основы модуляционных преобразований звуковых сигналов. СПб.: СПбГУКиТ, 2004. 102 с.
8. Smith III J.O. Efficient Synthesis of Stringed Musical Instruments // Proceedings of the Conference on International Computer Music (ICMC, Tokyo, Japanese). San Francisco: Computer Music Association, 1993. PP. 64–71.
9. Roads C. Microsound. Boston: MIT Press, 2004. 424 p.
10. Рогозинский Г.Г., Подольский Д.А., Горячев Н.В. Wintermute v.0.2.8. Свидетельство о регистрации программы для ЭВМ RU 2019613755 от 19.03.2019. Опубл. 22.03.2019.
11. Рогозинский Г.Г., Горячев Н.В. Метод моделирования синтезаторов звуковых сигналов // Труды учебных заведений связи. 2019. Т. 5. № 1. С. 25–30. DOI:10.31854/1813-324X-2019-5-1-25-30
12. Smith III J.O. Spectral Audio Signal Processing. Stanford: W3K Publishing, 2011. 674 p.
13. Schottstaedt B. An Introduction to FM. URL: https://ccrma.stanford.edu/software/snd/snd/fm.html (дата обращения 06.02.2022)

Статья поступила в редакцию 06.02.2022; одобрена после рецензирования 03.06.2022; принята к публикации 06.06.2022.
The article was submitted 06.02.2022; approved after reviewing 03.06.2022; accepted for publication 06.06.2022.

Информация об авторах:

РОГОЗИНСКИЙ Глеб Гендрихович — доктор технических наук, начальник НОЦ «Медиацентр» Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича,
https://orcid.org/0000-0001-5698-2347

ЧЕСНОКОВ Михаил Александрович — кандидат технических наук, ведущий инженер-электроник НТЦ «ЯФИ»
https://orcid.org/0000-0001-7951-8627

КУТЛЫЯРОВА Александра Александровна — ассистент кафедры звукорежиссуры Санкт-Петербургского государственного института кино и телевидения
https://orcid.org/0000-0002-8621-1313