UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

WYDZIAŁ TELEKOMUNIKACJI
I ELEKTROTECHNIKI

ZESZYTY NAUKOWE NR 249

TELEKOMUNIKACJA I ELEKTRONIKA
10

BYDGOSZCZ – 2007

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

ZESZYTY NAUKOWE NR 249

TELEKOMUNIKACJA I ELEKTRONIKA 10

BYDGOSZCZ – 2007

ISSN 0209-0589

Wydawnictwa Uczelniane Uniwersytetu Technologiczno-Przyrodniczego
ul. Ks. A. Kordeckiego 20, 85-225 Bydgoszcz, tel. (052) 3749482, 3749426
e-mail: wydawucz@utp.edu.pl http://www.utp.edu.pl/~wyd

Wyd. I. Nakład 100 egz. Ark. aut. 5,40. Ark. druk. 5,25. Zamówienie nr 9/2007
Oddano do druku i druk ukończono w październiku 2007 r.
Zakład Małej Poligrafii UTP Bydgoszcz, ul. Ks. A. Kordeckiego 20

Contents

FOREWORD

This publication is the tenth issue of the University Scientific Journal series Electronics and Telecommunications, and at the same time the first one after the change of our University name that is now the University of Technology and Life Sciences.

There are five articles published in this issue. Two of them came from abroad.

Fabio Bagarello from the University of Palermo, Italy reviews in his paper the results concerning multiplication of distributions using a pedagogical approach.

In Tadeus Uhl (Flensburg University of Applied Sciences, Germany) paper, the concepts, standards and properties associated with the new service Voice over IP (VoIP) are discussed. The paper includes a comparison of the Internet telephony with the Public Switched Telephone Network (PSTN) telephony. The results of this comparison, in author's opinion, throw new light on the areas in which a great deal of work is still to be done.

Igor Jawors'kyj from the University of Technology and Life Sciences in Bydgoszcz, Poland presents in his paper the description of properties of stochastic oscillations in terms of probabilistic characteristics of the periodically correlated stochastic processes and their generalizations. This article is supported by an extensive bibliography.

Adam Dąbrowski, Piotr Kardyś and Tomasz Marciniak from Poznań University of Technology, Poland present an article dealing with the short-range wireless systems for transmission of information in biomedical devices. The authors are concentrated on consideration of the Bluetooth digital wireless connectivity system.

In the second article coming from Poznań University of Technology, Paweł Pawłowski and Adam Dąbrowski discuss the extended precision method for accumulation of floating point numbers in digital signal processors. The idea presented in this paper is used to overcome the phenomenon of reduction of accuracy of computations in the case of long series of additions of small numbers successively added to a relatively large intermediate result.

I invite you to read this, in my opinion, very interesting issue, and to consider submissions of your manuscripts for publication in the incoming issues of our University Scientific Journal series Electronics and Telecommunications.

With friendly greetings

Editor of Electronics and Telecommunications Series
Sławomir Cieślik

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 249
TELEKOMUNIKACJA I ELEKTRONIKA 10 (2007), 5-26

A PEDAGOGICAL APPROACH TO MULTIPLICATION OF DISTRIBUTIONS IN ANY SPATIAL DIMENSIONS

F. Bagarello

Dipartimento di Metodi e Modelli Matematici. Facoltà di Ingegneria,
Università di Palermo, I – 90128 Palermo, Italy
e-mail: bagarell@unipa.it
web page: www.unipa.it/~bagarell

Abstract

We review some results concerning multiplication of distributions using a pedagogical approach: we will first sketch the problems which may arise trying to multiply two distributions and, after discussing some mathematical tools, we will introduce several definitions of multiplication. In particular we give the details of a recent definition which works in any spatial dimension. As a particular application, we prove that delta functions and their derivatives can be multiplied.

Keywords: distributions; delta functions; functional analysis

1 Introduction

This paper was originally motivated by the request of writing something mathematically rigorous and, at the same time, clear for an audience mainly of engineers. For this reason, after this short introduction, we will discuss first few general facts about distributions, showing in particular that multiplying two such objects may be a rather dangerous operation. After that we will set up a mathematical framework where distributions *take their place*, discussing explicitly some mathematical tools, generally well known to any mathematical audience, that will be used in the rest of the paper. Finally we will introduce and discuss few different multiplications of distributions which we have originally defined elsewhere, and we will show that in this way we bypass the difficulties previously discussed.

It should be noticed that in the literature several examples of multiplication of distributions exist. In [1] and [2] we have proposed our own definition of multiplication in one spatial dimension, $d = 1$, and we proved that two or more delta functions can be multiplied. However, this definition does not admit a *natural* extension to $d > 1$ and this is a strong limitation for concrete applications. Extensions of our procedure to higher spatial dimensions were strongly urged by a large number of engineers and it has produced some new results in [3]. We will review these results in the last part of this paper.

2 Some preliminary mathematics

We divide this section in three parts. In the first subsection we will discuss some heuristic results on distributions, focusing in particular on some problems arising when dealing with the Dirac delta distribution. In the second part we will construct the rigorous mathematical framework in which the distribution theory *lives* and we discuss few important results. In the third part we will briefly review few mathematical facts which are useful in the rest of the paper.

2.1 Problems with $\delta(x)$

We begin this section by recalling the famous integral expression for the Dirac delta function $\delta(x)$: $\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ipx}dp$. To a deeper analysis this object looks very strange. Indeed we have, if $x \neq 0$,

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ipx}dp = \frac{1}{2\pi}\lim_{R,\infty}\int_{-R}^{+R} e^{-ipx}dp = \frac{1}{2\pi}\lim_{R,\infty}\frac{e^{-ipx}}{ix}\Big|_{-R}^{+R} =$$

$$= \lim_{R,\infty}\frac{1}{2\pi}\frac{e^{iRx}-e^{-iRx}}{ix} = \lim_{R,\infty}\frac{1}{\pi}\frac{\sin Rx}{x},$$

and this limit does not exist since $\sin(Rx)$ is an oscillating function. If, on the contrary, $x = 0$ we get

$$\delta(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dp = \frac{1}{2\pi}p\,|_{-\infty}^{+\infty} = \infty.$$

Therefore we don't really know what $\delta(x)$ is if $x \neq 0$, while $\delta(x)$ is infinite if $x = 0$. This makes quite hard to consider $\delta(x)$ as a *canonical* function, which associates to each value of a certain domain one and only one finite value of the related codomain.

Nevertheless, it is quite useful in a huge number of applications to have *mathematical object* which behaves in the following way:

$$\int_{\mathbb{R}} \delta(x - x_0) f(x) dx = f(x_0). \tag{2.1}$$

for all functions $f(x)$ which are at least continuous in x_0. This is a second well known property of the delta functions, and it is actually taken quite often as its definition. It can be proven that no locally integrable positive function exists which satisfies (2.1), see Appendix. However we can still try to define $\delta(x)$ as a suitable limit of certain functions.

With this in mind we define the sequence $\delta_n(x) = n\, rect(nx)$, where

$$rect(x) = \begin{cases} 1 & -\frac{1}{2} \leq x \leq \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that $\{\delta_n(x)\}$ does not converge pointwise in $\mathbb{R}$. In particular, it is quite easy to understand that it cannot be converging in $x = 0$, since the sequence $\{\delta_n(0) = n\}$ diverges to $+\infty$. On the contrary if $x_0 \neq 0$ the sequence $\{\delta_n(x_0)\}$ converges to 0.

However it is easy to check that for all continuous functions $f(x)$ the following holds:

$$\lim_{n,\infty} \int_{\mathbb{R}} \delta_n(x) f(x) dx = f(0) = \int_{\mathbb{R}} \delta(x) f(x) dx. \tag{2.2}$$

Indeed we have, using well known elementary analysis to compute the integral

$$\int_{\mathbb{R}} \delta_n(x) f(x) dx = n \int_{-\frac{1}{2n}}^{\frac{1}{2n}} f(x) dx = n \left[\left(\frac{1}{2n} + \frac{1}{2n} \right) f(\xi_n) \right] = f(\xi_n),$$

where $\frac{1}{2n} + \frac{1}{2n}$ is the measure of the integration interval and ξ_n is a certain point in this same interval. Of course when $n \to \infty$ the interval $[-\frac{1}{2n}, \frac{1}{2n}]$ becomes smaller and smaller and therefore ξ_n tends to zero. Using now the continuity at the origin of the function $f(x)$ and formula (2.1) we conclude that

$$\lim_{n,\infty}\int_{\mathbb{R}}\delta_n(x)f(x)dx=\lim_{n,\infty}f(\xi_n)=f\left(\lim_{n,\infty}\xi_n\right)=f(0)=\int_{\mathbb{R}}\delta(x)f(x)dx.$$

The meaning of what we have seen so far can be summarized as follows: *the sequence $\delta_n(x)$ does not converge for all $x \in \mathbb{R}$ but it converges in a weak sense, i.e. when smeared with a continuous function. Moreover, the weak limit of $\delta_n(x)$ behaves exactly as $\delta(x)$ in (2.1).* In other words: we can introduce a delta function as the weak limit of a certain sequence!

We want to remark two important aspects of this procedure: first of all there exist many other possible sequences $\{\delta_n(x)\}$ converging weakly to $\delta(x)$. These are the so-called *delta-sequences*, and can be constructed for instance starting from an even, non-negative, C^∞-function $\Phi(x)$ with support in $[-1,1]$ and defining $\delta_n(x)=n\Phi(nx)$, $\forall n\in\mathbb{N}$. Again, for each $f(x)$ continuous in $x=0$, formula (2.2) holds. The second remark, which is crucial for what we want to discuss in the rest of the paper, is that this same approach does not work if we try to define $\delta^2(x)$. Indeed if we repeat the same steps as above it is clear that $\lim_{n,\infty}\int_{\mathbb{R}}\delta_n^2(x)f(x)dx$ does not exist for all continuous functions $f(x)$ but only, at most, for those which go to zero fast enough when $x\to 0$. This is the main reason why we say that the square of a delta function cannot be defined!

2.2 The Schwartz framework

Defining the delta function via a weak limit is not the most convenient way to proceed. For this reason we discuss here a rigorous framework, first proposed by Laurent Schwartz, [9] and references therein, where the distributions appear as natural objects and no weak limiting procedure is needed, in principle.

We begin by defining two spaces of regular functions which will play a major role in the treatment of the distributions. Let

$$\mathcal{D}(\mathbb{R})=\{h(x)\in C^\infty \text{ compactly supported in } \mathbb{R}\}. \tag{2.3}$$

and

$$\mathcal{S}(\mathbb{R})=\left\{g(x)\in C^\infty:\lim_{|x|,\infty}|x|^k g^{(l)}(x)=0,\quad \forall k,l\in\mathbb{N}_0\right\}, \tag{2.4}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. They satisfy the following inclusion $\mathcal{D}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \subset \mathcal{L}^2(\mathbb{R})$ and, more than this, it is possible to check that they are both dense in the space

$$\mathcal{L}^p(\mathbb{R}) = \left\{ f(x) \text{ L-measurable} : \|f\|_p^p := \int_{\mathbb{R}} |f(x)|^p \, dx < \infty \right\}$$

in the norm $\|.\|_p$, for each $p \geq 1$. They are also closed with respect to their own natural topologies, $\tau_{\mathcal{D}}$ and $\tau_{\mathcal{S}}$, whose definition can be found for instance in [9]. This means that given for instance a sequence of functions $f_n(x) \in \mathcal{D}(\mathbb{R})$ which is $\tau_{\mathcal{D}}$-Cauchy then it converges in the same topology $\tau_{\mathcal{D}}$ to a function of $\mathcal{D}(\mathbb{R})$. Analogously the sequence $f_n(x) \in \mathcal{S}(\mathbb{R})$ is $\tau_{\mathcal{S}}$-Cauchy iff it converges in the topology $\tau_{\mathcal{S}}$. In this case the limit belongs to $\mathcal{S}(\mathbb{R})$. Finally, we will say that $f_n(x)$ is $\|.\|_p$-convergent if it converges in the norm $\|.\|_p$. In this case its limit belongs to $\mathcal{L}^p(\mathbb{R})$.

The possibility of introducing different possible definitions of convergence is a standard fact in mathematics: it is well known, already at an elementary level, the difference between the *pointwise* and the *uniform* convergence of a sequence of functions. In particular, for instance, it is well known that any uniformly convergent sequence is also pointwise convergent, while the converse is not true. In our present settings it is possible to prove, for instance, that if a sequence converges in the $\tau_{\mathcal{S}}$ topology, then it is also $\|.\|_2$-convergent but not vice-versa.

Let now $\mathcal{H}$ be an Hilbert space and F a map with domain $D(F) \subseteq \mathcal{H}$.

Definition 2.1 *A functional on $\mathcal{H}$ is a linear map $F : D(F) \to \mathbb{C}$: $F[\alpha_1 f_1 + \alpha_2 f_2] = \alpha_1 F[f_1] + \alpha_2 F[f_2]$, for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and for all $f_1, f_2 \in D(F)$.*

A first relevant class of examples of functionals is constructed in the following way: let $\varphi(x)$ be a fixed function in $\mathcal{L}^2(\mathbb{R})$ and let us define the map

$$F_\varphi[f] := \int_{\mathbb{R}} \overline{\varphi(x)}\, f(x)\, dx =< \varphi, f >, \tag{2.5}$$

where $< .,. >$ is the usual scalar product in $\mathcal{L}^2(\mathbb{R})$. It is clear that, for any fixed choice of $\varphi(x)$, F_φ is a functional with domain $D(F_\varphi) = \mathcal{L}^2(\mathbb{R})$.

A second example is the following: $\forall\, f(x) \in \mathcal{S}(\mathbb{R})$ we put

$$F_{x_0}[f] := f(x_0) = \int_{\mathbb{R}} \delta(x - x_0) f(x) dx. \tag{2.6}$$

F_{x_0} is the functional associated to $\delta(x-x_0)$, and is defined on a proper subset of $\mathcal{L}^2(\mathbb{R})$, $\mathcal{S}(\mathbb{R})$, (and therefore also on $\mathcal{D}(\mathbb{R})$). Of course definition (2.6) cannot be extended to all of $\mathcal{L}^2(\mathbb{R})$ since in the Lebesgue theory the value of a function in a single point (or in a set of zero measure) is a meaningless notion.

It is worth stressing that the *rigorous* definition of the delta functional in equation (2.6) is $F_{x_0}[f] := f(x_0)$, while the other part of the same equation has only to be considered as a convenient extension of formula (2.5) where, however, we are introducing a intrinsically not well-defined object, $\delta(x-x_0)$, as if it was a regular function.

Definition 2.2 *A functional T on $\mathcal{D}(\mathbb{R})$ is* continuous *if, for each sequence $\{\varphi_j\} \subset \mathcal{D}(\mathbb{R})$ $\tau_{\mathcal{D}}$-convergent to a certain $\varphi \in \mathcal{D}(\mathbb{R})$, then $T[\varphi_j] \to T[\varphi]$ in $\mathbb{C}$. Any such functional is a* **distribution**. *The set of all the distributions is called $\mathcal{D}'(\mathbb{R})$.*

In complete analogy we have the following:

Definition 2.3 *A continuous functional on $\mathcal{S}(\mathbb{R})$ is a* **tempered distribution**. *The set of all the tempered distributions is called $\mathcal{S}'(\mathbb{R})$.*

The following inclusion holds: $\mathcal{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R})$. Now it is a trivial exercise to check that both F_φ in (2.5) and F_{x_0} in (2.6) are continuous functionals on $\mathcal{D}(\mathbb{R})$, for each $\varphi(x) \in \mathcal{L}^2(\mathbb{R})$ and for each $x_0 \in \mathbb{R}$. Therefore they are both distributions. Moreover, it is possible to check that they are also tempered distributions.

Let us now prove, as an exercise, that F_φ is continuous on $\mathcal{S}(\mathbb{R})$. For this we recall that if $\{f_j\}$ $\tau_{\mathcal{S}}$-converges to f, then we also have $\|f_j - f\|_2 \to 0$. Therefore, using the linearity of F_φ and the Schwarz inequality, we have $|F_\varphi[f_j] - F_\varphi[f]| = |<\varphi, f_j> - <\varphi, f>| = |<\varphi, f_j - f>| \leq \|\varphi\|_2 \|f_j - f\|_2 \to 0$, which is what we had to prove.

On the distributions we can operate almost as for ordinary functions. More in details: two distributions T_1 and T_2 are equal if they act in the same way on any test function, i.e. if $T_1[\varphi] = \int_{\mathbb{R}} f_1(x)\,\varphi(x)\,dx = \int_{\mathbb{R}} f_2(x)\,\varphi(x)\,dx = T_2[\varphi]$, for each $\varphi(x) \in \mathcal{S}(\mathbb{R})$ or for each $\varphi(x) \in \mathcal{D}(\mathbb{R})$.

Moreover, given two distributions T_1 and T_2, their sum is defined as $(T_1+T_2)[\varphi] = T_1[\varphi] + T_2[\varphi]$, $\forall\,\varphi(x) \in \mathcal{D}(\mathbb{R})$. Also, if $\alpha \in \mathbb{C}$ and $T \in \mathcal{D}'(\mathbb{R})$, we define a new distribution (αT) as $(\alpha T)[\varphi] = \alpha(T[\varphi])$, $\forall\,\varphi(x) \in \mathcal{D}(\mathbb{R})$. Finally, we can also define the derivative of such an extremely singular object via the following integration by parts-like definition:

$$\int_{\mathbb{R}} \frac{d^kT(x)}{dx^k}\,\varphi(x)\,dx := (-1)^k \int_{\mathbb{R}} T(x)\,\frac{d^k\varphi(x)}{dx^k}dx, \tag{2.7}$$

for $k = 0, 1, 2, \ldots$ and for all $\varphi(x) \in \mathcal{D}(\mathbb{R})$. This is surely well defined since the right hand side exists because, since $\varphi(x) \in \mathcal{D}(\mathbb{R})$, then $\frac{d^k\varphi(x)}{dx^k} \in \mathcal{D}(\mathbb{R})$ as well for all $k \geq 0$.

Let now $\Omega \subset \mathbb{R}$ be an open set. We say that the distribution T *vanishes in* Ω if $T[\varphi] = 0$ for all $\varphi(x)$ with support in Ω. The support of T is the complement in $\mathbb{R}$ of the largest open set on which T vanishes.

2.3 More mathematics

In what follows we will need few notions on analytic functions. We begin with the definition itself of an analytic function $f(z) = u(x,y) + iv(x,y)$, with $z = x+iy \in \mathbb{C}$. Here we have introduced the real $(u(x,y))$ and the complex $(v(x,y))$ parts of $f(z)$, and we have observed that each complex number $z = x+iy$ can be put in one-to-one correspondence with $(x,y) \in \mathbb{R}^2$. Such a complex function $f : \mathbb{C} \to \mathbb{C}$ is said to be analytic if it admits derivative for each z in its domain, $D(f)$. Necessary and sufficient conditions for this to be true are the well known *Cauchy-Riemann conditions*, $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$, $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$, with u and v differentiable in $\tilde{D}(f) := \{(x,y) \in \mathbb{R}^2 \,:\, x+iy \in D(f)\}$. For analytic functions many results exists in the literature, starting with the Cauchy theorem which states that $\oint_\gamma f(z)\,dz = 0$ if γ is a closed curve all contained in a simply connected region, passing through the integral formula for $f(z)$ analytic in a region $\mathcal{R}$: $f(z) = \frac{1}{2\pi i}\oint_\gamma \frac{f(z')\,dz'}{z'-z}$ for each z internal to $\gamma \subset \mathcal{R}$ and arriving to the Taylor and the Laurent expansion. We refer to [10] for further readings.

Analytic functions are useful here because of the following regularization procedure, which was introduced in [7], and where it is proven that, given a distribution T with compact support, the function

$$\mathbf{T}^0(z) \equiv \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{T(x)}{x-z}\,dx \tag{2.8}$$

exists and is analytic in z in the whole z-plane minus the support of T. Moreover, if $T(x)$ is a continuous function with compact support, then $T_{red}(x,\epsilon) \equiv \mathbf{T}^0(x+i\epsilon) - \mathbf{T}^0(x-i\epsilon)$ converges uniformly to $T(x)$ on the whole real axis for $\epsilon \to 0^+$. Finally, if T is a distribution in $\mathcal{D}'(\mathbb{R})$ with compact support then $T_{red}(x,\epsilon)$ converges to T in the following weak sense

$$T(\phi) = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} T_{red}(x,\epsilon)\,\phi(x)\,dx$$

for every test function $\phi \in \mathcal{D}(\mathbb{R})$.

As discussed in [7], this definition can be extended to a larger class of one-dimensional distributions with support not necessarily compact, $\mathcal{V}'(\mathbb{R})$. while it is much harder to extend the same definition to more than one spatial dimension.

The next mathematical result we want to discuss here, and which will be very useful in the following, is a second regularization procedure known as *the sequential completion*. This is related to well known results on the regularity of the convolution of distributions and test functions. Let $\phi \in \mathcal{D}(\mathbb{R})$ be a given function with support in $[-1,1]$, $supp(\phi) \subseteq [-1,1]$, and $\int_{\mathbb{R}} \phi(x)\,dx = 1$. As already discussed before. we call $\delta-$sequence the sequence δ_n, $n \in \mathbb{N}$, defined by $\delta_n(x) \equiv n\,\phi(nx)$. Then, $\forall T \in \mathcal{D}'(\mathbb{R})$, the convolution $T_n(x) \equiv (T * \delta_n)(x) = \int_{\mathbb{R}} T(y)\delta_n(x-y)\,dy$ is a $C^\infty-$function for any fixed $n \in \mathbb{N}$. The sequence $\{T_n\}$ converges to T in the topology of $\mathcal{D}'$, when $n \to \infty$. Moreover, if $T(x)$ is a continuous function with compact support, then $T_n(x)$ converges uniformly to $T(x)$. A detailed proof of these convergence properties can be found, for instance, in [4]. We will return on these and other features of the sequential completion later on.

One of the most important tool used in [1, 2, 3] is the so called *Lebesgue dominated convergence theorem* (LDCT) which gives sufficient conditions to exchange a limit with an integral. More in details, it is well known that given a sequence of continuous functions $\{f_n(x)\}$ uniformly convergent to $f(x)$ for each $x \in [a,b]$, with $0 < b-a < \infty$, then $f(x)$ is Riemann-integrable and

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \lim_{n\to\infty} f_n(x)\,dx = \int_a^b f(x)\,dx.$$

Of course, uniform convergence is a very strong requirement. Moreover, the above result cannot be proven in general if the interval $[a,b]$ is replaced by a more general set whose Lebesgue measure is not necessarily finite. However, if we replace the Riemann with the Lebesgue integral, see [10] for instance, then it is possible to weaken quite a bit the requirements on the sequence $\{f_n(x)\}$. For that it is necessary to introduce first the following definition:

Definition 2.4 *A sequence $\{f_n(x)\}$ converges* almost everywhere *(a.e.) to a function $f(x)$ in A, A a Lebesgue measurable subset of $\mathbb{R}$, if we have $\lim_{n\to\infty} f_n(x) = f(x)$ for each $x \in A$ but, at most, for a set of zero measure.*

As an example the sequence $\{f_n(x) = x^n\}$ converges to $f(x) \equiv 0$ a.e. in $[0,1]$. Indeed, it converges pointwise to $f(x)$ in $[0,1[$, while it does not converge for $x = 1$. However, since the Lebesgue measure of the set $\{1\}$ is zero, the previous definition is satisfied.

Now the LDCT looks like:

Teorema 2.5 *Let $\{f_n(x)\}$ be a sequence of functions converging a.e. to $f(x)$ on the Lebesgue-measurable set A and let $\varphi(x)$ be a non-negative and L-integrable function such that $|f_n(x)| \leq \varphi(x)$, for all $n \in \mathbb{R}$ and a.e. in A. Then both $f_n(x)$ and $f(x)$ are Lebesgue-integrable on A and*

$$\lim_{n\to\infty} \int_A f_n(x)\, dx = \int_A f(x)\, dx. \tag{2.9}$$

The proof of this theorem can be found in many textbooks of functional analysis, see for instance [10], and we will omit it here. Other results related again to the exchange of a limit with a Lebesgue integral also exist, but will play no role in this paper. Just for completeness's sake, we cite the *Lebesgue monotone convergence theorem* and the *Fatou's Lemma*, for which again we refer to [10].

3 A first attempt

In [1, 2] we have introduced a class of multiplications of distributions making use of the two different regularizations discussed above and adapted to our purposes.

As already said, the first ingredient is the regularization procedure proposed in [7], which produces the function $T_{red}(x,\epsilon) = \mathbf{T}^0(x+i\epsilon) - \mathbf{T}^0(x-i\epsilon)$. $T_{red}(x,\epsilon)$ converges, under certain conditions also discussed before, to T as ϵ goes to zero.

The second ingredient is the method of sequential completion, which produces the C^∞ function $T_n(x) \equiv (T*\delta_n)(x) = \int_{\mathbb{R}} T(y)\delta_n(x-y)\,dy$ starting from any given distribution $T \in \mathcal{D}'$ and from any $\delta-$sequence.

For any pair of distributions $T, S \in \mathcal{D}'(\mathbb{R})$, $\forall\,\alpha, \beta > 0$ and $\forall\,\Psi \in \mathcal{D}(\mathbb{R})$ we define the following quantity:

$$(S\otimes T)_n^{(\alpha,\beta)}(\Psi) \equiv \frac{1}{2}\int_{-\infty}^{\infty}\left[S_n^{(\beta)}(x)\,T_{red}(x,\frac{1}{n^\alpha}) + T_n^{(\beta)}(x)\,S_{red}(x,\frac{1}{n^\alpha})\right]\,\Psi(x)\,dx \tag{3.1}$$

where

$$S_n^{(\beta)}(x) \equiv (S * \delta_n^{(\beta)})(x), \tag{3.2}$$

with $\delta_n^{(\beta)}(x) \equiv n^\beta \Phi(n^\beta x)$. This integral is surely well defined for any choice of α, β, T, S and Ψ. Moreover, if the limit of the sequence $\left\{(S\otimes T)_n^{(\alpha,\beta)}(\Psi)\right\}$ exists for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$, we define $(S\otimes T)_{(\alpha,\beta)}(\Psi)$ as:

$$(S\otimes T)_{(\alpha,\beta)}(\Psi) \equiv \lim_{n\to\infty}(S\otimes T)_n^{(\alpha,\beta)}(\Psi) \tag{3.3}$$

Of course, as already remarked in [1], the definition (3.3) really defines many multiplications of distributions. In order to obtain one definite product we have to fix the positive values of α and β and the particular function Φ which is used to construct the δ-sequence. Moreover, if $T(x)$ and $S(x)$ are two continuous functions with compact supports, and if α and β are any pair of positive real numbers, then: (i) $T_n^{(\beta)}(x)\,S_{red}(x,\frac{1}{n^\alpha})$ converges uniformly to $S(x)\,T(x)$; (ii) $\forall\,\Psi(x) \in \mathcal{D}(\mathbb{R}) \Rightarrow (T\otimes S)_{(\alpha,\beta)}(\Psi) = \int_{-\infty}^{\infty} T(x)\,S(x)\,\Psi(x)\,dx$.

It is furthermore very easy to see that the product $(S\otimes T)_{(\alpha,\beta)}$ is a linear functional on $\mathcal{D}(\mathbb{R})$ due to the linearity of the integral and to the properties of the limit. The continuity of such a functional is, on the contrary, not obvious at all, but, as formulas (3.5)-(3.10) will show, this is a free benefit of our definition.

We now list, with no detail, few results obtained in [1, 2]. We postpone to the next section the detailed analysis of the role of the LDCT in the computation of the limits arising from our strategy. In Section IV we will also comment on the case $\alpha = \beta$, which essentially reproduces what is assumed in [8].

If we assume Φ to be of the form

$$\Phi(x)=\begin{cases}\frac{x^m}{N_m}\cdot\exp\{\frac{1}{x^2-1}\}, & |x|<1\\ 0, & |x|\geq 1.\end{cases} \tag{3.4}$$

where m is an even natural number and N_m is a normalization constant which gives $\int_{-1}^{1}\Phi(x)\,dx=1$, and we put $A_j\equiv\int_{-\infty}^{\infty}\frac{\Phi(t)}{t^j}\,dt$, whenever it exists, we have proved that:
if $m>1$ then

$$(\delta\otimes\delta)_{(\alpha,\beta)}=\begin{cases}\frac{1}{\pi}A_2\delta, & \alpha=2\beta\\ 0, & \alpha>2\beta;\end{cases} \tag{3.5}$$

if $m>2$ then

$$(\delta\otimes\delta')_{(\alpha,\beta)}=0 \qquad \forall\alpha\geq 3\beta; \tag{3.6}$$

if $m>3$ then

$$(\delta'\otimes\delta')_{(\alpha,\beta)}=\begin{cases}\frac{-6}{\pi}A_4\delta, & \alpha=4\beta\\ 0, & \alpha>4\beta.\end{cases} \tag{3.7}$$

Also, if $m>3$ then

$$(\delta\otimes\delta'')_{(\alpha,\beta)}=\begin{cases}\frac{6}{\pi}A_4\delta, & \alpha=4\beta\\ 0, & \alpha>4\beta.\end{cases} \tag{3.8}$$

If $m>4$ then

$$(\delta'\otimes\delta'')_{(\alpha,\beta)}=0 \qquad \forall\alpha\geq 5\beta. \tag{3.9}$$

Finally, if $m>5$ then

$$(\delta''\otimes\delta'')_{(\alpha,\beta)}=\begin{cases}\frac{120}{\pi}A_6\delta, & \alpha=6\beta\\ 0, & \alpha>6\beta.\end{cases} \tag{3.10}$$

It is worth stressing that, because of our technique which strongly relies on the LDCT, the above formulas only give sufficient conditions for the product between different distributions to exist. In other words: formula (3.5) does not necessarily implies that $(\delta\otimes\delta)_{(\alpha,\beta)}$ does not exist for $\alpha<2\beta$. In this case, however, different techniques should be used to check the existence or the non-existence of $(\delta\otimes\delta)_{(\alpha,\beta)}$.

More remarks on this approach can be found in [1] where, in particular the above results are extended to the product between two distributions like $\delta^{(l)}$ and $\delta^{(k)}$ for generic $l,k=0,1,2,\ldots$. Again in [1] we have also discussed in some details two applications of our multiplication to quantum mechanical systems. The first model is described by the hamiltonian

$H_d = -\frac{1}{2}\frac{d^2}{dx^2} + V_o\delta(x)\delta(x-d)$, where d is a fixed length. This hamiltonian, depending on the value of d, describes different physics and we have discussed how our regularization procedure may help in giving a meaning to H_d also when $d = 0$. A second application concerns the analysis of the wave-function of a two-particles quantum system, [1].

In [2] we have introduced possible extensions of the definition of our multiplication to more and to not commuting distributions. Both these extensions are of a crucial relevance in quantum field theory. In particular we have seen that the attempt of multiplying more distributions open various inequivalent possibilities.

For instance, if S_1, S_2 and S_3 are three commuting distributions for which both the regularization procedures discussed before make sense, it is quite natural to consider the following quantity

$$(S_1 \otimes S_2 \otimes S_3)_n^{(\alpha,\beta)} \equiv \frac{1}{3}[(S_1 \otimes S_2)_n^{(\alpha,\beta)}\, S_3 + (S_1 \otimes S_3)_n^{(\alpha,\beta)}\, S_2 + \\ +(S_2 \otimes S_3)_n^{(\alpha,\beta)}\, S_1], \tag{3.11}$$

which is certainly well defined for any fixed n, since each term above is the product of a C^∞ function for a distribution. As usual, what may or may not exist is the limit for $n \to \infty$ of $(S_1 \otimes S_2 \otimes S_3)_n^{(\alpha,\beta)}(\Psi)$, for any $\Psi \in \mathcal{D}(\mathbb{R})$. If this limit exists we say that the distributions can be multiplied and we put

$$(S_1 \otimes S_2 \otimes S_3)_{(\alpha,\beta)}(\Psi) \equiv \lim_{n\to\infty} (S_1 \otimes S_2 \otimes S_3)_n^{(\alpha,\beta)}(\Psi). \tag{3.12}$$

We refer to [2] for further extensions to more distributions. In particular, in [2] we show that an arbitrary number of delta distributions can be simultaneously multiplied. Also, we apply our regularization procedure to a non-trivial problem in quantum field theory, and more specifically to a 1+1 dimensional Klein-Gordon model, but unfortunately we conclude that our definition does not allow to *cure* the divergences arising in that model, and which in the standard treatment are usually taken care of by the so-called *renormalization procedure*.

4 A more convenient definition in $d = 1$

In a very recent paper, [3], we have proposed a different definition of multiplication of two distributions. The idea behind this new definition is very simple: since the regularization $T \to T_{red}$ cannot be easily generalized to higher spatial dimensions, $d > 1$, we use twice the convolution procedure in (3.2), $T \to T_n^{(\alpha)} = T * \delta_n^{(\alpha)}$, with different values of α as we will see.

Let $\Phi(x) \in \mathcal{D}(\mathbb{R})$ be a given non negative function, with support in $[-1,1]$ and such that $\int_{\mathbb{R}} \Phi(x)\,dx = 1$. In the rest of this section, as in [1, 2, 3], we will mainly consider the choice of $\Phi(x)$ in (3.4). As we have discussed in [1] the sequence $\delta_n^{(\alpha)}(x) = n^\alpha \Phi(n^\alpha x)$ is a *delta-sequence* for any choice of $\alpha > 0$. This means that: (1) $\delta_n^{(\alpha)}(x) \to \delta(x)$ in $\mathcal{D}'(\mathbb{R})$ for any $\alpha > 0$; (2) for all $\alpha > 0$, if $T(x)$ is a continuous function with compact support then the convolution $T_n^{(\alpha)}(x) = (T * \delta_n^{(\alpha)})(x)$ converges to $T(x)$ uniformly; (3) $\forall \alpha > 0$, if $T(x) \in \mathcal{D}(\mathbb{R})$, then $T_n^{(\alpha)}(x)$ converges to $T(x)$ in the topology $\tau_{\mathcal{D}}$ of $\mathcal{D}(\mathbb{R})$; (4) if $T \in \mathcal{D}'(\mathbb{R})$ then $T_n^{(\alpha)}(x)$ is a C^∞ function and it converges to T in $\mathcal{D}'(\mathbb{R})$ as n diverges, independently of $\alpha > 0$.

We remark here that all these results can be naturally extended to larger spatial dimensions, and this will be useful in the next section.

Our next step is to replace definitions (3.3) with our alternative multiplication. To begin with, let us consider two distributions $T, S \in \mathcal{D}'(\mathbb{R})$, and let us compute their convolutions $T_n^{(\alpha)}(x) = (T * \delta_n^{(\alpha)})(x)$ and $S_n^{(\beta)}(x) = (S * \delta_n^{(\beta)})(x)$ with $\delta_n^{(\alpha)}(x) = n^\alpha \Phi(n^\alpha x)$, and with $\alpha, \beta > 0$ to be fixed in the following. $T_n^{(\alpha)}(x)$ and $S_n^{(\beta)}(x)$ are both C^∞ functions, so that for any $\Psi(x) \in \mathcal{D}(\mathbb{R})$ and for each fixed $n \in \mathbb{N}$, all the following integrals surely exist:

$$(S \odot T)_n^{(\alpha,\beta)}(\Psi) \equiv \frac{1}{2} \int_{\mathbb{R}} \left[S_n^{(\alpha)}(x)\, T_n^{(\beta)}(x) + S_n^{(\beta)}(x)\, T_n^{(\alpha)}(x) \right] \Psi(x)\, dx, \quad (4.1)$$

$$(S \odot_d T)_n^{(\alpha,\beta)}(\Psi) \equiv \int_{\mathbb{R}} S_n^{(\alpha)}(x)\, T_n^{(\beta)}(x)\, \Psi(x)\, dx, \quad (4.2)$$

$$(S \odot_{ex} T)_n^{(\alpha,\beta)}(\Psi) \equiv \int_{\mathbb{R}} S_n^{(\beta)}(x)\, T_n^{(\alpha)}(x)\, \Psi(x)\, dx. \quad (4.3)$$

The suffix d stands for *direct*, while ex stands for *exchange*, and they have a physical genesis. It is clear that if $S \equiv T$ then the three integrals above coincide: $(S \odot S)_n^{(\alpha,\beta)}(\Psi) = (S \odot_d S)_n^{(\alpha,\beta)}(\Psi) = (S \odot_{ex} S)_n^{(\alpha,\beta)}(\Psi)$, for all α, β, Ψ, n. In general, however, they are different and we will discuss an example in which they really produce different results when $n \to \infty$. We say that the distributions S and T are $\odot$-multiplicable, and we indicate with $(S \odot T)_{(\alpha,\beta)}$ their product, if the following limit exists for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$:

$$(S \odot T)_{(\alpha,\beta)}(\Psi) = \lim_{n,\infty} (S \odot T)_n^{(\alpha,\beta)}(\Psi). \quad (4.4)$$

Analogously we put, when they exist,

$$(S \odot_d T)_{(\alpha,\beta)}(\Psi) = \lim_{n,\infty} (S \odot_d T)_n^{(\alpha,\beta)}(\Psi). \tag{4.5}$$

and

$$(S \odot_{ex} T)_{(\alpha,\beta)}(\Psi) = \lim_{n,\infty} (S \odot_{ex} T)_n^{(\alpha,\beta)}(\Psi). \tag{4.6}$$

Whenever they exist we have $(S \odot S)_{(\alpha,\beta)}(\Psi) = (S \odot S)_{(\beta,\alpha)}(\Psi) = (S \odot_d S)_{(\alpha,\beta)}(\Psi) = (S \odot_{ex} S)_{(\alpha,\beta)}(\Psi)$, for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$ while they are different, in general, if $S \neq T$. Of course, the existence of these limits in general will depend on the values of α and β and on the particular choice of $\Phi(x)$. For this reason, as in [1, 2], we are really defining a *class of multiplications* of distributions more than a single one.

We have:

1. if $S(x)$ and $T(x)$ are continuous functions with compact support then

$$(S \odot T)_{(\alpha,\beta)}(\Psi) = (S \odot_d T)_{(\alpha,\beta)}(\Psi) =$$
$$= (S \odot_{ex} T)_{(\alpha,\beta)}(\Psi) = \int_{\mathbb{R}} S(x)\, T(x)\, \Psi(x)\, dx, \tag{4.7}$$

for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$ and for all $\alpha, \beta > 0$.

2. for all fixed n, for all $\alpha, \beta > 0$ and for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$ we have

$$(S \odot_d T)_n^{(\alpha,\beta)}(\Psi) = (S \odot_{ex} T)_n^{(\beta,\alpha)}(\Psi) \tag{4.8}$$

3. for all n, $\alpha, \beta > 0$ and for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$ we find

$$(S \odot_d T)_n^{(\alpha,\beta)}(\Psi) = (T \odot_d S)_n^{(\beta,\alpha)}(\Psi), \text{ and} \tag{4.9}$$
$$(S \odot_{ex} T)_n^{(\alpha,\beta)}(\Psi) = (T \odot_{ex} S)_n^{(\beta,\alpha)}(\Psi) \tag{4.10}$$

4. given $S, T \in \mathcal{D}'(\mathbb{R})$ such that the following quantities exist we get

$$(S \odot T)_{(\alpha,\beta)}(\Psi) = \frac{1}{2}\left\{(S \odot_d T)_{(\alpha,\beta)}(\Psi) + (S \odot_{ex} T)_{(\alpha,\beta)}(\Psi)\right\} \tag{4.11}$$

We will now discuss few examples of these definitions, beginning with the maybe most relevant one for concrete applications, i.e. with the multiplication of two delta functions.

Example nr.1: $(\delta \odot \delta)_{(\alpha,\beta)}$

First of all we remind that, in this case, the $\odot$, $\odot_d$ and $\odot_{ex}$ multiplications all coincide. If the following limit exists for some $\alpha, \beta > 0$, we have

$$(\delta \odot \delta)_{(\alpha,\beta)}(\Psi) = (\delta \odot \delta)_{(\beta,\alpha)}(\Psi) = \lim_{n,\infty} \int_{\mathbb{R}} \delta_n^{(\alpha)}(x)\, \delta_n^{(\beta)}(x)\, \Psi(x)\, dx =$$

$$= \lim_{n,\infty} n^{\alpha+\beta} \int_{\mathbb{R}} \Phi(n^\alpha x)\, \Phi(n^\beta x)\ \Psi(x)\, dx,$$

for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$. It is an easy exercise to check that this limit does not exist, if $\Psi(0) \neq 0$, if $\alpha = \beta$. Therefore we consider, first of all, the case $\alpha > \beta$. In this case we can write

$$(\delta \odot \delta)_n^{(\alpha,\beta)}(\Psi) = n^\beta \int_{-1}^{1} \Phi(x)\, \Phi(xn^{\beta-\alpha})\, \Psi(xn^{-\alpha})\, dx,$$

and the existence of its limit can be proved using the LDCT. Indeed, if we choose $\Phi(x)$ as in (3.4) and we define $B_m = \frac{1}{eN_m} \int_{-1}^{1} x^m\, \Phi(x)\, dx$, which is surely well defined and strictly positive for all fixed even m, it is easy to deduce that

$$(\delta \odot \delta)_{(\alpha,\beta)}(\Psi) = \begin{cases} B_m \Psi(0) = B_m\, \delta(\Psi), & \alpha = \beta\left(1 + \frac{1}{m}\right) \\ 0, & \alpha > \beta\left(1 + \frac{1}{m}\right). \end{cases} \tag{4.12}$$

The proof of this claim goes as follows: let us introduce the function $f_n^{(\alpha,\beta)}(x) = n^\beta\, \Phi(x)\, \Phi(xn^{\beta-\alpha})\, \Psi(xn^{-\alpha})$, with $x \in [-1,1]$. The first remark is that the role of α and β is crucial here (and everywhere in this kind of computations): the case $\alpha = \beta$ corresponds to what discussed in [8] in the attempt of the author to define the square of $\delta(x)$. As it is discussed there, this is not a good choice: $\delta^2(x)$ cannot be defined in this way and the reason is essentially the one we have discussed in Section 2.1 where we have taken $\alpha = \beta = 1$. This is in agreement with what we have stated above, that is that $(\delta \odot \delta)_{(\alpha,\alpha)}(\Psi)$ does not exist if $\Psi(0) \neq 0$. However, since here α and β do not need to coincide, we can try to see if something different happens. And in fact this is the case! We start taking here $\alpha > \beta > 0$. Since $\Phi(x)$ is defined as in (3.4), we deduce that

$$f_n^{(\alpha,\beta)}(x) = \frac{x^m}{N_m} n^{\beta(1+m)-\alpha m} \Phi(x) \exp\left\{\frac{1}{((xn^{\beta-\alpha})^2 - 1)}\right\} \Psi\left(\frac{x}{n^\alpha}\right)$$

Of course, if $\beta(1+m) - \alpha m \leq 0$ then $n^{\beta(1+m)-\alpha m} \leq 1$ for all $n \geq 1$. Moreover, for each $x \in]-1,1[$, for each even $m \in \mathbb{N}$ and for $\beta > \alpha$ we deduce that

$|x|^m \le 1$ and $\exp\left\{\frac{1}{((xn^{\beta-\alpha})^2-1)}\right\} \le \frac{1}{e}$. Also, since $\Psi(x)$ is continuous on a compact set, it surely admits maximum: $M_\Psi := \max_{x\in\mathbb{R}}|\Psi(x)|$. Therefore, at least if $\alpha > \beta$ and $\beta(1+m) - \alpha m \le 0$, we find that

$$\left|f_n^{(\alpha,\beta)}(x)\right| \le \frac{\Phi(x)M_\Psi}{eN_m} =: \varphi(x)$$

and is clear that $\varphi(x)$ is integrable in $[-1,1]$. To apply the LDCT we also have to find the limit of the sequence $f_n^{(\alpha,\beta)}(x)$, which must be convergent a.e. in $[-1,1]$ to a certain function $f^{(\alpha,\beta)}(x)$. Suppose therefore that $\beta(1+m) = \alpha m$ which, incidentally, automatically implies $\alpha > \beta$. With this choice of α and β we find that

$$\lim_{n,\infty} f_n^{(\alpha,\beta)}(x) = \frac{x^m\Phi(x)\Psi(0)}{N_m e} =: f(x),$$

pointwise in $[-1,1]$. If, on the other way, $\beta(1+m) < \alpha m$, then $\lim_{n,\infty} f_n^{(\alpha,\beta)}(x) = 0$ in $[-1,1]$. This implies that, in both cases, the LDCT can be applied and, as a consequence, that (4.12) holds true.

As we can see, this result is quite close to the one in (3.5) but for some minor differences here and there. Analogously to what already stressed before, formula (4.12) does not imply that $(\delta\odot\delta)_{(\alpha,\beta)}(\Psi)$ does not exist if $\alpha < \beta\left(1+\frac{1}{m}\right)$ because using the LDCT we only find sufficient conditions for $(\delta\odot\delta)_{(\alpha,\beta)}(\Psi)$ to exist. Here we can say a bit more, because we now know that $(\delta\odot\delta)_{(\alpha,\beta)}(\Psi) = (\delta\odot\delta)_{(\beta,\alpha)}(\Psi)$, which was not true in general for the multiplication $\otimes_{(\alpha,\beta)}$ introduced in [1]. Therefore we find that

$$(\delta\odot\delta)_{(\alpha,\beta)}(\Psi) = \begin{cases} B_m\,\delta(\Psi), & \alpha = \beta\left(1+\frac{1}{m}\right)^{-1}, \text{ or } \alpha = \beta\left(1+\frac{1}{m}\right) \\ 0, & \alpha < \beta\left(1+\frac{1}{m}\right)^{-1} \text{ or } \alpha > \beta\left(1+\frac{1}{m}\right), \end{cases} \tag{4.13}$$

while nothing can be said in general if $\alpha \in \left]\beta\left(1+\frac{1}{m}\right)^{-1}, \beta\left(1+\frac{1}{m}\right)\right[$.

Example nr.2: $(\delta\odot\delta')_{(\alpha,\beta)}$

In this case the three multiplications $\odot$, $\odot_d$ and $\odot_{ex}$ do not need to coincide, and, as we will see, they do not.

First of all we concentrate on the computation of $(\delta\odot_d\delta')_{(\alpha,\beta)}$. Because of (4.8) this will also produce $(\delta\odot_{ex}\delta')_{(\beta,\alpha)}$. We have

$$(\delta\odot_d\delta')_n^{(\alpha,\beta)}(\Psi) = n^{\alpha+2\beta}\int_{\mathbb{R}}\Phi(n^\alpha x)\,\Phi'(n^\beta x)\,\Psi(x)\,dx,$$

where Φ' is the derivative of Φ. Again, it is easy to check that the limit of this sequence does not exist, if $\alpha = \beta$, for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$ but only for those $\Psi(x)$ which go to zero fast enough when $x \to 0$. Of course, this is not what we want. Let us then consider $(\delta \odot_d \delta')_n^{(\alpha,\beta)}(\Psi)$ for $\alpha > \beta$. In this case we can write

$$(\delta \odot_d \delta)_n^{(\alpha,\beta)}(\Psi) = n^{2\beta} \int_{-1}^{1} \Phi(x)\, \Phi'(xn^{\beta-\alpha})\, \Psi(xn^{-\alpha})\, dx,$$

and by the LDCT we deduce that

$$(\delta \odot_d \delta')_{(\alpha,\beta)}(\Psi) = \begin{cases} K_m\, \delta(\Psi), & \alpha = \beta\, \frac{m+1}{m-1} \\ 0, & \alpha > \beta\, \frac{m+1}{m-1}, \end{cases} \tag{4.14}$$

where $K_m = \frac{m}{N_m e} \int_{-1}^{1} x^{m-1}\, \Phi(x)\, dx$.

The situation is completely different for the $(\delta \odot \delta')_{(\beta,\alpha)}(\Psi)$. Indeed, it is not difficult to understand that the LDCT cannot be used to prove its existence. The reason is quite general and is the following:

suppose that for two distributions T and S in $\mathcal{D}'(\mathbb{R})$ $(T \odot_d S)_{(\alpha,\beta)}$ exists for α and β such that $\alpha > \gamma\beta$, where γ is some constant larger than 1 (which appears because of the LDCT). For instance $\gamma = \frac{m+1}{m-1}$ in this example, while in Example nr.1 we had $\gamma = 1 + \frac{1}{m}$. Therefore, since if $(S \odot_{ex} T)_{(\beta,\alpha)}$ and $(S \odot_d T)_{(\alpha,\beta)}$ both exist then they coincide, using (4.11) we have

$$(S \odot T)_{(\alpha,\beta)}(\Psi) = \frac{1}{2}\left\{(S \odot_d T)_{(\alpha,\beta)}(\Psi) + (S \odot_d T)_{(\beta,\alpha)}(\Psi)\right\} \tag{4.15}$$

Of course, $(S \odot T)_{(\alpha,\beta)}(\Psi)$ exists if $(S \odot_d T)_{(\alpha,\beta)}(\Psi)$ and $(S \odot_d T)_{(\beta,\alpha)}(\Psi)$ both exist, which in turn implies that $\alpha > \gamma\beta$ and, at the same time, that $\beta > \gamma\alpha$. These are clearly incompatible. Therefore in order to check whether $(S \odot T)_{(\alpha,\beta)}(\Psi)$ exists or not it is impossible to use the LDCT: some different strategy should be considered.

Example nr.3: $(\delta' \odot \delta')_{(\alpha,\beta)}$

As for Example nr.1 we remark that also now the $\odot$, $\odot_d$ and $\odot_{ex}$ multiplications coincide. If the following limit exists for some $\alpha, \beta > 0$, we have

$$(\delta' \odot \delta')_{(\alpha,\beta)}(\Psi) = (\delta' \odot \delta')_{(\beta,\alpha)}(\Psi) = \lim_{n,\infty} \int_{\mathbb{R}} \delta_n'^{(\alpha)}(x)\, \delta_n'^{(\beta)}(x)\, \Psi(x)\, dx =$$

$$= \lim_{n,\infty} n^{2\alpha+2\beta} \int_{\mathbb{R}} \Phi'(n^\alpha x)\, \Phi'(n^\beta x)\, \Psi(x)\, dx,$$

for all $\Psi(x) \in \mathcal{D}(\mathbb{R})$. As before, it is quite easy to check that this limit does not exist, if $\Psi(0) \neq 0$, when $\alpha = \beta$. Therefore we start considering the case $\alpha > \beta$. In this case we can write

$$(\delta' \odot \delta')_n^{(\alpha,\beta)}(\Psi) = n^{\alpha+2\beta} \int_{-1}^{1} \Phi'(x)\, \Phi'(xn^{\beta-\alpha})\, \Psi(xn^{-\alpha})\, dx,$$

and again the existence of its limit can be proved using the LDCT. Choosing $\Phi(x)$ as in (3.4) and defining $\tilde{B}_m = \frac{m}{\epsilon N_m} \int_{-1}^{1} x^{m-1}\, \Phi'(x)\, dx$, which surely exists for all fixed even m, we deduce that

$$(\delta' \odot \delta')_{(\alpha,\beta)}(\Psi) = \begin{cases} \tilde{B}_m\, \delta(\Psi), & \alpha = \beta\, \frac{m+1}{m-2} \\ 0, & \alpha > \beta\, \frac{m+1}{m-2}. \end{cases} \tag{4.16}$$

However this is not the end of the story, because we still can use the symmetry $(\delta' \odot \delta')_{(\alpha,\beta)}(\Psi) = (\delta' \odot \delta')_{(\beta,\alpha)}(\Psi)$ mentioned before. We find

$$(\delta' \odot \delta')_{(\alpha,\beta)}(\Psi) = \begin{cases} \tilde{B}_m\, \delta(\Psi), & \alpha = \beta\, \frac{m-2}{m+1}, \text{ or } \alpha = \beta\, \frac{m+1}{m-2} \\ 0, & \alpha < \beta\, \frac{m-2}{m+1}, \text{ or } \alpha > \beta\, \frac{m+1}{m-2}, \end{cases} \tag{4.17}$$

while nothing can be said in general if $\alpha \in \left]\beta\, \frac{m-2}{m+1}, \beta\, \beta\, \frac{m+1}{m-2}\right[$. Needless to say, here we have to restrict the computation to the following even values of m: $m = 4, 6, 8, \ldots$.

5 More spatial dimensions

While the definition of the multiplication given in [1], as we have stressed before, cannot be extended easily to $\mathbb{R}^d$, $d > 1$, definitions (4.4), (4.5) or (4.6) admit a natural extension to any spatial dimensions. We concentrate here only on the symmetric definition, (4.4), since it is the relevant one for multiplying two deltas.

The starting point is again a given non negative function $\Phi(\underline{x}) \in \mathcal{D}(\mathbb{R}^d)$ with support in $I_1 := [-1, 1]\times\cdots\times[-1, 1]$, and such that $\int_{I_1} \Phi(\underline{x})\, d\underline{x} = 1$. In this case the delta-sequence is $\delta_n^{(\alpha)}(\underline{x}) = n^{d\alpha}\Phi(n^\alpha \underline{x})$, for any choice of $\alpha > 0$. The same results stated in the previous section again hold in this more general situation. For instance, if $T \in \mathcal{D}'(\mathbb{R}^d)$ then $\{T_n^{(\alpha)}(\underline{x}) = (T*\delta_n^{(\alpha)})(\underline{x})\}$ is a sequence of C^∞ functions and it converges to T in $\mathcal{D}'(\mathbb{R}^d)$ as n diverges, independently of $\alpha > 0$.

Therefore, let T, S be two distributions in $\mathcal{D}'(\mathbb{R}^d)$, and let us consider their convolutions $T_n^{(\alpha)}(\underline{x}) = (T * \delta_n^{(\alpha)})(\underline{x})$ and $S_n^{(\beta)}(\underline{x}) = (S * \delta_n^{(\beta)})(\underline{x})$ with

$\delta_n^{(\alpha)}(\underline{x}) = n^{d\alpha}\Phi(n^\alpha \underline{x})$, for $\alpha, \beta > 0$. As usual, $T_n^{(\alpha)}(\underline{x})$ and $S_n^{(\beta)}(\underline{x})$ are both C^∞ functions, so that the following integral surely exists:

$$(S \odot T)_n^{(\alpha,\beta)}(\Psi) \equiv \frac{1}{2}\int_{\mathbb{R}^d}\left[S_n^{(\alpha)}(\underline{x})\, T_n^{(\beta)}(\underline{x}) + S_n^{(\beta)}(\underline{x})\, T_n^{(\alpha)}(\underline{x})\right]\,\Psi(\underline{x})\,d\underline{x}, \tag{5.1}$$

$\forall\,\Psi(\underline{x}) \in \mathcal{D}(\mathbb{R}^d)$. As before the two distributions S and T are $\odot$-multiplicable if the following limit exists independently of $\Psi(\underline{x}) \in \mathcal{D}(\mathbb{R}^d)$:

$$(S \odot T)_{(\alpha,\beta)}(\Psi) = \lim_{n,\infty}(S \odot T)_n^{(\alpha,\beta)}(\Psi). \tag{5.2}$$

We consider in the following the $\odot$-multiplication of two delta functions, considering two different choices for the function $\Phi(\underline{x})$ both extending the one-dimensional choice in (3.4).

The starting point of our computation is the usual one: if it exists, $(\delta \odot \delta)_{(\alpha,\beta)}(\Psi)$ must be such that

$$(\delta \odot \delta)_{(\alpha,\beta)}(\Psi) = \lim_{n,\infty}\, n^{d\alpha+d\beta}\int_{\mathbb{R}^d}\Phi(n^\alpha\,\underline{x})\,\Phi(n^\beta\,\underline{x})\;\Psi(\underline{x})\,d\underline{x}.$$

Again, this limit does not exist if $\alpha = \beta$, but for very peculiar functions $\Psi(\underline{x})$. If we consider what happens for $\alpha > \beta$ then the limit exists under certain extra conditions.

For instance, if we take $\Phi(\underline{x}) = \prod_{j=1}^d \Phi(x_j)$, where $\Phi(x_j)$ is the one in (3.4), the computation factorizes and the final result, considering also the symmetry of the multiplication, is a simple extension of the one in (4.13):

$$(\delta \odot \delta)_{(\alpha,\beta)}(\Psi) = \begin{cases} B_m^d\,\delta(\Psi), & \alpha = \beta\left(1+\frac{1}{m}\right)^{-1}, \text{ or } \alpha = \beta\left(1+\frac{1}{m}\right) \\ 0, & \alpha < \beta\left(1+\frac{1}{m}\right)^{-1} \text{ or } \alpha > \beta\left(1+\frac{1}{m}\right), \end{cases} \tag{5.3}$$

A different choice of $\Phi(\underline{x})$, again related to the one in (3.4), is the following:

$$\Phi(\underline{x}) = \begin{cases} \frac{\|\underline{x}\|^m}{N'_m}\exp\{\frac{1}{\|\underline{x}\|^2-1}\}, & \|\underline{x}\| < 1 \\ 0, & \|\underline{x}\| \geq 1, \end{cases} \tag{5.4}$$

where N'_m is a normalization constant and $\|\underline{x}\| = \sqrt{x_1^2 + \cdots x_d^2}$. With this choice, defining $C_m = \frac{1}{N'_m e}\int_{\mathbb{R}}\|\underline{x}\|\,\Phi(\underline{x})\,d\underline{x}$ and using the symmetry property of $\odot_{(\alpha,\beta)}$, we find

$$(\delta \odot \delta)_{(\alpha,\beta)}(\Psi) = \begin{cases} C_m\,\delta(\Psi), & \alpha = \beta\left(1+\frac{d}{m}\right)^{-1}, \text{ or } \alpha = \beta\left(1+\frac{d}{m}\right) \\ 0, & \alpha < \beta\left(1+\frac{d}{m}\right)^{-1} \text{ or } \alpha > \beta\left(1+\frac{d}{m}\right). \end{cases} \tag{5.5}$$

We see that also with this choice of $\Phi(\underline{x})$ the limit defining the product of two delta functions can be defined, at least under certain conditions. The main difference between the above results is, but for the value of the constants, that d explicitly appears in the result in (5.3), while it appears only in the conditions on α and β in (5.5).

We have shown here that using twice the sequential method in the attempt of multiplying distributions is convenient because it allows a natural extension of the multiplication to any spatial dimensions. We hope that our new definition can be used in the analysis of the Klein-Gordon model, which was not *cured* by our *old* definition. Moreover we are planning a collaboration with a group of engineers to analyze some three-dimensional engineering structures, trying to extend the results in [5, 6].

Acknowledgments

This work has been partially supported by M.U.R.S.T. The author wishes to thank Dr. Andrzej Borys for his kind invitation to write this pedagogical review paper.

Appendix: $\delta(x)$ is not a function!

We prove here that no locally integrable positive function exists which satisfies (2.1).

Suppose this is not so and that such a function, $t(x)$, does exists. Therefore, for instance, we must have $\int_{\mathbb{R}} t(x)\,\psi(x)\,dx = \psi(0) = 0$ if $\psi(x) = x^2e^{-x^2}$. Since $\psi(x)$ is even, strictly positive if $x \neq 0$ and it acquires its maximum values for $x = \pm 1$, we can have $\int_{\mathbb{R}} t(x)\,\psi(x)\,dx = 0$ only if the function $t(x)\,\psi(x)$ is equal to zero almost everywhere in $\mathbb{R}$, which is only possible if $t(x)$ is by itself equal to zero almost everywhere in $\mathbb{R}$. Since by assumption $t(x)$ belongs to the set

$$\mathcal{L}^1_{loc}(\mathbb{R}) = \Big\{h(x) \text{ L-measurable such that} \int_\chi |h(x)|\,dx < \infty$$

$$\forall \chi \subset \mathbb{R} : mis(\chi) < \infty\Big\},$$

then $t(x)$ is Lebesgue-equivalent to the function $T(x) = 0$, $\forall x \in \mathbb{R}$. Let now take $\varphi(x) = e^{-x^2}$. Then we can conclude that

$$\int_{\mathbb{R}} t(x)\,\varphi(x)\,dx = \int_{\mathbb{R}} T(x)\,\varphi(x)\,dx = 0.$$

But, on the other hand, we should also have $\int_{\mathbb{R}} t(x)\,\varphi(x)\,dx = \varphi(0) = 1$, and these are incompatible.

References

[1] F. Bagarello, *Multiplication of Distributions in One Dimension: Possible Approaches and Applications to δ-function and its Derivatives*, J. Math. Anal. Appl., **196**, 885-901 (1995)

[2] F.Bagarello, *Multiplication of Distributions in one dimension and a First Application to Quantum Field Theory*, J. Math. Anal. Appl., **266**, 298-320, (2002)

[3] F.Bagarello, *Multiplication of distributions in any dimension: applications to δ-function and its derivatives*, J. Math. Anal. Appl., in press

[4] F. Bagarello, *Fisica Matematica per la Matematica, la Fisica e l'Ingegneria*, Zanichelli (2007)

[5] B. Biondi, S. Caddemi, *Closed form solutions of Euler-Bernoulli beams with singularities*, Int. J. Sol. and Struct., **42**, (2005) 3027-3044

[6] B. Biondi, S. Caddemi, *Euler-Bernoulli beams with multiple singularities in the flexural stiffness*, Europ. J. of Mech., in press

[7] H.J.Bremermann, and L. Durand III, J. Math. Phys. **2** (1961), 240-257.

[8] J.F.Colombeau, *New Generalized Functions and Multiplication of Distributions* North-Holland, Amsterdam (1984).

[9] M. Reed e B. Simon, *Methods of modern mathematical physics*, vol. I: Functional Analysis, Academic Press, New York and London

[10] W. Rudin, *Real and complex analysis*, McGraw-Hill (1970)

PODEJŚCIE DYDAKTYCZNE DO PROBLEMU MNOŻENIA DYSTRYBUCJI W PRZESTRZENIACH O DOWOLNYM WYMIARZE

Streszczenie

W artykule przedstawiono podejście dydaktyczne do zagadnienia mnożenia dystrybucji. Na początku pokazano na poziomie podstawowym, jakie problemy pojawiają się, gdy mnożone są dystrybucje. Następnie wprowadzono zaawansowane narzędzia matematyczne do analizy problemu i podano szereg użytecznych definicji mnożenia dystrybucji, które mogą mieć zastosowanie w naukach inżynieryjnych. W szczególności przeanalizowano dogłębnie możliwości jednej z zaproponowanych definicji dla mnożenia dystrybucji i jej pochodnych w przestrzeniach o dowolnym wymiarze.

Słowa kluczowe: dystrybucje, funkcje delta, analiza funkcjonalna

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 249
TELEKOMUNIKACJA I ELEKTRONIKA 10 (2007), 29-39

VOIP AS A PRECURSOR OF NEXT GENERATION NETWORKS

Tadeus Uhl

Flensburg University of Applied Sciences
Institute of Communications Technology
Kanzleistr. 91-93, D 24943 Flensburg
tadeus.uhl@fh-flensburg.de

The aim of this paper is to sketch out the concepts, standards and properties associated with the new service Voice over IP (VoIP). The paper also includes a comparison of Internet telephony and Public Switched Telephone Network (PSTN) telephony. This comparison throws light on areas in which there is still a great deal of work to be done. The paper will then assess whether VoIP can genuinely be considered to be a precursor of Next Generation Networks (NGNs). For a more practical discussion of this issue, the definition of an NGN will be used as a starting-point. A commentary of the unsolved problems arising from this discussion then follows. The paper concludes with a summary.

Keywords: communication networks, communication services, communication protocols, network architecture, multimedia applications, IP, VoIP, NGN

1. INTRODUCTION

Telecommunications technology has developed enormously in the last hundred years or so. Among the basic inventions that were to become the main driving forces behind this development are: pulse code modulation (1939; Reeves), semiconductor technology (1947; Bardeen, Brattain and Shockley), laser diode (1962; Hall) and optical fibre (1966; Kao and Hockham). Thanks to these inventions, digital transmission (1960), then digital switching (1965) and finally optical transmission (1973) became a reality, and analog communications networks were replaced by digital networks.

The networks of the past and even some present-day networks belong to the group of networks known as dedicated networks (each service requiring its own network). This changed towards the end of the nineties, however, with the introduction of ISDN (Integrated Services Digital Networks) and a new brand of networks arose known as integrated networks, the word integrated referring to the multiple services that these networks support. This process of integration has continued to this day: modern GSM, UMTS and TV networks, and the World Wide Web are all integrated service networks.

The services themselves that the networks transmit have also undergone an equally rapid development, especially in the last twenty years. The variety of services offered today is very wide indeed. The most recent developments include high-speed data transfer, VoIP, SMS, MMS, home banking, WWW, audio/video on demand, teleconferencing, messaging and, most recently, an increasing trend towards IPTV. Many of these applications have become very popular indeed. Parallel to the development of these new services, we have witnessed the recent appearance of innovative systems designed to exploit the impressive multimedia capabilities of the IP platform. A good example of such systems are the so-called e-learning systems, that are widely used in education. In the near future, we can definitely expect a variety of other innovative services and high-capacity communications systems to flood the market.

One new technology that has evolved rapidly in the last seven years is Voice over IP (VoIP). VoIP requires an IP transport platform. Several standards (e.g. H.323 [7], SIP [14]) have been published in the last few years, allowing this service to become a reality. There are several VoIP systems on the market, (Siemens' HiPath 5000 and OpenScape, Innovaphone PBX from TLK Computer GmbH, SIPSTAR IP-PBX from Nero AG, OmniPCX Enterprise from Alcatel AG, Cisco CallManager from Cisco Systems GmbH, ASTERISK from DIGIUM, etc.) that work in accordance with the standards named above. VoIP is becoming incredibly popular. There are already about half a million subscribers in Germany alone that have opted for this Internet-based form of telephony. The technology is particularly popular among subscribers who frequently make international calls. Among other things, this paper will show what opportunities this innovative, multimedia application offers but also where its limitations lie.

2. ARCHITECTURES FOR VOIP

The architectures for VoIP in fixed-line networks and mobile communications networks have already been decided upon. Here, first, a presentation of the architecture for VoIP in fixed-line networks (Fig. 1).

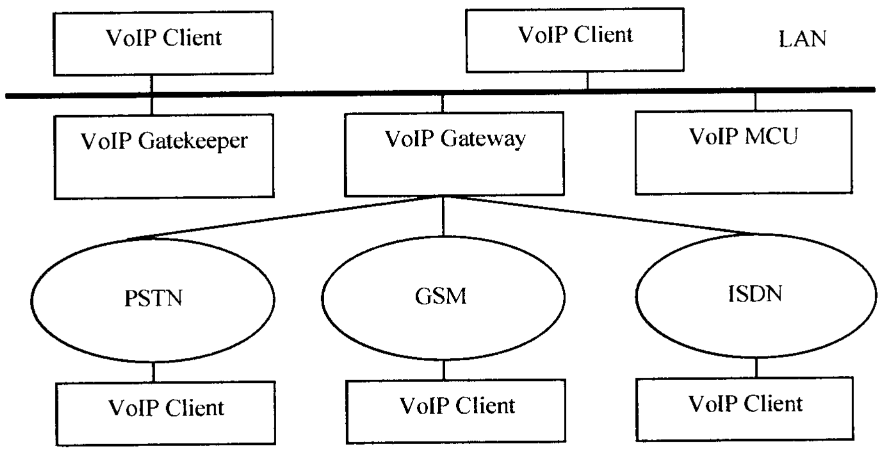


Fig. 1. Architecture for VoIP in Fixed-line Networks
Rys. 1. Architektura VoIP dla sieci stacjonarnych

Fig. 1 shows clearly the four main constituents of the VoIP architecture: terminal, gatekeeper, gateway and multipoint control unit. The terminal comprises the user terminal equipment and supports voice, video and data communication. The gatekeeper provides the terminals and gateways with services for authentication and call control. The gateway is the interface between heterogeneous networks such as the LAN and the ISDN net. The multipoint control unit is needed to support conferencing.

Fig. 2 represents the architecture for VoIP in mobile communications networks. The figure clearly shows the four main constituents: PoC client, PoC server, presence server and GLS. The PoC client is a user terminal and supports the PoC service (Push to Talk over Cellular – the first implementation of the VoIP service in mobile communication). The PoC server provides the PoC service on the basis of the protocol SIP and the GLMS controls the management and administration tasks of the mobile wireless network. Last but not least, the PoC clients gain access to the PoC service over the radio channels of the GSM/GPRS or the UMTS. It is also possible to access external PoC networks via a gateway.

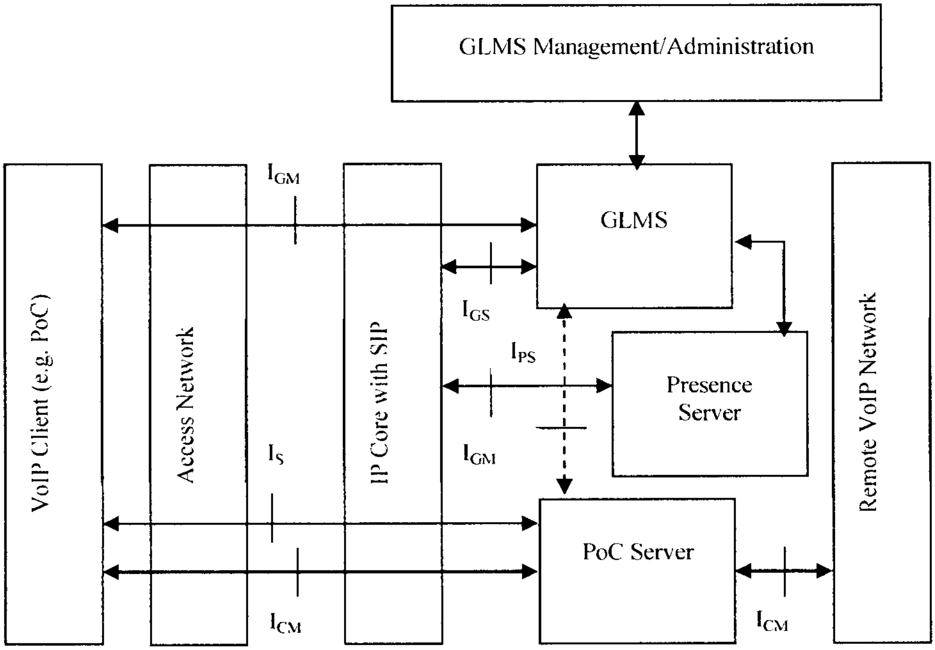


Fig. 2. Architecture for VoIP in Mobile Communications Networks
Rys. 2. Architektura VoIP dla sieci mobilnych

Two signaling protocols – a) H.323 from ITU-T and b) SIP from IETF – can be used in both architectures. These protocols are to be found in the application layer of the logical model of the Internet. They draw on the TCP an UDP protocols located within the transport layer and the IP protocol in the network layer. This will be dealt with in more detail later.

For the moment, let us turn our attention to the H.323 Standard. Fig. 3 shows where the H.323 standards for VoIP are located within the layer model. Signaling in H.323 is done using the three protocols H.225.0, H.225.0 RAS (used optionally) and H.445. There are two possible configurations of the signaling protocol: either with or without a gatekeeper. In the case of the more frequently used option using a gatekeeper, an H.323 connection can be broken down into five distinct phases [1, 18]:

– the signaling phase through H.225.0 RAS (client A calls client B via the gatekeeper).

- the connection control of the logical voice channels through H.245 (done directly between client A und client B),
- the conversation phase during which not only voice packets but also control packets are constantly exchanged,
- the release of the logical voice channels by H.245,
- the release of the bandwidth used by the gatekeeper and confirmation to the subscriber that the connection has been terminated.

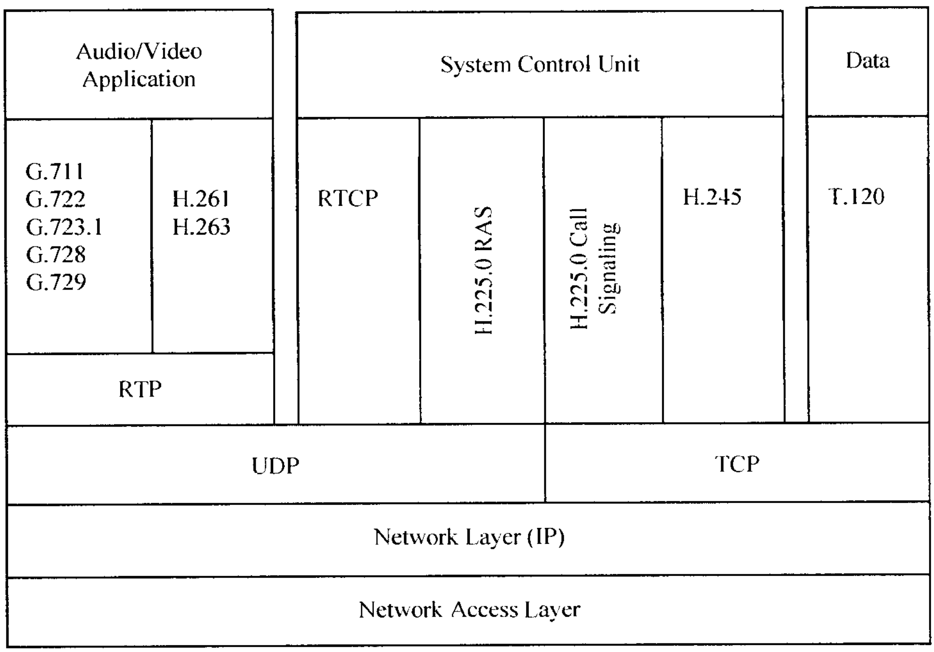


Fig. 3. Overview of the VoIP Protocol Suite according to H.323
Rys. 3. Stos protokolarny dla usługi VoIP zgodnie ze standardem H.323

Fig. 4 shows where the SIP standards are to be found within the layer model. It is immediately clear that there is only one signaling protocol. Two different configurations can be used for the establishment of an SIP connection: the redirect mode and the proxy mode. And the connection behaves differently as well, depending on which mode is chosen. In both cases, however, a so-called location server supports the SIP server with the addressing. When the more frequently used option – the proxy mode configuration – – is used, there are three phases to the SIP connection (in contrast to the five of H.323!) [1, 19]:

- the signaling phase used for the establishment of the connection (client A calls client B via the proxy server),
- the conversation phase during which not only voice packets but also control packets are constantly exchanged, (directly between client A and client B),
- the signaling phase during connection release (also done via the proxy server).

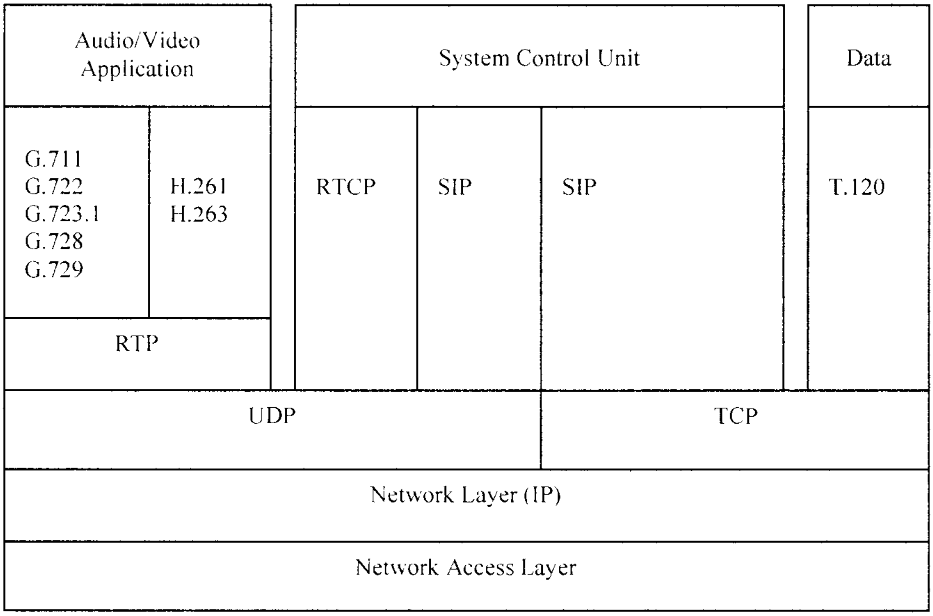


Fig. 4. Overview of the VoIP Protocol Suite with the SIP
Rys. 4. Stos protokolarny dla usługi VoIP z protokołem SIP

The following presents a comparison of the two signaling protocols described above (cf. Table 1).

Table 1. H.323 and SIP in comparison
Tabela 1. Porównanie protokołów H.323 i SIP

Property	H.323 (+/–)	SIP (+/–)
Coding	ASN.1-based. complicated (–)	Text-based, simple (+)
Handling	involved procedures in several phases (–)	quite straightforward procedures in few phases (+)
Name Lookup	well-suited to PSTN numbering (+)	problematic conversion to IP numbering (–)
Compactness	complex due to extensive program code (–/+)	relatively compact initially, but complexity increases steeply with increased ease of telephony (+/–)
Expandability	clear-cut, well-conceived method of expandability (+)	expandable in principle, but no management has been implemented as yet for any ensuing expansions (–)

It can be seen that there are advantages but also drawbacks to both protocols. It is difficult to choose a clear favorite for VoIP. And this is how it turns out in practice; several companies that design and build VoIP systems have decided to put their eggs in both baskets. Slowly, however, SIP seems to be getting the edge.

Both signaling protocols can be used to good effect on IP transport platforms, which, as is generally known, were not developed for hard real-time services. So the questions arise: Just how efficient are IP platforms, and can they be used for time-critical services such as VoIP? The following section is dedicated to these issues.

3. CAPACITY OF AN IP ENVIRONMENT

Fig. 5 shows the analysis environment used in the analyses. It consists of the real-time communications system HiPath 5000 [13], the measurement system DA-360 (incorporating a load generator) [3], and the analysis software application TraceView VoIP (implementing a PESQ algorithm) [2].

In order to conduct a quantitative and qualitative analysis of quality of speech, the DA-360 measurement system was added to the net to propagate 511-byte packets and simulate a varying network load. A series of speech tests was conducted for each of the varying net loads. Ten tests were considered to be an appropriate number for each series. The reference file used for the series of tests was the file Or105.wav in ITU-T [5]. The size of the user data packets transmitted was 480 byte and the jitter buffer was adjusted to 100 ms. The objective values gained for quality of speech using the analysis software TraceView VoIP were then compared with subjective assessments. Table 2 shows the results of the analyses.

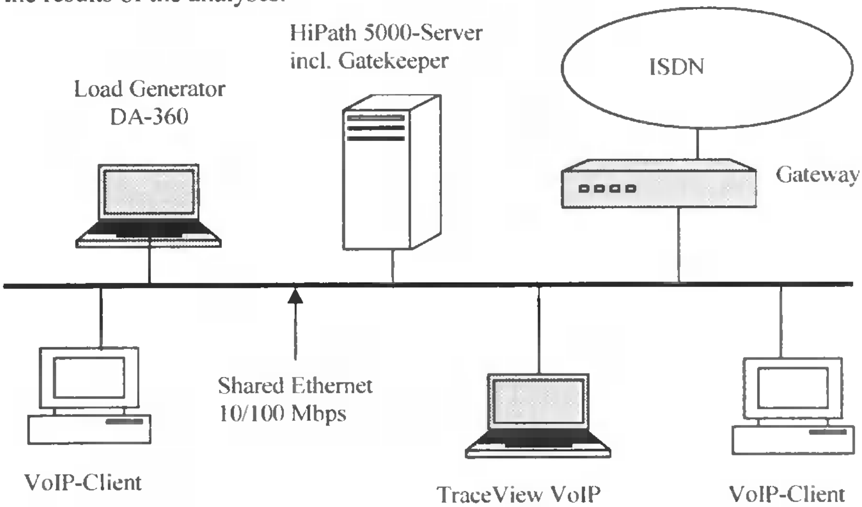


Fig. 5. Analysis environment
Rys. 5. Analizowany LAN system

With respect to the results of the analyses, it must be said that speech quality remains excellent until net load reaches 85 percent. It is also noticeable that as the net load increases, the average net delay also increases. The increase in jitter delay (not shown in Table 2) was countered by the size of the jitter buffer. It becomes clear that the objective measurement of speech quality coincides to a large extent with the subjective assessment of speech quality.

Table 2. Analysis results alongside PESQ values
Tabela 2. Wyniki analizy w wartościach PESQ

Generator load [%]	PESQ value (mean)	Subjective evaluation	Average delay [ms]
0	4.402	excellent	160.4
50	4.402	excellent	167.2
75	4.402	excellent	167.6
85	4.402	excellent	168.4
95	2.746	poor	220.5

Fig. 6 shows the results gained from qualitative and quantitative analyses using classical performance values (communication patter and loss probability). In this experiment packet size, too, was varied during the production of network load.

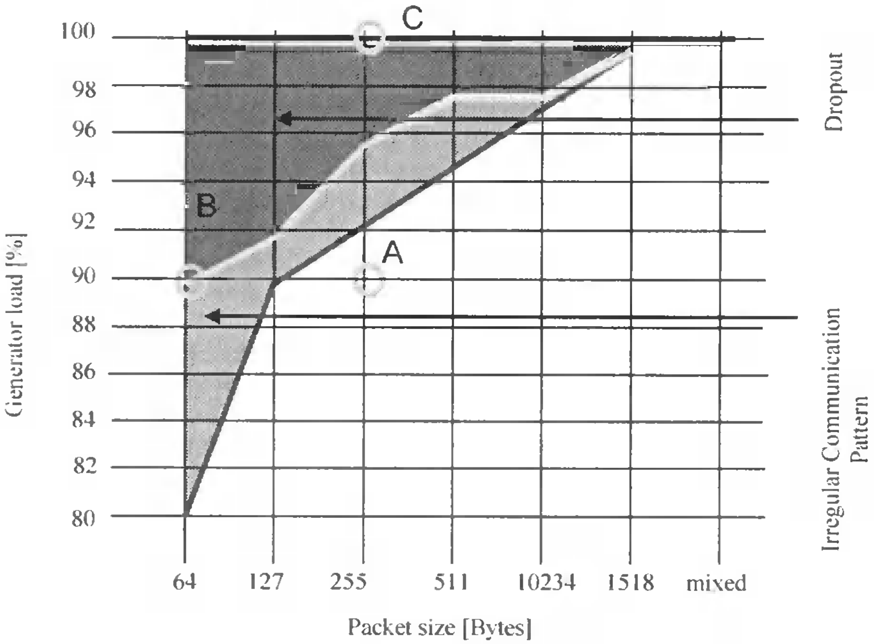


Fig. 6. Measurement Results using Classic Parameters
Rys. 6. Wyniki pomiarowe klasycznych parametrów sieciowych

When it came to evaluating the quality of speech, three areas were picked out for individual consideration: area A (best quality possible), Area B (disturbed communication) and area C (communication impossible). It is to be noted that the quality of speech is influenced only when net load reaches 90 per cent (using the smallest packets admissible in Ethernet). This area increases when load packet size is increased. In the grey area, it was possible to detect irregular communication patters by metrological means (i.e. objectively). Yet these had no noticeable effect on the subjective assessment of the quality of speech. The top, left-hand triangle contains dropouts that were reflected in a bad quality of speech and could also be detected, i.e. proved metro logically, through packet loss.

All in all, it must be recognized that the IP platform transports the hard real-time service VoIP astoundingly well and offers sufficient capacity even when network load is heavy. This has been confirmed by several studies published elsewhere (e.g. in switched networks with other VoIP systems such as OpenScape [20]).

4. A COMPARISON OF INTERNET TELEPHONY AND PSTN TELEPHONY

Subscribers of the PSTN (Public Switched Telephone Network) are used to the high quality of its voice service. Dedicated channels are used that assure a sufficient bandwidth. Maximum values of the network parameters for QoS (such as delay ≤ 150 ms according to ITU G.114, or packet loss ≤ 2%) have been laid down and are guaranteed by network providers. This is not the case with Internet telephony (a variant of VoIP). What is more: the transmission of voice packets on an IP platform does not take place over dedicated channels and is done according to the principle of "best-effort delivery". It is also to be noted that QoS in VoIP is greatly influenced by the codec's used that are frequently heterogeneous and incompatible with one another. So the lack of a guaranteed QoS in VoIP must be seen as a great drawback.

In PSTNs, the voice service comes with a range of handy features, e.g. call diversion, call hold, call park and pickup, call waiting, message waiting indication, name identification, call completion, call offer, and more. Besides allowing the subscriber to make a call Internet telephony offers very few other service features, and these vary from provider to provider. This is a further shortcoming of VoIP. To be fair, however, VoIP does offer quick and easy conferencing; so that is one thing in its favour.

But PSTNs also offer the so-called intelligent services such as emergency response service, universal number, green number service, etc. The regulatory authorities in many countries stipulate that PSTNs must maintain a number of services, not least of which is the emergency response service. Services like these are not yet possible in Internet telephony and this does not seem likely to change in the foreseeable future: yet another disadvantage of VoIP. At this point it seems fitting to mention the IETF specification known as SPIRITS (Service in the PSTN/IN Requesting Internet Service) [11]. Perhaps this will eventually solve the problem.

PSTNs are some of the most reliable communications systems. To prevent eavesdropping and misuse of information, only the network providers' technicians are permitted to make measurements in the network (except, that is, at the subscribers' terminal equipment) and so it is not easy to gain access to the transport platform. IP transport platforms are another matter entirely. Currently available protocol analysers (e.g. Ethereal) are capable of receiving the RTP packets [10] used in VoIP sessions and decoding their contents - a further weakness of Internet telephony. Firewalls can offer the companies that use VoIP a little security in this respect, but their use does sometimes make billing a tricky matter. In any event, the accounting systems must have access to the accounting data stored within the protected intranet, and that it where the problems arise.

Regulatory authorities demand that all PSTN subscribers be uniquely identifiable and traceable. In Internet telephony dsl connections are often used, and this means that IP addresses are allocated dynamically (by the DHCP [12]). What is more: they are forever being changed, and identifying the subscriber requires a great deal of effort and is only possible when services providers are prepared to divulge data. The terminal equipment (and that includes mobiles) may use different platforms to gain access to the VoIP service. And that being the case, it cannot be guaranteed that the subscriber can be traced. But this must be a basic requirement of a number of services, *especially* the emergency response service: yet another weak point of Internet telephony.

Telecommunication legislation also stipulates that authorised eavesdropping be possible on PSTNs. This is a considerable challenge to their providers. Internet telephony is currently marketed by several providers. It is frequently the case that the systems they use implement different communications protocols, coding methods, management systems, etc. In some systems (e.g. OpenScape) signaling is encrypted for instance. This all makes for a quite heterogeneous Internet telephony landscape. Neither do VoIP providers have systems to store connection data and conversations, so eavesdropping is impossible. This can be considered as both an advantage (from the subscribers' point of view) and a disadvantage (from the legislator's point of view).

As already mentioned above, compatibility between the various VoIP systems is currently practically non-existent. Bridging the networks is difficult, and the number of gateways that have to be implemented to interwork the networks is therefore considerable. This hampers any attempt to integrate the networks. Standardisation work (e.g. MEGACO Standard [9]) is one way to solve the problem, but it will be a long and arduous task – another reason for not switching to VoIP.

It can be said in conclusion that it will be a long time before Internet telephony has the properties and features that one has grown to expect from PSTN telephony. A considerable number of problems have to be solved as quickly as possible. This will continue to be a great challenge to designers and manufacturers of VoIP systems and networks. There is, however, one overriding advantage to VoIP: it effectively uses state-of-the-art technology in its transport system, i.e. packet-switched communication and thanks to the Internet, this technology is already available worldwide! This technology has also been designated as the prerequisite for Next Generation Networks [17]. So the question arises: Is VoIP a precursor of NGNs? The next section provides an answer.

5. IS VOIP A PRECURSOR OF NGN?

First of all a definition of NGN in the terms of ITU-T Y.2001 [6]:

"A Next Generation Network is a packet-based network able to provide telecommunication services and able to make use of multiple, broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It offers its users unrestricted access to different service providers. It supports generalized mobility, which will allow consistent and ubiquitous provision of service to users".

This definition describes several distinct characteristics of an NGN. They are realised in VoIP in the following ways:

1. Packet-based transfer. VoIP fulfils this requirement without question.
2. Separation of control functions. A complete separation of signalling data and user data transfer has not yet been achieved in VoIP; the two are handled together on an IP platform.
3. Broadband capability. VoIP uses a broadband platform. Unfortunately, bottlenecks occur all too often at the access points; so recent attempts have been made to introduce new access methods such as ADSL2, ADSL2+, VDSL with the aim to make innovative, broadband services like TV over IP a reality and so ensure that VoIP continues to be a trendsetter.
4. End-to-end quality of service. This is one of VoIP's greatest deficiencies. For the time being, there is no mechanism for reserving bandwidth in the IP platform or its controlling. There is a great deal of work to catch up on in this area of VoIP (as the-re is with all time-critical services being transported on this platform). A solution could be found in the near future with the introduction of the new Version 6 of the IP protocol (that allows prioritising of data transmission). As long as the current IP platform is used with IPV4, it is essential to use suitable methods for assessing QoS. And that in turn assumes the availability and use of effective measuring equipment.

5. Interworking with legacy networks. Thanks to the implementation of specialised gateways it is already possible to interwork VoIP, PSTN and mobile cellular networks. Much effort is being put into the development of new gateways that will be able to handle multiple signalling protocols and do this in a separate signalling layer (i.e. separate to the transport layer).
6. Generalised mobility. VoIP already fulfils this requirement thanks to recent advances in the areas of WLAN and mobile communication.
7. Security. Unfortunately, VoIP offers inadequate security of communicated data, as was pointed out in Section 4. One way to solve this problem would be, of course, by encrypting the data streams. But let us not forget that data encryption can only be achieved at the cost of a sharp increase in the capacity of any resources used.
8. Unrestricted access to different service providers. VoIP cannot fulfil this requirement either. Access is limited to particular service providers whose systems are often incompatible with one another (e.g. SIPGate [15] and Skype [16]). Only a clear separation of application layer and transport layer can provide a solution to this problem, and a great deal of work is being done in this direction. The first results are already to be seen in the form of the specification Parlay/OSA from ETSI [8] (in the Unified Modelling Language UML) or JAIN from Sun Microsystems [4] (in Java).
9. Compliancy with regulatory requirements. There is a great deal of work to catch up on in this area of VoIP, too.

All things considered, there is a lot to be said for the new service VoIP. It uses a modern, high-performance technology – packet switching. VoIP is one of the so-called multimedia services that are particularly important these days. The use of this service in many intranets and in the Internet has given both academics and practitioners the opportunity to discover the strengths and weaknesses of services like these in practical environments. This is of great importance when it comes to designing and developing the brand of new, universal communications systems that NGNs will be. And that is why the service VoIP can be considered as one of the first precursors of NGNs.

6. SUMMARY

In the course of this study, VoIP was considered in detail from both theoretical and practical points of view. First, current VoIP standards were discussed. Then concrete measurements of VoIP were made with the aim of assessing the suitability of the IP transport platform for the new service. The investigation showed that the IP platform is actually an extremely competent transport system. Internet telephony and PSTN telephony were then compared. The study shows that VoIP is currently a long way from matching the features of PSTN telephony. There is a lot of work to be done before these discrepancies are eliminated. It will be a huge challenge to system developers. The study concluded with an enquiry into whether VoIP can be considered as a precursor of NGNs. A comparison of the features of NGN with the properties of VoIP showed that, despite many provisos, this question can be answered in the affirmative. It became increasingly clear in the course of the study, however, that academics are faced with many interesting tasks that will have to be solved swiftly.

Any analysis of the communications market shows that the new service VoIP has actually become very popular; growth rates are increasing almost exponentially. The main reason for this is that since the service was introduced, telephoning has suddenly become very cheap. It is also quite easy to integrate the service in the terminal equipment with other services such as data services (Email, WWW, MMS, etc.). And the worldwide availability of the Internet simply cannot be overlooked. All these are reasons for VoIP's unprecedented proliferation.

However: As this study has shown, there is still a lot to do. Let's get to it!

BIBLIOGRAPHY

[1] Badach A., 2005. Voice over IP. Die Technik. Hanser-Verlag, Munich-Vienna.
[2] http://trafficlyser.de – Protocol analyzer TraceView VoIP.
[3] http://www.acterna.com – Protocol analyzer DA-360.
[4] http://www.bcs-oops.org.uk/resources/20050202_JAIN%20SLEE.pdf – Java API for Integrated Networks.
[5] http://www.itu.int/rec/T-REC-P.862/en – ITU-T Recommandation, P.862 (PESQ).
[6] http://www.itu.int/rec/T-REC-Y.2001/en – ITU-T Recommandation Y. 2001 (NGN).
[7] http://www.openh323.org/standards.html – H.323-Protocol suite.
[8] http://www.parlay.org/en/specifications – (Parlay Open Service Architecture).
[9] http://www.rfc-editor.org – RFC 3015 (MEGACO).
[10] http://www.rfc-editor.org – RFC 3550 (RTP).
[11] http://www.rfc-editor.org – RFC 3910 (SPIRITS).
[12] http://www.rfc-editor.org – RFC 4477 (DHCP).
[13] http://www.siemens.de – HiPath 5000 system.
[14] http://www.sipcenter.com/files/sip2.pdf – SIP-Protocol suite.
[15] http://www.sipgate.de – SipGate provider in Germany.
[16] http://www.skype.com/intl/de – Skype provider in Germany.
[17] Sigmund G., 2002. Next Generation Networks. Hüthig-Verlag Heidelberg.
[18] Nowicki K., Uhl T., 2004. Protokollanalyse bei VoIP-Systemen. Proc. of the II Conference "Technologie Informacyjne" Gdansk, vol. 5, 1011-1022.
[19] Uhl T., Nowicki K., 2005. Push-to-Talk als eine der ersten VoIP Applikationen in Mobilfunknetzen. Proc. of the III Conference "Technologie Informacyjne" Gdansk, vol. 7, 377-384.
[20] Uhl T., Nowicki K., 2006. Messmethoden zur Beurteilung der Qualität von VoIP-Verbindungen. Proc. of the IV Conference "Technologie Informacyjne" Gdansk, vol. 11, 753-762.

VOIP JAKO PILOT SIECI NASTĘPNEJ GENERACJI

Streszczenie

Praca zawiera krótki opis architektury sieciowej i standardów dla usługi VoIP. Zawiera ona również wyniki badań dotyczących oceny przydatności platformy IP do tej usługi. W dalszej części pracy przedstawiono porównanie usługi telefonicznej w internecie z tą usługą w sieciach telefonii stałej. Wyniki tego porównania wskazują na otwarte problemy do dalszych prac badawczych i rozwojowych. Końcową część pracy poświęcono odpowiedzi na pytanie: Czy VoIP może być uznany za pilota nowej generacji sieci, tj. NGN? Przedstawione oszacowania generują wiele otwartych problemów, które powinny być wkrótce rozwiązane. Pracę kończy podsumowanie.

Słowa kluczowe: sieci komunikacyjne, usługi komunikacyjne, protokoły komunikacyjne, architektura sieciowa, multimedialne usługi, IP, VoIP, NGN

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 249
TELEKOMUNIKACJA I ELEKTRONIKA 10 (2007), 41-53

BLUETOOTH VERSUS OTHER SHORT-RANGE WIRELESS SYSTEMS FOR TRANSMISSION OF INFORMATION IN BIOMEDICAL DEVICES

Adam Dąbrowski, Piotr Kardyś, Tomasz Marciniak

Chair of Control and System Engineering
Division of Signal Processing and Electronic Circuits
Poznań University of Technology
Piotrowo 3a, 60-965 Poznan, Poland

In this article the Bluetooth digital wireless connectivity system is compared with other contemporary techniques of short-range communications as regards their application in biomedical devices. Furthermore some interesting proposals are presented for Bluetooth implementation in supporting systems for handicapped people. One of such systems supporting speech has been drawn up by the authors.

Keywords: Bluetooth, IrDA, ZigBee, HomeRF, Wi-Fi, biomedical systems

1. INTRODUCTION

The main purpose for introducing Bluetooth – a digital, wireless communication standard exploiting radio waves and operating on short-range distances, was an idea of easy, wireless, remote, and automatically organized cooperation of modern, commonly used electronic equipment such as audio, telecommunication, IT devices, etc [10]. The Bluetooth, whose name comes from the Danish king Harald Blåtand, has become a real alternative for other wireless, short-range communication standards, which have typically been used up till now, mainly for the infrared radiation standard (IrDA), which is characterized not only by a very limited distance, but primarily by the necessity of the proper placement of devices cooperating with each other due to a narrow propagation angle. The Bluetooth is obviously deprived of this kind of flaws.

1.1. SHORT DESCRIPTION OF THE BLUETOOTH TECHNOLOGY

The Bluetooth telecommunication system provides 3 $Mb{\cdot}s^{-1}$ (in total) wireless data and voice transfer among various devices in the range of up to 100 meters. The radio modules operate in the license-free ISM band, intended for industrial, scientific and medical applications. The radio modules use the frequency range between 2.402 and 2.480 GHz and the band frequency division (FDMA) with 79 carrier frequencies defined in 1 MHz intervals. Transmission of signals is based on the Gaussian frequency shift keying (GFSK) modulation, while the time division multiple access (TDMA) ensures full duplex connectivity. The radiation power of the Bluetooth transmitters ranges from -30 to 20 dBm. The maximum output power level of a specific transmitter may be 0, 4, or 20 dBm depending on the assumed power protection class. Receiving modules have been designed to minimize the impact of interference and noise at the range end. The cyclic redundancy check (CRC) control code and the forward error correction (FEC) algorithm have been applied. Fast changes of the carrier frequency up to 1600 times per second (frequency hopping) and monitoring of the power level prevent the interference effects [1, 2].

1.2. COMMON APPLICATIONS

Wireless communication between electronic devices undoubtedly brings a number of improvements, first of all, increases the comfort of their usage. Applications of the Bluetooth technology are almost unlimited [7]. The most typical include wireless Internet access from mobile computers through a mobile phone, automatic, wireless data transfer to mobile computers (files, e-mails) received by means of a telephone, wireless data transmission between computers, and wireless connection with peripheral equipment. The Bluetooth specification remains open, which allows permanent searching for its new applications. It can also have an impact on introduction of completely new devices. C-Pen – wireless pocket scanner with the OCR signs recognizing function, transferring data directly to a computer, Info-Wear – a device resembling a watch, functionally corresponding with a palmtop are already prototypes; Info-Wear II – a "wireless wallet", enables customers to pay at the Bluetooth till without taking a credit card out of a wallet!

In the authors' opinion, apart from a number of applications of the Bluetooth system partly enumerated above, which raise the usage efficiency of electronic equipment for healthy people, the Bluetooth technology might also turn out to be helpful in the construction of devices intended for handicapped people (for example the blind, hard of hearing, or mute) and can considerably improve the standard of their life.

2. COMPETITIVE SHORT-RANGE WIRELESS SYSTEMS RELATED TO THEIR APPLICATION IN BIOMEDICAL DEVICES

Nowadays there are quite a few telecommunication systems (including the Bluetooth) available on the market, ensuring wireless short-range data transmission between devices. They can be distinguished by many parameters, such as achieved channel bit rate, operating range and current consumption, if considering only the most consumer-oriented parameters. Systems of short-distance cordless connectivity of low power consumption, which couple various common, quite often mobile devices, serve for WPANs, i.e., wireless personal area networks realization, while longer-range systems of higher power consumption create WLANs, i.e., wireless local area networks. Certainly, we cannot definitely declare, which telecommunication system is better. Each of them was optimized taking into account different applications. Thus they meet various requirements. Although in some cases they can seem to be competitive, they of course complement each other. This also takes place while implementing them in medical

Depending on operating specificity of a given device, a choice of some technological solutions can turn out more accurate, while others – pointless.

2.1. IRDA (INFRARED DATA ASSOCIATION)

Wireless IrDA connectivity technique making use of infrared radiation of 875 ±25 nm wavelength as a transmission medium is currently, perhaps, the most widespread system of short-distance wireless data transmission though over the last years rival short-range radio systems have been gradually pushing it aside. Although IrDA link throughput amounting to 4 $Mb \cdot s^{-1}$ (according to IrDA 1.3 specification [11]) offers slightly better possibilities than the Bluetooth system (3 $Mb \cdot s^{-1}$ according to 2.0 +EDR specification [1]), the necessity of placing cooperating devices in front of each other, severely limited range, typically to one meter, as well as ±15° propagation angle [12] undermine its usefulness. Furthermore the IrDA standard ensures data exchange between only two devices, so simultaneous cooperation of even a few apparatus and connecting them into a local mini-network are not possible in this case. These features exclude IrDA implementation in medical devices primarily because of too short range and narrow propagation angle, which can lead to transmission cutoffs. As far as file transfer between mobile computers is concerned, we accept the requirement of reestablishing the connection but while examining a patient by means of some diagnostic equipment it is out of the question. In some examinations, especially invasive, a patient should not undergo them more than once in a short time. Unusually short range and the necessity of placing IrDA transceivers in front of each other rule out the possibility of remote monitoring of examination process even from an assistant's next-door room. Bearing features mentioned above in mind, the IrDA technique should be viewed upon as undoubtedly inadequate for biomedical applications.

Recently introduced Area Infrared standard offers better parameters than the IrDA interface. First of all, propagation angle got widened to ±60°, a number of cooperating devices was increased to 10, while the maximum transmission range – up to 8 meters. However, extending the transmission range, significantly lowers the bit rate– even to 250 $kb \cdot s^{-1}$ [12]. Nevertheless, despite better properties of the Area Infrared standard in comparison to the IrDA, its implementation in biomedical equipment still remains questionable – primarily because of an absolute necessity of direct "eyeing" between cooperating units. Any disturbance of this condition leads to transmission latency or transmission holding and in the worst case – to its complete cutoff, which, at least as far as medical equipment is concerned, is unacceptable.

2.2. ZIGBEE (IEEE 802.15.4 STANDARD)

ZigBee is an interesting communication standard, which has been introduced quite recently. Its radio range achieves 75 meters but its most useful properties are ultra low current consumption and the resulting long battery working time exceeding even two years in case of alkaline batteries. The ZigBee technology uses the following frequency bands 868 ... 870 MHz, 902 ... 928 MHz as well as 2.4 ... 2.4835 GHz, offers, however, rather low bit rates, respectively 20, 40 and 250 $kb \cdot s^{-1}$ [6], thus one order of magnitude less than the Bluetooth. It is a consequence of the imposed assumptions for this standard specified only for low data rate applications. Various steering and control systems, in which signal transmission occurs sporadically or periodically and commands are encoded by means of some simple patterns, are a good example. Of course, in such conditions, limiting power consumption becomes possible but a range of applying the ZigBee narrows down drastically. Taking into account the role of the ZigBee standard in medical equipment, only very specific telemetric devices, in which measurements are cyclic, e.g., every few seconds or even less frequent, can be considered. The ZigBee standard is not suitable for transmission of biological signals sampled at higher rates, so applications connected for example with precise monitoring of human being's functions can be troublesome.

2.3. HOMERF TECHNOLOGY

HomeRF and Bluetooth technologies have a lot in common. Both systems appeared around the same time, use 2.4 GHz ISM (industrial, scientific, medical) frequency band and are applied both for data and voice transfer. The HomeRF ensures, however, higher bit rate – 10 $Mb{\cdot}s^{-1}$ (according to HomeRF 2.0 specification [11]) exceeding Bluetooth achievements a few times. Considering operating range adjustment to some specified needs as well as possibilities of limiting power consumption by transmitter units in advance, the HomeRF technology turns out to be less flexible than the Bluetooth. This results from the basic assumption that the HomeRF network was to guarantee 50 m operating range tantamount to a typical flat. Employing the Bluetooth technology in a given biomedical device we have got an opportunity to pick a transmitter of an adequate power class that meets requirements for a specified application and reduces power consumption simultaneously. For instance, for medical aid devices, whose components are placed close to the patient, 10 m operating range is long enough. Energy saving is especially important for portable battery-powered medical devices, in which either one-battery working time or how often accumulators need charging decide upon their attractiveness.

2.4. WI-FI NETWORKS

Wi-Fi technology comprises, at this moment, three IEEE standards, namely: 802.11a (used in the USA), as well as 802.11b and 802.11g responsible for wireless local area networks (WLANs). Data transmission takes place in the following frequency bands 5.1 ... 5.825 GHz (802.11a), 2.4 ... 2.4835 GHz (802.11b) and 2.4 ... 2.497 GHz (802.11g), reaching 54 $Mb{\cdot}s^{-1}$ bit rate in (a) and (g), and 11 $Mb{\cdot}s^{-1}$ in (b) standard respectively [6, 10, 11]. Among already enumerated systems, the WLAN networks ensure the largest operating range (a few hundreds meters) but they require quite a lot of energy. The Wi-Fi technology implemented in medical equipment does not have to be most appropriate. It results from its original designing, i.e., organizing networks in charge of Internet standards together with all typical LAN applications. The WLAN networks were optimized differently than wireless personal area networks (WPANs) and they offer slightly other possibilities. Their basic task is, first of all, to cooperate with computer platforms.

2.5. IEEE 802.15.3 STANDARD

802.15.3 standard based on the UWB (ultra wide band) technique gives us more opportunities. It makes use of 3.1 ... 10.6 GHz frequency band, in which very short-time pulses of only a few picoseconds are sent. Lately the UWB technique has been applied in WLAN as well as in WPAN networks ensuring high data rates. The achieved bit rates are approximately inversely proportional to the transmission range and, for example, they equal 22 $Mb \cdot s^{-1}$ at 100 m distance, 55 $Mb \cdot s^{-1}$ at 50 meters and 125 $Mb \cdot s^{-1}$ at 20 m (based on Pulse-Link experience [13]). High channel throughput will allow moving picture transfer (e.g., operation monitoring), which the Bluetooth telecommunication system certainly will never be able to guarantee. The raised 802.15.3 parameters result in unprecedented high power consumption in comparison to the Bluetooth technology, thus a device designer ought to consider 802.15.3 application carefully.

3. BENEFITS OF THE BLUETOOTH TECHNOLOGY

Despite the co-existence of many alternative wireless connectivity techniques for short-range distances, the Bluetooth technology has already reached its stable position on the market of commonly used electronic equipment. Apart from its very differentiated applications dedicated to wide range of consumers, the Bluetooth can also be employed in other fields to construct specialized devices, including biomedical devices. Analyzing telecommunication systems described earlier, we can find at least a few advantages of using the Bluetooth technology, rendering it especially adequate for implementation in some situations.

The Bluetooth, as a typical radio system, overcomes easily wireless IrDA connectivity technique exploiting infrared radiation as a transmission medium, which is very inconvenient in using, because this technique requires maintaining direct “eyeing” between transceivers. Considering the application of the IrDA technique in medical equipment, this condition seems to be really inconvenient. In comparison with the modern ZigBee communication system designed mainly for steering and control systems, the Bluetooth technology turns out undoubtedly more energy consuming, however, it offers one order higher throughput than the ZigBee. In turn the HomeRF technique, long-distance Wi-Fi networks as well as new IEEE 802.15.3 standard triumph over the Bluetooth technology with respect to achievable bit rates. Unfortunately, incomparably large power consumption of these systems, optimized from a completely different perspective, is a price we pay for high throughput, so their implementation in portable biomedical devices would call for some discussion. From this point of view the Bluetooth technique is definitely more flexible as a device designer has the possibility of choosing transmitter power class accurately for a specific case.

The Bluetooth standard cannot be regarded as closed. It undergoes constant improvements, which are proved to be at the cutting-edge of specifications, the number of which has already reached nine. The most significant changes have been introduced in the last one as the Bluetooth link throughput has been considerably raised with the aid of optional differential phase modulations – $\pi/4$-DQPSK and 8-DPSK. Thanks to this undertaking the throughput of asynchronous channel increased 3 times, while of synchronous – almost 14. In this way a range of the Bluetooth system applications is still widening, with the exception of moving pictures transmission in real time.

4. EXAMPLES OF THE BLUETOOTH EMPLOYMENT IN AID SYSTEMS FOR HANDICAPPED PEOPLE

4.1. DIGITAL HEARING AIDS

An interesting concept of the Bluetooth technology application would be its appliance to digital hearing aids cooperating with mobile phones. Currently produced, modern hearing aids are usually equipped with a set of functions, which are designed to improve speech perception. These functions are realized via sophisticated algorithms working in real time thanks to a digital signal processor. Among the most practical consumer-oriented functions of the hearing aid are: noise reduction, signal separation along with directional filtering as well as digital audio-zooming [14]. A simplified block scheme of a modern hearing aid is presented in Fig 1. It comprises a microphone matrix, a remote controller and an earphone now quite often having a cordless connection with the remote controller by means of FM modulation instead of a regular cable.

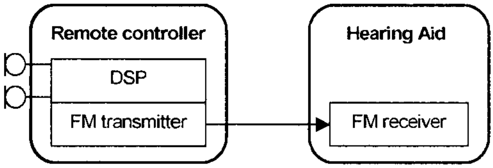


Fig. 1. Main blocks of a modern hearing aid
Rys. 1. Główne bloki współczesnego aparatu słuchowego

Relatively often hearing impaired people, just like anyone else, long for routine usage of a cell phone. Unfortunately this is especially tough for them and even troublesome because they may need to have an earphone in each ear and anyway, hearing their interlocutors through a piezoelectric loudspeaker is almost an unachievable goal. Just then the Bluetooth telecommunication system may turn out to be helpful. Employing the Bluetooth technology in the hearing aids would allow for their remote cooperation with mobile phones equipped with the Bluetooth interface (Fig. 2), and furthermore, what is especially essential, it would ensure a proper processing of telephone signal adequately to hearing loss. In the discussed application the Headset and Handsfree profiles of the Bluetooth specification would be certainly employed [1].

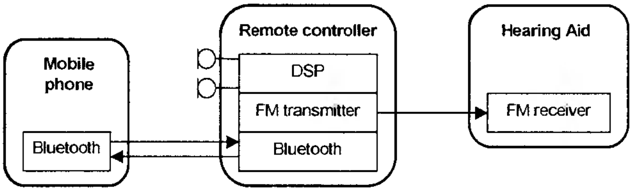


Fig. 2. Communication between digital hearing aid and a mobile phone
Rys. 2. Komunikacja cyfrowego aparatu słuchowego z telefonem komórkowym

Undoubtedly, the system presented in Fig. 2 could be strongly simplified. Assuming that nowadays everyone has a mobile phone, including people with impaired hearing as they often use the SMS technique for communication, a possibility of incorporating the hearing aid into the phone should be considered. From a technical point of view it is certainly feasible without any considerable increase in production costs of mobile phones because processors of such phones can handle quite easily processing of signals from the microphone matrix. In this way the remote controller, typical for the hearing aid, could be entirely eliminated. Of course integrating the hearing aid with a cell phone would cause many software changes but only one hardware modification, namely adding an advanced microphone matrix to the phone.

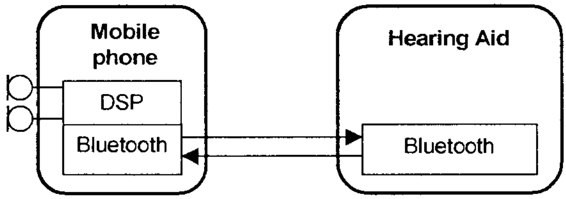


Fig. 3. Integrating a hearing aid with a mobile phone
Rys. 3. Integracja aparatu słuchowego z telefonem komórkowym

Moreover, the transfer of adequately formed acoustic signal to an earphone, which could occur by means of the Bluetooth technique as the FM transmission, has also some limitations. For instance, it has only scores of programmable channels at disposal. This feature can generate interference when a lot of hearing-impaired people use the same type of supporting system within the same area. An application of the Bluetooth technology would resolve this problem, because it allows theoretically up to 79 Bluetooth devices to operate in the vicinity. Fig. 3 illustrates the proposed concept of the direct communication of digital hearing aid (the earphone) with a mobile phone. Obviously, the hearing aid has to be equipped with the Bluetooth radio module. Proposals of the Bluetooth technology employment in the hearing systems can be found in several projects, e.g., in the European 5th Framework "BlueEar" project [15].

4.2. ORIENTATION-AID SYSTEM FOR THE BLIND

The Bluetooth technology may turn out to be unequalled in supporting the blind people moving within the city area. Because of the range limited to ten meters (in the case of transmitters produced according to the 3rd power class) the Bluetooth communication technique is perfectly suited for this kind of purpose and soon may become the key element of remote orientation systems of the blind in the city area. The main task of such systems is providing the blind people with information about their current position, and also about the possible barriers appearing nearby.

The concept of orientation-aid system based on the Bluetooth technology assumes placement of a number of special Bluetooth transmitting units, emitting previously recorded voice announcements in significant town locations, namely at crossroads, at pedestrian crossings, at public transport stops, as well as near more important facilities and even shops (Fig. 4). A blind person equipped with a special Bluetooth receiver with a headphone (instead of a loudspeaker) or with a device miniaturized to a typical wireless Bluetooth headset, while moving in the city, would be listening to voice announcements about her/his current position, sent from subsequent transmitters distributed along particular streets. Apart from the information about names of streets, the voice announcements could be enriched with other very useful information for a blind person, e.g., about pavement surface (asphalt, cobbled), its condition, and the distance to the nearby bus stop, etc. Moreover, public institutions and shopkeepers could install independent transmitters at the entrance doors, which would inform blind people about the places they have to pass by. In case of too many independent transmitters at one place, it would be indispensable to limit their range by decreasing transmission power below 0° dBm.

Aid system transmitters should not be operating in constant work mode, because ordinary passers-by, who use Bluetooth headsets for their mobile phones, would practically become receivers of announcements useless for them, therefore all kinds of the city Bluetooth transmitters included in this aid systems should be slave units and connections should have to be initiated exclusively by devices belonging to the orientation-aid system users. Only then, the voice messages would be sent to the receivers, which initiate the connection. The transmission would be performed through the synchronous channel, dedicated just for sending the sound (mainly the voice) with the limited frequency range to 4 kHz.

Systems similar to the proposed above already function in some cities (e.g., in England) [9, 16], however up till now the Bluetooth technology has not been used for this purpose. Obviously the orientation-aid systems for the blind will never entirely replace a stick, or a guide dog. Undoubtedly, they will make blind people's movements easier in urbanized areas, especially in big and small cities.

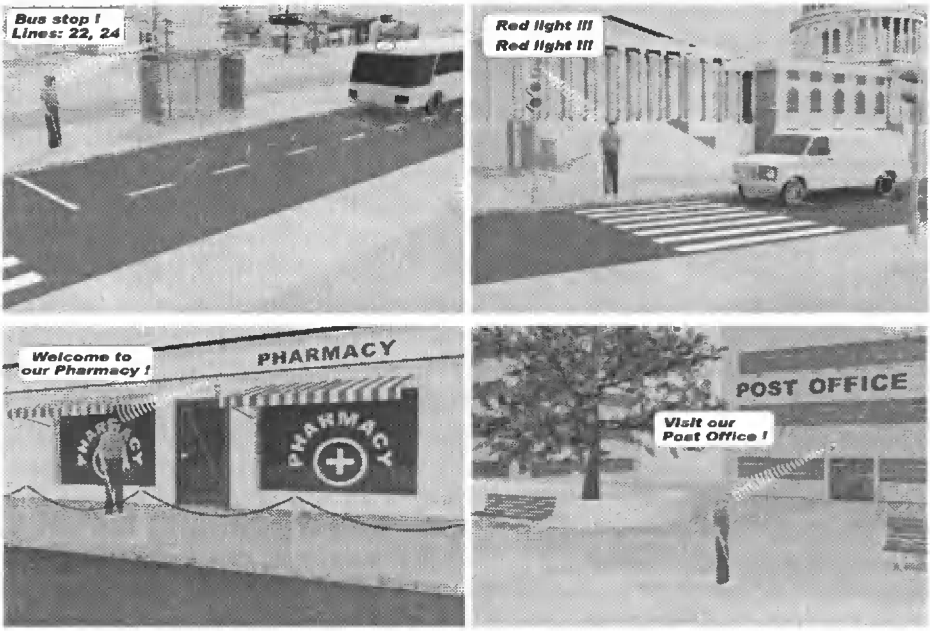


Fig. 4. Functioning of orientation-aid system for the blind in urban area
Rys. 4. Działanie systemu wspomagania orientacji osób niewidomych w przestrzeni miejskiej

4.3. SPEECH-AID SYSTEMS

The analysis of the Bluetooth employment in the realization of speech-aid systems is justifiable for obvious reasons exclusively in the case of digital systems. The aim of incorporation the Bluetooth technique into speech-aid systems is improvement of their functioning or comfort of their usage. Because of the complexity of such systems and inclusion of many components (e.g., a microphone, a proper processing unit, a loudspeaker with an amplifier), the incorporation of the Bluetooth technology can prove to be very practical and undoubtedly making the usage of such devices easier.

From the point of view of properties and transmission parameters required for the speech-aid systems, the Bluetooth technique, although not unlimited, is perfectly useful. Its minimal range up to 10 meters for the transmitters produced in the 3rd class of protection, is perfectly enough for stable connectivity between the parts of the aid device as in the majority of cases, they are very close to each other. The minimum bit rate of the Bluetooth synchronous channel used for transferring sound (mainly voice) amounting 64 $kb \cdot s^{-1}$ may seem to be some kind of a limit. However, it is not a critical parameter, as in the assumptions, the task of speech-aid systems is not to emit sounds of a very high quality but the sound comparable with a voice in a telephone conversation or from some wireless. The bit rate of 64 $kb \cdot s^{-1}$ meets exactly those requirements and results from the limited acoustic band up to 4 kHz, and consequently from sampling rate of 8 $kSamples \cdot s^{-1}$ and 8-bit representation of every sample [8]. In case this bit rate was not enough, Bluetooth modules produced according to 2.0 +EDR specification, ensuring several times higher throughput of the synchronous link (864 $kb \cdot s^{-1}$ for 3-EV5 packets [1]), should be used.

Another issue is the choice of couplings, which could be replaced by a wireless link. Therefore, some general cases including the most probable technological solutions should be considered here. If the method of the speech supporting bases on residual or distorted acoustic signal emitted by the patient, received with the use of a microphone placed near the mouth, or if in the processing other information coming from other measuring sensors placed near the mouth or the larynx are used (e.g., microwave or ultrasonic sensors, which monitor the setting of vocal tract elements), it seems to be reasonable to eliminate the link between the microphone or other sensors and the main processing unit itself. Elimination of this cable link will undoubtedly make the usage of speech-aid systems easier, because similar links, for example of a mobile phone and headset, are particularly troublesome for their users.

Replacement of a cable microphone with a wireless one is relatively easy. All we have to do is to reach for a wireless Bluetooth headset, which is designed especially for mobile phones. The examples of such headsets are shown in Fig. 5. However, if in practice, the headset microphone moved a little off the mouth does not live up to the expectations, it is advisory to change its position or even the type of the microphone. If other kinds of sensors, for example microwave or ultrasonic sensors, are assigned, it will be necessary to adjust a miniature circuit encoding the received signals and cooperating with the Bluetooth transmitting module. Perhaps in such a case, it will be necessary to transmit data through the asynchronous channel because of the required higher bit rate. Figure 6 presents an example of the complete Bluetooth radio module. It has two high-rate communication ports (USB, UART), a PCM codec interface and could be used in a similar application as well. It is worth mentioning here, that currently produced Bluetooth radio modules are still being minimized.

Fig. 5. Exemplary Bluetooth headsets (BT400 and BT400-G3, Bluetake)
Rys. 5. Przykładowe zestawy Bluetooth (BT400 oraz BT400-G3, Bluetake)

Fig. 6. Complete Bluetooth radio module (ROK101 007, Ericsson)
Rys. 6. Kompletny moduł radiowy Bluetooth (ROK101 007, Ericsson)

The indispensable element of the electronic speech-aid systems is the loudspeaker unit including an amplifier and a speaker. Unfortunately, the size and weight of this part of the device are first of all conditioned by the shape and size of the speaker, which because of the proper sound parameters should not be too small. Therefore, this bigger and heavier element of the system is often placed in a separate casing, which is worn for instance on a belt around the upper-body or is attached to the trouser belt while the processing unit itself altogether with all the switches and regulators constitutes a separate and easily accessible part. In this sort of aid device structures, the Bluetooth technology could replace the cable link of the main processing unit with the loudspeaker unit. Owing to this, either at home or at work a mute person would not have to wear the loudspeaker unit all the time. It would be enough to be within the operating range of the Bluetooth radio module placed in its casing. Certainly, it would be a great convenience and comfort, though the sound would not come from a handicapped person but from the loudspeaker unit left nearby or even from additionally attached loudspeaker columns.

Undoubtedly, it has to be admitted that the Bluetooth technology provides new opportunities for the realization of modern, electronic speech-aid systems. It can improve their functionality and convenience. The parameters of this system meet the demands of the bit rate for the voice transmission required by these devices as well as of an expected operating range. Nevertheless, it has to be taken into consideration, that a galvanic (wireless) split of any device means the necessity of introducing additional supply sources for each subunit. In practice, in many cases it can turn out to be very troublesome.

5. PROTOTYPE SPEECH-AID SYSTEMS BASED ON THE BLUETOOTH TECHNOLOGY

It was the authors' intention to design a complex speech-aid system addressed to possibly large group of users – people mute from infancy as well as patients with voice problems resulting from larynx surgical operations. Surgeries like these are most frequently applied in tumor cases. Then, extraction of the infected tissues is unfortunately usually unavoidable. The operation range depends on a tumor size and position, which is different for each patient. Despite this operation the patient is able to utter certain speech-like sounds, however, they are very difficult to comprehend. Although a project of such universal speech-aid system brings about a lot of theoretical and technical problems to resolve we dealt with this successfully owing to the Bluetooth technology employment.

The complete speech-aid system consists of a few components: a microphone, a central module worn on the hand (Fig. 7), a loudspeaker unit fitted on a belt around the waist and a portable computer installed nearby, yet not requiring any wire connection with the other components. The microphone and the loudspeaker are connected with the central module using soft cables, whereas electronic circuits of the central module communicate with the portable computer in a cordless way using the Bluetooth technology. The detailed description of this speech-aid system can be found in [3], and two of the implemented speech improvement method in [4] and [5].

Fig. 7. Central module of the prototype speech-aid system
Rys. 7. Moduł centralny prototypowego systemu wspomagania mowy

6. CONCLUSIONS

The Bluetooth telecommunication system, competing with other modern wireless techniques of connectivity, can be effectively employed to construct some selected biomedical devices increasing their functionality and comfort of usage. The suggested

Bluetooth implementations in supporting systems dedicated to handicapped people, suffering from hearing, sight, and speech dysfunctions, show a completely new area of its applications. The prototype speech-aid system drawn up by the authors, tested successfully by patients after surgical larynx operations, undoubtedly proves the very essence of this thesis.

This work was supported by 93/1582 and DS-93/152/07 projects.

BIBLIOGRAPHY

[1] Bluetooth specification, ver. 2.0 + EDR.
[2] Dąbrowski A., Kardyś P., 2001. Bluetooth wireless transmission system (in Polish). KKRRiT Conference Poznań, 0.4.1-0.4.8.
[3] Dąbrowski A., Kardyś P. et al., 2004. BlueVoice – a specialized speech-aid system for hardly speaking people. Foundations of Control and Management Sciences 1, 21-33.
[4] Dąbrowski A., Kardyś P., Meyer A., 2003. Application of the method of spectral contrast enhancement for esophageal speech (in Polish). 10th International Symposium of Sound Engineering and Tonmeistering Wrocław, 153-157.
[5] Dąbrowski A., Kardyś A., Kardyś P., Meyer A., 2006. Vowels insertion into pseudowhisper of laryngectomized patients. 11th Symposium "New Trends in Audio and Video Technology" Białystok (to be published).
[6] Dąbrowski A., Marciniak T., Cetnarowicz D., 2005. Contemporary interfaces for digital data transmission (in Polish). [In:] Nowoczesne przedsiębiorstwo. Poznan University of Technology, 226-233.
[7] Kardyś P., Dąbrowski A., 2001. Bluetooth. New possibilities of modern electronic equipment (in Polish). Elektronizacja 4-5-6, 21-23, 7-9, 14-15.
[8] Kardyś P., Dąbrowski A., 2001. Audio and voice processing in the Bluetooth system (in Polish). KKRRiT Conference Poznań, 16.2.1-16.2.4.
[9] Kemmerling M., Schliepkorte H., 1999. An orientation and information system for blind people based on RF speech beacons. TIDE.
[10] McDermott-Wells P., XII 2004 – I 2005. What is Bluetooth? 23(5), 33-35.
[11] Miller B., Biskidian Ch., 2003. Bluetooth revealed. The insider's guide to an open specification for global wireless communication. Prentice Hall PTR.
[12] Nawrocki W., 2005. Measurement Systems and Sensors. Artech House.

WEB SITES

[13] From the WLAN world. http://www.interline.pl/portal/modules/news/article.php?storyid=77
[14] Hearing systems. http://www.phonak.com
[15] http://www.blueear.org
[16] The orientation-assisting device. http://www.rnib.org.uk

BLUETOOTH A INNE SYSTEMY ŁĄCZNOŚCI BEZPRZEWODOWEJ KRÓTKIEGO ZASIĘGU ZASTOSOWANE DO TRANSMISJI SYGNAŁÓW W URZĄDZENIACH BIOMEDYCZNYCH

Streszczenie

W artykule dokonano porównania cyfrowego systemu łączności bezprzewodowej Bluetooth z innymi współczesnymi technikami komunikacyjnymi krótkiego zasięgu pod względem możliwości ich zastosowania w urządzeniach biomedycznych, wykazując przewagę technologii Bluetooth do tych zastosowań. Ponadto przedstawiono kilka ciekawych propozycji implementacji technologii Bluetooth w systemach wspomagających osoby niepełnosprawne. Jeden z takich systemów – wspomagający mowę – został opracowany przez autorów.

Słowa kluczowe: Bluetooth, IrDA, ZigBee, HomeRF, Wi-Fi, urządzenia biomedyczne

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 249
TELEKOMUNIKACJA I ELEKTRONIKA 10 (2007), 55-76

MATHEMATICAL MODELS AND ANALYSIS OF STOCHASTIC OSCILLATIONS

Igor Javors'kyj

Institute of Telecommunication, University of Technology and Life Sciences, Bydgoszcz, Poland
Karpenko Physic-Mechanical Institute of National Academy of Sciences of Ukraine, Lviv

Description of the properties of stochastic oscillation in terms of probabilistic characteristics of the periodically correlated stochastic processes and their generalizations is presented. General approach to the problems of estimation of these characteristics on the basis of natural data is considered. possible estimation methods are analyzed and the results of investigation of hidden periodicities are given.

Keywords: stochastic oscillation, periodically correlated random processes, estimation methods, hidden periodicities

1. INTRODUCTION

Both rough recurrence and stochastic are characteristic feature of time changeability of many physical processes. Recurrence of the oscillations property - rhythmic – can be caused by both the effect of external forces (forced oscillations) on a given system and results from the internal interrelations (eigen oscillations) existing in systems. Rhythmical processes are encountered in many spheres of science and engineering including radiophysics, geophysics, oceanology, meteorology, climatology, vibrodiagnostics, hydroacoustics, biology, seismology, economics, telecommunication etc.

Processes which are caused by astrophysical factors, should be noted among all the diversity of forced oscillations occurring in nature. These are: the annual and diurnal oscillations of geophysical, oceanological and meteorological quantities [1], which are the result of the earth's revolution around the sun and rotation of earth round its axis of the equator; the tidal oscillations of the sea level, the earth's crust, sea currents, internal waves whose polyrhythmic are caused by the polyharmonic character of the potential of tideforming forces. Among the autooscillation processes we should note the oscillations in autogenerators of various physical nature [2-5], signals of geomagnetic pulsations [6, 7], vibrations [8], oscillations in biological systems (biorhythmic) [9].

A certain model of mathematical oscillations is the methodological base for investigation of the oscillation processes structure on the basis of experimental data. In the pioneering investigations of rhythmic, the evident advantages were given to deterministic conception which is based on models in the form of periodical functions. The aspiration to take into account random features of oscillations encouraged the development of probabilistic methods that consider phenomena as stationary random processes. Within the framework of such an approach, rhythmic features of physical processes manifest themselves in an oscillatory behavior of correlogram and existence of several peaks of the estimates of the spectral density. However, these characteristics describe the average properties of processes and do not provide information on their temporal structure, which can be determined, though in an idealized form, with use of deterministic models. Natural combination and development of these two approaches give rise to a concept that represents probabilistic models as periodically correlated random processes (PCRP) and PCRP-related processes (bi-, poly-, and almost periodically correlated). Such models generalize the notion of the recurrence to situations where stochasticity plays a significant role. These models provide an opportunity to describe the structure of rhythmic variations more thoroughly and objectively and include the abovementioned models as particular cases. The PCRP model allow us not only to analyze processes using special methods characteristic of each model, but also to investigate the different phenomena in common terms for all models.

In this paper we consider probabilistic characteristics of periodically nonstationary probabilistic models, analyze the properties of such models and develop the general approach for the estimation of such characteristic from the empirical data. We analyze the methods of statistical estimation and present radically new results of the investigation of hidden periodicities.

The manuscript presents a survey of results of the author and his colleagues' investigations realized at Karpenko Physico-Mechanical Institute of National Academy of Sciences of Ukraine in Lviv, and at University of Technology and Life Sciences in Bydgoszcz.

2. PERIODICALLY CORRELATED RANDOM PROCESSES AND THEIR GENERALIZATIONS AS PROBABILISTIC MODEL OF STOCHASTIC OSCILLATIONS

The methods of periodically correlated random processes and their generalizations extend our capabilities for understanding the regularities of stochastic oscillations. These methods provide additional opportunities associated with the description of the properties of oscillations in terms of time-varying probabilistic characteristic. Within the framework of the second order theory, the hidden periodicity of physical processes is manifested as periodic temporal variations of the mean and the correlation function:

$$
\begin{gathered}
m(t) = E\xi(t), \\
b(t,u) = E\overset{\circ}{\xi}(t)\overset{\circ}{\xi}(t+u) = b(t+T,u), \qquad (1) \\
\overset{\circ}{\xi}(t) = \xi(t) - m(t).
\end{gathered}
$$

The properties of (1) define the class of PCRP.

Apparently, the first mathematically correct definition of PCRP was given by L.I. Koronkevich [10]. This definition is applicable to the description of the properties of solutions for differential equations with periodic coefficients and random right-hand sides. The possibility of using such processes to describe stochastic oscillations was also mentioned by R.L. Stratonovich (11) (see also the review [12]). Several papers [13-15] consider transformations of periodically nonstationary and periodically correlated signals. Correlation and spectral properties of PCRP were studied by S.M. Rytov [5], E.G.

Gladyshev [16], L.I. Gudzenko [17], H. Ogura [18], A. Papoulis [12], L. Franks [19], W. Gardner and H. Hurd [80, 81]. Gudzenko employed PCRP to analyze fluctuations in autooscillation systems. Rytov [5] indicated the possibility for applying PCRP to study noise in cyclic remagnetization. A.F. Romanenko and G.A. Sergeev [22] pointed to the adequacy of employing the PCRP methods to describe a turbulent flow of water near a ship screw propeller and to investigate temporal variations in the electric-power consumption and physiological changes. Ya.P. Dragan [23-27] developed the fundamentals of the PCRP theory with a limited average power. In collaboration with K.S. Voichishin, Dragan also applied the PCRP model to formulate certain general properties of the stochastic model of rhythmic [28-32]. Voichishin [33] made the first attempts for using this model to analyze daily rhythmic variations in certain geophysical processes. Dragan and I.N. Javors'kyj extended this approach to study the properties of wind waves [34]. In both cases, the estimates of the mean and variance were analyzed. Note that, in earlier studies, the PCRP model, which was understood even in a more comprehensive sense (with use of histograms and estimates of the correlation function), was applied to investigate daily variations in the temperature and humidity of the air and the soil temperature [35-40]. An analogous modification of this model was used to study diurnal and seasonal changes in meteorological processes [35]. P.Ya. Groisman [41] considered the periodic correlation of precipitation series.

However, the systematic application of the PCRP model for analysis of rhythmic signals was limited mainly because of the absence of an appropriate procedure of data processing. Papers [2, 6, 7, 43-115] were devoted to the development of such a procedure and application of this method for the investigation of the structure of rhythmic variations in physical processes.

Within the framework of the PCRP model of rhythmic, the mean describes the regular periodical oscillations, the variance characterizes the periodicity of the power of fluctuations around this regular behaviour and the correlation function describes the character of periodic changes in correlations between the values of fluctuating parameters at various moments of time separated by equal time intervals.

If we assume that the PCRP mean and the correlation function are absolutely integrable within the interval [0, T], then these characteristics can be represented in Fourier series:

$$m(t)=\sum_{k\in Z} m_k e^{ik\frac{2\pi}{T}t},\ b(t,u)=\sum_{k\in Z} B_k(u)e^{ik\frac{2\pi}{T}t}, \tag{2}$$

where $|m_k \to 0|$ and $|B_k(u)\to 0|$ as $k\to\infty$ .The coefficients m_k and $B_k(u)$ (the latter coefficients are referred to as correlation components) quantitatively characterize the waveforms of the periodic curves representing the mean and the correlation function. The correlation components $B_k(u)$ satisfy the equations:

$$B_k(-u)=B_k(u)e^{ik\frac{2\pi}{T}u},\ B_k(u)=\overline{B_{-k}(u)}. \tag{3}$$

The zero[th] correlation component is an even $B_0(-u)=B_0(u)$ and positive-definite function. In other words, this component have all the properties of the correlation function of a stationary random process.

The representation of the PCRP in terms of stationary connected components [18, 23, 25]:

$$\xi(t)=\sum_{k\in Z}\xi_k(t)e^{ik\frac{2\pi}{T}t} \tag{4}$$

is important for understanding the structure of PCRP as a model of stochastic oscillations. As can be seen from expression (4), PCRP can be represented as a sum of amplitude and phase-modulated harmonics, whose frequencies are multiple of the fundamental oscillation frequency $\omega_0=\frac{2\pi}{T}$. Comparing (2) and (4), we find that the components m_k coincide with the mean of stationary random processes $\xi_k(t)$. The autocorrelation functions of these processes $D_{kk}(u)=E\overset{\circ}{\xi}_k(t+u)\overset{\circ}{\xi}_k(t)$ determine the PCRP zero[th] correlation component:

$$B_0(u)=\sum_{k\in Z}D_{kk}(u)e^{-ik\frac{2\pi}{T}u}. \tag{5}$$

The cross-correlation functions of components whose numbers are shifted by l determine the l th correlation components:

$$B_l(u)=\sum_{k\in Z}D_{k+l,k}(u)e^{-ik\frac{2\pi}{T}u}. \tag{6}$$

As can be seen from the representation (4), the PCRP model covers various simpler models of rhythmic variations, including the additive model $\xi(t)=\eta(t)+f(t)$, where $\eta(t)$ is a stationary random process and $f(t)$ is a periodic function; the multiplicative model $\xi(t)=\eta(t)f(t)$; and the additive-multiplicative model $\xi(t)=g(t)+\eta(t)f(t)$, where $g(t)$ is a periodic function. The first model is often used in hydrometeorological studies. The second and third models are mainly applied to describe variations in the variance, that is, in the power of fluctuation oscillations [116, 117].

The PCRP variable spectral density $f(\omega,t)$ is a complex function: $f(\omega,t)=\mathrm{Re}\,f(\omega,t)-i\,\mathrm{Im}\,f(\omega,t)$. Its real part is determined by the cosine transform of the even part of the correlation function:

$$\mathrm{Re}\,f(\omega,t)=\frac{1}{2\pi}\int_0^\infty b^{`}(t,u)\cos\omega u\,du, \tag{7}$$

whereas the imaginary part is given by the sine transform of the odd part of the correlation function:

$$\mathrm{Im}\,f(\omega,t)=\frac{1}{2\pi}\int_0^\infty b^{c}(t,u)\sin\omega u\,du. \tag{8}$$

The real and imaginary parts of the spectral density are even and odd functions of the frequency, respectively:

$$\operatorname{Re} f(-\omega,t)=\operatorname{Re} f(\omega,t), \qquad \operatorname{Im} f(\omega,t)=-\operatorname{Im} f(-\omega,t).$$

Since

$$b(t,u)=2\int_0^\infty\left[\operatorname{Re} f(\omega,t)\cos\omega u-\operatorname{Im} f(\omega,t)\sin\omega u\right]du,$$

for $u=0$ we have:

$$b(t,0)=2\int_0^\infty \operatorname{Re} f(\omega,t)\,d\omega. \qquad (9)$$

This formula allows us to provide a physical interpretation of the function $\operatorname{Re} f(\omega,t)$. Since $b(t,0)$ characterizes the instantaneous power of the process, $\operatorname{Re} f(\omega,t)$ describes the distribution of this power in the (ω,t) plane. The integration of this quantity with respect to all the relevant frequencies yields the value of the power at a given moment of time t. However, we cannot interpret this quantity as the power spectral density because this quantity is not necessarily nonnegative for all (ω,t), although, similarly to the spectral density of a stationary random process, this function is even. For $b(t,0)=const$, expression (9) is reduced to the well-known relation for stationary random processes. Then $\operatorname{Re} f(\omega,t)=f(\omega)$ is the power spectral density. This function can be interpreted in terms of energy characteristics for the so-called quasi-stationary random process, when the rate of the variation of the correlation function, due to the lag is much higher than the rate of temporal variation of this correlation function. Such a quasi-stationary behaviour may also occur in the case of PCRP. Then, we have $\operatorname{Re} f(\omega,t)\geq 0$, and we can consider this function as the power spectral density. Generally, we cannot use this physical interpretation. As can be seen from the formula $b^c(t,u)=\frac{1}{2}\left[b(t,u)-b(t-u,u)\right]$, if the correlation function rapidly decays with the growth in the lag and displays only small variations in the argument t within the interval $[t-u,t]$, then $b^c(t,u)$ is small. Hence, the function $b^c(t,u)$ can be used to characterize transient processes, and $\operatorname{Im} f(\omega,t)$ describes the properties of such processes in the frequency domain.

The variable spectral density $f(\omega,t)$ is a periodic function of time. The amplitudes of harmonics of this function determine the spectral components $f_k(\omega)$:

$$f(\omega,t)=\sum_{k\in Z} f_k(\omega)e^{ik\frac{2\pi}{T}t}.$$

These components are defined by means of the Fourier transform of the correlation components:

$$f_k(\omega)=\frac{1}{2\pi}\int_0^{\infty}B_k(u)e^{-i\omega u}du. \tag{10}$$

If the correlation components are absolutely integrable, then $f_k(\omega)$ is continuous for all $\omega\in R$ and $|f_k(\omega)\to 0|$ as $\omega\to\pm\infty$.

Using expressions (3) we derive:

$$f_{-k}(\omega)=\overline{f_k(-\omega)},\ f_k(-\omega)=f_k\left(\omega+k\frac{2\pi}{T}\right).$$

The zeroth spectral component $f_0(\omega)$ is a real even function and $f_0(\omega)\geq 0$ for all $\omega\in R$. This conclusion is rather natural because $f_0(\omega)$ is the Fourier transform of the zeroth correlation component $B_0(u)$, which coincides with the correlation function corresponding to the stationary approximation of PCRP. The zeroth spectral component describes the frequency distribution of the average power of oscillations, whereas higher order spectral components characterize frequency properties of cross-correlations $\xi_k(t)$ of modulating processes.

Vector PCRP are natural probabilistic models of rhythmic variations of vector physical quantities. The mean of vector PCRP are described by periodic vectors $m_{\vec{v}}(t)=m_{\vec{v}}(t+T)$, whereas correlation functions $b_{\vec{v}}(t,u)$ and spectral densities $f_{\vec{v}}(\omega,t)$ are given by periodic dyad tensors. Similarly to vector stationary random processes [1, 118], the properties of vector PCRP can be described in terms of the invariants of tensors $b_{\vec{v}}(t,u)$ and $f_{\vec{v}}(\omega,t)$ [58]. These invariants unambiguously characterize the correlation and spectral structure of vector random processes regardless of the choice of the coordinate frame.

More sophisticated models of rhythmic variations are based on bi-and poly-PCRP, which describe both the interference and nonlinear interaction of oscillations with different periods. To analyze modulation effects in the PCRP model, we should represent a PCRP as a series (4). Then, the interaction of two rhythms gives rise to a process:

$$\xi(t)=\sum_{l,k\in Z}\xi_{k,l}(t)e^{i\Lambda_{kl}t},$$

where $\Lambda_{kj}=k\frac{2\pi}{T_1}+j\frac{2\pi}{T_2}$. The mean and the correlation function of such process, which is referred to as the bi-PCRP, are written as:

$$m(t)=\sum_{l,k\in Z}m_{kl}e^{i\Lambda_{kl}t},\ b(t,u)=\sum_{l,k\in Z}B_{kl}(u)e^{i\Lambda_{kl}t}.$$

The mixed components of the mean m_{kj} and the correlation function $B_{kj}(u)$ characterize the modulation interaction of the k^{th} and j^{th} harmonics. The structure of matrices representing these quantities is indicative of the presence of bi-rhythmic variations of a certain type.

Generalizing the model of bi-PCRP to additive-multiplicative interaction involving many rhythms, we arrive at the notion of poly-PCRP, which can be represented in the following form:

$$\xi(t)=\sum_{l_1,\ldots,l_\nu\in Z}\xi_{l_1,\ldots,l_\nu}(t)e^{it\sum_{j=1}^{\nu}l_j\Lambda_j}.$$

Components $\xi_{l_1,\ldots,l_\nu}(t)$ characterize modulation interactions of N rhythms with periods T_j.

Poly-PCRP, in their turn, form a subclass of almost-PCRP, which can be represented as

$$\xi(t)=\sum_{j\in Z}\xi_j(t)e^{i\omega_j t},$$

where $\xi_j(t)$ are the stationary connected processes. Almost-PCRP are reduced to poly-PCRP if the Fourier exponents ω_j, can be written as:

$$\omega_j=\sum_{k=1}^{N}r_{jk}\Lambda_k, \tag{11}$$

where r_{jk} are integers. In this case, the basis $\left\{\Lambda_k, k=\overline{1,N}\right\}$ is referred to as finite and integer basis of the set $M=\{\omega_k, k\in Z\}$. If the set M is an arithmetic progression, i.e., $\omega_j=j\omega_0$, $\omega_0=const$ is the difference of this progression, then the class of almost-PCRP is reduced to the class of PCRP. We can describe the latter class using the invariance of its characteristics with respect to shifts by $T=\frac{2\pi}{\omega_0}$.

In contrast to the subclasses that include poly- and bi-PCRP models, the general model of rhythmic oscillations in the form of almost-PCRP allow us to investigate only the interference of rhythms with frequencies ω_j. We can reveal nonlinear properties of the relevant processes and phenomena by studying the structure of elementary oscillations in greater detail with the use of relationship (11), which is satisfied for bi- and poly-PCRP models.

3. ESTIMATION OF PROBABILISTIC CHARACTERISTICS OF OSCILLATIONS

As can be seen from the aforesaid, the requirements to the model of rhythmic variations in the form of PCRP and their generalizations are formulated in terms of the type of temporal variations of probabilistic characteristics. Importantly, these requirements are also applicable to the second-order moment functions. Within the framework of such an approach, we should abandon the assumption that rhythmic variations are reduced to a visible recurrence of values, and perturbations distort a strictly periodic pattern. This assumption provides the basis for a wide use of the additive model for the description of seasonal rhythmic variations of geophysical processes. However, processes that occur in the atmosphere and ocean are essentially nonlinear. Therefore, it is doubtful whether we could represent the temporal variations of the quantities under consideration as mutually independent regular seasonal changes and random fluctuations. The PCRP model is based on the assumption that the relevant processes are interdependent. Obviously, the substantiation of the concept related to this model should be based on empirical data, and we should examine PCRP methods of statistical analysis adequate to the problem under study. Initial approaches to this problem were indicated by Gudzenko [17, 119], who demonstrated that it is possible to estimate the mean and the correlation function either by evaluating appropriate Fourier components or by analyzing the counts sampled through a time interval equal to the correlation period. The first method is referred to as component and the second is called coherent. Dragan [31] thoroughly investigated the properties of PCRP sequences of counts. Based on the periodicity on the average, Dragan also substantiated the applicability of such sequences for statistical estimation of PCRP characteristics [26]. Subsequent stages in the development of the methods of PCRP statistics [2, 6, 7,43-115] involved the solution of problems of estimation theory in the context of the systematic development of the means for statistical analysis of empirical data.

Evidently, the coherent and component methods of estimation can be deduced from the periodicity of the probabilistic characteristics of PCRP. Coherent estimates for the mean and the correlation function are written as:

$$\hat{m}(t)=\frac{1}{N}\sum_{n=0}^{N-1}\xi(t+nT),$$

$$\hat{b}(t,u)=\frac{1}{N}\sum_{n=0}^{N-1}\left[\xi(t+u+nT)-\hat{m}(t+u+nT)\right]\left[\xi(t+nT)-\hat{m}(t+nT)\right].$$

Component estimates for the mean and the correlation function are based on the estimates for the components of the relevant Fourier series, i.e.:

$$\hat{m}(t)=\sum_{l=-N_1}^{N_1}\hat{m}_l e^{il\frac{2\pi}{T}t}, \tag{12}$$

$$\hat{b}(t,u)=\sum_{l=-N_2}^{N_2}\hat{B}_l(u)e^{il\frac{2\pi}{T}t}, \tag{13}$$

where:

$$\hat{m}_l=\frac{1}{\theta}\int_0^\theta \xi(t)e^{-il\frac{2\pi}{T}t}dt, \tag{14}$$

$$\hat{B}_l(u)=\frac{1}{\theta}\int_0^\theta\left[\xi(t)-\hat{m}(t)\right]\left[\xi(t+u)-\hat{m}(t+u)\right]e^{-il\frac{2\pi}{T}t}dt. \tag{15}$$

In formulas (12) and (13), N_1 and N_2 are the numbers of components estimated for the mean and the correlation function, respectively.

The estimates for the mean and the correlation function of PCRP based on coherent averaging of the counts sampled with a time interval equal to the correlation period T employ only one value of the process within the period. Statistics (12) and (13) are based on all the values of the relevant continuous realization. Therefore, if the correlation function of a PCRP displays considerable changes within the interval equal to the correlation period, then, for a given the realization length, component estimates are characterized by a smaller variance that coherent estimates. This advantage of component estimates is especially important if the number of components of characteristics being evaluated is small because the variance of the component estimate increases with the growth of this number.

Coherent and component methods of estimation of probabilistic PCRP characteristics can be considered as particular cases of a more general method of estimation – the method of linear filtration (2, 51, 68):

$$\hat{m}(t)=\int_0^{\theta}\xi(t-\tau)h(\tau)d\tau, \tag{16}$$

$$\hat{b}(t,u)=\int_0^{\theta}\overset{\circ}{\xi}(t+u-\tau)\overset{\circ}{\xi}(t-\tau)h(\tau)d\tau. \tag{17}$$

If the condition:

$$\int_0^{\theta}h(\tau)e^{-ik\frac{2\pi}{T}\tau}d\tau=1,\ k=\overline{-N_1,N_1} \tag{18}$$

is satisfied, then $E\hat{m}(t)=m(t)$, i.e., (16) provides an unbiased estimate. Suppose that $h(\tau)$ is a periodic function, $h(\tau+T)=h(\tau)$. Then, we can write:

$$h(\tau)=\sum_{l\in Z}h_l e^{il\frac{2\pi}{T}\tau}$$

Hence, with allowance for (17) with $\theta=NT$, we find that $h_l=\theta^{-1}$. Expressions (16) and (17) in this case are transformed into the formulae for coherent estimates. If the number of harmonics is finite, the function $h(\tau)$ describes the impulse response of a component filter. Thus, the class of estimates described by (16) and (17) with a periodic weight function $h(\tau)$ is completely covered with coherent and component estimates.

The frequency characteristics of the coherent and component comb filters are given by:

$$H(\omega)=e^{-i\frac{\omega T}{2}(N-1)}\sin\omega\frac{N}{2}T\left(2\pi N\sin\frac{\omega T}{2}\right)^{-1}$$

and

$$H(\omega)=\frac{1}{2\pi}\sum_{k=-M}^{M}e^{-i\left(\omega-k\frac{2\pi}{T}\right)\frac{\theta}{2}}\sin\left(\omega-k\frac{2\pi}{T}\right)\frac{\theta}{2}\left[\left(\omega-k\frac{2\pi}{T}\right)\frac{\theta}{2}\right]^{-1},$$

respectively. The component filter is characterized by a lower level of side lobes as compared with the coherent filter and by a finite number of transmission bands. The number of these bands is determined by the number M of components being estimated. The above-specified properties of the filters mainly account for a higher reliability of component estimates. The difference between component and coherent estimates vanishes as $M\to\infty$.

The better quality of component estimates is due to the use of *apriori* data concerning the number of components of characteristics being estimated. One can employ *apriori* data concerning the correlation structure of the PCRP to find more efficient estimates. One of the possible ways to improve the estimation efficiency is to choose optimal linear filter [70]. In particular, the minimum, on the average, variance of estimates can be achieved with the use of a linear invariant filter, and the minimum variance for an arbitrary moment of time is achieved by means of a filter with periodically varying parameters.

The spectral properties of the PCRP are characterized by a two-frequency spectral density. For the considered class of processes, the spectral density is concentrated only along the straight lines $\omega_2=\omega_1-k\frac{2\pi}{T}$. Therefore, we can reduce the empirical spectral analysis of PCRP to the estimation of components $\hat{f}_k(\omega)$ or the variable spectral density $\hat{f}(\omega,t)$ whose Fourier expansion involves these components [2, 81, 85].

We can form the statistics of the variable spectral density and spectral components using expressions (7), (8), and (10). Similarly to the spectral analysis of stationary random processes, we can derive consistent estimates by smoothing correlograms:

$$\hat{f}(\omega,t)=\frac{1}{2\pi}\int_{-u_m}^{u_m}\hat{b}(t,u)k(u)e^{-i\omega u}du,$$
$$\hat{f}_k(\omega)=\frac{1}{2\pi}\int_{-u_m}^{u_m}\hat{B}_k(u)k(u)e^{-i\omega u}du. \quad (19)$$

The correlation window $k(u)$ is an even function $k(-u)=k(u)$ and $k(0)=1$. For $|u|\geq u$, where u_m is the point of correlogram truncation, $k(u_m)=0$. For a given realization length θ, the variances of the estimates (19) decrease with the narrowing of the correlation window. Fluctuation components of estimate biases should display a similar behaviour. However, by increasing u_m we can reduce the bias components that determine the resolving power of spectral analysis. Such contradictory tendencies in variations of estimate characteristics impede the choice of parameters θ and u_m. For PCRP with known or preset characteristics, we can ensure a substantiated choice of these parameters from evaluated characteristics of the statistical accuracy of estimation. The employed approach allows us to provide recommendations for processing PCRP realizations of specific types. This approach is favourable for revealing general properties of spectral estimates. Note that the results of comprehensive investigations of the properties of spectral estimates for stationary random processes (e.g., see [90-94]) provide the basis for the empirical spectral analysis of PCRP at the initial stage of studies.

The methods of coherent and component estimation are also suitable for the statistical analysis of PCRP-related processes, such as vector-PCRP, bi-PCRP, and poly-PCRP [60, 67, 69,77]. However, generally, characteristics of bi- and poly-PCRP can be estimated only with using of the component method. Coherent averaging is applicable only when data sampling is performed with a small time interval that contains an integer number of periods of other rhythms. If this condition is not satisfied, then we can employ coherent data sampling, where the realization length cannot be matched with time scales of variations in probabilistic characteristics, only to estimate additive components. If we can separate time intervals that include a sufficient number of smaller periods and temporal variations of characteristics corresponding to the larger period are insignificant within these intervals of time, then coherent averaging within such time intervals allow us to evaluate sliding estimates for the characteristics of bi- and poly-PCRP with a satisfactory accuracy. Such an estimation procedure assumes that these generalizations of PCRP can be represented as PCRP with slowly varying characteristics.

Both coherent and component estimates of the characteristics of polyrhythmic processes are biased, and their biases are determined, to a great extent, by the difference of the relevant correlation periods. If this difference is small, then, to ensure the required resolving power, we should choose the processed realization fragment in such a way that its length should be much greater than that required to ensure the smallness of biases when each rhythm is processed separately. It is not always the case that we can meet this requirement. Then, we encounter an urgent problem of using processing methods with a higher selectivity [72, 78, 107, 111-114], in particular, the least-squares method. For PCRP the class of least-squares estimates with $\theta = NT$ coincides with the class of component estimates.

Above, we considered the properties of continuous estimates of probabilistic characteristics. In the overwhelming majority of problems of practical importance, we deal with time series that consist of discrete sequences of values $\xi(nh)$, where h is the sampling interval. We should analyze the influence of the sampling interval on the estimation quality both from the viewpoint of revealing reliable changes in the estimates for the mean, correlation function, and the spectral density in time domain and in the context of the possibility to make reliable conclusions concerning the dependences of the correlation function and correlation components on the lag and the sensitivity of spectral characteristics to the frequency. To obtain intermediate time-domain estimates with using the coherent method, it would be appropriate to employ a trigonometric interpolation [59]. Interpolated estimates are unbiased if the sampling step h satisfies the condition $h \leq \frac{T}{2M+1}$, where M are the numbers of higher order components of characteristics being evaluated. Simultaneously, this condition ensures the unbiasedness of sample estimates for the components [57]. If this condition is not met, component estimates are perturbed by the effect of overlapping. The variances of estimates depend on the correlation properties of PCRP. The sampling interval should be chosen in such a manner as to ensure the smallness of the differences between the reliability of discrete estimates and the reliability of the corresponding continuous estimates.

If the correlation period is not multiple of the sampling interval h, then we should apply the component method of estimation, because coherent data sampling in such a situation would result in the accumulation of errors. The use of the component method simultaneously solves the interpolation problem. For $h \le \frac{T}{2M+1}$, the relation between the qualities of coherent and component discrete estimates is close to that characteristic of continuous processing [86].

If the sampling interval is not matched properly with the temporal structure of a PCRP, then the sample estimates of spectral components are also distorted. In this case, the estimates of the components $f_l(\omega)$ take the values $f_{l+qL}(\omega)$, $q \in Z$, $L = T/h$. The estimation of the spectral components can be also accompanied by additional errors due to the overlapping of the values of the same component at the frequencies $\omega \pm \frac{2k\pi}{\Delta u}$, $k \in Z$, where Δu is the lag discretization interval. To reduce these errors, we should choose the corresponding step Δu. These errors are insignificant if Δu is reduced to such an extent that the values of the spectral components at the frequencies $\omega \pm \frac{2k\pi}{\Delta u}$ are sufficiently small. The step Δu also influences the variance of the estimates. If the values of the spectral characteristics are large outside the interval $\left[-\frac{\pi}{\Delta u}, \frac{\pi}{\Delta u}\right]$, then the statistical accuracy of discrete estimates is considerably lower than the accuracy of continuous estimates.

To eliminate overlapping errors arising when we process time series using the bi-PCRP methods, we should match the sampling step with the numbers of highest order components for both correlation periods. Naturally, evaluating the estimates for the mean and the correlation function, we should satisfy different requirements to the sampling step. If this step is such that $h \le \frac{T_1}{2N_1+1}$ and $h \le \frac{T_1}{2N_1+1}$, where N_l are the numbers of highest order components for each of the periods, then the errors that arise when continuous averaging is replaced by discrete averaging are determined by the difference between the relevant integral transforms of the correlation components and the corresponding integral sums for the given sample step.

In many situations that occur in reality, time intervals that separate different values of series of empirical data are not only variable but also random. This circumstance necessitates the construction and investigation of PCRP statistics in the case of stochastic discretization. Analysis of such series can be based on the modified coherent and component methods (59,66). If the values of a time series are obtained in the so-called stationary discretization regime, then both of these methods provide estimates whose quality is only slightly lower than the quality of estimates evaluated from equidistant time series. In such a situation, the component method displays a higher sensitivity to the deviations of the properties of discretization flows from the stationary flow.

4. DETECTION OF HIDDEN PERIODICITIES

To apply the methods of statistical analysis based on PCRP and their generalizations, we should preliminarily determine the correlation periods for the processes under investigation. In many cases, we can make a decision concerning the values of these periods based on the analysis of the physical nature of the phenomenon under study. Specifically, when studying daily and annual variations of geophysical processes, we can find an obvious solution to the problem associated with the correlation period. Astronomical factors determine a polyrhythmic character of tidal oscillations of the sea level, sea currents, and internal waves. In this case, the correlation periods are determined from the well-known polyharmonic representation of the potential of tide-generating forces. When we investigate rhythmic variations of stochastic intrinsic oscillations that occur in various dynamical systems, the problem of determining the correlation period ceases to be trivial. In such situations, to solve this problem, we should develop methods of estimation that would be adequate to the accepted models [55, 58, 61, 80, 83, 87, 104, 105].

The problem of the detection of hidden periodicities, which is a classical problem of mathematical statistics of random processes, was formulated back in the 19th century [95, 120]. Initially, the solution of this problem was reduced to the evaluation of parameters of a periodic or nearly periodic deterministic function. Subsequently, the detection of hidden periodicities was transformed into the problem of searching for reliable peaks of the power spectral density for stationary random processes.

In terms of the PCRP model, hidden periodicities do not necessarily manifest themselves as peak values of the spectral density [2, 121]. Within the framework of the PCRP model, the detection of periodicities is formulated as a problem of determining the period of temporal variations in the estimates of probabilistic characteristics of stochastic oscillations, such as the mean, correlation function and variable spectral density. Such an approach separates the search for the period of regular oscillations from the determination of the recurrence period of correlations.

Based on the representation (4), we can reduce the problem of the determination of the PCRP correlation period to the problem of parametric estimation. In the case of a Gaussian PCRP, we can employ the maximum likelihood method. The corresponding estimate can be represented in the form of a power series in a certain small parameter ε.

$$\hat{T} = T_0 + \varepsilon T_1 + \varepsilon^2 T_2 + \ldots \quad (20)$$

In the first-order approximation for ε, this estimate is unbiased, and the variance of this estimate coincides with the variance of the efficient estimate. If we increase the realization length being processed, the parameter ε decreases simultaneously with the accuracy of the first-order approximation. Therefore, the maximum-likelihood estimate of the period is asymptotically unbiased and asymptotically efficient. The efficiency of this estimate is due to the use of *apriori* data concerning the probabilistic structure of the PCRP. At the initial stage of investigation, such data are usually lacking, which makes us employ less efficient methods that do not require any *apriori* information. Development of such methods can be based on the concepts of the above-considered methods of coherent and component estimation. Both coherent and component statistics are characterized by certain resonant properties with respect to the correlation period. These statistics reach their extreme at points T that asymptotically tend to the true values of the period. Specifically, the functionals of the mean can be written as:

$$\hat{m}(t)=\frac{1}{2N+1}\sum_{n=-N}^{N}\xi(t+n\tau), \tag{21}$$

$$\begin{Bmatrix}\hat{m}_k^c\\ \hat{m}_k^s\end{Bmatrix}=\frac{1}{\theta}\int_0^{\theta}\xi(t)\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\Lambda_k t dt,\ \Lambda_k=\frac{2\pi}{\tau}, \tag{22}$$

$$m(t,\tau)=\sum_{n=-N_1}^{N_1}\hat{m}_k(\tau)e^{i\Lambda_k t}. \tag{23}$$

Analogously to relations (21)-(23), we can also define the functionals of the correlation function [76].

Similarly to the maximum likelihood method, the estimates of the period are determined from the nonlinear equation:

$$\frac{dS(\tau)}{d\tau}+\varepsilon\frac{dN(\tau)}{d\tau}=0,$$

where $S(\tau)$ and $N(\tau)$ are the regular and fluctuation components of the functionals (21)-(23) respectively and ε is a small parameter. Solutions to this equation can be represented in the form (20). If the correlation function decreases with the growth in the lag, then the small parameter, which is defined by the expression $\varepsilon=\frac{\sqrt{\left[EN^2(T)\right]}}{S(T)}$, tends to zero as $\theta\to\infty$. Therefore, the estimates of the period are asymptotically unbiased and consistent.

The advantage of the statistics (21)-(23), similarly to the statistics of correlation characteristics defined by analogy with (21)-(23), is associated with the fact that, in addition to the correlation period, these statistics allow us to simultaneously determine the characteristics of PCRP.

The developed procedure for the detection and analysis of hidden periodicities was employed in the investigation of wind waves, swell, Wolf number series, geomagnetic pulsations Pcl and Pc3, annual and interyearly rhythmic variations of hydrometeorological processes and vibroacoustic signals. Based on this procedure, we revealed qualitatively new features of the probabilistic structure of rhythmic variations of the above-specified processes [2, 6, 7, 51, 83].

The results of our studies clearly demonstrate that the methods for the detection of hidden periodicities based on functionals (21)-(23) allow us also to make conclusions concerning the applicability boundaries of models of rhythmic variations in the form of PCRP and their generalizations. The proposed approach naturally extends the problem of hidden periodicities searching on the basis of a more logical and comprehensive notion of rhythmic of stochastic oscillations, which, apart from other advantages, removes a severe limitation associated with the requirement that the considered oscillation process should be stationary.

5. CONCLUSIONS

Relying on the results of the performed investigations, we developed, in collaboration with our colleagues, the fundamentals of the theory of the statistical analysis of rhythmic signals on the basis of models in the PCRP form and their generalizations. Within the framework of the spectral-correlation theory of nonstationary random processes of these classes, we substantiated a general approach to the investigation of the stochastic recurrence in phenomena of different nature. It is shown that, using the first- and second-order characteristics of such processes, we can reveal substantial and closely interrelated features of rhythmic variations, such as the periodicity and randomness. Eventually, these features reflect stochastic amplitude and phase modulation of signals. We developed and investigated the methods for the estimation of probabilistic characteristics of PCRP and their generalizations, including the coherent and component methods, the least-squares method and methods of linear filtration. We developed a general approach to the problems of estimation and construction of optimal estimates for PCRP characteristics. On the basis of the PCRP model, we developed methods for the detection and analysis of hidden periodicities. We also created software for the statistical processing of signals with the stochastic recurrence. With the use of these means, we revealed previously unknown properties of rhythmic variations of several physical processes that occur on the Earth, in the atmosphere, in the ocean, in the ionosphere and in various technological systems. The proposed parametric models of stochastic oscillations provide the background for the performing and designing of experimental statistical investigations, simulation and forecasting of various processes and recognition and diagnostics of the states of dynamic systems that generate these processes.

BIBLIOGRAPHY

[1] Rozhkov V., 1979. The probabilistic analysis of oceanological processes methods. Gidrometeoizdat Leningrad (in Russian).

[2] Dragan Ya., Rozhkov V., Javors'kyj I., 1987. The methods of probabilistic analysis of oceanological rhythmic. Gidrometeoizdat Leningrad (in Russian).

[3] Gudzenko L., 1959. The small fluctuation in essentially nonlinear autooscillation system. Dokl. Akad. Nauk USSR 125(l), 62-65 (in Russian).

[4] Malakhov A., 1968. The fluctuation in autooscillation systems. Nauka Moscow (in Russian).

[5] Rytov S., 1976. An introduction to statistical radiophysics. P. I. Nauka Moscow (in Russian).

[6] Mikhailyshyn V., Fligel S., Javors'kyj I., 1990. The probabilistic model of the signal periodicity of geomagnetic pulsations. Pcl. The investigation of the structure and wave properties near-earth plasma. Nauka Moscow, 76-88 (in Russian).

[7] Mikhailyshyn V., Fligel S., Javors'kyj I., 1990. Statistical analysis of wave packets of geomagnetic pulsations Pcl type by the method of periodically correlated random processes. Geomagnetizm i Aeronomia 30(5), 757-764 (in Russian).

[8] Karmalita V., 1986. The digital processing of the random oscillations. Mashinostroenie Moscow (in Russian).

[9] Aschoff Yu.-Ed., 1981. The biological rhythms. Plenum Press New York.

[10] Koronkevich O., 1957. The linear dynamic systems under action of the random forces. Naukovi Zapysky Lvivskoho Universytetu 44(8), 175-183 (in Ukrainian).

[11] Stratonovich R., 1961. Selected problems of the fluctuation theory in radio-engineering. Sov. Radio Moscow (in Russian).
[12] Papoulis A., 1983. Random modulation: a review. IEEE Trans. Acoust., Speech and Signal Processing 31(l), 96-105.
[13] Kozel S., 1959. The transformation of the periodically nonstationary fluctuations by linear filter. Trudy Moskovskoho Fiziko-tekhnicheskoho Instituta 4, 10-16 (in Russian).
[14] Myslovich M., Pryimak N., Shcherbak L., 1980. Periodically correlated random processes in problems of acoustic information processing. Znaniye Kiyev (in Russian).
[15] Tikhonov V., 1956. When nonstationary random process can be substituted for stationary process. Zhurnal Teoreticheskoi i Eksperimentalnoi Fiziki 31(9), 2057--2059 (in Russian).
[16] Gladyshev E., 1959. Periodically and almost periodically correlated processes with continuous time. Teoriya Veroyatnostei i yeyo Primenenie 3(2), 84-189 (in Russian).
[17] Gudzenko L., 1959. On periodically nonstationary processes. Radiotekhnika i Elektronika 6(6), 1026-1046 (in Russian).
[18] Ogura H., 1971. Spectral representation of periodic nonstationary random processes. IEEE Trans-Inform. Theory 17(2), 143-149.
[19] Franks L., 1969. Signal theory. Englewood Cliffs Prentice Hall New York.
[20] Gardner W., 1986. The spectral correlation theory of cyclostationary time-series. Signal Processing 11(l), 3-36.
[21] Cyclostationarity in Communications and Signal Processing, 1994. W. Gardner (ed.). IEEE New York.
[22] Romanenko A., Sergeyev G., 1968. The problems of the applied analysis of random processes. Sov. Radio Moscow (in Russian).
[23] Dragan Ya., 1969. On periodically correlated random processes and systems with periodic parameters. Otbor i Peredacha Informatsiyi 22, 27-33 (in Russian).
[24] Dragan Ya., 1970. On spectral properties of periodically correlated random processes. Otbor i Peredacha Informatsiyi 30, 16-24 (in Russian).
[25] Dragan Ya., 1975. On representation of periodically correlated random processes by stationary components. Otbor i Peredacha Informatsiyi 45, 7-20 (in Russian).
[26] Dragan Ya., 1978. The harmonizability and the decomposition of the random processes with finite average power. Dokl. Akademiyi Nauk USSR, ser. A, no. 8, 679-684 (in Russian).
[27] Dragan Ya., 1980. The structure and representation of the stochastic signal models. Naukova Dumka Kiyev (in Russian).
[28] Dragan Ya., 1985. Periodic and periodically nonstationary random processes. Otbor i Peredacha Informatsiyi 72, 3-17 (in Russian).
[29] Voichishyn K., Dragan Ya., 1971. On simple stochastic model of the natural rhythmic processes. Otbor i Peredacha Informatsiyi 29, 7-15 (in Russian).
[30] Dragan Ya., 1972. About basing of rhythmic stochastic model. Otbor i Peredacha Informatsiyi 31, 12-21 (in Russian).
[31] Dragan Ya., 1972. The properties of the periodically correlated random process samples. Otbor i Peredacha Informatsiyi 33, 9-12 (in Russian).
[32] Dragan Ya., 1972. On biperiodically correlated random processes. Otbor i Peredacha Informatsiyi 34, 12-14 (in Russian).

[33] Voichyshyn K., 1975. The problems of statistical analysis of nonstationary (rhythmic) phenomena conformable to some geophysical object. Cand. Phys. and Math. Sci. Dissertation, Institute of Earth Physics, Moscow (in Russian).
[34] Dragan Ya., Javors'kyj I., 1975. About representation of sea waves by periodically correlated random processes and their statistical processing methods. Methods of Representation and Apparatus Analysis of Random Processes and Fields. XIII All-Union Symposium, Leningrad, 29-33 (in Russian).
[35] Zhukovsky E., 1969. The investigation of the statistical characteristics of relative air humidity. Sbornik Trudov po Agronom. Fizike 20, 3-28 (in Russian).
[36] Kiselyeva T., Chudnovsky A., 1968. Statistical investigation of the diurnal cycle of the air temperature. Bull. Nauchno-tekhnicheskoi Informatsiyi po Agronom. Fizike 11, 17-38 (in Russian).
[37] Mamontov N., 1968. The standard deviation and coefficients of skewness of air temperature in south-east of West-Siberia plain. Trudy NIIAK 54, 29-34 (in Russian).
[38] Mamontov N., 1968. The investigation of distribution statistics of the relative air humidity. Trudy NIIAK 54, 18-28 (in Russian).
[39] Mishchenko Z., 1960. The air temperature diurnal cycle and plant termoperiodicity. Trudy GGO 91, 15-28 (in Russian).
[40] Mishchenko Z., 1966. The air temperature diurnal cycle and its agroclimatic significance. Gidrometeoizdat Leningrad (in Russian).
[41] Kolyesnikova V., Monin A., 1965. On spectrum of meteorological field oscillations. Izvestiya AN USSR, Ser. Fizika Atmosfery i Okeana 1(7), 653-669 (in Russian).
[42] Groisman P., 1977. The estimate of autocorrelation matrixes of the precipitation time series considered as periodically correlated random processes. Trudy GGL 247, 119-127 (in Russian).
[43] Dragan Ya., Javors'kyj I., 1975. About description of sea wave rhythmic. Proc. of VI All-Union Workshop on Statistical Hydroacoustics, Novosibirsk, 197-206 (in Russian).
[44] Dragan Ya., Javors'kyj I., 1977. The description of sea wave rhythmic. Visnyk AN USSR 2, 26-36 (in Russian).
[45] Javors'kyj I., 1978. The investigation of rhythmic of sea waves and their influence on transfer of underwater acoustic signals. Cand. Phys. and Math. Sci. Dissertation. Far-Eastern Scientific Centre of AN USSR, Vladivostok (in Russian).
[46] Javors'kyj I., 1978. The estimates of probabilistic characteristics of sea waves considered as periodically correlated random processes. All-Union Institute of Scientific and Technical Information Moscow.
[47] Dragan Ya., Javors'kyj I., 1980. Statistical processing of space and time signals with rhythmic structure. Signal Spacetime Processing Voronezh: Vor. State University, 77-85 (in Russian).
[48] Dragan Ya., Mezentsev V., Javors'kyj I., 1980. The problem of verification of stochastic rhythmic model. Otbor Peredacha Informatsiyi 59, 3-12 (in Russian).
[49] Javors'kyj I., 1981. The problems of correlation analysis of periodically correlated random processes. Proc. of X All-Union Workshop on Statistical Hydroacoustics, Novosibirsk, 52-54 (in Russian).

[50] Dragan Ya., Mezentsev V., Javors'kyj I., 1981. The problems of spectral processing of rhythmic signals. Thesis of Report of XIII All-Union Conf. on Coding and Transfer of Information, P. IV, Moscow-Kuybyshev, 91-95 (in Russian).
[51] Dragan Ya., Javors'kyj I., 1982. Rhythmic of Sea Waving and Underwater Acoustic Signals. Naukova Dumka Kyiv (in Russian).
[52] Javors'kyj I., 1984. The properties of mean and correlation function estimates of periodically correlated random processes. Otbor i Peredacha Informatsiyi 67, 22-28 (in Russian).
[53] Dragan Ya., Mezentsev V., Javors'kyj I., 1984. The computation of the spectral characteristic estimates of periodically correlated random processes. Otbor i Peredacha Informatsiyi 69, 29-35 (in Russian).
[54] Dragan Ya., Rozhkov V., Javors'kyj I., 1984. The application of methods of periodically correlated random processes theory for probabilistic analysis of oceanological time series. Probabilistic Analysis and Modelling of Oceanological Processes. Gidrometeoizdat Leningrad, 4-23 (in Russian).
[55] Javors'kyj I., 1984. The application of Buys-Ballot scheme for statistical analysis of rhythmical signals. Izvestiya Vysshykh Uchebnykh Zavedenij, ser. Radioelektronica 27(11), 31-37 (in Russian).
[56] Rozhkov V., Trapeznikov Yu., Javors'kyj I., 1984. Statistical analysis of seasonal survey data. Proc. XIV Conf. Baltic Oceanogr. – Gdynia, Polish Academy of Sciences, 370-374.
[57] Javors'kyj I., 1985. The component sample estimates of probabilistic characteristics of periodically correlated random processes. Otbor i Peredacha Informatsiyi 72, 17-27 (in Russian).
[58] Javors'kyj I., 1985. On statistical analysis of periodically correlated random processes. Radiotekhnika i Electronika 6, 1096-1104 (in Russian).
[59] Javors'kyj I., 1985. The computation of characteristics of periodically correlated random processes by data random selection. Thesis of Reports of I Acoustical Seminar on Modelling, Algorithms and Reception of Decision, Moscow, 44-49 (in Russian).
[60] Javors'kyj I., 1986. Statistical analysis of biperiodically correlated random processes. Otbor i Peredacha Informatsiyi 73, 12-21 (in Russian).
[61] Javors'kyj I., 1986.On period estimate of periodically correlated random processes Otbor i Peredacha Informatsiyi 74, 6-13 (in Russian).
[62] Dragan Ya., Javors'kyj I., 1986. On characteristics of the system for analysis of periodically correlated random processes. Statistical Processing of Experimental Data, Novosibirsk Electro-tekhnical Inst., 53-58 (in Russian).
[63] Javors'kyj I., 1986. The component estimates of probabilistic characteristics of periodically correlated random processes. Avtomatika 4, 44-48 (in Russian).
[64] Javors'kyj I., 1987. The statistical analysis of vector periodically correlated random processes. Otbor i Peredacha Informatsiyi 76, 3-12 (in Russian).
[65] Javors'kyj I., 1987. The interpolation of the estimates of periodically correlated random processes. Avtomatika l, 36-41 (in Russian).
[66] Kostjukov Ya., Mezentsev V., Uljanich I., Javors'kyj I., 1987. The analysis of nonequidistant time series of Gulf of Riga water temperature and salinity. Proc. of Ail-Union Scientific Research Institute of Hydrometeorological Information – World Data Centre 139, 87-97 (in Russian).

[67] Javors'kyj I., 1987. Biperiodically correlated random processes as model of bi-rhythmic signals. All-Union Conf. on Information Acoustics, Moscow, 6-10 (in Russian).
[68] Mezentsev V., Javors'kyj I., 1984. The estimation of rhythmic signal probabilistic characteristics as a problem of linear filtration. Izvestia Vysshykh Uchebnykh Zavedenij, ser. Radioelectronika 27(11), 31-37 (in Russian).
[69] Javors'kyj I., 1988. The statistical properties of biperiodically correlated random processes sample sequences. Otbor i Obrabotka Informatsiyi 1(77), 16-23 (in Russian).
[70] Mezentsev V., Javors'kyj I., 1988. The properties of the optimum estimates of rhythmic signal probabilistic characteristics. Thesis of the Reports of All-Union Workshop On Signal Processing, Uljanovsk Polytechnical Institute, 16-18 (in Russian).
[71] Javors'kyj I., 1988. The methods of statistical analysis of periodically correlated and allied random processes and their application in oceanology. Perspective Methods of Experimental Planning and Analysis for Investigation of Random Processes and Fields, Moscow, 102-103 (in Russian).
[72] Javors'kyj I., 1988. The least squares method for statistical analysis of signals with rhythmical structure. II All-Union Acoustical Seminar on Modelling, Algorithms and Reception of Decisions, Moscow, 5-6 (in Russian).
[73] Kostjukov Ya., Mezentsev V., Uljanich I., Javors'kyj I., 1988. The annual cycle structure of the temperature, salinity and level time series at the shore station of the Gulf of Riga. The Regime-forming Factors, Hydrometeorological and Hydrochemical Processes on USSR Seas., Gidrometeoizdat Leningrad, 113-124 (in Russian).
[74] Mjakisheva N., Rozhkov V., Uljanich I., Javors'kyj I., 1989. The estimates of probabilistic characteristics of hydrometeorological processes annual and diurnal changeability. The Regime-forming Factors, Information Base and Methods of its Analysis, Gidrometeoizdat Leningrad, 47-77 (in Russian).
[75] Rozhkov V., Chemyshov S., Trapeznikov Ya., Javors'kyj I., 1989. The comparison of the probabilistic characteristics estimation methods of the oceanological process annual changeability. The Regime-forming Factors, Information Base and Methods of its Analysis. Gidrometeoizdat Leningrad, 138-149 (in Russian).
[76] Javors'kyj I., 1989. The statistical analysis methods of the oceanological process rhythmics. Dokt. Phys. and Math. Sci. Dissertation, Arctic and Antarctic Institute Leningrad (in Russian).
[77] Javors'kyj I., 1989. Statistical analysis of poly- and almost periodically correlated random processes. Otbor i Obrabotka Informatsiyi 3(79), 3-10 (in Russian).
[78] Mikhailyshyn V., Javors'kyj I., 1990. LSM-analysis while identifying poly-rhythmic structure of stochastic signals. Proc. of the First Intern. Conf. on Information Technologies for Image Analysis and Pattern Recognition, Lviv, 321-325.
[79] Mikhailyshyn V., Javors'kyj I., 1991. Probabilistic model of oceanological process rhythmic and its analysis methods. Perspective Methods of Experimental Planning and Analysis for Investigation of Random Processes and Fields, Moscow, 24-25 (in Russian).
[80] Mikhailyshyn V., Javors'kyj I., 1992. The hidden periodicities: models, detecting, analysis. Probabilistic Models and Processing of the Random Signals and Fields, Kharkiv, 107-110 (in Ukrainian).

[81] Panasjuk O., Javors'kyj I., 1992. The properties of periodically correlated random process spectral characteristic estimates by coherent analysis. Ibid., 111-114 (in Ukrainian).

[82] Mikhailyshyn V., Javors'kyj I., 1992. Parametric models of stochastic oscillations and their statistical processing methods. Signal and Image Processing and Pattern Recognition. Proc. of the First All-Ukrainian Conf., Kyiv, 52-54 (in Ukrainian).

[83] Mikhailyshyn V., 1992. The statistical methods of hidden periodicities detecting and their application for analysis of geomagnetic pulsation oscillation structure. Cand. Phys. and Math. Sci. Dissertation, Lviv State University (in Russian).

[84] Mikhailyshyn V., Javors'kyj I., 1994. The probabilistic structure of the season variability of air temperature. Meteorologia i Hydrologia 2, 20-35 (in Russian).

[85] Panasyuk O., Javors'kyj I., 1994. Spectral analysis of signals with stochastic recurrence. Izvestia Vysshykh Uchebnykh Zavedeniy, ser. Radioelektronika 37(12), 47-57 (in Russian).

[86] Isayev I., Javors'kyj I., 1995. Component analysis of the time series with rhythmical structure. Izvestia Vysshykh Uchebnykh Zavedeniy, ser. Radioelektronika 38(1), 34-45 (in Russian).

[87] Mikhailyshyn V., Javors'kyj I., 1993. The component estimation of the period of stochastic recurrence of rhythmical signals. Avtomatika 4, 44-48 (in Russian).

[88] Mikhailyshyn V., Fligel D., Javors'kyj I., 1990. The probabilistic model of the signals periodicity of geomagnetic pulsation Pc1. Investigation of structure and wave properties near-earth plasma. Nauka Moscow, 76-88 (in Russian).

[89] Mikhailyshyn V., Javors'kyj I., 1996. Modulated non-stationary random process. Dopovidi NAN Ukrainy 2, 70-74 (in Ukrainian).

[90] Javors'kyj I., Mikhailyshyn V., 1996. Probabilistic models and investigation of hidden periodicities. Appl. Math. Lett. 9(2), 21-23.

[91] Javors'kyj I., Mikhailishin V., 1996. Probabilistic models and statistical analysis of stochastic oscillations. Pattern Recognition and Image Analysis 6(4), 749-763.

[92] Isayev I., Javors'kyj I., 1996. Component spectral analysis of signals with stochastic recurrence. Problemy Upravlenia i Informatiki 4, 104-113 (in Russian).

[93] Panasyuk O., Javors'kyj I., 1996. Properties of spectral component estimates of periodically correlated random processes. Elektronnoe Modelirovanie 18(2), 17-24 (in Russian).

[94] Mikhailyshyn V., Javors'kyj I., 1996. Statistical analysis of radiophysical processes with hidden periodicity. Radiofizika i Radioastronomia 2, 243-254 (in Russian).

[95] Javors'kyj I., Mikhailyshyn V., 1996. Probabilistic analysis of electromagnetic phenomena with stochastic periodicity. Proc. of VI Intern. Conf. on Mathematical Methods in Electromagnetic Theory, Lviv, 252-255 (in Ukrainian).

[96] Javors'kyj I., Mikhailyshyn V., 1996. Detecting of hidden periodicities of time series generated by non-linear processes in magnetoplazma. Ibid., 596-605 (in Russian).

[97] Mikhailyshyn V., Javors'kyj I., 1997. Narrow-band non-stationary random processes. Radiotekhnika i Elektronika 42(5), 596-605 (in Russian).

[98] Panasyuk O., Javors'kyj I., Karpenko G., 1997. Properties of spectral component estimates in periodically correlated random processes. Engineering Simulation 14, 191-204.

[99] Mikhailyshyn V., Javors'kyj I., Vasilina Yu., Drabich O., Isayev I. Probabilistic models and statistical methods of vibration signal analysis for machine and constructor diagnostics. Fizyko-Khimichna Mekhanika Materialiv 5, 61-74 (in Ukrainian).

[100] Panasyuk O., Javors'kyj I., 1997. Coherent spectral analysis on time series with periodical probabilistic structure. Izvestia Vysshykh Uchebnych Zavedenij, ser. Radioelektronika 40(8), 51-59 (in Russian).

[101] Isayev I., Javors'kyj I., 1998. Spectral analysis of time series with rhythmical structure. Problemy Upravlenia i Informatiki 2, 115-122 (in Russian).

[102] Javors'kyj I., Drzycimski Z., Majewski J., Zakrzewski Z., 1998. Correlation and spectral properties of non-stationary modulated signals. Kwartalnik Elektrotechniki i Elektroniki 2, 21-50 (in Polish).

[103] Javors'kyj I., Zakrzewski Z., 1998. Coherent analysis of non-stationary amplitude modulated signals. Otbor i Peredacha Informacii 12(88), 71-76 (in Ukrainian).

[104] Drabych O., Mykhajlyshyn V., Javors'kyj I., 1999. Detection of regular periodical changes of signal and their periods estimation. Vidbir i Obrobka Informatsiji 13(89), 26-31 (in Ukrainian).

[105] Javors'kyj I., Isayev I., Majewski J., 1999. The properties of component estimators of non-stationary amplitude modulated signal characteristics. Vidbir i Obrobka Informatsiji 13(89), 32-37 (in Ukrainian).

[106] Javors'kyj I., Zakrzewski Z., 2000. Coherent covariance analysis of non-stationary modulated signals. Kwartalnik Elektroniki i Telekomunikacji 46(3), 355--400 (in Poland).

[107] Zabolotnyj O., Mykhajlyshyn V., Javors'kyj I., 2000. The least squares method of polyrhythmic statistical analysis. Dopovidi NAN Ukrainy 8, 93-100 (in Ukrainian).

[108] Javors'kyj I., Isayev I., Majewski J., 2000. The properties of discrete component estimator characteristics of non-stationary modulated signals. Vidbir i Obrobka Informatsiji 14(90), 41-46 (in Ukrainian).

[109] Drabych O., Mykhajlyshyn V., Javors'kyj I., 2000. Determination of correlation period of the periodically correlated random processes using covariation transformations. Vidbir i Obrobka Informatsiji 14(90), 47-52 (in Ukrainian).

[110] Javors'kyj I., Mykhajlyshyn V., 2001. Mathematical models, detection and analysis of hidden periodicities. Vidbir i Obrobka Informatsiji 15(91), 87-95 (in Ukrainian).

[111] Zabolotnyj O., Mykhajlyshyn V., Javors'kyj I., 2002. The likelihood estimate analysis of probability characteristics of signal with polyrhythmic structure. Vidbir i Obrobka Informatsiji 16(92), 47-52 (in Ukrainian).

[112] Zabolotnyj O., Mykhajlyshyn V., Javors'kyj I., 2002. Mean estimate of polyrhythmic time series by the least squares method. Vidbir i Obrobka Informatsiji. 17(93), 65-70 (in Ukrainian).

[113] Javors'kyj I., Mykhajlyshyn V., Zabolotnyj O., 2003. Least Squares Method in Statistical Analysis of Polyrythmics. Appl. Math. Let. 16, 1217-1222.

[114] Zabolotnyj O., Mykhajlyshyn V., Javors'kyj I., 2004. The least squares of probabilistic characteristics of polyrhythmic signals. Izvestija Vysshykh Uchebnych Zavedenij, ser. Radioelektronika 8, 67-75 (in Russian).

[115] Isayev I., Kravets I., Javors'kyj I., 2004. To parametric modelling of periodically correlated random processes. Vidbir i Obrobka Informatsiji, 21(97), 11-165 (in Ukrainian).
[116] Gruza G., 1982. Some general problems of time series statistical analysis in climatology. Proc. of All-Union Scientific Research Institute of Hydrometeorological Information – World Data Centre 83, 3-9 (in Russian).
[117] Poljak I., 1978. Methods of Analysis of Random Processes and Fields in Climatology. Gidrometeoizdat Leningrad (in Russian).
[118] Bjelyshev A., Klevantsov Yu., Rozhkov V., 1983. Probabilistic analysis of sea currents. Gidrometeoizdat Leningrad (in Russian).
[119] Gudzenko L., 1961. The generalization of the ergodic theorem for nonstationary random processes. Izvestia Vysshikh Uchebnykh Zavedenij, ser. Radiofizika IV(2), 265-274 (in Russian).
[120] Serebrjennikov M., Pervozvansky A., 1965. The hidden periodicities detecting. Nauka Moscow (in Russian).
[121] Yaglom A., 1981. Correlation theory of stationary random functions. Gidrometeoizdat Leningrad (in Russian).

MODELE MATEMATYCZNE I ANALIZA OSCYLACJI STOCHASTYCZNYCH

Streszczenie

W pracy przedstawiono właściwości oscylacji stochastycznych przy wykorzystaniu charakterystyk probabilistycznych okresowo skorelowanych procesów stochastycznych i ich uogólnień. Zaprezentowano ogólne podejście do problemów estymacji powyższych charakterystyk na bazie tzw. danych naturalnych. Przeanalizowano również możliwe do zastosowań metody estymacji oraz wyniki badań dotyczących znajdywania ukrytych okresowości.

Słowa kluczowe: oscylacje stochastyczne, okresowo skorelowane procesy stochastyczne, metody estymacji, ukryte okresowości

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 249
TELEKOMUNIKACJA I ELEKTRONIKA 10 (2007), 77-84

EXTENDED PRECISION METHOD FOR ACCUMULATION OF FLOATING-POINT NUMBERS IN DIGITAL SIGNAL PROCESSORS

Paweł Pawłowski, Adam Dąbrowski

Chair of Control and System Engineering
Division of Signal Processing and Electronic Circuits
Poznań University of Technology
Piotrowo 3a, 60-965 Poznań, Poland

In this paper a novel approach for realization of the floating-point arithmetic using an original idea referred to as the two-accumulator concept is proposed. The considered arithmetic operations are typical for various tasks of digital signal processing. The idea presented in this paper overcomes the phenomenon of reduction of accuracy of computations in the case of a long series of additions of small numbers successively added to a relatively large intermediate result. This phenomenon is a substantial drawback of the classic floating-point arithmetic.

Keywords: floating-point arithmetic, digital signal processor, accumulator

1. INTRODUCTION

Digital signal processing is a quickly growing part of the modern electronic world. However, the fast growth and new facilities bring also many problems to the designers and programmers. One of them is the need of assurance of the maximal accuracy of computations at the lowest computation cost [7] for both fixed-point as well as floating-point number representation formats [4, 9, 10]. The IEEE standardized them and defined the floating-point number representation format, which, among other features, guarantees an extremely large dynamic range. Floating-point computations are widely used in software and hardware if a high precision of data processing is required [8, 12]. Unfortunately, this representation has also drawbacks. One of the most important of them is a significant loss of accuracy of the result of long series additions of small numbers successively summed to relatively large intermediate results.

2. NUMBERS IN THE FLOATING-POINT FORMAT

The well-known and widely used IEEE-754 standard was established in 1985 and two years later the generalized IEEE-854 standard appeared [5, 6]. Both of them specify the floating-point number representation formats. Although they are widely used by companies producing processors and software, non-standard realizations can also be found (e.g., in Texas Instruments TMS320C 3x, 4x families [14]). IEEE-754 standard brought the so-called single precision (32-bit) (Fig. 1) and the double precision (64-bit) formats.

31	30							23		22	21		1	0
s	e_7	e_6	e_5	e_4	e_3	e_2	e_1	e_0	**1.**	f_{22}	f_{21}	. . .	f_1	f_0

Fig. 1. Floating-point single precision format
Rys. 1. Format zapisu liczb zmiennoprzecinkowych o pojedynczej precyzji

Single precision numbers consists of a sign bit s, 8-bit biased exponent e and the 23 bits of fraction f of the 24-bit mantissa. If the mantissa is normalized to range <1,2) the most significant bit (MSB) of the mantissa is equal to 1. It is a hidden bit, i.e., it is not represented as the physical bit (see Fig. 1).

Apart from special cases the value of the represented number can be computed as follows [5]:

$$v = (-1)^s 2^{e-127}(1.f) \text{ if } 0 < e < 255 \tag{1}$$

Normalization of the mantissa in the IEEE-754 standard to range <1,2) ensures unique number representations, but lowers precision of the representation of numbers near to zero and brings a problem of representation of zero as the lowest absolute value of the normalized single precision number is equal to:

$$L_{\min\_32bit\_n} = 2^{1-127} \cdot 1 = 2^{-126}. \tag{2}$$

It can be noticed that L_{min_32bit} does not depend on the mantissa length. If we allow the denormalized numbers (i.e., actually fixed-point numbers) defined as:

$$v = (-1)^s 2^{-126}(0.f) \tag{3}$$

we make it possible to represent zero (for $f = 0$) and also extend the precision of the number representation near to zero. The lowest absolute value of the single precision number greater than zero is then equal to:

$$L_{\min\_32bit\_nn} = 2^{-126} \cdot 2^{-23} = 2^{-149} \tag{4}$$

This equation is valid for single precision numbers only, because it depends on the mantissa length. If we extend the mantissa to 32-bit (a 40-bit extension of the 32-bit single precision format) we obtain (5)

$$L_{\min\_40bit\_nn} = 2^{0-126} \cdot 2^{-31} = 2^{-157}. \tag{5}$$

Most of the contemporary DSPs realize in hardware the single precision arithmetic at the normalized mantissas. If we want to allow for the denormalized numbers, the arithmetic unit should be more complicated and it would consume more power. On the contrary, the commonly used PC software (e.g., Matlab) realizes the double precision arithmetic and uses (certainly only if needed) the denormalized numbers near to zero.

3. ADDITION OF FLOATING-POINT NUMBERS

3.1. ALGORITHM OF FLOATING-POINT ADDITION

Floating-point addition is a much more complicated operation than the fixed-point addition. The floating-point addition can be split into four steps:

a) comparison of exponents of the numbers to be added,

b) equalization of exponents of both numbers. An exponent of the lower number should be increased together with the right shift of the mantissa. It should be noticed that during this shifting the hidden 1 (MSB of the fraction) appears as a physical bit,
c) addition of the mantissas,
d) normalization of the result. If a mantissa of the result remains in range <1,2), it should be unchanged. If the mantissa is greater than or equal to 2 it should be right shifted by one bit to the normalized value together with the decrease by one of the exponent.

In case of a negative number or subtraction, the whole process should be realized as in the addition case except for replacing the addition of the mantissas with the subtraction of them. Normalization of the mantissa, which is less than 1 consists in left shifting together with incrementing the exponent.

3.2. SOURCES OF ERRORS

Sources of errors in the floating-point addition described above are in steps b) and d). An important error is made at the correction step of the fraction of the smaller number after the equalization of the exponents. In the extreme case, if we assume m-bit of the fraction and the difference between exponents greater than m+1, the lower number will be zeroed. The result will stay equal to the greater number and the error of such addition will be equal to the lower number. Thus we loose as many bits of the fraction as they are needed to represent the difference between the exponents of the added numbers.

Mainly because of this, Analog Devices (in the Sharc DSP family) but also other companies extended the single precision IEEE-754 standard and offer fractions longer by 8 bits, i.e., 32-bit long fractions (including the hidden bit) [2, 3].

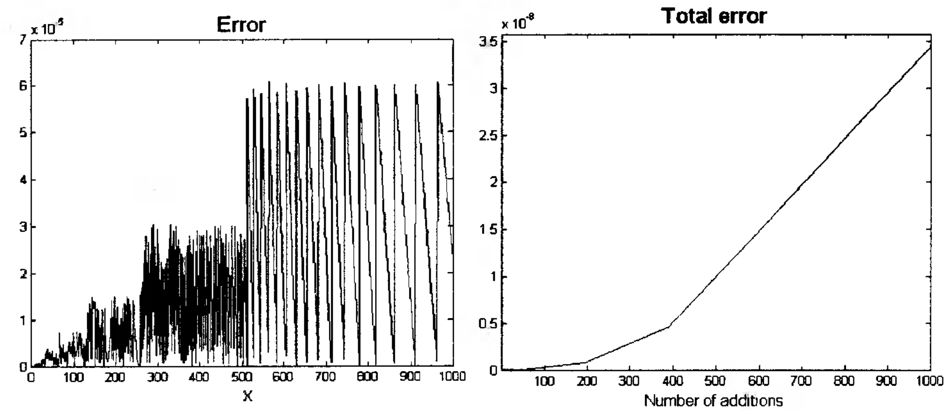
The second error source is the fraction normalization of the result but this can bring at most a loss of merely one bit.

Rarely overflow and underflow errors can occur, if the result exceeds the dynamic range, but these situations should not appear in the right functioning algorithms.

Figure 2a shows an error of a single addition of a small constant to the variable x, which changes from 0 to 1000. It can be noticed that at each power of 2 (i.e., 128, 256, 512) the error rises. The reason for this phenomenon results from the sources described above. Figure 2b depicts the total error of a long series (1 000) of additions (accumulations) of a small constant $x = 1.25\cdot 10^{-6}$.

3.3. TWO-ACCUMULATOR METHOD

To reduce errors produced during a long series of additions (accumulations) [11], the authors proposed to use two accumulators. The proposed modified algorithm of the floating-point addition computes a difference between exponents of the numbers to be added. If this difference is lower than the given threshold Tr, the input number is accumulated to the first accumulator. Otherwise two following steps are made: the value of the first accumulator is added to the number stored in the second accumulator and the input number is moved to the first accumulator. Variations of contents of both accumulators during multiple accumulation of number $W_m = 1.25 \cdot 10^{-5}$ are presented in Fig. 3. Accumulator 1 works periodically with small numbers, while accumulator 2 stores large portions of the results. If intermediate results are needed (in most cases they are not) the contents of both accumulators should be added.



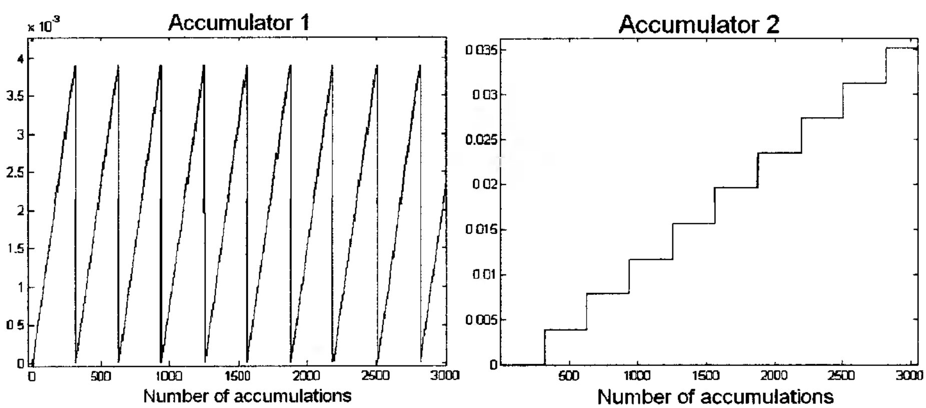
(a) (b)

Fig. 2. (a) Error of the addition of x plus a constant $1.25\cdot10^{-6}$
(b) Total error of a long series of additions of a constant $1.25\cdot10^{-6}$

Rys. 2. (a) Błąd dodawania zmiennej x i stałej $1.25\cdot10^{-6}$
(b) Błąd wielokrotnej akumulacji stałej $1.25\cdot10^{-6}$

4. DIGITAL GENERATOR AS APPLICATION EXAMPLE

To verify the proposed method the authors designed a precise digital generator of sine waveforms with slow frequency modulation (FM). The design was based on the ADSP-21061 Sharc DSP by Analog Devices [1, 2]. Simulations were made with the prepared software realizing the considered DSP arithmetical unit using the Matlab computing environment.



(a) (b)

Fig. 3. Variations of accumulator 1 – (a) and 2 – (b) contents during the multiple accumulation of number $W_m = 1.25\cdot10^{-5}$

Rys. 3. Zmienność zawartości akumulatora 1 – (a) i 2 – (b) podczas wielokrotnej akumulacji liczby $W_m = 1,25\cdot10^{-5}$

4.1. DIGITAL GENERATOR OF SINE WAVEFORMS WITH FM

Real-time digital generation of sine waveforms with frequency modulation by a triangle signal can be realized as follows

$$x_n = A \cdot \sin(\varphi_n) \tag{6}$$

where:

$$\varphi_n = \varphi_{n-1} + W_g \cdot (M_n + 1) \tag{7}$$

$$M_n = M_{n-1} + W_m \tag{8}$$

$$W_g = 2\pi \cdot T_s \cdot f_g \tag{9}$$

$$W_m = \pm \frac{4 \cdot g \cdot f_m}{f_s} \tag{10}$$

A – amplitude,
f_g – frequency of the sine signal (without modulation),
f_m – frequency of the modulation signal,
f_s – sampling rate,
g – modulation factor [13].

It can be noticed that accumulations occur in two places: during the computation of the phase of the sine and during the computation of M_n (accumulation of W_m).

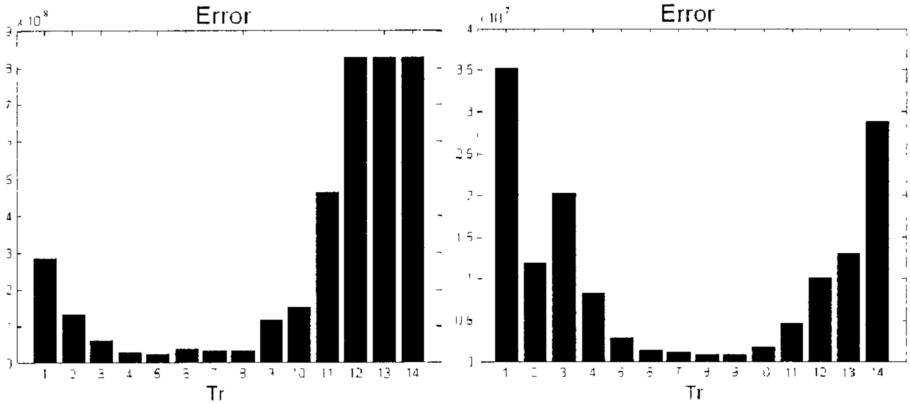
If we assume: $f_s = 48kS/s$, $f_m = 0.5Hz$, $g = 0.3$ we get

$$W_m = 1.25 \cdot 10^{-5} \tag{11}$$

The maximal value of W_m is equal to g and the difference of exponents of the accumulator and of W_m is equal to 15. Thus the classic addition brings the loss of 15 or 14 bits of the fraction of the lower number. These errors of addition lead to a not acceptable mismatch in parameters of the generated signal. A partial solution of this problem is simply to extend the fraction size. The usage of 40-bit extended single precision format brings significant improvement of the signal quality (see Tab. 1) [1, 2, 13]. However, a much better solution is to use two 40-bit accumulators as it is explained below.

4.2. SIMULATIONS AND MEASUREMENT RESULTS

On the basis of the software experiments with the arithmetical unit realizing accumulation with the presented two-accumulator method the authors made a number of simulations to check effectiveness of the proposed idea. The designed arithmetical unit utilizes two 40-bit accumulators. Fig. 4a presents the maximal error of accumulation of number $W_m = 1.25 \cdot 10^{-4}$ in the modulation process (cf., Section 4.1) and $W_m = 1.25 \cdot 10^{-5}$ (Fig. 4b) at different thresholds Tr. In the modulation process we have 2 400 accumulations in the case presented in Fig. 4a and 24 000 accumulations in Fig. 4b. A significant lowering of the total error could be noticed at thresholds equal to 5 and 9, respectively, at the considered experiment conditions.

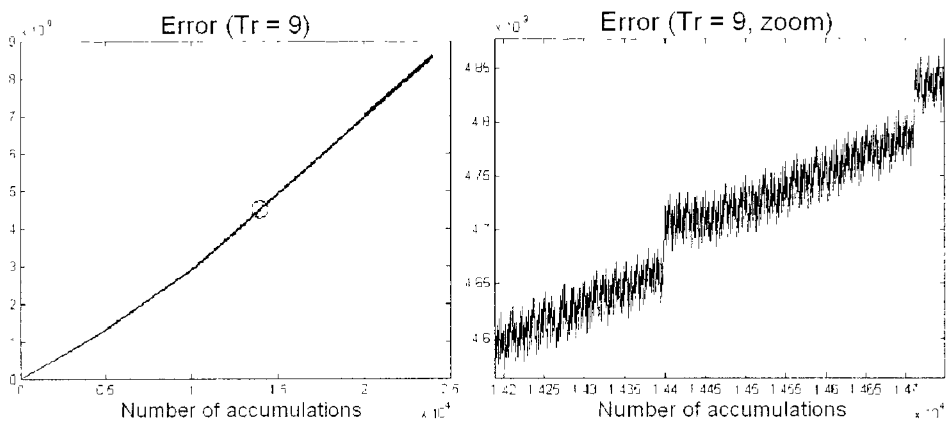


(a) (b)

Fig. 4. Maximal error of accumulation of number $W_m = 1.25\cdot10^{-4}$ (a), $W_m = 1.25\cdot10^{-5}$ (b) at different Tr

Rys. 4. Maksymalny błąd akumulacji liczby $W_m = 1.25\cdot10^{-4}$ (a), $W_m = 1.25\cdot10^{-5}$ (b) dla różnych wartości progu Tr

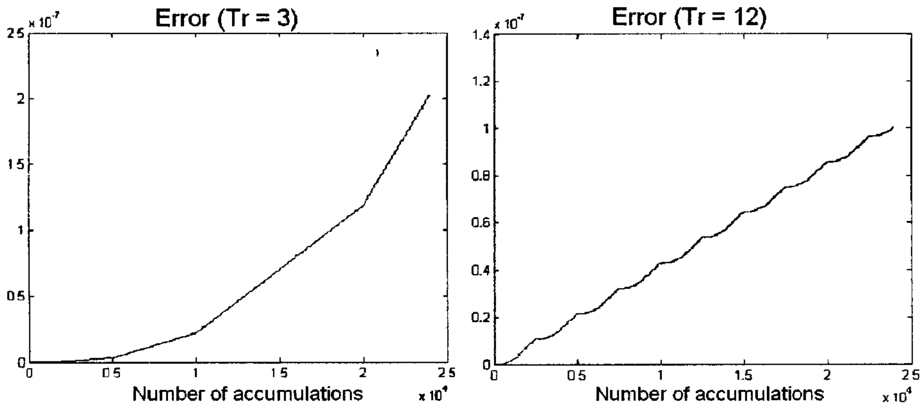
The plot of the accumulation error at the optimal threshold Tr is shown in Fig. 5. The curve is close to linear, but it contains oscillations visible in the zoomed plot. At the low threshold (Fig. 6a) the error looks like in the classic adder with only one accumulator (cf., Fig. 2) and the broken line is visible. Large threshold (Fig. 6b) brings more errors in accumulator 1, visible as the local growth of the slope of the error curve. The three mentioned methods are compared in Table 1.



(a) (b)

Fig. 5. (a) Maximal error of accumulation of number $W_m = 1.25\cdot10^{-5}$ at Tr = 9 and the zoomed fragment (b)

Rys. 5. (a) Maksymalny błąd akumulacji liczby $W_m = 1.25\cdot10^{-5}$ dla Tr = 9 oraz powiększony fragment (b)



(a) (b)

Fig. 6. Error of accumulation of number $W_m = 1.25\cdot10^{-5}$ at Tr = 3 (a) and Tr = 12 (b)
Rys. 6. Błąd akumulacji liczby $W_m = 1{,}25\cdot10^{-4}$ dla Tr = 9 i Tr = 12 (b)

Table 1. Accumulation of number W_m
Tabela 1. Porównanie metod akumulacji liczby W_m

Accumulation method	Accumulator length	Maximal computation error
Single accumulator	32 bit	$2.2\cdot10^{-4}$
Single accumulator	40 bit	$4.3\cdot10^{-7}$
Two accumulators (Tr = 9)	40 bit	$8.6\cdot10^{-9}$

5. SUMMARY

The experiments have shown that the proposed method of two accumulators substantially extends the precision of the data accumulation. Considerations also point to the DSP realization of the modified accumulation algorithm. It does not bring important complications to the digital signal generation software running at the DSP platforms but brings significantly greater accuracy of parameters. The authors plan to continue this research and to start the design project in order to produce a hardware arithmetical unit on chip, which will use the presented two-accumulator method on the fly without any intervention of the user.

BIBLIOGRAPHY

[1] Analog Devices, 2004. ADSP-2106x SHARC Processor, User's Manual. Analog Devices, Inc., Rev 2.1.
[2] Analog Devices, 2000. ADSP-2106x SHARC DSP Microcomputer Family, ADSP-21061/ADSP-21061L. Analog Devices, Inc., Rev B.
[3] Analog Devices, 2001. ADSP-21160M SHARC DSP Microcomputer. Analog Devices, Inc., Rev. 0.
[4] Fang C.E., Rutenbal R.A., Chen T., 2003. Fast, accurate static analysis for fixed-point finite-precision effects in DSP designs. ICCAD'O3, November 11-13. Sanhose, California, USA, 275-282.

[5] IEEE Standard Board, 1985. "IEEE Standard for Binary Floating-Point Arithmetic". IEEE Std 754.
[6] IEEE Standard Board, 1987. "IEEE Standard for Radix-Independent Floating-Point Arithmetic". IEEE Std 854.
[7] Kolagotla R.K., Fridman J., Aldrich B.C., Hoffman M.M., Anderson W.C., Allen M.S., Witt D.B., Dunton R.R., Booth L.A. Jr., 2002. High Performance Dual-MAC DSP Architecture. IEEE Signal Processing Magazine, July 2002, 42-53.
[8] Kramer W., 1998. A priori worst case error bounds for floating-point computations. IEEE Trans. Comp. 47, July 1998, 750-756.
[9] Lewis D.M., 1995. 114 MFLOPS logarithmic number system arithmetic unit for DSP applications. IEEE Journal of Solid-State Circuits 30(12), 1547-1553.
[10] Mahesh M.N., Mehendale M., 1999. Improving performance of high precision signal processing algorithms on programmable DSPs. Proc. of the 1999 IEEE International Symposium on Circuits and Systems 3, 488-491.
[11] Olejniczak M., 1993. Digital filters realizations with the use of floating-point arithmetic (in Polish). Ph. D. Thesis. Poznan University of Technology (unpublished).
[12] Paliouras V., Karagianni K., Stouraitis T., 2000. A Floating-Point Processor for Fast and Accurate Sine/Cosine Evaluation. IEEE Trans. on Circuits and Systems - II: Analog and Digital Signal Processing 47(5), 441-451.
[13] Portalski M., Pawlowski P., Dabrowski A., 2005. Synthesis of some class of non-harmonic tones. XXVIII, IC-SPETO International Conference on Fundamentals of Electrotechnics and Circuit Theory 2, 439-442.
[14] Texas Instruments, 1994. TMS320C4x User's Guide," Texas Instruments Inc.

This work was supported by the grant No. 93-1580 and BW 93-41/07.

METODA ZWIĘKSZANIA PRECYZJI AKUMULACJI LICZB ZMIENNOPRZECINKOWYCH W CYFROWYCH PROCESORACH SYGNAŁOWYCH

Streszczenie

W pracy zaprezentowano metodę poprawy precyzji akumulacji liczb zmiennoprzecinkowych z wykorzystaniem dwóch akumulatorów. Rozważane operacje arytmetyczne są typowym zadaniem wielu algorytmów cyfrowego przetwarzania sygnałów. Zaprezentowane przez autorów rozwiązanie znacząco zmniejsza wpływ utraty precyzji wyniku wielokrotnych dodawań małych liczb do dużych względem nich wyników pośrednich. Własność ta jest istotną wadą klasycznych arytmetyk zmiennoprzecinkowych.

Słowa kluczowe: arytmetyka zmiennoprzecinkowa, cyfrowy procesor sygnałowy, akumulator

ISSN 0209-0589