# COMMUNICATIONS

*Dear reader,*

*Although this volume of the Scientific Letters of the University of Zilina is devoted mostly to mathematics, there are also other scientific areas. Mathematics has been a very important tool used for investigations of various problems in technical and natural sciences for a long time. Currently also the humanities and social sciences consider mathematics to be useful and helpful for scientific work.*

*Cooperation of mathematicians and physicists with their colleagues from technical departments has a long tradition and rich history at our university. The topic of this issue was chosen taking into account the goal – to point out the link between mathematics and technical sciences in research not only at our university, but also at other universities both in Slovakia and abroad. Apart from some authors from the University of Zilina we succeeded to get papers written by scientific workers from other institutions as well. The papers cover mathematics in technical sciences and also in some natural sciences and the humanities.*

*In this issue you will find articles supporting the idea that mathematics will certainly not lose its important role at research in a wide spectrum of sciences.*

*Mariana Marcokova*

Tomas Kisela *

# APPLICATIONS OF THE FRACTIONAL CALCULUS: ON A DISCRETIZATION OF FRACTIONAL DIFFUSION EQUATION IN ONE DIMENSION

*The paper discusses the problem of classical and fractional diffusion models. It is known that the classical model fails in heterogeneous structures with locations where particles move at a large speed over a long distance. If we replace the second derivative in the space variable in the classical diffusion equation by a fractional derivative of order less than two, we obtain the fractional diffusion equation (FDE) which is more useful in this case. In this paper we introduce a discretization of FDE based on the theory of the difference fractional calculus and we sketch a basic numerical scheme of its solution. Finally, we present some examples comparing classical and fractional diffusion models.*

## 1. Introduction

Fractional calculus is a mathematical discipline dealing with derivatives and integrals of non-integer orders. Its story started around 1695 when Leibniz first suggested the idea of half-derivative in his correspondence with l'Hospital. During centuries many mathematicians have contributed to the discussion on the "right" definition of a fractional derivative and various approaches have been proposed (see [13]). However, this concept of fractional derivatives and integrals has been still considered only as a pure theoretical matter because no real applications were known at that time.

The situation changed in last decades when many physical and engineering problems turned out to be described more precisely using mathematical tools of the fractional calculus. We focus on one of such phenomena – anomalous diffusion which is faster than the classical model predicts.

The main purpose of this paper is to obtain a numerical solution of the fractional diffusion equation (FDE). In general, considering the discrete approximations of the fractional differential equations, a lot of questions still remain open. One of them is the problem of nonzero initial or boundary conditions of the Riemann-Liouville type (i.e. when the values of fractional derivatives and integrals are given at point zero). Although we present only the example of the FDE with boundary conditions, we believe that the numerical scheme proposed in this paper allows to involve initial or boundary conditions of any type and any values.

The paper is organized as follows. In Section 2 we recall some basic notions of the continuous fractional calculus. Its discrete analogue is considered in the next section. Section 4 summarizes some known facts about the FDE and outlines the background of this model. In Section 5 we introduce a discretization of the FDE based on the theory presented in Section 3. Finally, we illustrate the method with several examples.

## 2. Fundamentals of the Fractional Calculus

As we foreshowed above, various definitions of a fractional derivative and integral appeared during the history of the fractional calculus and many of them are used up to this day. More information can be found in relevant monographs, e.g. [10], [11], [13], [14]. We are going to recall and employ the Riemann-Liouville approach which is the most frequent one.

It originates from the Cauchy formula for $m$-multiple integral

$$I_a^m f(x) = \int_a^x \frac{(x-\xi)^{m-1}}{(m-1)!} f(\xi) d\xi, \quad m = 1, 2, 3 \ldots \tag{1}$$

which can be naturally generalized. Considering the Euler Gamma function

$$\Gamma(t) = \lim_{n\to\infty} \frac{n! n'^{-1}}{t(t+1)\ldots(t+n)}, \quad t \neq -j, j = 0, 1, 2\ldots$$

with the property $\Gamma(m) = (m-1)!$, one can easily generalize the formula (1) for noninteger $m$.

Fractional derivatives are then usually constructed by the repeated differentiation of an appropriate fractional integral.

On this account we formulate the following definition.

* **Tomas Kisela**
Institute of Mathematics, Brno University of Technology, Brno, Czech Republic, E-mail: kisela.tomas@gmail.com

**Definition 2.1** Let $\alpha \in \boldsymbol{R}^+$, $a,b \in \boldsymbol{R}$ be such that $a < b$. Let $f(x)$ be an integrable function on $[a, b)$. We define the left fractional integral of $f(x)$ by

$$ {}_a\boldsymbol{D}_x^{-\alpha} f(x) = \int_a^x \frac{(x-\xi)^{\alpha-1}}{\Gamma(\alpha)} f(\xi) d\xi\,, \quad x \in [a, b). $$

Further, let $m \in Z^+$ be such that $m - 1 < \alpha \leq m$ and let $f(x)$ be $m$-times differentiable on $[a, b)$ except for a set of the zero measure. Then we define the left fractional derivative of $f(x)$ by

$$ {}_a\boldsymbol{D}_x^{\alpha} f(x) = \frac{d^m}{dx^m}\, {}_a\boldsymbol{D}_x^{-(m-\alpha)} f(x)\,, \quad x \in [a, b). $$

Finally, we put ${}_a\boldsymbol{D}_x^0 f(x) = f(x)$ for $x \in [a, b)$.

The left fractional derivative and integral (jointly called the left differintegral) are obviously global operators employing only points from the left of the point $x$ (if the variable $x$ has the meaning of time, the operator depends on past). If $x$ has the meaning of a space variable, there is no reason to prefer the left-side definition to the following right-side one.

**Definition 2.2.** Let $\alpha \in \boldsymbol{R}^+$, $a,b \in \boldsymbol{R}$ be such that $a < b$. Let $f(x)$ be an integrable function on $(a, b]$. We define the right fractional integral of $f(x)$ by

$$ {}_x\boldsymbol{D}_b^{-\alpha} f(x) = \int_x^b \frac{(\xi-x)^{\alpha-1}}{\Gamma(\alpha)} f(\xi) d\xi\,, \quad x \in (a, b]. $$

Further, let $m \in Z^+$, be such that $m - 1 < \alpha \leq m$ and let $f(x)$ be $m$-times differentiable on $(a, b]$ except for a set of the zero measure. Then we define the right fractional derivative of $f(x)$ by

$$ {}_x\boldsymbol{D}_b^{\alpha} f(x) = \left(-\frac{d}{dx}\right)^m {}_x\boldsymbol{D}_b^{-(m-\alpha)} f(x)\,, \quad x \in (a, b]. $$

Finally, we put ${}_x\boldsymbol{D}_b^0 f(x) = f(x)$ for $x \in (a, b]$.

If we need to consider all points of the interval $(a, b)$, we have to utilize both left and right differintegral. Sometimes it may be useful to employ the so-called Riesz fractional integral.

**Definition 2.3** Let $\alpha \in \boldsymbol{R}^+$, and $a,b \in \boldsymbol{R}$ be such that $a < b$. Let $f(x)$ be an integrable function on $[a, b]$. Then we define the Riesz fractional integral of $f(x)$ by

$$ {}_a\boldsymbol{D}_{b,x}^{-\alpha} f(x) = \frac{{}_a\boldsymbol{D}_x^{-\alpha} f(x) + {}_x\boldsymbol{D}_b^{-\alpha} f(x)}{2}\,, \quad x \in [a, b]. $$

Properties of the Riemann-Liouville and Riesz differintegrals are discussed in more details in [10], [11], [13], [14]. Among numerous interesting properties of these operators we emphasize at least their linearity.

## 3. Discrete Fractional Calculus

Contrary to the continuous case, the theory of discrete fractional calculus is much less developed. Its concept starts from the time-scale theory, where the notions of delta and nabla differences are generalized to other discrete settings ([7], [12]). We are not going to recall here related definitions of the time-scale theory (see [4]) because we work only with a special setting anyway, namely with a finite set of equidistant points with a stepsize $h \in \boldsymbol{R}^+$. Therefore we introduce the set of mesh points $\boldsymbol{T} = \{x_n = nh, n = 0,1, \ldots, N\}$, where the research of fractional differences is advanced (for the methods allowing to solve fractional difference equations see [1]).

In our conception, the notion of the $h$-backward (nabla) difference is a starting point for the left fractional difference operators. Note that some authors study the $h$-forward (delta) difference calculus. However, this approach causes that the fractional differences are defined at points which may not belong to the original mesh (see [1], [5]) and therefore it is not suitable from the numerical point of view.

Before we proceed to definitions of fractional differences and sums, we remind the definitions of nabla and delta differences on the mesh $\boldsymbol{T}$.

**Definition 3.1** Let $f$ be a function defined on the mesh $\boldsymbol{T}$ and put $f_n = f(x_n)$. Then the nabla difference is defined by

$$ \nabla_n f_n = \frac{f_n - f_{n-1}}{h}\,, \quad 1 \leq n \leq N, $$

and the delta difference by

$$ \Delta_n f_n = \frac{f_{n+1} - f_n}{h}\,, \quad 0 \leq n \leq N-1. $$

Further, let $m \in Z^+$. Then the $m$-th differences are defined by the recursion formulas $\nabla_n^m f_n = \nabla_n(\nabla_n^{m-1} f_n)$ and $\Delta_n^m f_n = \Delta_n(\Delta_n^{m-1} f_n)$ respectively.

In the following definition we use the binomial coefficient where

$$ \binom{v}{n} = \frac{\Gamma(v+1)}{\Gamma(v-n+1)\Gamma(n+1)}\,, \quad v \in \boldsymbol{R}, n \in Z^+ \cup \{0\}. $$

**Definition 3.2** Let $\alpha \in \boldsymbol{R}^+$, $x_n \in \boldsymbol{T}$. We define the left fractional sum of $f$ at $x_n$ by

$$ {}_0\nabla_n^{-\alpha} f_n = h^\alpha \sum_{k=1}^{n} \binom{n-k+\alpha-1}{n-k} f_k\,, \quad 0 \leq n \leq N. $$

Further, let $m \in Z^+$ be such that $m - 1 < \alpha \leq m$ and $m < N$. Then we define the left fractional difference of $f$ at $x_n$ by

$$ {}_0\nabla_n^{\alpha} f_n = \nabla_n^m \left({}_0\nabla_n^{-(m-\alpha)} f_n\right), \quad m \leq n \leq N. $$

Finally, we put ${}_0\nabla_n^0 f_n = f_n$.

Since our mesh $\boldsymbol{T}$ is finite, it is meaningful to define the right fractional sum and difference. Because of the consistency with the previous Definition 3.2 we have to utilize the notation based on delta differences.

**Definition 3.3** Let $\alpha \in \boldsymbol{R}^+$, $x_n \in \boldsymbol{T}$. We define the right fractional sum of $f$ at $x_n$ by

$$ {}_n\Delta_N^{-\alpha} f_n = h^\alpha \sum_{k=n}^{N-1} \binom{k-n+\alpha-1}{k-n} f_k, \ 0 \le n \le N. $$

Further, let $m \in Z^+$ be such that $m - 1 < \alpha \le m$ and $m < N$. Then we define the right fractional difference of $f$ at $x_n$ by

$$ {}_n\Delta_N^{\alpha} f_n = (-\Delta_n)^m \left( {}_n\Delta_N^{-(m-\alpha)} f_n \right), \ 0 \le n \le N - m. $$

Finally, we put ${}_n\Delta_N^0 f_n = f_n$.

Last we introduce a discrete analogy of the Riesz fractional integral.

**Definition 3.4** Let $\alpha \in \boldsymbol{R}^+$, $x_n \in \boldsymbol{T}$. We define the Riesz fractional sum of $f$ at $x_n$ by

$$ {}_0\diamondsuit_{N,n}^{-\alpha} f_n = \frac{ {}_0\nabla_n^{-\alpha} f_n + {}_n\Delta_N^{-\alpha} f_n}{2}, \ 0 \le n \le N. $$

## 4. The Origin of the Fractional Diffusion Equation

Roughly speaking, the diffusion is a process when particles spread from areas of a high concentration to areas of a low concentration. The classical diffusion model in one dimension is described by the well-known partial differential equation

$$ \frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t), \qquad (2) $$

where $u(x,t)$ is an unknown concentration of particles and $f(x,t)$ is a known source term. Besides, we need an initial condition and, moreover if the region is bounded, then also boundary conditions.

However, this model is insufficient in the case of fast diffusion, i.e. in the process describing the spreading of particles according to different concentration, but with a significantly greater number of very fast particles. The fast diffusion occurs, e.g., in porous media and it is traditionally described by adding a nonlinearity into the equation (2).

The fractional calculus allows us to model fast diffusion too, but the equation remains still linear. The corresponding FDE has the form (see [2], [3], [14])

$$ \frac{\partial u(x,t)}{\partial t} = \frac{\partial^2}{\partial x^2} \, {}_{-\infty}\boldsymbol{D}_{\infty,x}^{\alpha-2} u(x,t) + f(x,t), \ \alpha \in (1,2], \qquad (3) $$

where we write the second derivative and the Riesz integral separately due to the discretization process which appears later. The order of differintegration $\alpha$ allows us to control the speed of spreading. Note that the Riesz integral is associated with the space variable $x$ and from the computational viewpoint the time variable $t$ plays a role of a parameter.

From the statistical point of view the derivation of the FDE (3) is quite analogical to the classical case. We start from the usual random walk but in our case we remove the assumption concerning a finite variance of particle jumps. In other words, we accept the power-law behaviour for long jumps (sometimes called Levy flights). The process of statistical derivation directly implies the fundamental solution of the corresponding fractional Cauchy problem for (3), which coincides with Levy $\alpha$-stable distributions (we emphasize that this solution cannot be expressed analytically and it is defined via its inverse Fourier transform). This is another interesting analogy because Levy $\alpha$-stable distributions play in the generalized central limit theorem the same role as the gaussian distribution in the classical central limit theorem. For more information see e.g. [3], [14] or more detailed [2]. An alternative derivation utilizing the eulerian model based on finite volumes approach was proposed in [17].

Interpreting the fractional term in (3), we note that in the classical model we assume that the particle flux $J(x, t)$ depends only on the neighbourhood of $x$ and, therefore, is local. On the other hand, the fractional model deals with situations when particles move so fast that we cannot neglect even the effect of particles being far from a particular point. Hence, the flux has a global character and is proportional to the derivative of the Riesz $(2 - \alpha)$-integral of the concentration $u(x, t)$, i.e.

$$ J(x,t) = -\frac{\partial}{\partial x} \, {}_{-\infty}\boldsymbol{D}_{\infty,x}^{\alpha-2} u(x,t). \qquad (4) $$

There are other ways of the generalization of equation (2). For example, we can introduce weight coefficients of the left and the right differintegral for modeling anisotropic media, or we may consider a fractional derivative with respect to time on the left side of the equation. However, in this paper we are focused just on the case described above, in particular, we intend to discuss the numerical solution of the equation (3) on a bounded domain which is not so well established topic like e.g. the Cauchy problem.

Now, we introduce the problem we are going to discuss. First, the studied equation should be slightly changed because the formula (3) was derived for an infinite line while we are going to deal with the compact interval [0, 1]. Hence, we assume that the concentration $u(x, t)$ is zero everywhere except for $x \in [0, 1]$ where the equation (3) holds. Consequently, we can change the bounds of the differintegration from $-\infty$ and $\infty$ to 0 and 1 respectively.

The initial condition is standard: we suppose that the distribution of the concentration $u(x, t)$ is known for all $x$ at the time beginning $t = 0$. Considering the boundary conditions, the situation turns out to be much more complicated. From the mathematical viewpoint there are two possible types of boundary conditions for this FDE. First, we may consider a generalization of the Dirichlet condition, i.e. to prescribe a value for $(2 - \alpha)$-integral, or a generalization of the Neumann condition which deals with a value of $(\alpha - 1)$-derivative.

However, we wish to model some real situation so that the question of physical interpretation of the boundary conditions arises. The solution of this issue is not widely accepted so far. In [8] the authors suggest the idea of inseparable twins which essentially means that we have to find another physical quantity which

corresponds to the needed fractional derivative and then we interpret the boundary condition via this new quantity. In our case, the inseparable twins are settled by the relation (4), which gives a clear physical meaning to the Neumann condition (flux). The sense of the Dirichlet condition still remains doubtful.

Considering these ideas, we arrive at the initial value problem with Neumann boundary conditions in the form

$$\begin{aligned}
\frac{\partial u(x,t)}{\partial t} &= \frac{\partial^2}{\partial x^2}\,{}_0\boldsymbol{D}_{1,x}^{\alpha-2}u(x,t)+f(x,t),\ x\in(0,1),\ t\in\boldsymbol{R}^+,\\
u(x,0) &= g(x),\ x\in(0,1),\\
\frac{\partial}{\partial x}\,{}_0D_{1,x}^{\alpha-2}u(x,t)\Big|_{x=0} &= b^0(t),\ t\in\boldsymbol{R}^+,\\
\frac{\partial}{\partial x}\,{}_0D_{1,x}^{\alpha-2}u(x,t)\Big|_{x=1} &= b^1(t),\ t\in\boldsymbol{R}^+,
\end{aligned}\tag{5}$$

i.e. we prescribe the flux on both ends of the interval [0, 1].

## 5. The Discretization of the Fractional Diffusion Equation

Before we start to deal with the problem (5), we should clarify our discrete fractional tools and perhaps mention other approaches. Some authors use, probably for the sake of simplicity, local approximation of the differintegral (see [2]). However, this approach is slightly rough because it suppresses the whole idea of fractional derivative as a global operator.

As a much better way it turns out to use fractional differences. This approach starts from the Grunwald-Letnikov definition of a differintegral which coincides under some assumptions with the Riemann-Liouville one (see [14]). As one can observe in many papers (e.g. [6], [9], [15] and [16]), the fractional differences are the most frequently used approximations of differintegrals of the Riemann-Liouville and the Caputo types (more information about the Caputo differintegrals in [14]). On the other hand, the fractional differences do not naturally work with initial and boundary conditions which are needed in the Riemann-Liouville approach. Therefore, all papers mentioned above deal either with the Caputo differintegral or with the Riemann-Liouville differintegral and zero initial conditions. If nonzero initial conditions occur in the Riemann-Liouville case, Podlubny introduces a change of variable which transforms the initial conditions into the zero ones (see [15]).

Being inspired by the fractional calculus on time-scales (Definitions 3.3 and 3.2), we consider a fractional difference to be an ordinary difference of an appropriate fractional sum. That is why the numerical scheme presented in this paper removes the problems with nonzero Riemann-Liouville initial or boundary conditions as we illustrate with problem (5).

We introduce an equidistant discretization of the space variable $x_n = nh$, where $n = 0, 1, ..., N$, $h = 1/N$ and of the time variable $t_i = i\tau$, where $i = 0, 1, ..., M$, $\tau = T/M$. $N$ is a number of space

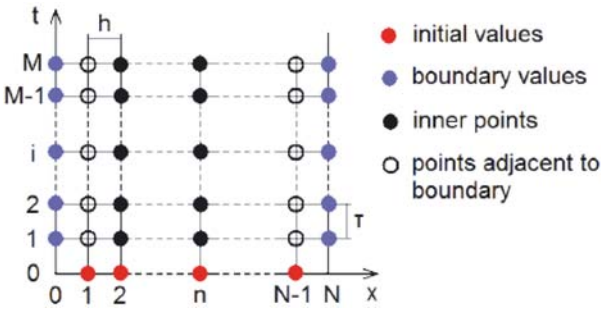



*Fig. 1 The scheme of the discretization*

mesh points, $M$ a number of mesh points in the time domain and $T$ is a maximum time we are interested in (see Fig. 1). The discretized problem (5) then looks like

$$\begin{aligned}
\nabla_i c_{n,i} &= \diamondsuit_n^2\,{}_0\diamondsuit_{N,n}^{\alpha-2}c_{n,i}+f_{n,i},\ 1\le n\le N-1,\ 1\le i\le M,\\
c_{n,0} &= g_n,\ 1\le n\le N-1,\\
{}_0\diamondsuit_{N,n}^{\alpha-1}c_{0,i} &= b_i^0,\ 1\le i\le M,\\
{}_0\diamondsuit_{N,n}^{\alpha-1}c_{N,i} &= b_i^1,\ 1\le i\le M,
\end{aligned}\tag{6}$$

where the symbol $\diamondsuit_n^2$ denotes the classical second central difference

$$\diamondsuit_n^2 f_n = \nabla_n(\Delta_n f_n) = \Delta_n(\nabla_n f_n) = \frac{f_{n+1}-2f_n+f_{n-1}}{h^2}.$$

The Riesz fractional difference has not been defined previously, because it is not clear which type of a difference should be applied to the corresponding Riesz integral. Considering our problem, we suggest the above introduction of $\diamondsuit_n^2$ with respect to its symmetry. This choice implies that the symbol ${}_0\diamondsuit_{N,n}^{\alpha-1}$ used in the boundary conditions is not defined uniquely because we do not determine whether the nabla or delta difference is supposed to be utilized.

This uncertainty allows to employ the delta difference for $n = 1$ and the nabla one for $n = N - 1$ which is necessary for involving the boundary conditions to the system of equations. Rewrite all integer-order differences and study the equations originated from (6).

At the inner points (black circles in Fig. 1), i.e for $1 \le i \le M$ and $2 \le n \le N - 2$ we arrive to the expected relation

$$\begin{aligned}
&\frac{1}{\tau}(u_{n,i}-u_{n,i-1}) =\\
&=\frac{1}{h^2}\left({}_0\diamondsuit_{N,n}^{\alpha-2}u_{n-1,i}-2\,{}_0\diamondsuit_{N,n}^{\alpha-2}u_{n,i}+{}_0\diamondsuit_{N,n}^{\alpha-2}u_{n+1,i}\right)+f_{n,i},
\end{aligned}\tag{7}$$

where for $i = 1$ the initial condition has to be used. At the points adjacent to the boundary points (white circles in Fig. 1) the situ-

ation changes. Close to the left boundary, i.e. for $1 \le i \le M$ and $n = 1$ we obtain the equation involving the boundary condition $b_i^0$

$$\frac{1}{\tau}(u_{1,i} - u_{1,i-1}) = \\ = \frac{1}{h^2}(-hb_i^0 - {}_0\Diamond_{N,n}^{\alpha-2}u_{1,i} + {}_0\Diamond_{N,n}^{\alpha-2}u_{2,i}) + f_{1,i}. \tag{8}$$

Similarly at the right boundary ($n = N - 1$) we utilize the second boundary condition $b_i^1$

$$\frac{1}{\tau}(u_{N-1,i} - u_{N-1,i-1}) = \\ = \frac{1}{h^2}({}_0\Diamond_{N,n}^{\alpha-2}u_{N-2,i} - {}_0\Diamond_{N,n}^{\alpha-2}u_{N-1,i} + hb_i^1) + f_{N-1,i}. \tag{9}$$

For an easier implementation it is necessary to rewrite the equations (7)-(9) into a compact matrix form.

Obviously, it is possible to write the Riesz fractional sum of a function $f$ (without the first and the last point) in the form of matrix product

$$\begin{pmatrix} {}_0\Diamond_{N,n}^{-\alpha}f_1 \\ {}_0\Diamond_{N,n}^{-\alpha}f_2 \\ \vdots \\ {}_0\Diamond_{N,n}^{-\alpha}f_{N-1} \end{pmatrix} = \frac{h^\alpha}{2}\begin{pmatrix} 2\binom{\alpha-1}{0} & \binom{\alpha}{1} & \cdots & \binom{\alpha+N-3}{N-2} \\ \binom{\alpha}{1} & 2\binom{\alpha-1}{0} & \cdots & \binom{\alpha+N-4}{N-3} \\ \vdots & \vdots & \ddots & \vdots \\ \binom{\alpha+N-3}{N-2} & \binom{\alpha+N-4}{N-3} & \cdots & 2\binom{\alpha-1}{0} \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{pmatrix},$$

where the matrix of the generalized binomial coefficients with the factor $h^\alpha/2$ will be denoted by ${}_0\Diamond_N^{-\alpha}$ and its dimension will always correspond with the dimension of the vector $f$.

After an analysis of the relations (7)-(9) we arrive at the matrix of coefficients

$$D_{NN} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

Now we denote $\boldsymbol{E}$ the unit matrix of the dimension $N - 1$ and define the vectors of all the involved quantities as $u_0 = (g_1 \dots g_{N-1})^T$ (the initial concentration), $u_i = (u_{1,i} \dots u_{N-1,i})^T$ (a concentration in $t = i\tau$), $b_i = (-hb_i^0 0 \dots 0\ hb_i^1)^T$ (the boundary vector in $t = i\tau$) and $f_i = (f_{1,i} \dots f_{N-1,i})^T$ (the source vector in $t = i\tau$). Then we can write the system of linear equations (7)-(9) for every $i$-th time level in the matrix form

$$\left(\frac{1}{\tau}\boldsymbol{E} - \frac{1}{h^2}D_{NN} \cdot {}_0\Diamond_N^{-(2-\alpha)}\right) \cdot u_i = f_i + \frac{1}{h^2}b_i + \frac{1}{\tau}u_{i-1},$$

$1 \le i \le M.$

Therefore we can calculate the solutions in the space region recursively.

Our scheme works with the generalized Neumann boundary conditions on both ends of the domain. Note that the use of the generalized Dirichlet boundary conditions on both ends of the domain would change the coefficients matrix $D_{NN}$ to the form

$$D_{DD} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{pmatrix}$$

and the boundary vector does not involve the $h$, i.e. $b_i = (b_i^0 0 \dots \dots 0\ b_i^1)^T$. Analogously, if the Neumann condition is prescribed on the left-side and the Dirichlet one on the right-side, the resulting coefficient matrix would be

$$D_{ND} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{pmatrix}$$

and the boundary vector $b_i = (-hb_i^0 0 \dots 0\ b_i^1)^T$.

## 6. Examples

We present some examples of the problem (5), where we consider the parameters $\alpha = 2, 1.9, 1.7, 1.4$. The initial state is depicted in all the figures with the red line. The space step is $h = 0.01$ (i.e. $N = 100$), the time step $\tau = 0.01$ and we follow the system until the time is $T = 0.2$ (i.e. $M = 20$).

First, we consider the insulated system with no source term, hence the total mass of the system is conserved. Translated to the mathematical language it means

$$f(x,t) = 0, \quad x \in (0,1), t \in \boldsymbol{R}^+,$$
$$\frac{\partial}{\partial x}{}_0D_{1,x}^{\alpha-2}u(x,t)\Big|_{x=0} = 0, \quad t \in \boldsymbol{R}^+,$$
$$\frac{\partial}{\partial x}{}_0D_{1,x}^{\alpha-2}u(x,t)\Big|_{x=1} = 0, \quad t \in \boldsymbol{R}^+.$$

The initial distribution of particles (i.e. the initial condition) is

$$u(x,0) = \begin{cases} 2, & x \in (0.2, 0.7), \\ 0, & x \notin (0.2, 0.7). \end{cases}$$

The solutions obtained via our discretization scheme are drawn in Fig. 2. The conservation of mass is evident but there is another interesting effect. With the decreasing value of α, the particles are accumulated close to the boundaries. Roughly speaking, the lower value of α means greater probability of very long particle jumps, hence due to the boundedness of the region more and more particles are forced to stop on the boundaries although their jump was supposed to be longer.

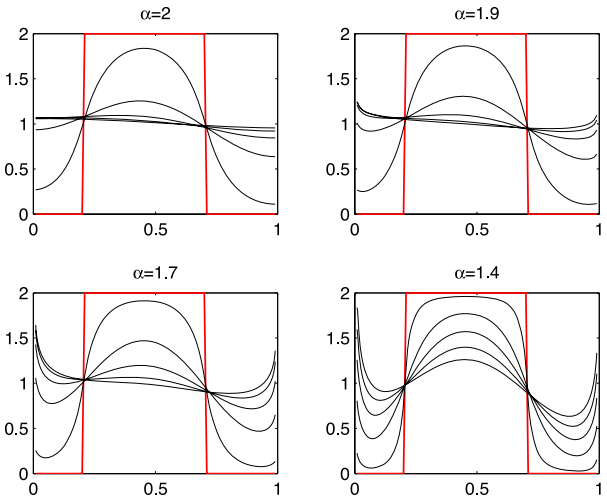



*Fig. 2 The insulated system with no source term*

The second example describes the constant inflow to the system from the right side. We assume that the region is empty at the beginning and there is no source term:

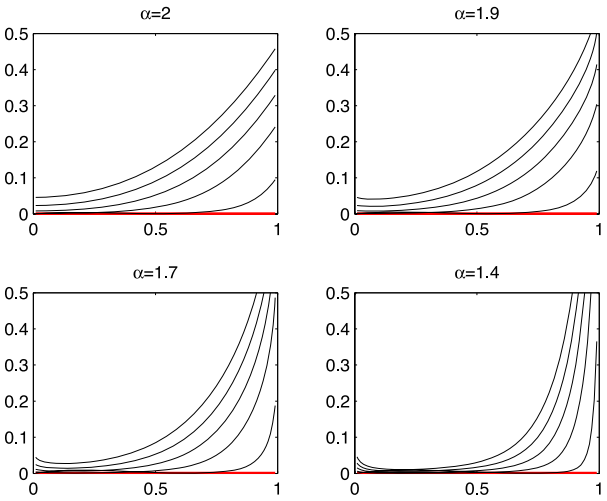



*Fig. 3 The inflow*

$$\begin{aligned} f(x,t) &= 0, \quad x \in (0,1),\ t \in \boldsymbol{R}^+, \\ u(x,0) &= 0, \quad x \in (0,1), \\ \frac{\partial}{\partial x}\,{}_0D^{\alpha-2}_{1,x}u(x,t)\Big|_{x=0} &= 0, \quad t \in \boldsymbol{R}^+, \\ \frac{\partial}{\partial x}\,{}_0D^{\alpha-2}_{1,x}u(x,t)\Big|_{x=1} &= 1, \quad t \in \boldsymbol{R}^+. \end{aligned}$$

The numerical solutions of this problem are depicted in Fig. 3. The amount of the influent particles is the same for all α, hence due to the boundary effect described in the previous example, there are fewer particles in the middle of the region for the decreasing α.

Last we present the system with insulated boundaries, but with nonzero source term. The source works for the first five time-steps (blue lines in Fig. 4) and then it is cut off:

$$\begin{aligned} f(x,t) &= \begin{cases} 2, & x \in (0.2, 0.7),\ t \in (0, 0.05), \\ 0, & otherwise, \end{cases} \\ u(x,0) &= 0, \quad x \in (0,1), \\ \frac{\partial}{\partial x}\,{}_0D^{\alpha-2}_{1,x}u(x,t)\Big|_{x=0} &= 0, \quad t \in \boldsymbol{R}^+, \\ \frac{\partial}{\partial x}\,{}_0D^{\alpha-2}_{1,x}u(x,t)\Big|_{x=1} &= 0, \quad t \in \boldsymbol{R}^+. \end{aligned}$$

The solutions drawn in Fig. 4 confirm the observation from the first example, i.e. the lower values of α emphasize sudden changes of the particle concentration.

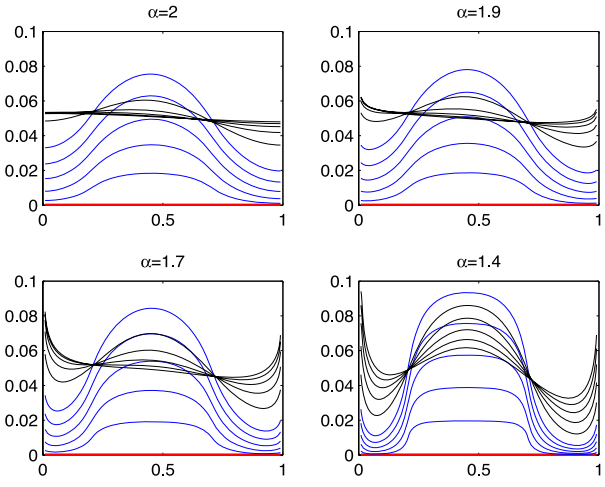



*Fig. 4 The system with a temporary source*

## 7. Summary

We develop a numerical scheme which provides numerical solutions for both initial and boundary value problems involving the Riemann-Liouville fractional differential equations. In particular, we can take also the case when these conditions are nonzero. The key idea is to consider the fractional difference strictly as a composition of the integer-order difference and the fractional

sum. The choice of the type of this integer-order difference (delta versus nabla) depends on the particular equation. We illustrate this choice by the discretization of the fractional diffusion equation.

**Acknowledgement**

The research was supported by the research plan MSM 0021630518 "Simulation modelling of mechatronic systems" of the Ministry of Education, Youth and Sports of the Czech Republic and by the grant # 201/08/0469 of the Czech Grant Agency.

## References

[1] ATICI, F. M., ELOE, P. W.: *Initial Value Problems in Discrete Fractional Calculus,* Proc of the Am Math Soc 137, 2008, pp. 981–989.
[2] BENSON, D. A.: *The Fractional Advection-Dispersion Equation: Development and Application,* 1998, p.144, University of Nevada, Reno, Ph.D. thesis.
[3] BENSON, D. A., WHEATCRAFT, S. W., MEERSCHAERT, M. M.: *The Fractional-order Governing Equation of Levy Motion,* Water Resources Research 36, 2000, pp. 1413–1423.
[4] BOHNER, M., PETERSON, A.: *Dynamic equations on time scales: An introduction with applications,* Birkhauer Boston Inc., Boston, MA, 2001.
[5] CERMAK, J., NECHVATAL, L.: *On (q,h)-Analogue of Fractional Calculus,* J Nonlin Math Phys - to appear.
[6] DIETHELM, K., et al.: *Algorithms for the Fractional Calculus: A Selection of Numerical Methods,* Comput Methods Appl Mech Engrg 194, 2005, pp. 743–773.
[7] GREY, H. L., ZHANG, N. F.: *On a New Definition of the Fractional Difference,* Math Comp 50, 1988, pp. 513–529.
[8] HEYMANS, N., PODLUBNY, I.: *Physical Interpretation of Initial Conditions for Fractional Differential Equations with Riemann-Liouville Fractional Derivatives,* Rheologica Acta 45 2005, 765–771.
[9] HILFER, R.: *Applications of Fractional Calculus in Physics, Singapore: World Scientific 2000,* p. 463.
[10] KILBAS, A. A., SRIVASTAVA, H. M., TRUJILLO, J. J.: *Theory and Applications of Fractional Differential Equations,* John van Mill, Netherlands, Elsevier, 2006, p. 523, ISBN 978-0-444-51832-3.
[11] MILLER, K. S., ROSS, B.: *An Introduction to the Fractional Calculus and Fractional Differential Equations,* John Wiley & Sons, New York, 1993.
[12] MILLER, K. S., ROSS, B.: *Fractional Difference Calculus,* Proc Int Symp Unival Funct, Frac Calc Appl, Koriyama, Japan, May (1988) 139-152; Ellis Horwood Ser Math Appl, Horwood, Chichester, 1989.
[13] OLDHAM, K. B., SPANIER, J.: *The Fractional Calculus,* London: Academic Press, 1974, p. 225
[14] PODLUBNY, I.: *Fractional Differential Equations.* United States: Academic Press, 1999, p. 340, ISBN 0-12-558840-2.
[15] PODLUBNY, I.: *Matrix Approach to Discrete Fractional Calculus,* Fractional Calculus and Applied Analysis 3, 2000, pp. 359–386.
[16] PODLUBNY, I., et al.: *Matrix Approach to Discrete Fractional Calculus II: Partial Fractional Differential Equations,* J of Computational Physics 228, 2009, pp. 3137–3153.
[17] SCHUMER, R., et al.: *Eulerian Derivation of the Fractional Advection-Dispersion Equation,* J. of Contaminant Hydrology 48, 2001, pp. 69–88.

Vladimir Guldan – Mariana Marcokova *

# ORTHOGONAL POLYNOMIALS AND RELATED SPECIAL FUNCTIONS APPLIED IN GEOSCIENCES AND ENGINEERING COMPUTATIONS

*In applications of mathematics involving either the Laplace or the Helmholtz equation in spherical coordinates the associated Legendre equation occurs. Its solutions are called associated Legendre functions. They have some relations to classical Legendre orthogonal polynomials and classical Jacobi orthogonal polynomials. The classical Legendre polynomials and the associated Legendre functions have been used by some authors at the approximation of length of curves and cylindrical surfaces and the approximation of Earth shape in geosciences and engineering computations. In the present paper we investigate some relations between these functions pointing out possible applications in geosciences.*

## 1. Definition and basic properties of orthogonal polynomials

Definition 1. Let $\langle a, b\rangle \subset R$ be a finite or infinite interval. A function $v(x)$ is called the weight function if at this interval it fulfills the following conditions:

(i) $v(x)$ is nonnegative at $\langle a, b\rangle$, i.e.

$$v(x) \geq 0$$

(ii) $v(x)$ is integrable at $\langle a, b\rangle$, i.e.

$$0 < \int_a^b v(x)dx < \infty$$

and

(iii) for every $n = 1, 2, 3 \ldots$

$$0 < \int_a^b |x|^n v(x)dx < \infty\,.$$

**Definition 2.** Let $\{P_n(x)\}_{n=0}^{\infty}$ be a system of polynomials, where every polynomial $P_n(x)$ has the degree $n$. If for all polynomials of this system

$$\int_a^b P_n(x)P_m(x)v(x)dx = 0,\ n \neq m\,,$$

then the polynomials $\{P_n(x)\}_{n=0}^{\infty}$ are called orthogonal in $\langle a, b\rangle$ with respect to the weight function $v(x)$. If moreover for every $n = 1, 2, 3 \ldots$

$$\left\|P_n(x)\right\|_{v(x)} = \left[\int_a^b P_n^2(x)v(x)dx\right]^{\frac{1}{2}} = 1\,,$$

then the polynomials are called orthonormal in $\langle a, b\rangle$ with respect to $v(x)$.

**Remark 1.** The condition of the orthonormality of the system $\{P_n(x)\}_{n=0}^{\infty}$ has the form

$$\int_a^b P_n(x)P_m(x)v(x)dx = \delta_{nm}\,,$$

where $\delta_{nm} = 1$ for $n = m$ and $\delta_{nm} = 0$ for $n \neq m$.

**Theorem 1.** For every weight function $v(x)$ there exists one and only one system of polynomials $\{P_n(x)\}_{n=0}^{\infty}$ orthonormal in $\langle a, b\rangle$, where

$$P_n(x) = \sum_{k=0}^{n} a_k^{(n)} x^{n-k};\quad a_0^{(n)} > 0\,.$$

*Proof.* E. g. in [12].

**Theorem 2.** A polynomial $P_n(x)$ is orthogonal in $\langle a, b\rangle$ with respect to the weight function $v(x)$, if and only if for arbitrary polynomial $S_m(x)$ of the degree $m < n$ the following condition is fulfilled

$$\int_a^b P_n(x)S_m(x)v(x)dx = 0\,.$$

*Proof.* E. g. in [12].

**Theorem 3.** If the interval of orthogonality is symmetric according to the origin of coordinate system and a weight function $v(x)$ is even function, then every orthogonal polynomial $P_n(x)$ fulfils the equality

$$P_n(-x) = (-1)^n P_n(x)\,.$$

* **Vladimir Guldan, Mariana Marcokova**
Department of Mathematics, Faculty of Science, University of Zilina, Slovakia, E-mail: vladimir.guldan@post.sk, mariana.marcokova@fpv.uniza.sk

*Proof.* E. g. in [12].

## 2. Jacobi polynomials, Legendre polynomials, Legendre associated functions

It is well-known that Jacobi polynomials $\{P_n(x; \alpha, \beta)\}_{n=0}^{\infty}$ are orthogonal in the interval $I = \langle -1, 1\rangle$ with respect to the weight function

$$J(x) = (1 - x)^{\alpha}(1 + x)^{\beta}, \ x \in (-1, 1), \tag{1}$$

where $\alpha > -1$, $\beta > -1$. Very important special case of Jacobi polynomials are classical Legendre polynomials $\{P_n(x; 0, 0)\}_{n=0}^{\infty}$, for which $\alpha = \beta = 0$ in the weight function $J(x)$. In the next we denote them by $\{P_n(x)\}_{n=0}^{\infty}$.

Classical orthogonal polynomials are solutions of the second order linear homogeneous differential equations of the form (cf. e.g. [12])

$$a(x)y_n'(x) + b(x)y_n''(x) + \lambda_n y_n(x) = 0,$$

where $a(x)$ is a polynomial of the degree at most 2, $b(x)$ is the polynomial of the degree 1 and $\lambda_n$ does not depend of $x$. For the classical Jacobi polynomials this equation has the form

$$(1 - x^2)y_n''(x) + [\beta - \alpha - (\alpha + \beta + 2)x]y_n'(x) +$$

$$+ n(n + \alpha + \beta + 1)y_n(x) = 0,$$

which in the case of classical Legendre polynomials is reduced to the form

$$(1 - x^2)y_n''(x) - 2xy_n'(x) + n(n + 1)y_n(x) = 0. \tag{2}$$

Associated Legendre equation occurs in applications characterized by Laplace or Helmholtz equation in spherical coordinates. For $m = 0, 1, 2, \ldots n$ it has the form

$$(1 - x^2)y_n''(x) - 2xy_n'(x) + \left[n(n + 1) - \frac{m^2}{1 - x^2}\right]y_n(x) = 0.$$

Observe that for $m = 0$ it reduces to Legendre equation (2). Its solutions are called associated Legendre functions of the first and second kind, respectively. They are defined by (cf. [1])

$$P_n^m(x) = (1 - x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}P_n(x) \tag{3}$$

and

$$R_n^m(x) = (1 - x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}R_n(x),$$

respectively. Here $P_n(x)$ and $R_n(x)$ are solutions of the Legendre equation (2), the first of them are the classical Legendre polynomials.

## 3. Classical Legendre polynomials and generalized Legendre polynomials

As it is seen from preliminaries, the Legendre classical polynomials $\{P_n(x)\}_{n=0}^{\infty}$ are orthogonal in $I = \langle -1, 1\rangle$ with respect to the weight function $L(x) = 1$.

Now we introduce the system of polynomials $\{Q_n(x)\}_{n=0}^{\infty}$ which will be the polynomials orthonormal in $I$ with respect to the weight function

$$Q(x) = (x^2)^{\gamma},$$

where $\gamma > 0$ and $Q_n(+\infty) > 0$. It is clear that these polynomials are generalization of the classical Legendre polynomials, which can be obtained by substituting $\gamma = 0$ in the weight function $Q(x)$.

Further, we introduce two classes of orthonormal polynomials:

(a) polynomials $\left\{P_n\left(x; 0, \gamma - \frac{1}{2}\right)\right\}_{n=0}^{\infty}$ orthonormal in $I$ with respect to the weight function

$$J_1(x) = (1 + x)^{\gamma - \frac{1}{2}}$$

and

(b) polynomials $\{P_n(x; 0, \gamma)\}_{n=0}^{\infty}$ orthonormal in $I$ with respect to the weight function

$$J_2(x) = (1 + x)^{\gamma}.$$

In both these cases we have classical Jacobi polynomials orthogonal with the weight function (1) for $\alpha = 0$, $\beta = \gamma - 1/2$ and $\alpha = 0$, $\beta = \gamma$, respectively. In the next we establish relations between them and the polynomials $\{Q_n(x)\}_{n=0}^{\infty}$.

## 4. Jacobi and Legendre classical polynomials with different weight functions

**Theorem 4.** In the notations introduced in the previous sections

$$Q_{2n}(x) = 2^{\frac{\gamma}{2} - \frac{1}{4}}P_n\left(2x^2 - 1; 0, \gamma - \frac{1}{2}\right) \tag{4}$$

and

$$Q_{2n+1}(x) = 2^{\frac{\gamma}{2}}xP_n(2x^2 - 1; 0, \gamma). \tag{5}$$

*Proof.* According to the Theorem 3, the function $Q_{2n}(x)$ is even function. Putting $t = x^2$ we denote $W_n(t) = Q_{2n}(x)$. The orthogonality of the polynomials $\{Q_n(x)\}_{n=0}^{\infty}$ for $r = 0, 1, \ldots, n - 1$ and $n > 0$ yields

$$0 = \int_0^1 x^{2r} Q_{2n}(x) x^{2\gamma} dx = \frac{1}{2}\int_0^1 t^r W_n(t) t^{\gamma-\frac{1}{2}}\, dt =$$
$$= \frac{1}{2^2}\int_{-1}^1 \left(\frac{\tau+1}{2}\right)^r W_n\left(\frac{\tau+1}{2}\right)\left(\frac{\tau+1}{2}\right)^{\gamma-\frac{1}{2}} d\tau =$$
$$= \frac{1}{2^{\gamma+\frac{3}{2}}}\int_{-1}^1 \left(\frac{\tau+1}{2}\right)^r W_n\left(\frac{\tau+1}{2}\right)(\tau+1)^{\gamma-\frac{1}{2}}\, d\tau.$$

From that it is clear that the polynomials $W_n\left(\frac{x+1}{2}\right)$ are orthogonal in $I$ with respect to the weight function $J_1(x)$. According to the Theorem 1, taking into account the uniqueness of these polynomials, we have

$$W_n\left(\frac{x+1}{2}\right) = k\, P_n\left(x; 0, \gamma-\frac{1}{2}\right),$$

where $k > 0$ in consequence of the fact that $P_n\left(\infty; 0, \gamma-\frac{1}{2}\right) > 0$ and $W_n(+\infty) > 0$.

From the orthonormality of the polynomials $W_n(t)$ we derive

$$\frac{1}{2} = \int_0^1 W_n^2(t) t^{\gamma-\frac{1}{2}} dt = k^2 \int_{-1}^1 P_n^2\left(\tau; 0, \gamma-\frac{1}{2}\right)\left(\frac{\tau+1}{2}\right)^{\gamma-\frac{1}{2}} \frac{1}{2} d\tau =$$
$$= \frac{1}{2^{\gamma+\frac{1}{2}}} k^2 \int_{-1}^1 P_n^2\left(\tau; 0, \gamma-\frac{1}{2}\right)(\tau+1)^{\gamma-\frac{1}{2}} d\tau$$

from where we have $k = 2^{\frac{\gamma}{2}-\frac{1}{4}}$ and the relation (4), i.e.

$$Q_{2n}(x) = 2^{\frac{\gamma}{2}-\frac{1}{4}} P_n\left(2t-1; 0, \gamma-\frac{1}{2}\right), \quad t = x^2.$$

Now we prove the relation (5). Putting $t = x^2$ we have

$$\overline{W}_n(t) = x^{-1} Q_{2n+1}(x),$$

where $\overline{W}_n(t)$ is the polynomial of the degree $n$ and $Q_{2n+1}(x)$ is odd function. For $r = 0, 1, ..., n-1$ and $n > 0$ the orthogonality of the polynomials $\{Q_n(x)\}_{n=0}^{\infty}$ yields

$$0 = \int_0^1 x^{2r+1} Q_{2n+1}(x) x^{2\gamma} dx = \frac{1}{2}\int_0^1 t^r \overline{W}_n(t) t^{\gamma+\frac{1}{2}}\, dt =$$
$$= \frac{1}{2^2}\int_{-1}^1 \left(\frac{\tau+1}{2}\right)^r \overline{W}_n\left(\frac{\tau+1}{2}\right)\left(\frac{\tau+1}{2}\right)^{\gamma+\frac{1}{2}} d\tau =$$
$$= \frac{1}{2^{\gamma+\frac{5}{2}}}\int_{-1}^1 \left(\frac{\tau+1}{2}\right)^r \left(\frac{\tau+1}{2}\right)^{\frac{1}{2}} \overline{W}_n\left(\frac{\tau+1}{2}\right)(\tau+1)^{\gamma}\, d\tau.$$

From there

$$\left(\frac{x+1}{2}\right)^{\frac{1}{2}} \overline{W}_n\left(\frac{x+1}{2}\right) = \overline{k}\, P_n(x; 0, \gamma),$$

where $\overline{k} > 0$ and from the orthonormality of the polynomials $t^{\frac{1}{2}} \overline{W}_n(t)$ we derive

$$\frac{1}{2} = \int_0^1 x^{-2} Q_{2n+1}^2(x) x^{2\gamma}\, dx = \int_0^1 t\, \overline{W}_n^2(t) t^{\gamma}\, dt =$$
$$= \frac{1}{2}\int_{-1}^1 \left(\frac{\tau+1}{2}\right)\overline{W}_n^2\left(\frac{\tau+1}{2}\right)\left(\frac{\tau+1}{2}\right)^{\gamma} d\tau =$$
$$= \frac{1}{2^{\gamma+1}}\overline{k}^2 \int_{-1}^1 P_n^2(\tau; 0, \gamma)(\tau+1)^{\gamma}\, d\tau.$$

Finally we get $\overline{k} = 2^{\frac{\gamma}{2}}$ and the relation (5) of the theorem.

## 5. Applications in geosciences

Classical Jacobi orthogonal polynomials and their special cases – Legendre polynomials as well as their generalizations and associated Legendre functions are important tools in investigating geoscientific problems. Such geosciences like geodesy, geography, geology, geophysics, meteorology and others need adequate mathematical tools called geomathematics (cf. [3, 9]). We point out some problems of geomathematics where the functions investigated in previous sections of this paper are used.

In [5] we suggested to compute the length of a junction curve of a railroad between two sections of a rail with different curvatures by means of classical Legendre polynomials. Junction curve is a curve inserted between a straight line and a circle arc. Junction curve enables fluent transition between two sections of a rail with different curvatures. At our railroads a cubic parabola is used as the junction curve. In [2] the length of such junction curve is computed by using binomial series, because the integral of the function

$$\sqrt{1+\left[f'(x)\right]^2}, \quad f(x) = c\,x^3,$$

(is the specific constant) had to be computed. We suggested to use Legendre polynomials, because the best approximation of a function by means of a polynomial is the approximation by means of an expansion of the function into a series of orthogonal polynomials followed by substituting the function by the partial sum of such series (cf. [12]). So we used the following expansion of a function $\sqrt{1+x}$ (cf. [12])

$$\sqrt{1+x} = \frac{4}{3\sqrt{2}} P_0(x) - \frac{4}{\sqrt{2}} \sum_{n=1}^{\infty} \frac{(-1)^n P_n(x)}{(2n-1)(2n+3)},$$

where $\{P_n(x)\}_{n=0}^{\infty}$ are classical Legendre polynomials.

Another application of classical Legendre polynomials can be found in [6–7]. The authors used Legendre polynomials for the approximation of cylindrical surfaces which was submitted as a strategy of measurement supported by experimental verification. The results were very good - the precision was high.

Legendre associated functions were used in [14-15] for the approximation of the Earth shape. The author uses the functions $P_n^m(\sin\varphi)$ ($\varphi$ is geocentric coordinate of geocentric radiusvector of a point of the Earth) defined by (3). The gravity potential of the

Earth is expressed by expansion into the series of spherical functions, where the functions $P_n^m(\sin\varphi)$ occur.

Geomathematics uses also expansions of functions in two variables into double series of orthogonal polynomials (cf. [8, 13]), orthogonal transforms (cf. [10–11]) and wavelets theory (cf. [16]).

## 6. Conclusion

In this paper we proved relations (4) and (5) which may be used as expressions of the classical Legendre polynomials of the argument $2x^2 - 1$ by means of generalized Legendre polynomials $[Q_n(x)]_{n=0}^{\infty}$ of the argument $x$. Other relations for the orthogonal polynomials in general in the forms of recurrence relations can be found in [4]. The Legendre associated functions connected with the above polynomials are the important part of geomathematics – mathematics developed for solutions of geoscientific problems.

**Acknowledgment:** This research has been supported by the Slovak Grant Agency VEGA through the project No. 1/0867/08.

## References

[1] ANDREWS, L.C.: *Special functions of mathematics for engineers,* McGraw – Hill, Inc. 1992.
[2] BITTERER, L.: *Geometry of rail (in Slovak),* University of Zilina 1997.
[3] FREEDEN, W., GERVENS, T., SCHREINER, M.: *Constructive Approximation on the Sphere.* With Applications to Geomathematics, Oxford: Clarendon Press 1998.
[4] FTOREK, B.: *Differential Recurrence Formula for the Second Derivatives of Orthogonal Polynomials,* Studies of University in Zilina - math.-phys. series, Vol. 13(2001), 75–79.
[5] GULDAN, V.: *Infinite Series, Special Functions and Study Programme Geodesy and Cartography (in Slovak),* 3rd Zilina's didactic conference DIDZA 2006 (2006), 5 p. on CD.
[6] JANECKI, D., STEPIEN, K.: *Legendre Polynomials Used for the Approximation of Cylindrical Surfaces,* Communications – Scientific Letters of the University of Zilina, 4/2005, 59–61.
[7] JANECKI, D., STEPIEN, K.: *Legendre Polynomials Used for the Approximation of Cylindrical Surfaces,* Proc. of International Conference Transcom 2005, Section 6, Zilina 2005 (2005), 269–272.
[8] KOVACIK, O.: *Once More on the Double Fourier-Haar Series,* Mathematica Slovaca 52 (2002), 215–220.
[9] KRCHO, J.: *Modelling of Georelief and its Geometrical Structure Using DTM: Positional and Numerical Accuracy,* Q111 Publishers 2001.
[10] MAMRILLA, D.: *On the Systems of First Order Quasi-linear Differential Equations,* Tribun EU 2008.
[11] MARCOKOVA, M., GULDAN, V.: *On one Orthogonal Transform Applied on a System of Orthogonal Polynomials in Two Variables,* Aplimat: Journal of Applied Mathematics, Vol.2., No.2 (2009), 239–245.
[12] SUJETIN, P. K.: *Classical Orthogonal Polynomials (in Russian),* Nauka Moskva 1979.
[13] SUJETIN, P. K.: *Orthogonal Polynomials in Two Variables,* Gordon and Breach Science Publishers 1999.
[14] TENZER, R.: *Geopotential Model of Earth - Approximation of Earth Shape,* Geopotential Model Testing Methods, Communications – Scientific Letters of the University of Zilina, 4/2001, 50–58.
[15] TENZER, R.: *Methodology for Testing of Geopotential Model,* Studies of University of Zilina – Civil Engineering series (2002), 113–114.
[16] WALTER, G. G.: *Wavelets and Other Orthogonal Systems with Applications,* CRC Press 1994.

Iveta Tomcikova – Dobroslav Kovac – Irena Kovacova *

# STRESS FIELD DISTRIBUTION IN MAGNETOELASTIC PRESSURE FORCE SENSOR

*Modelling of the stress field distribution in a magnetoelastic sensor core is proposed. The components of stress field are carried out by a numerical solution of boundary-value problem for stress field. The obtained results are presented in graphical form.*

## Introduction

The magnetoelastic pressure force sensor operates on the Villari effect principle. The Villari effect, also called a stress-magnetization, is based on the change of permeability of a ferromagnetic body exposed to a mechanical force, which results in mechanical stress within the body. The permeability increment is proportional to that component of mechanical stress, which is collinear to magnetic field vectors [1].

The modelling of the magnetic field distribution in the core of such a sensor with respect to the variation of the permeability due to mechanical stress requires to determine the stress field distribution.

In the general case, the planar stress field is unambiguously determined by three components, two perpendicular components and a shear component [2]. These components can be appointed by various methods. In this paper a method based on a numerical solution of boundary-value problem is proposed. This task was solved by professional code PDE Toolbox in package MATLAB.

## 1. Magnetoelastic pressure force sensor

The arrangement of the magnetoelastic pressure force sensor is depicted in Fig. 1. The sensor core consists of fifty ferromagnetic transformer sheets, thickness of each being 0.5mm. Each sheet contains four holes of the radius of 1 mm. The conductors of primary (10 turns/0.35mm) and secondary windings (8 turns/0.25mm) are led through these holes collaterally. The sensor is assigned for the measuring of nominal pressure force 120 kN that is equivalent to 100 MPa. A simplifying representation of the sensor core is in Fig. 2.

If Cartesian coordinate system is applied to such a simplified form of sensor core (see Fig. 2), the current density vector has

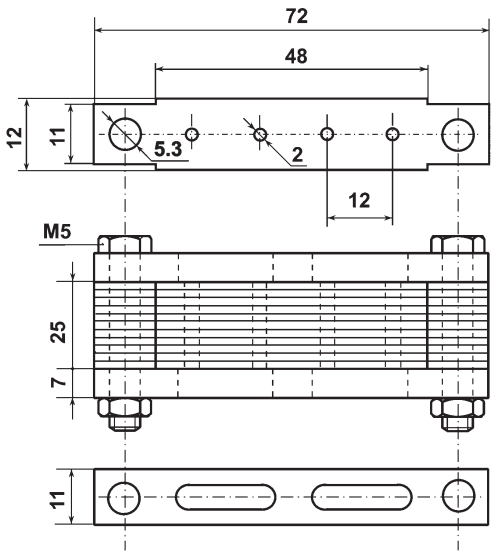



*Fig. 1 The magnetoelastic pressure force sensor*

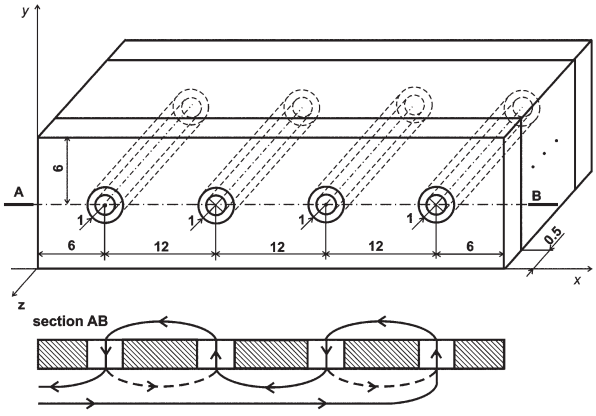



*Fig. 2. The magnetoelastic pressure force sensor core in simplified form*

* **Iveta Tomcikova[1], Dobroslav Kovac[1], Irena Kovacova[2]**
[1] Department of Theoretical Electrotechnics and Electrical Measurement, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia, E-mail: iveta.tomcikova@tuke.sk
[2] Department of Electrical Engineering, Mechatronics and Industrial Engineering, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia

a nonzero component only in the z-direction. This component produces magnetic field whose vectors have nonzero components only in the directions $x$ and $y$ [3].

Moreover, the dimensions of the sheet are designed in such a way that the sheet can be divided into four squared shapes, each with a circle hole in the middle. A conductor (the cross-section of which is the same as the cross-section of the primary windings conductors) replaces the primary winding. One of such shapes is called the integrative sensor element (Fig. 3*left*). With respect to the symmetry of the sheet it is sufficient to determine the magnetic field distribution in the integrative sensor element only. The stress field distribution is necessary to determine only in a part of the integrative sensor element that is made of a ferromagnetic material (see Fig. 3*right*).

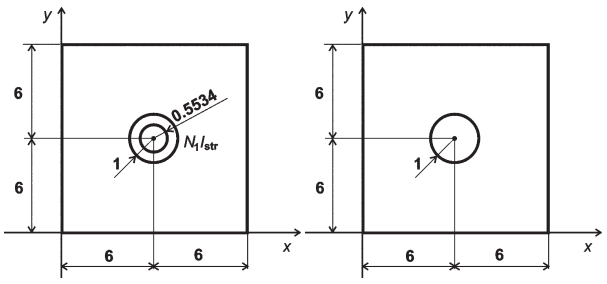



*Fig. 3 The integrative sensor element; left: for determination the magnetic field distribution, right: for determination the stress field distribution*

For the modelling of the magnetic field distribution in the sensor core with respect to the variation of material permeability due to mechanical stress it is necessary to investigate only the $x$-direction stress ($\sigma_x$) and y-direction stress ($\sigma_y$) in the integrative sensor element (see Fig. 4).

## 2. PDE system formulation for plane stress

The determination of stress field in the integrative sensor element can be formulated as a boundary–value problem in terms of the displacement components. The first of all it is essential to formulate the PDE system for balance of force in terms of the displacement components with the boundary conditions.

For plane stress, in the body made of the homogeneous isotropic material, the balance of force equations are [2]:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + X = 0\,, \tag{1}$$

$$\frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} + Y = 0\,, \tag{2}$$

where $\sigma_x$ is the $x$-direction stress, $\sigma_y$ being the y-direction stress, $\tau_{xy}$ being the shear stress and $X$, $Y$ being the volume forces.

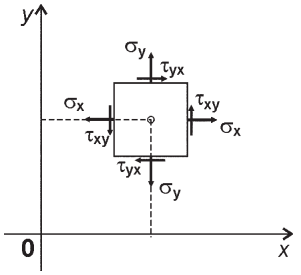



*Fig. 4 The stress components in the xy plane*

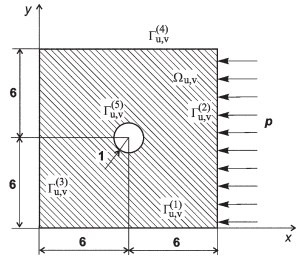



*Fig. 5 The domain $\Omega_{u,v}$*

The stress components are closely related to the strain components and these relations are defined by Hooke law

$$\begin{bmatrix}\sigma_x\\ \sigma_y\\ \tau_{xy}\end{bmatrix} = \frac{E}{1-\nu^2}\begin{bmatrix}1 & \nu & 0\\ \nu & 1 & 0\\ 0 & 0 & \dfrac{1-\nu}{2}\end{bmatrix}\begin{bmatrix}\varepsilon_x\\ \varepsilon_y\\ \gamma_{xy}\end{bmatrix} \tag{3}$$

where $E$ is Young's modulus, $v$ being Poisson's ratio,

The strain components can be expressed by the displacements, $u$ the displacement in the $x$-direction and $v$ the displacement in the $y$-direction

$$\varepsilon_x = \frac{\partial u}{\partial x},\ \varepsilon_y = \frac{\partial v}{\partial y},\ \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}, \tag{4}$$

where $\varepsilon_x$ is the $x$ – direction strain, $\varepsilon_y$ being the $y$ – direction strain, $\gamma_{xy}$ being the shear strain.

Combining the equations (1) – (4) and assuming that there are no volume forces, we arrived at the PDE system for balance of force in terms of the displacement components. The system takes form

$$\frac{E}{1-\nu^2}\frac{\partial^2 u}{\partial x^2} + \frac{E}{2(1+\nu)}\frac{\partial^2 u}{\partial y^2} + \nu\frac{E}{1-\nu^2}\frac{\partial^2 v}{\partial x \partial y} + \\ + \frac{E}{2(1+\nu)}\frac{\partial^2 v}{\partial y \partial x} = 0, \tag{5}$$

$$\frac{E}{2(1+\nu)}\frac{\partial^2 v}{\partial x^2} + \frac{E}{1-\nu^2}\frac{\partial^2 v}{\partial y^2} + \frac{E}{2(1+\nu)}\frac{\partial^2 u}{\partial x \partial y} + \\ + \nu\frac{E}{1-\nu^2}\frac{\partial^2 u}{\partial y \partial x} = 0, \tag{6}$$

From mathematical point of view, the equations (5) and (6) represent a system of two PDE of elliptic type.

The following boundary conditions for the system PDE can be given:

- For each point on the boundary is specified the value of the displacement

$$u = g_1,\ v = g_2. \tag{7}$$

This boundary condition is called Dirichlet boundary condition.

- For each point on the boundary is known the surface load (the derivative of the displacement by the outward normal vector)

$$\boldsymbol{e}_n\left(\begin{bmatrix} \frac{E}{1-\nu^2} & 0 \\ 0 & \frac{E}{2(1+\nu)} \end{bmatrix}\nabla u\right) + \boldsymbol{e}_n\left(\begin{bmatrix} 0 & \frac{\nu E}{1-\nu^2} \\ \frac{E}{2(1+\nu)} & 0 \end{bmatrix}\nabla v\right) = p_x\,, \quad (8)$$

$$\boldsymbol{e}_n\left(\begin{bmatrix} 0 & \frac{E}{2(1+\nu)} \\ \frac{\nu E}{1-\nu^2} & 0 \end{bmatrix}\nabla u\right) + \boldsymbol{e}_n\left(\begin{bmatrix} \frac{E}{2(1+\nu)} & 0 \\ 0 & \frac{E}{1-\nu^2} \end{bmatrix}\nabla v\right) = p_y\,, \quad (9)$$

where $\boldsymbol{e}_n$ is the outward normal vector of the boundary, $\nabla$ being Hamilton operator, $p_x$ being the pressure in the direction of $x$-axis, $p_y$ being the pressure in the direction of $y$-axis.

This boundary condition is called Neumann boundary condition.

## 3. Boundary-value problem formulation

The boundary-value problem formulation also requires determining the domain, in which the stress field will be solved. The integrative sensor element has the squared shape with a circle hole in the middle (see Fig. 4). It can be considered to be a planar body with the following properties: dimension 12-by-12 mm, thickness 0.5 mm and the radius of the circle hole 1 mm. The element is made of ferromagnetic transformer sheet, whose parameters are: $E = 1.86.105$ MN/m$^2$, $v = 0.3$. The element is exposed to a constant continuous pressure stress $\boldsymbol{p} = -p_x\,\boldsymbol{e}_x$ ($\boldsymbol{e}_x$ being the unit vector in the direction of x-axis).

The domain $\Omega_{u,v}$, in which the stress field will be solved, is surrounded by five boundaries $\Gamma^{(1)}_{u,v}$, $\Gamma^{(2)}_{u,v}$, $\Gamma^{(3)}_{u,v}$, $\Gamma^{(4)}_{u,v}$, a $\Gamma^{(5)}_{u,v}$ (see Fig. 5). The boundary conditions for the displacement components are following:

- The boundary $\Gamma^{(2)}_{u,v}$ is subjected to a pressure $\boldsymbol{p} = -p_x\,\boldsymbol{e}_x$:

$$\frac{E}{1-\nu^2}\frac{\partial u}{\partial x} + \frac{\nu E}{1-\nu^2}\frac{\partial v}{\partial y} = -p_x\,, \quad (10)$$

$$\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = 0\,. \quad (11)$$

- There is no displacement on the boundary $\Gamma^{(3)}_{u,v}$:

$$u = 0, \quad v = 0, \quad (12)$$

- The remaining boundaries $\Gamma^{(1)}_{u,v}$, $\Gamma^{(4)}_{u,v}$ a $\Gamma^{(5)}_{u,v}$ are free (no normal stress):

$$\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = 0\,, \quad (13)$$

$$\nu\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0\,. \quad (14)$$

The investigated boundary-value problem was solved by using a professional code PDE Toolbox. The toolbox is based on the finite element method (FEM). The solution was made by using the adaptive mode option, which enables to refine the mesh in areas where the gradient of the solution is large in order to increase the accuracy of the solution [7]. The obtained results (for generated triangular mesh consisting of 42 112 nodes and 83 456 triangles) at the pressure force $p_x = 8{,}23$ MPa are depicted in Figs. 6 – 11. In these figures are plotted the contour lines, e.g. the lines connected points of equal value of the depicted quantity in the domain $\Omega_{u,v}$. In Figs. 6 and 7 are visualised the displacements $u(x, y)$ in the $x$ – direction and $v(x, y)$ in the $y$ – direction. The $x$ – direction strain and the $y$ – direction strain are depicted in Figs. 8 and 9. The $x$ – direction stress and the $y$ – direction stress are shown in Figs. 10 and 11.

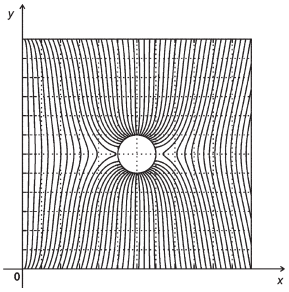



*Fig. 6 The displacement u in the x - direction*

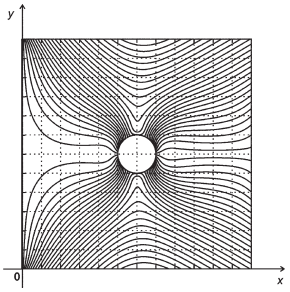



*Fig. 7 The displacement v in the y - direction*

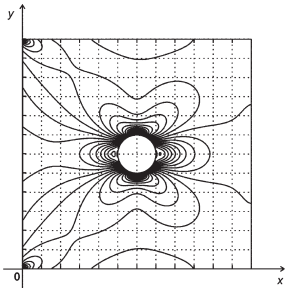



*Fig. 8 The strain $\varepsilon_x$ in the x - direction*

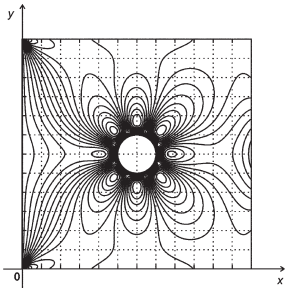



*Fig. 9 The strain $\varepsilon_y$ in the y - direction*

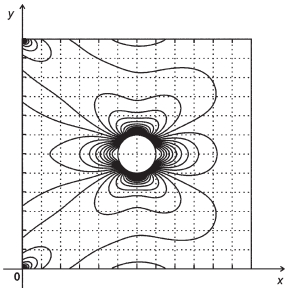



*Fig. 10 The stress $\sigma_x$ in the x - direction*

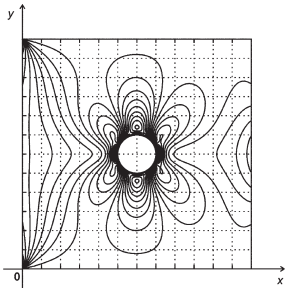



*Fig. 11 The stress $\sigma_y$ in the y - direction*

## 4. Conclusion

The paper is aimed at assignment of stress field distribution in the magnetoelastic pressure force sensor. The task was solved as the boundary-value problem in terms of the displacement components by the professional code PDE Toolbox in package MATLAB. The toolbox provides a powerful and flexible environment not only for the solution of elliptic, parabolic and hyperbolic partial differential equations, but it also handles the solution of a system of partial differential equations. The normal components of the stress field and also the components of the strain and displacement were visualised.

**Acknowledgements**

The paper was prepared under the support of Slovak grant projects VEGA No. 1/4174/07, VEGA No. 1/0660/08, KEGA No. 3/5227/07, KEGA No. 3/6386/08 and KEGA No. 3/6388/08.

## References

[1] ZEHNULA, K.: *Sensors of Non-electric Quantities,* SNTL, Prague 1977.
[2] TREBUNA, F., JURICA, V., SIMCAK, F.: *Elasticity and Strength I.,* Edition of scientific and special literature. Košice 2000.
[3] MAYER, D., ULRYCH, B.: *Fundamentals of Electric and Magnetic Fields Numerical Calculations,* SNTL, Prague 1988.
[4] TREBUNA, F., JURICA, V., SIMCAK, F.: *Elasticity and Strength II,.* Edition of Scientific and Special Literature, Košice 2000.
[5] MUSCHELISVILI, N. I.: *Some Fundamental Problems of Mathematical Stress Theory,* Publishing House of Science Academy USSR, Moscow 1954.
[6] SAVIN, G.: *Stress Field Distribution around Holes,* Nauka dumka, Kiev 1968.
[7] Partial Differential Equation Toolbox User's Guide.

Ondrej Kovacik – Pavol Orsansky *

# PARTIAL DIFFERENTIAL EQUATION FOR HEAT CONDUCTION AND ITS SOLVABILITY

*This paper deals with the heat conduction equation which is a model of thermomechanical processes in the special case. The considered case is based on the nonstandard type of coefficients in this equation.*

## 1. Introduction and notations

Let $\Omega$ be a physical body which is from the material point of view some bounded domain in the Euclidean space E3. Let $\partial\Omega$ be its surface, i.e. the boundary of the domain of $\Omega$. Let this boundary have the lipschitz piecewise smooth property. As an example we can take surfaces of polyhedrons, cones and other usual geometric configurations. Suppose this body to be nonhomogenous and nonanisotropic. Its heat conduction is in a relation to the external and internal influence sources, e.g. chemical reactions, electrical resistances etc. and to the structure of a material of the body $\Omega$. Let us know these properties in any point $x = (x_1, x_2, x_3)$ of $\Omega$ defined by the following functions

$$a_i = a_i(x, u_0, u_1, u_2, u_3), \quad i = 0, 1, 2, 3. \tag{1}$$

These functions depend on the location of a point $x$ in $\Omega$, on the searched function

$$u_0 = u(x,t)$$

and on its derivatives

$$u_i = \frac{\partial u(x,t)}{\partial x_i}, \quad i = 1, 2, 3.$$

Suppose that the function $f(x,t)$ describes an existence and actions of interior heat sources in our body $\Omega$. The dependence of the coefficients $a_i$ on the function $u = u_0$ follows from the well known fact that the heat conduction depends on the heat (see e.g. [1]). Then the equation for the heat conduction has the general form

$$\frac{du}{dt} - \sum_{i=1}^{3} \frac{\partial}{\partial x_i} a_i(x, u_0, u_1, u_2, u_3) - a_0(x, u_0, u_1, u_2, u_3) = f(x,t). \tag{2}$$

For $a_i = \frac{\lambda}{c\rho} u_{ii}$, $i = 1, 2, 3$, we have the following equation

$$\frac{du}{dt} = \frac{\lambda}{c\rho}(u_{11}, + u_{22} + u_{33}) = f.$$

i. e. the classical heat conduction equation for a 3-dimensional body in the form

$$\frac{du}{dt} - \frac{\lambda}{c\rho}\Delta u = f, \tag{3}$$

where c, ρ, λ are specific constants depending on the properties of a material and $\Delta$ is the Laplace operator (for using such operator see e.g. [5, 6]) .

A nonstationary heat conduction depends on the other input data, i.e. for our differential equations (2) or (3) from physical point of view it is necessary to suppose some boundary conditions and initial conditions. The boundary conditions characterize the heat exchange of $\Omega$ with its neighbourhood in any point of boundary $\partial\Omega$. The initial condition in the starting point of time $t_0 = 0$ describes the initial statement of heat in the body $\Omega$. Let us have the boundary conditions for $x \in \partial\Omega$ and for $t > 0$ in the form

$$u(x,t) = g_1(x,t) \tag{4}$$

and

$$\frac{du}{d\vec{n}}(x,t) = g_2(x,t), \tag{5}$$

where $\vec{n}$ is the normal vector to the boundary $\partial\Omega$. In the time $t_0 = 0$ we have the initial condition on $\Omega$ represented by the function

$$u(x,0) = h(x). \tag{6}$$

Classical problems of this type are simply solved by Fourier methods and by orthogonal function series (see e.g. [2, 5]).

* **Ondrej Kovacik, Pavol Orsansky**
Department of Mathematics, Faculty of Science, University of Zilina, Slovakia, E-mail: Ondrej.Kovacik@fpv.uniza.sk

## 2. Coefficients of heat conduction equation in nonstandard spaces

In the paper [4] there are introduced new function spaces denoted by $L^{p(x)}$ and $W^{k,p(x)}$. The theory of these spaces is inspired by the problems of mathematical physics. The basic problem is in solvability of stationary problems of heat equations in nonstandard space, but we continue this study by dynamical problems of heat equation in nonstandard environment.

Suppose that the functions mentioned above have the growth of the type of variable powers in the variables $u_i$, $i = 0, 1, 2, 3$ Suppose that there is given some function $p(x)$ and there exist real numbers $p_1$ and $p_2$ for which

$$1 < p_1 \le p(x) \le p_2 < \infty$$

almost everywhere on $\Omega$. The equality

$$q(x) = \frac{p(x)}{p(x) - 1}$$

defines the so called conjugate function $q(x)$ for the function $p(x)$ on $\Omega$.

Now we assume that for the coefficients $a_i$ the below estimation holds

$$\left|a_i(x, u_0, u_1, u_2)\right| \le a_i(x) + \sum_{j=1}^{3} c_{ij} \left|u_j\right|^{p(x)-1}, \; i = 1, 2, 3.$$

Here $c_{ij}$ are some nonnegative constants and $a_i(x) \ge 0$ are the functions integrable in the power $q(x)$ over $\Omega$ for $i = 1, 2, 3$. That means that

$$\int_\Omega \left|a_i(x)\right|^{q(x)} dx = \iiint_\Omega \left|a_i(x_1, x_2, x_3)\right|^{q(x_1, x_2, x_3)} dx_1 dx_2 dx_3 < \infty .$$

For example, we can assume that

$$a_i(x, u_0, u_1, u_2, u_3) = \left|u_i\right|^{p(x)-1}, \; i = 0, 1, 2, 3. \qquad (7)$$

The function $p(x)$ here we assume to be piecewise constant on $\Omega$.

For the other example we can assume that the function $p(x)$ is defined on two parts of $\Omega$, i.e. $\Omega_1$ and $\Omega_2$. In the first part for the function $p(x)$ we assume $p(x) = 2$ and on the second part there is a function $p(x) = 3$. Then, the relation (7) obtains the form

$$a_i(x, u_0, u_1, u_2, u_3) = \begin{cases} |u_i|, & x \in \Omega_1 \\ |u_i|^2, & x \in \Omega_2 = \Omega \backslash \Omega_1 \end{cases}, \; i = 0, 1, 2, 3.$$

Now we note that the theory of linear parabolic equations in the first part and on the other side in the second part of $\Omega$ is discussed in many papers and books with the problem of crossing conditions on boundary of domains $\Omega_1$ and $\Omega_2$. Our problem is defined in a very global situation. That means that we can investigate some another problems using the methods of time discretisation, so called Rothe's method.

## 2. Rothe's method

This method is based on discretisation in the time. In every moment of our time we can find the solution for stationary partial differential equation. This method is based on the decomposition of some time interval $\langle 0,T\rangle$ by the time sequence in the form $0 \le t_0 < t_1 \ldots < t_n \le T$ for every available natural $n$. In this time we can reformulate some property from the paper [4].

Suppose that the function $p(x)$ is defined on $\Omega$ and satisfies the conditions mentioned above. Let the functions (coefficients) $a_i(x, u_0, u_1, u_2, u_3)$ fulfill the next inequalities

$$\sum_{i=0}^{3} \left[a_i(x, u_0, u_1, u_2, u_3) - a_i(x, v_0, v_1, v_2, v_3)\right]\left[u_i - v_i\right] \ge 0$$

and

$$\sum_{i=0}^{3} \left[a_i(x, u_0, u_1, u_2, u_3)\right] u_i \ge c_1 \sum_{i=1}^{3} \left|u_i\right|^{p(x)-1} - c_2$$

for any $u = (u_0, u_1, u_2, u_3)$, $v = (v_0, v_1, v_2, v_3)$.

Here we assume that these functions are defined for every $x$ from some domain $\Omega$ in $\boldsymbol{R^4}$ and for some positive constants $c_1$, $c_2$. Suppose that the first inequality is strict for all $u \ne v$. Let the boundary conditions (4) and (5) and the initial condition (6) be fulfilled. Without loss of generality we can assume that $g_1 = g_2 = 0$ For nonhomogenous conditions on the boundary of $\Omega$ we can transform this problem into the problem with homogenous conditions.

Define a function operator

$$H(x) = -\sum_{i=1}^{3} \frac{\partial}{\partial x_i} a_i(x, u_0, u_1, u_2, u_3) - a_0(x, u_0, u_1, u_2, u_3)$$

and assume that the functions $u_i$, $H$, $f$, $\dfrac{\partial f}{\partial t}$ are integrable in the sense presented above on the set $\Omega$. Then, there exists a unique essentially bounded weak solution $u(x,t)$ of our problem. This solution is the classical (strictly) solution if the functions $a_i$, $f$, $H$, $g_1$, $g_2$ are at least continuous and differentiable on the product set in the form $\Omega \times \langle 0,T\rangle$.

The mentioned integration is in the Lebesgue sense, where the set of zero measure is not essential. Therefore, we call the solution to be "weak", which is more general than the classical solution.

The mentioned Rothe's method consists in some decomposition of the time interval $\langle 0,T\rangle$ with the sufficiently large numbers into small intervals with the step $T/n$. Then we have this decomposition in the form

$$t_0 = 0 < t_1 = \frac{T}{n} < t_2 = \frac{2T}{n} < \text{É} \; < t_n = T .$$

Now we use the approximation of the type

$$\frac{du}{dt} = \frac{u(x, t_i) - u(x, t_{i-1})}{T/n}, \; t \in \langle t_{i-1}, t_i\rangle, \; i = 1, 2, \ldots, n$$

Then we have the equation in the discrete form

$$-\left[a\left(x,u(x,t_i),u'(x,t_i)\right)\right]' + u(x,t_i)\frac{n}{T} = f(x,t_i) + u(x,t_{i-1})\frac{n}{T},$$

which is, in principle, an elliptic partial differential equation for some fixed $t_i$. This is based on the value of $u(x,t_{i-1})$ which is the known function on $\Omega$, because we can go by the step by step principle from the point $t_0 = 0$ and this is an initial condition for this equation.

Therefore, we can use so called Galerkin's method with respect to the finite elements method to solve an elliptic boundary problem. Then we obtain for the time moment $t_i$ the solution $u(x,t_i)$ of the corresponding elliptic problem. In this way we obtain a set of the functions $\{u(x,t_i)\}$, $i = 1, 2, ..., n$ which are defined on the set $\Omega$. Finally, we take the following function

$$u^*(x,t) = u(x,t_{i-1}) + \frac{t - t_{i-1}}{T/n}\left[u(x,t_i) - u(x,t_{i-1})\right].$$

Here we get the linear approximation of the solution of our initial and boundary problem for $i \in \langle t_{i-1},t_i\rangle$, $i = 1, 2, ..., n$ of the form $u^*(x,t) \approx u(x,t_i)$. That means that we approximately know the solution for more complicated bodies in the sense of heat distribution. For more information on numerical methods the reader can refer to [1, 5, 7, 8, 9].

**Acknowledgement:** This research has been supported by the Slovak Grant Agency VEGA through the projects No. 1/0867/08 and No. 2/0097/08.

## References

[1] BARTAK, J. et al.: *Partial Differential Equations - part II. (in Czech)*, SNTL Praha, 1988.
[2] GULDAN, V., MARCOKOVA, M.: *Orthogonal Polynomials and Related Special Functions Applied in Geosciences and Technical Computations* (submitted).
[3] KACUR, J.: *Method of Rothe in Evolution Equations,* Teubner, Leipzig 1985.
[4] KOVACIK, O., RAKOSNIK, J.: *On the Spaces $L^{p(x)}$* and *$W^{k,p(x)}$* , Czechoslovak Math. J., 41(116) 1991, 592–618.
[5] MARCOKOVA, M., GULDAN, V.: *On One Orthogonal Transform Applied on a System of Orthogonal Polynomials in Two Variables,* J. Appl. Math., 2/2009, 239–245.
[6] PERICHTA, P., AC, V., ORSANSKY, P.: *The Method for Calculation of the Thermomechanical Deformation,* Proc. of Int. Conf Appl. Phys.Condensed Mater., 2007, 263–266.
[7] REKTORYS, K. et al.: *Review of Applied Mathematics (in Czech),* Academia Praha, 1995.
[8] REKTORYS, K.: *Method of Time Discretization and Partial Differential Equations (in Czech)*, SNTL, Praha 1985.
[9] VITASEK, E.: *Numerical Methods (in Czech)*, SNTL Praha, 1987.

Maria Zaskalicka – Pavel Zaskalicky – Mariana Benova – Mahmud A.R. Abdalmula – Branislav Dobrucky *

# ANALYSIS OF COMPLEX TIME FUNCTION OF CONVERTER OUTPUT QUANTITIES USING COMPLEX FOURIER TRANSFORM/SERIES

*The paper deals with the complex- and discrete Fourier transform which has been considered for both three- and two phase orthogonal voltages and currents of systems. The investigated systems are power electronic converters supplying alternating current motors. Output voltages of them are strongly non-harmonic ones, so they must be pulse-modulated due to requested nearly sinusoidal currents with low total harmonic distortion. Modelling and simulation experiment results of half-bridge matrix converter for both steady- and transient states are given under substitution of the equivalence scheme of the electric motor by resistive-inductive load and back induced voltage. The results worked-out in the paper confirm a very good time-waveform of the phase current and results of analysis can be used for fair power design of the systems.*

## Nomenclature

| | |
|---|---|
| $U$ | *root-mean-square (RMS) value of phase voltages* |
| $U^{L-L}$ | *root-mean-square value of line-to-line voltages* |
| $U_{AC}$ | *magnitude of AC interlink voltage* |
| $U_{DC}$ | *direct current supply voltage magnitude* |
| $U_{emf}$ | *induced count-vise voltage* |
| $A_v$ | *amplitude of v-harmonic quantities* |
| $v$ | *order of harmonics* |
| $f(t)$ | *time function* |
| $\boldsymbol{f(t)}$ | *complex time function* |
| $C_v$ | *complex amplitude of v-harmonic* |
| $C^*$ | *complex conjugated amplitude* |
| $f_s$ | *switching frequency* |
| *PWM* | *pulse-width-modulation* |
| $f_1$ | *fundamental frequency* |
| $m_a$ | *amplitude modulation ratio* |
| $m_f$ | *frequency modulation ratio* |
| $\boldsymbol{a}$ | *exp(2π/3)* |
| *i, k* | *indexes* |
| *DC/AC* | *voltage sourced inverter* |
| *AC/AC* | *direct matrix converter* |
| *T* | *time period* |

## 1. Introduction

Time domain waveforms of electrical quantities can be either *continuous* or *discrete*, and they can be either *periodic* or *aperiodic*. This defines four types of Fourier transforms: the Fourier series (continuous, periodic), and the Fourier transform (continuous, aperiodic) and discrete versions: the Discrete Fourier Transform – DFT (discrete, periodic), the Discrete Time Fourier Transform (discrete, aperiodic) [1]–[3]. All four members of the Fourier transform family above can be carried out with either real- or complex input data. In spite of complex amplitudes of harmonic components is notation of Fourier series in complex form more compact and easier than pure real expressions. This holds true also for complex Fourier transform which is very close to complex Fourier series [3]. Both of them, Fourier transform and series, operate usually with real time functions [4], [5]. The method of complex conjugated amplitudes has been used for solving of electrical circuits, and electrical machines, too [6], [7].

However, the output quantities of real power electronic converters can be transformed into complex time functions using Park or Clarke transform, respectively, as vectors rotating in complex Gauss plain [8]–[10]. The most advantage of this form of presentation is – in the case of symmetrical sys-tem – that periodicity of the waveforms in the complex plain is 2*m*-times higher than in a real time domain. So, the Fourier analysis, also integral values calculation, can be done more quickly. An-other benefit is possibility of direct use of complex Fourier transform/series because quantity func-tions present complex input data for continuous or digital processing.

## 2. Non-harmonic waveforms of power converter output quantities

Output voltage of a power electronic converter is strongly non-harmonic because of its switching-, pulse nature. The single-

* **Maria Zaskalicka[1], Pavel Zaskalicky[1], Mariana Benova[2], Mahmud A.R. Abdalmula[3], Branislav Dobrucky[2],**
[1] Faculty of Electrical Engineering, Technical University of Kosice, Slovakia
[2] Faculty of Electrical Engineering, University of Zilina, Slovakia
[3] University of 7th April, Zawia, Libya

phase voltage inverter with full pulse-width has rectangular waveform output with high content of harmonic components (more than 45 %). Three-phase inverters produce three line-to-zero (phase) and three line-to-line non-harmonic voltages, Fig. 1.

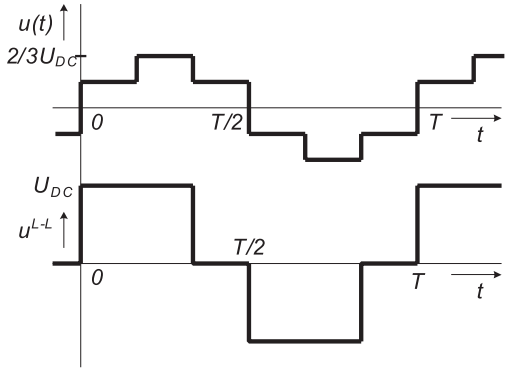



*Fig. 1 Output voltages of three-phase inverter with full-width pulses: phase - (a) and line-to line voltage (b)*

It is known that the conversion coefficient between phase- and line-to-line harmonic voltages is $\sqrt{3}/3$, and the magnitude of fundamental harmonic of phase voltage is by $\sqrt{3}/3$ less than line-to-line voltage.

Based on classical Fourier transform the amplitude of fundamental harmonic of phase voltage can be calculated:

$$\begin{aligned} A_1 &= \frac{8}{T}\int_0^{T/4} f(t)\cdot\sin\omega t\,dt = \\ &= \frac{4}{\pi}U_{DC}\left\{\int_0^{\pi/3}\frac{1}{3}\cdot\sin\omega t\,d\omega t + \int_{\pi/3}^{\pi/2}\frac{2}{3}\cdot\sin\omega t\,d\omega t\right\} = \\ &= \frac{2}{\pi}U_{DC} \end{aligned} \quad (1)$$

where $U_{DC}$ is supply voltage of the inverter.

Fundamental harmonic of line-to-line voltage (see Fig. 1) is similarly

$$\begin{aligned} A_1^{L-L} &= \frac{8}{T}\int_0^{T/4} f(t)\cdot\sin\omega t\,dt = \\ &= \frac{4}{\pi}U_{DC}\int_{\pi/6}^{\pi/2}\cdot\sin\omega t\,d\omega t = \frac{2\sqrt{3}}{\pi}U_{DC} \end{aligned} \quad (2)$$

so, the amplitudes are the same taking in account the relation coefficient.

Also, the total harmonic distortion factor (THD) of both voltages is the same. Let's calculate first the THD of phase voltage

$$\begin{aligned} &\sqrt{\frac{\sum U_v^2}{U_1^2}} = \sqrt{\frac{U^2-U_1^2}{U_1^2}} = \sqrt{\left(\frac{U}{U_1}\right)-1} = \\ &= \sqrt{\left(\frac{\sqrt{2}/3}{\sqrt{2}/\pi}\right)-1} = 0.3108 \end{aligned} \quad (3)$$

where $U_1$ RMS value of fundamental harmonic of the phase voltage ($= A_1/\sqrt{2}$),

$$U = \sqrt{\frac{4}{T}\int_0^{T/6}\left(\frac{1}{3}U_{DC}\right)^2 dt + \frac{4}{T}\int_{T/6}^{T/4}\left(\frac{2}{3}U_{DC}\right)^2 dt\,dt} = \frac{\sqrt{2}}{3}U_{DC},$$

and $U$ is RMS value of phase voltage.

The total harmonic distortion of line-to-line voltage will be the same

$$\begin{aligned} &\sqrt{\frac{\sum\left(U_v^{L-L}\right)^2}{\left(U_v^{L-L}\right)^2}} = \sqrt{\frac{\left(U^{L-L}\right)^2-\left(U_1^{L-L}\right)^2}{\left(U_1^{L-L}\right)^2}} = \\ &= \sqrt{\left(\frac{U^{L-L}}{U_1^{L-L}}\right)^2-1} = \sqrt{\left(\frac{\sqrt{2/3}}{\sqrt{6}/\pi}\right)^2-1} = 0.3108 \end{aligned} \quad (4)$$

where
$U$ and $U_1$ have the same meaning as above,
$U^{L-L} = \sqrt{(2/3)}$.

Finally, based on total mathematical induction, we can show that ***both voltages, phase- and line-to-line, comprise the same harmonic components.***

Proof #1:
General relation for $v$-harmonic calculation of line-to-line voltage

$$\begin{aligned} \left|A_v^{L-L}\right| &= \frac{8}{T}\int_0^{T/4} f(t)\cdot\sin\left(v\frac{2\pi}{T}\right)dt = \\ &= \frac{4}{\pi}U_{DC}\int_{\pi/6}^{\pi/2}\sin(v\omega t)d\omega t = \\ &= \frac{4}{\pi}\cdot\frac{1}{v}U_{DC}\left[\cos\left(v\frac{\pi}{6}\right)-\cos\left(v\frac{\pi}{2}\right)\right] \end{aligned} \quad (5)$$

Similarly for phase voltage

$$\begin{aligned} \left|A_v\right| = \frac{4}{3\pi}\cdot\frac{1}{v}U_{DC}\left\{\left[1-\cos\left(v\frac{\pi}{3}\right)\right]+\right. \\ \left.+2\left[\cos\left(v\frac{\pi}{3}\right)-\cos\left(v\frac{\pi}{2}\right)\right]\right\} \end{aligned} \quad (6)$$

From the initial condition for equivalence of the amplitudes

$$\left|A_v\right| = \frac{\left|A_v^{L-L}\right|}{\sqrt{3}} \quad (7)$$

$$\begin{aligned} &\frac{1}{3}\left[1+\cos\left(v\frac{\pi}{3}\right)-2\cos\left(v\frac{\pi}{2}\right)\right] = \\ &= \frac{1}{\sqrt{3}}\left[\cos\left(v\frac{\pi}{6}\right)-\cos\left(v\frac{\pi}{2}\right)\right] \end{aligned} \quad (8)$$

For odd $v$-numbers the terms of $[\cos(v.\pi/2)]$ will be equal to zero. Then

$$\frac{1}{3}\left[1+\cos\left(v\frac{\pi}{3}\right)\right] = \frac{1}{\sqrt{3}}\left[\cos\left(v\frac{\pi}{6}\right)\right] \quad (9)$$

After substitution $[1 + \cos(v.\pi/3)] = 2[\cos(v.\pi/6).\cos(v.\pi/6)]$ we can obtain

$$\cos\left(v\frac{\pi}{6}\right)=\pm\frac{\sqrt{3}}{2}\ \longrightarrow\ |U_v|=\frac{1}{\sqrt{3}}|U_v^{L-L}| \tag{10}$$

Eq. (10) implies that for odd *v*-numbers the condition (7) is fulfilled.

***A paradox of different shape of both voltages is possible to explain so that the phase-spectra of the voltages are not the same, they are different.***

## 3. Using orthogonal output voltages and complex Fourier analysis

Applying Park/Clarke transform the complex time function of output phase voltage in three-phase system is

$$\boldsymbol{u}(t)=\frac{2}{3}\left[u_1(t)+\mathbf{a}\cdot u_2(t)+\mathbf{a}^2\cdot u_3(t)\right]=u_\alpha+j\cdot u_\beta \tag{11}$$

where after adapting

$$\begin{aligned} u_\alpha &= \frac{1}{3}\left[2u_a(t)-u_b(t)-u_c(t)\right]\\ u_\beta &= \frac{\sqrt{3}}{2}\left[u_b(t)-u_c(t)\right] \end{aligned} \tag{12a,b}$$

It deals with the voltage vectors rotating in Gauss α,β-plain at an angular speed ω which can be also non-constant.

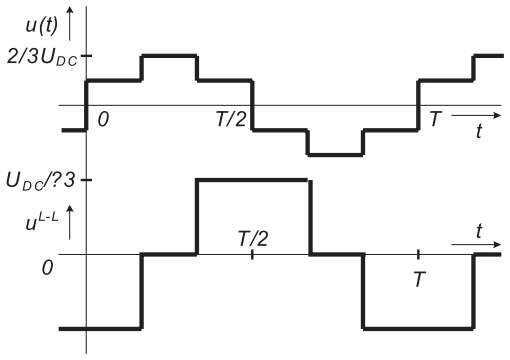



*Fig. 2 Orthogonal voltage systems of three-phase inverter with full-width pulses: direct $u_\alpha$ (a) - and quadrature voltage $u_\beta$ (b)*

Now, the voltages $u_\alpha$ and $u_\beta$ create an orthogonal system, and complex Fourier transform can be used.

So, then the complex Fourier transform or/and complex Fourier coefficients can be calculated

$$\boldsymbol{U}(v\omega t)=\int_0^T\boldsymbol{u}(t)\cdot\exp(-j\omega t)dt \tag{13}$$

or, respectively

$$\boldsymbol{C}_v=\frac{1}{T}\int_0^T\boldsymbol{u}(t)\cdot\exp\left(-jv\frac{2\pi}{T}t\right)dt \tag{13a}$$

whereby their mutual relation is $\boldsymbol{C}_v=\frac{1}{T}\boldsymbol{U}(v\omega t)$ (13b)

where $\omega = 2\pi/T$.

The discrete Fourier transformation has been used for calculation of individual harmonics coefficients [3]:

$$\boldsymbol{U}[v]=\frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{u}[n]\cdot\exp\left(-jv\frac{2\pi}{N}n\right) \tag{14}$$

Alternatively, Euler's relation can be used to rewrite the forward transform in a rectangular form:

$$\boldsymbol{U}[v]=\frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{u}[n]\cdot\left(\cos(2\pi vn/N)-j\cdot\sin(2\pi vn/N)\right) \tag{14a}$$

Real and imaginary part of $\boldsymbol{U}(v)$ can also be expressed:

$$\begin{aligned} \mathrm{Re}\{U(v)\} &= \frac{2}{N}\sum_{n=0}^{N-1}u(n)\cos\left(2\pi\frac{nv}{N}\right);\\ \mathrm{Im}\{U(v)\} &= \frac{-2}{N}\sum_{n=0}^{N-1}u(n)\sin\left(2\pi\frac{nv}{N}\right) \end{aligned} \tag{15a,b}$$

Based on the above definition the relation for complex Fourier coefficients of complex voltage function yields:

$$C_v=\frac{1}{T}\int_0^T\left[u_\alpha(t)+ju_\beta(t)\right]\cdot\exp\left(-jv\frac{2\pi}{T}t\right)dt \tag{16}$$

Eq. (16) can be decomposed into two scalar equations for $C_v^\alpha$ and $C_v^\beta$, if needed:

$$C_{v\alpha}=\frac{2}{T}\int_0^T u_\alpha(t)\cdot\exp\left(-jv\frac{2\pi}{T}t\right)dt \tag{16a}$$

$$C_{v\beta}=\frac{2}{T}\int_0^T u_\beta(t)\cdot\exp\left(-jv\frac{2\pi}{T}t\right)dt \tag{16b}$$

Such a Fourier series is developed on the system of orthogonal functions $\exp(j.n.2\pi.t/T)$, $n = 0, \pm1, \pm2$ ..., for which the integral

$$\int_0^T\exp\left(-jn\frac{2\pi}{T}t\right)\exp\left(-jm\frac{2\pi}{T}t\right)dt \tag{16c}$$

is equal to 0 for $m \neq -n$, and equal to $T$ for $m = -n$.

***The system of voltages is ortho-normal, too. Since $u_\alpha$ voltage will contain sin-terms only, the second one $u_\beta$ cos-terms.***

**Orthogonal voltage system for a two phase converter system consisting of two single-phase matrix converters**

The orthogonal two-phase converter system comprises two single-phase converters. Supposing bridge connection single-phase matrix converter [7], [14] then for its output voltage can be applied pulse-width modulation (PWM). The second voltage of the 2-phase orthogonal system, generated by a second single-phase bridge matrix converter has the same waveform but it is shifted by 90 degrees (or π/2 respectively) to the right. Thus the complex Fourier transform and analysis described above can be used. ***However, based on proof #1 the amplitude spectra $C_{v\alpha}$ and $C_{v\beta}$ of $u_\alpha$ and $u_\beta$ will be the same, so it is enough to calculate only one of them.***

The sinusoidal PWM is mature technology which can be classified into Natural sampled (analogue sinetriangle), regular sampled (digital sine-triangle) and space vector modulation. The natural Sampled and regular Sampled also can be divided into two-level (bipolar) or three-level (unipolar) PWM [5], [11]. There are two major concerns for generating sinusoidal PWM. The first is to minimize the creation of low order harmonics in the output voltage. The second is the mitigation of switching frequency harmonics. Both these concerns are controlled by the shape of the PWM pattern which should be controlled to minimize the generation of switching harmonics and maximum the harmonic cancellation between the line-line voltages.

The unipolar PWM was chosen as the best PWM for voltage control of single-phase inverters considering the output voltage harmonic spectrum. Normal unipolar sinusoidal PWM is shown in Fig. 3.

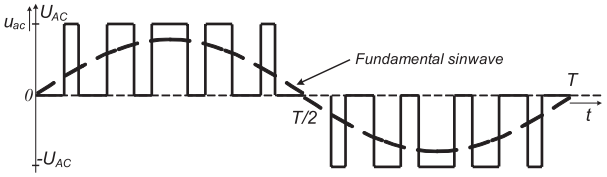



*Fig. 3 Output voltages of single-phase bridge matrix converter with PWM technique*

It can be observed that the area of each pulse corresponds approximately to the area under the sine-wave between the adjacent mid-points of the off-periods. The pulse-width modulated wave has much lower low/order harmonic content than the other waveforms. Generally the synchronous modulation is used. In synchronous modulation the modulation frequency is an integer multiple of the frequency of reference sine-wave [5], [13]. The turn on ($\alpha$) and turn off ($\beta$) angles are calculated by the discreet substitution of the reference sine-wave.

Both amplitude- and frequency modulation ratios $m_a$ and $m_f$ are defined as:

$$m_a = \frac{U_{1m}}{U_{AC}}; \quad m_f = \frac{f_s}{f_1}, \tag{17a,b}$$

where

$U_{1m}$ is reference amplitude of fundamental harmonic,
$U_{AC}$ magnitude of supply voltage,
$f_s$ switching frequency,
$f_1$ fundamental frequency.

So, the peak amplitude of the fundamental harmonic component (equal to reference voltage) is $m_a$ times $U$, and varies linearly with $m_a$ (providing $m_a \geq 1$). If the frequency modulation ratio $m_f$ is sufficiently great, the difference between real values and discrete values is negligible.

The converter's output voltage (Fig. 3) can be mathematically expressed as a Fourier series of the form [5], [12], [13].

$$u(t) = \sum_{v=1}^{\infty} \frac{2U}{v\pi} \sum_{k=1}^{(m_f/4)-1} (\sin 2v\alpha_k - \sin 2v\beta_k)\sin(v\omega t) \tag{18}$$

where $\alpha_k$ and $\beta_k$ - are turn-on and turn-off angles calculated for each $k$-interval by [13]:

$$\alpha_k = \frac{2k\pi}{m_f} - m_a \frac{\pi}{m_f} \sin \frac{2k\pi}{m_f} \tag{18a}$$

$$\beta_k = \frac{2k\pi}{m_f} + m_a \frac{\pi}{m_f} \sin \frac{2k\pi}{m_f} \tag{18b}$$

Calculated output voltage for $m_a = 1$, $m_f = 18$ and finite number of harmonics is depicted in Fig. 4 .

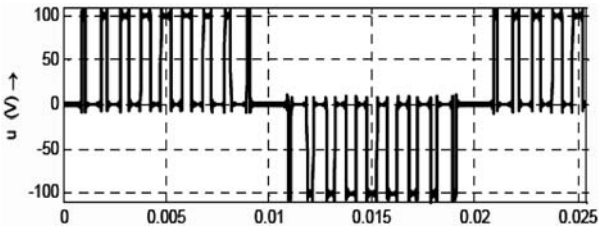
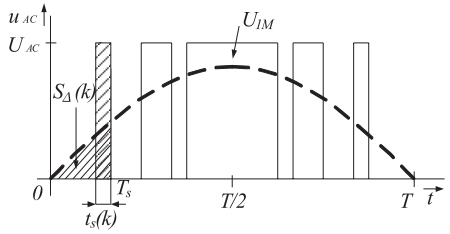
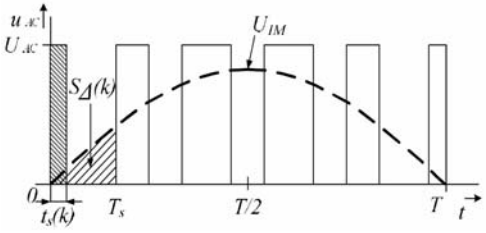



*Fig. 4 Output voltages of single-phase bridge matrix converter with PWM technique*

**Considering single-phase inverter and unipolar PWM with even $m_f$.**




*a)*




*b)*

*Fig. 5 Output voltages of single-phase bridge matrix converter with even $m_f$ a) or b)*

Then total voltage time waveform will be:

$$u(t) = \sum_{v=1}^{\infty} \frac{4 \cdot U_{DC}}{v \cdot \pi} \sum_{k=0}^{(m_f/4)-1} \left[ \cos\left( v \cdot k \cdot \frac{2\pi}{m_f} \right) - \cos\left( v \cdot k \cdot \frac{2\pi}{m_f} + v \cdot \frac{2\pi}{T} \cdot t_s(k) \right) \right] \cdot \sin(v\omega t) \tag{19}$$

where $t_s(k)$ is the switching instant at $k$-interval:

$$t_s(k) = \frac{1}{U_{DC}} \cdot S_\Delta(k) = \frac{m_a \cdot m_f}{2\pi} \left[ U_m \cdot \frac{m_f}{2\pi} \cdot \left[ \cos\left( \frac{2\pi}{m_f} \cdot k \right) - \cos\left( \frac{2\pi}{m_f} \cdot k + 1 \right) \right] \right] \tag{20}$$

and $S_\Delta(k)$ is the area under sinewave during $k$-switched interval:

$$S_\Delta(k) = U_m \cdot \frac{m_f}{2\pi} \cdot \left[ \cos\left( \frac{2\pi}{m_f} \cdot k \right) - \cos\left( \frac{2\pi}{m_f} \cdot (k+1) \right) \right] \tag{21}$$

## 4. Complex Fourier analysis of the voltage of AC/AC half-bridge matrix converter system

Due to rather high number of power semiconductor switches of both converters (totally 8), the half-bridge connection of the matrix converter is a better solution. The equivalent circuit diagram of one half-bridge single phase converter (one of two-phase orthogonal systems) is depicted in Fig. 6. Since the voltages of the matrix converter system are orthogonal, the second phase converter is the same and its voltage is shifted by 90 degree. Due to half-bridge connection the bipolar PWM should be used.

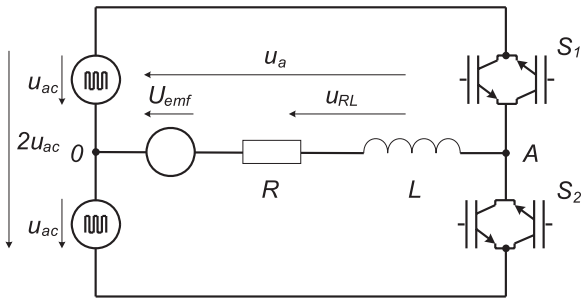



*Fig. 6 Circuit diagram of single-phase half-bridge matrix converter*

Contrary to the bridge matrix converter the half-bridge connection doesn't provide unipolar PWM control, so the bipolar pulse switching technique should be used. Based on bipolar PWM control the output orthogonal voltages will be presented in Figs. 8a and 8b. This type of control technique is more complicated than unipolar type.

**Considering bipolar PWM with switching frequency equal to odd multiply of fundamental frequency.**

Consequently, the harmonics in the converter output voltage waveform appears as sidebands, centered on the switching frequency $f_s$ and its multiples, that is, around harmonics $m_f$, 2 $m_f$, 3 $m_f$, and so on. This general pattern holds true for all $m$a smaller than (or equal to) 1. For a frequency modulation ratio $m$f >= 9 (which is our case), the harmonic amplitudes are almost independent on $m$f, though $m$f defines the frequencies at which they occur. Theoretically, the frequencies at which voltage harmonics occur can be defined as

$$f_v = \left( x \cdot m_f \pm y \right) \cdot f_1 , \tag{22}$$

that is, the harmonic order $v$ corresponds to the $y$-th sideband of the $x$-times the frequency modulation ratio $m_f$

$$v = \left( x \cdot m_f \pm y \right) \cdot f_1 , \tag{22a}$$

where the fundamental harmonic frequency corresponds to $v = 1$. For odd values of $x$, the harmonics exist only for even value of $y$, and opposite, for even values of $x$, the harmonics exist only for odd value of $y$.

Choosing the frequency modulation ratio mf as odd integer results in an odd symmetry [$u(-t) = -u(t)$] as well as half-wave symmetry [$u(-t) = -u(t + T_s/2)$] with the time origin shown in Figs. 3 or 4. Therefore, only odd harmonics are present and the even harmonics disappear from the wave form of $u$a. Moreover, only the coefficients of the sine series in Fourier analysis are finite; those for the cosine series are zero.

The harmonic spectrum is plotted in Fig. 7, which is plotted for $m_f = 39$.

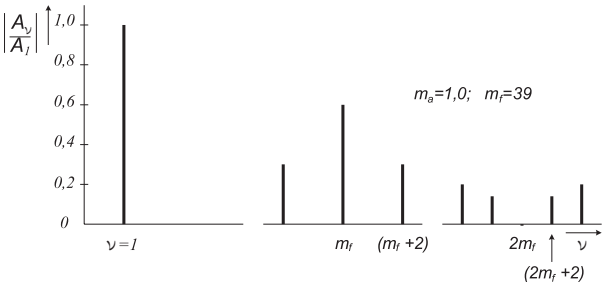



*Fig. 7 Harmonic amplitude spectrum of bipolar PWM with odd $m_f$*

For the parameters (the same as in [11] to be compared):

$Z_x$ $U_{DC}$ = 300 V – input voltage,
$f_{IN} = f_S$ = 39 kHz – switching frequency,
$f_{OUT}$ = 50 Hz – fundamental output frequency,
$m_a$ = 1; $m_f$ = 39 – amplitude and frequency ratios,

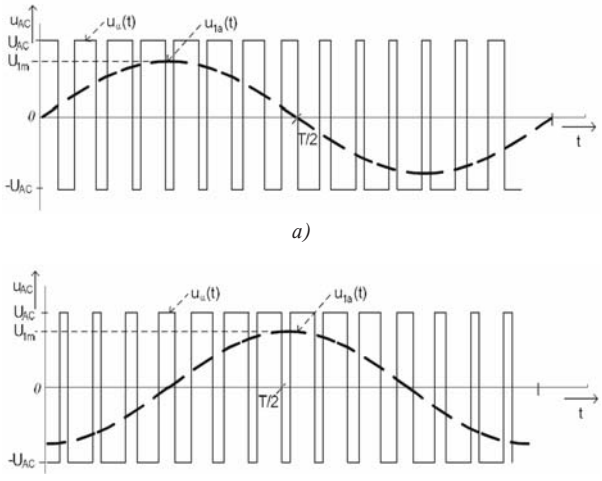
the amplitudes of the first 30 voltage harmonics (by 165$^{th}$-harmonic) were calculated:

$A_1 = 150.m_a = 150$ V; $A_{39} = 90.16$ V; $A_{39-2} = A_{39+2} = 47.70$ V;
$A_{39-4} = A_{39+4} = 2.70$ V;
$A_{78-1} = A_{78+1} = 27.15$ V; $A_{78-3} = A_{78+3} = 31.80$ V;
$A_{78-5} = A_{78+5} = 4.95$ V; $A_{117} = 16.95$ V;
$A_{117-2} = A_{117+2} = 9.30$ V; $A_{117-4} = A_{117+4} = 23.55$ V;
$A_{117-6} = A_{117+6} = 6.60$ V; $A_{156-1} = A_{156+1} = 10.20$ V;
$A_{156-3} = A_{156+3} = 1.35$ V; $A_{156-5} = A_{156+5} = 17.85$ V;
$A_{156-7} = A_{156+7} = 7.50$ V;

Note: The carried-out results are identical with those given in [11] for DC/AC inverter with bipolar sinusoidal PWM.

**Considering bipolar PWM with switching frequency equal to even multiply of fundamental frequency.**

The orthogonal voltages with bipolar PWM control are depicted in Figs. 8a, b.




*a)*

*b)*

*Fig. 8 Output orthogonal voltages of the half-bridge matrix converter system with bipolar pulse-width-modulation: direct (a) and quadrature one (b)*

It deals with sinusoidal bipolar pulse-width-modulation contrary to unipolar regular PWM [5], [11]. Switching-pulse-width can be determined based on equivalence of the average values of reference waveform and resulting average value of positive and negative switching pulses area during a switching period (see Figs. 5a, b and 8a, b).

Then, the total voltage time waveform will be:

$$u(t) = \sum_{v=1}^{\infty} \frac{4 \cdot U_{DC}}{v \cdot \pi} \sum_{k=0}^{(m_f/4)-1} \left\{ \left[ \cos\left( v \cdot k \cdot \frac{2\pi}{m_f} \right) - \cos\left( v \cdot k \cdot \frac{2\pi}{m_f} + v \cdot \omega \cdot t_s(k) \right) \right] - \left[ \cos\left( v \cdot (k+1) \cdot \frac{2\pi}{m_f} - v \cdot \omega \cdot t_s(k) \right) - \cos\left( v \cdot k \cdot \frac{2\pi}{m_f} \right) \right] \right\} \cdot \sin(v \cdot \omega \cdot t) \quad (23)$$

where the switching instant is equal to:

$$t_s(k) = \frac{1}{2U_{DC}} \cdot S_\Delta(k) + \frac{T_s}{2} \quad (24)$$

and the area under sinewave is the same as using unipolar PWM (see Eq. 21).

## 5. Current harmonics investigation under resistive-inductive load with *Uemf*

Current time-waveforms for a fundamental harmonic component in steady-state $i_{S1}(t)$ is [12]:

$$i_{S1}(t) = \frac{A_1 - U_{emf}}{Z_1} \cdot \sin(\omega t - \varphi_1) = \\ = \frac{U}{Z_1} C_I(1) \cdot \sin(\omega t - \varphi_1) \quad (26)$$

Current time-waveforms for higher-harmonic components in steady-state $i_{Sv}(t)$ are given [12]

$$i_{Sv}(t) = \frac{A_v}{Z_v} \cdot \sin(v\omega t - \varphi_v) = \\ = \frac{U}{Z_v} C_I(v) \cdot \sin(v\omega t - \varphi_v) = I_v \cdot \sin(v\omega t - \varphi_v) \quad (27)$$

where:
$A_v = A_1 \cdot C_v(v)$ – amplitude of $v$-harmonic voltage component extracted from Eq, (23),
$A_1 = m_a \cdot U_{AC}$ – amplitude of 1. harmonic voltage component,
$U_{emf}$ – counter-voltage (of electromagnetic force)
$|Z_v| = Z_v = \sqrt{R^2 + (v \cdot \omega \cdot L)^2}$ – module of complex impedance of resistive-inductive load
$\varphi_v = \arctan(v.\omega.L/R)$ – argument of complex impedance of resistive-inductive load
$C_I(v)$ – Fourier coefficient of $v$-harmonic current component,
$I_v$ – amplitude of $v$-harmonic current component.

Harmonic current components can be computed similarly using above methodology and work [12]. The accurate calculation of $U_{emf}$ can be obtained by using the motor circle diagram. The total current in steady-state will be summarisation of single harmonics.

The Fourier analysis can be used also for investigation of behaviour of the system in transient state. The total current of $v$-harmonic component $i_v$ will be summarisation of current in steady-state $i_{Sv}$ and current in the transient phenomenon $i_{Tv}$

$$i_v(t) = i_{Sv}(t) + i_{Tv}(t) = \frac{A_v}{Z_v} \cdot \sin(v\omega t - \varphi_v) + \\ + \frac{A_v}{Z_v} \cdot \sin\varphi_v \cdot \exp(-t/\tau) \quad (28)$$

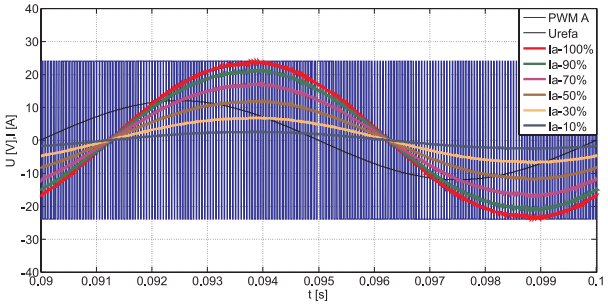
where $i_v$ is total current waveform
$i_{Sv}$ steady-state component of total current
$i_{Tv}$ transient component of total current
$\tau = R/L$ – time constant of resistive-inductance load.

Total current as well as both components should be calculated for each harmonic.

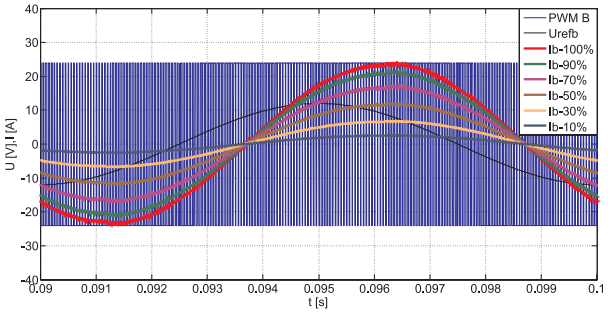
## 6. PC simulation in MatLab programming environment

The MatLab programming environment was used for simulation. Simulation experiments were done for the parameters: $R$ = = 0.1275 Ohm, $L$ = 21.6 mH, $U_{AC}$ = 12 V, $f$ = 100 Hz at $m_a$ = 1, $m_f$ = 100, $U_{emf}$ = 0.1–0.9 of $A_1$, time increment $\Delta t$ = 0.5 μs.

The graphic interpretation of steady-state is shown in Figs. 9a, b.




*a)*




*b)*

*Fig. 9 Output orthogonal voltages of the half-bridge matrix converter system with bipolar pulse-width-modulation and $m_a$ = 0.2, 0.4, 0.6, 0.8, direct phase-(a) and quadrature phase voltage (b)*

Time wave-forms of currents for transient-state and various time constants are depicted in Fig. 10.

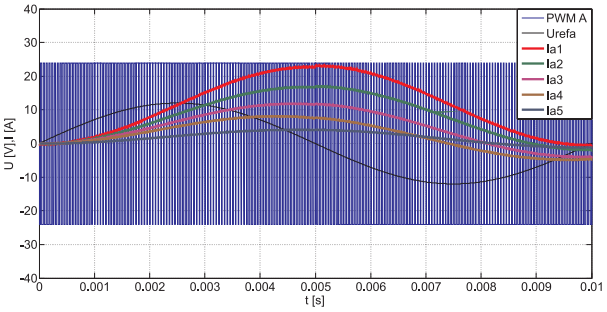



*Fig. 10 Voltage of the half-bridge matrix converter and time waveforms of transient currents for L/R = 0.17 s, 0.04 s, 0.01 s, 52 ms, 14 ms*

## 7. Conclusions

The complex Fourier transformation was considered for three- and two phase orthogonal systems of converter output voltages, strongly non-harmonic ones. The solution given in the paper makes it possible to analyse more exactly the effect of each harmonic component comprised in the total waveform on resistive-inductive load or induction motor quantities. The proposed system with AC interlink in comparison with currently used conventional systems uses two single phase half bridge matrix converters with bipolar pulse-width modulation. The advantage is then less number of semiconductor devices of the converters. However, in practice, the necessary imposition of a dead-band, or blanking time, results in some distortion of the output voltage. Then, the dead-band, its symptoms and related remedies, are necessary to take into account for solutions.

**Acknowledgement**

The authors wish to thank for support to R&D operational program Centre of excellence of power electronics systems and materials for their components No. OPVaV-2008/2.1/01-SORO, ITMS 26220120003 funded by European regional development fund (ERDF), and to VEGA Agency for the project No. 1/0470/09, and APVV project No-0510-06.

## References

[1] SMITH, S.W.: *The Complex Fourier Transform,* The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing, 1997–2007, pp. 567–580, ISBN 0-9660176-3-3.
[2] HOFFMANN, M.: *Fourier transform and DFT*, Digital Signal Processing Maths, DESY, Hamburg, Germany, 2007, pp. 33–36; 42–56.
[3] CIZEK, V.: *Discrete Fourier Transformation and Its Applications (in Czech),* Mathematical Seminar Edition, SNTL Publisher, Prague (CZ), 1981.

[4] BIRINGER, P. P., SLONIM, M.A.: *Determination of Harmonics of Converter Current and/or Voltage Waveforms (New Method for Fourier Coefficient Calculations)*, IEEE Transactions on Industry Applications, Vol. IA-16, 2/April 1980, pp. 248–253.
[5] DEHBONEI, H., L. BORLE, NAYAR, C.V.: *Optimal Voltage Harmonic Mitigation in Single-Phase Width Modulation,* Proc. of AUPEC'01 Australasian Universities Power Engineering Conference, Perth (AUS), 2001, pp. 296.
[6] TAKEUCHI, T. J.: *Theory of SCR Circuit and Application to Motor Control. Electrical Engineering.* College Press, Tokyo (JP), 1968 - Russian Edition, Energija Publisher, Leningrad-St. Petersburg (RF), 1973, pp. 248, T 051(01)-73 130-73.
[7] DOBRUCKY, B., SUL, R., BENOVA, M.: *Mathematical Modelling of Two-Stage Frequency Converter using Special Methods: Complex Conjugated Magnitudes and Orthogonal Park/Clarke Transformation Ones, J. of Applied Mathematics, Vol. II,* 3/2009, ISSN 1337-6365.
[8] JARDAN, R. J., DEVAN, B. S., SLEMON, R. G.: *General Analysis of Three Phase Inverters,* IEEE Trans. on Industry and General Applications, 6/1969, pp. 672–679.
[9] JARDAN, R. J.: *Modes of Operation of Three Phase Inverters,* IEEE Trans. on Industry and General Applications, 6/1969, pp. 680–685.
[10] SOLIK, I., VITTEK, J., DOBRUCKY, B.: *Time-Optimal Analysis of Characteristic Values of Periodical Waveforms in Complex Domain,* J. of Modelling, Simulation, and Control, AMSE Press, Vol. A 28, 3/1990, pp. 49–64.
[11] MOHAN, N., UNDELAND, T. M., ROBBINS, W. P.: *Power Electronics: Converters, Applications, and Design,* John Wiley & Sons, Inc., ISBN 0-471-42908-2.
[12] ZASKALICKY, P., ZASKALICKA, M.: *Calculation of the Current Waveform of an AC Converter by Fourier series,* Proc. of ISEM'05 Int'l Symposium on Electrical Machines, Prague, 2005, pp.194–199.
[13] ZASKALICKY, P., ZASKALICKA, M.: *Mathematical Model of the Single-Phase Inverter with Pulse-Width Modulation,* Proc. of Mechatronika'09 Int'l Conf., Tren. Teplice (SK), 2009, pp. CD-ROM.
[14] BENOVA, M., DOBRUCKY, B., SZYCHTA, E., PRAZENICA, M.: *Modelling and Simulation of HF Half-Bridge Matrix Converter System in Frequency Domain,* Logistyka (PL), No. 6/2009, ISSN 1231-5478.

Ivana Pobocikova *

# BETTER CONFIDENCE INTERVALS FOR A BINOMIAL PROPORTION

*Interval estimation of a binomial proportion is one of the basic problems in statistics. In technical practice a binomial proportion is often used in statistical quality control. The standard Wald interval and the exact Clopper-Pearson interval are the most common and frequently used intervals. They are presented in the majority of statistical literature. It is known that the Wald interval performs poorly and this interval should not be used. In this paper we recommend the alternatives of confidence intervals that have a better performance and are appropriate for practical use. We compare the performance of six alternatives of confidence intervals for a binomial proportion: the Wald interval, the Clopper-Pearson interval, the Wilson score interval, the Wilson score interval with continuity correction, the Agresti-Coull interval and the Jeffreys interval in terms of the coverage probability, the interval length and the root mean square error.*

*Key words: binomial distribution, binomial proportion, confidence interval, coverage probability, interval length, root mean square error*

## 1. Introduction

Interval estimation of a binomial proportion is one of the basic problems in statistics. In technical practice the binomial proportion is often used in statistical quality control.

Let random variable $X$ follow a binomial distribution with parameters $n \in N$ and $\pi \in (0,1)$, abbreviated $X \sim Bi(n,\pi)$. The probability that a random variable $X$ is equal to the value $x$ is given by

$$P(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x},\ x = 0,1,\ldots,n \qquad (1)$$

The parameter $\pi$ is also called binomial proportion. In practice the value of the parameter $\pi$ is usually unknown and must be estimated from a sample. Let $X$ be a number of successes in a random sample of size $n$. The maximum likelihood estimator for from the sample is $p = X/n$. This estimator is unbiased and consistent. The $100\cdot(1-\alpha)\%$ two-sided confidence interval for parameter $\pi$ is an interval $\langle p_L, p_U\rangle$ such as $P(p_L \le \pi \le p_U) \ge 1-\alpha$, where $(1-\alpha)$ is the desired confidence coefficient, $\alpha \in (0,1)$.

Due to the discrete nature of the binomial distribution the interval estimation of binomial proportion is a complicated problem. The standard Wald interval (Laplace, 1812) and the exact Clopper-Pearson interval (Clopper - Pearson, 1934) are the most common and most frequently used intervals. They are presented in the majority of statistical literature. The standard Wald interval (Wald) is based on the standard normal approximation to the binomial distribution. This interval is simple to compute, it is narrow, but the interval has a poor performance. It is known that its coverage probability behaves irregularly even when is not close to 0 and 1. The coverage probability is below a nominal level even for very large sample sizes. It is known that Wald interval has a problem with the zero width interval and overshoot (the lower bound can be below 0 and the upper bound can be above 1). Many autors have pointed out that this interval should not be used (Vollset, 1993, Newcombe, 1998, Brown, Cai, DasGupta, 2001, Pires, Amado, 2008).

The exact Clopper-Pearson interval is based on the exact binomial distribution. This interval eliminates overshoot and zero width intervals and it is known that this interval is strictly conservative and too wide (Newcombe, 1998, Brown, Cai, DasGupta, 2001, Pires, Amado, 2008). Its coverage probability is always equal to or above the nominal level.

In this paper we recommend the alternatives of confidence intervals for binomial proportion that have better performance and are often used in practice, but they are presented sporadically in the basic statistical literature. Here we consider the confidence intervals methods that are based on the standard normal approximation: Wilson score interval (Wilson), Wilson score interval with continuity correction (Wilson+CC), Agresti-Coull interval (Agresti-Coull), and finally the interval that is based on the Bayesian approach: Jeffreys interval (Jeffreys).

We demonstrate the performance of selected alternatives of confidence intervals. For comparison we use common criteria: coverage probability, average coverage probability, conservatism,

* **Ivana Pobocikova**
Department of Applied Mathematics, Faculty of Mechanical Engineering, University of Žilina, Slovakia, E-mail: ivana.pobocikova@fstroj.uniza.sk

interval length, average expected length, root mean square error. We summarize the results for the coverage probability in terms of the observed minimum coverage probability and the average coverage probability and we classify the alternatives of confidence intervals into two classes of acceptable intervals- strictly conservative intervals and intervals that are not strictly conservative, but conservative on average.

Our recommendation of these selected alternatives of confidence intervals is based on our investigations of these intervals and on the existing comparative studies that were presented in recent statistical literature, see e. g. Newcombe (1998), Brown, Cai, DasGupta (2001) and Pires, Amado (2008).

## 2. Alternatives of Confidence Intervals

**Clopper – Pearson interval.** The exact Clopper – Pearson interval (Clopper – Pearson, 1934) is based on inverting two-sided binomial tests on the null hypothesis $H_0 : \pi = \pi_0$ against the alternative $H_1 : \pi \neq \pi_0$. If $X = x$ is observed, the lower and upper bounds are the solutions of the equations

$$\begin{aligned} \sum_{k=0}^{x}\binom{n}{k}p_U^k(1-P_U^k)^{n-k} &= \frac{\alpha}{2} \\ \sum_{k=x}^{n}\binom{n}{k}p_L^k(1-P_L^k)^{n-k} &= \frac{\alpha}{2} \end{aligned} \tag{2}$$

The lower and upper bounds of $100\cdot(1-\alpha)\%$ Clopper – Pearson interval for $0 < X < n$ are

$$\begin{aligned} P_L &= \frac{x}{x+(n-x+1)F_{1-\frac{\alpha}{2}}(2(n-x+1),2x)}, \\ P_U &= \frac{(x+1)F_{1-\frac{\alpha}{2}}(2(x+1),2(n-x))}{n-x+(x+1)F_{1-\frac{\alpha}{2}}(2(x+1),2(n-x))}, \end{aligned} \tag{3}$$

where $F_\alpha(k_1, k_2)$ is the $\alpha$-quantile of $F$-distribution with $k_1$ and $k_2$ degrees of freedom. For $X = 0$ is $p_L = 0$ and $p_U = 1 - \left(\frac{\alpha}{2}\right)^{\frac{1}{n}}$. For $X = n$ is $p_L = \left(\frac{\alpha}{2}\right)^{\frac{1}{n}}$ and $p_U = 1$.

**Wald interval.** Wald interval (Laplace, 1812) is based on inverting Wald test and is obtained by using the Central Limit Theorem that the random variable $U = (p-\pi)/\sqrt{\frac{p(1-p)}{n}}$ is approximately standard normally distributed $N(0,1)$. Thus the lower and upper bounds of $100\cdot(1-\alpha)\%$ Wald interval are

$$\begin{aligned} p_L &= p - k_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}, \\ p_U &= p + k_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}, \end{aligned} \tag{4}$$

where $k_\alpha$ is the $\alpha$-quantile of standard normal distribution $N(0,1)$.

**Wilson score interval.** Wilson score interval (Wilson, 1927) is obtained by noting that the random variable $U = (p-\pi)/\sqrt{\frac{\pi(1-\pi)}{n}}$ is approximately standard normally distributed $N(0,1)$. The lower and upper bounds are the solutions of the equation $(p-\pi)^2 = k_\alpha^2\frac{\pi(1-\pi)}{n}$. The lower and upper bounds of $100\cdot(1-\alpha)\%$ Wilson scoreinterval for $0 < X < n$ are

$$\begin{aligned} p_L &= \frac{2np + k_{1-\frac{\alpha}{2}}^2 - k_\alpha\sqrt{4np(1-p)+k_{1-\frac{\alpha}{2}}^2}}{2\left(n+k_{1-\frac{\alpha}{2}}^2\right)}, \\ p_U &= \frac{2np + k_{1-\frac{\alpha}{2}}^2 + k_\alpha\sqrt{4np(1-p)+k_{1-\frac{\alpha}{2}}^2}}{2\left(n+k_{1-\frac{\alpha}{2}}^2\right)}, \end{aligned} \tag{5}$$

where $k_\alpha$ is the $\alpha$-quantile of standard normal $N(0,1)$ distribution.

For $X = 0$ is $p_L = 0$, for $X = n$ is $p_U = 1$.

**Wilson score interval with continuity correction.** The continuity correction suggested by Blyth and Still (1983). The lower and upper bounds for $0 < X < n$ are

$$\begin{aligned} p_L &= \frac{np - \frac{1}{2} + \frac{1}{2}k_{1-\frac{\alpha}{2}}^2 - k_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4}k_{1-\frac{\alpha}{2}}^2 + n\left(p-\frac{1}{2n}\right)\left(1-p+\frac{1}{2n}\right)}}{n+k_{1-\frac{\alpha}{2}}^2}, \\ p_U &= \frac{n\hat{p} + \frac{1}{2} + \frac{1}{2}k_{1-\frac{\alpha}{2}}^2 + k_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4}k_{1-\frac{\alpha}{2}}^2 + n\left(p+\frac{1}{2n}\right)\left(1-p-\frac{1}{2n}\right)}}{n+k_{1-\frac{\alpha}{2}}^2}, \end{aligned} \tag{6}$$

where $k_\alpha$ is the $\alpha$-quantile of standard normal $N(0,1)$ distribution .

For $X = 0$ is $p_L = 0$, for $X = n$ is $p_U = 1$.

**Agresti – Coull interval.** Agresti and Coull (1998) introduced a slight modification of the Wald interval by adding two successes

and two failures. The point estimator of $\pi$ is then $p_W = \frac{X+2}{n+4}$.

The lower and upper bounds of $100\cdot(1-\alpha)$% Agresti – Coull interval are

$$p_L = p_W - k_{1-\frac{\alpha}{2}}\sqrt{\frac{p_W(1-p_W)}{n}},$$
$$p_U = p_W + k_{1-\frac{\alpha}{2}}\sqrt{\frac{p_W(1-p_W)}{n}}, \quad (7)$$

where $k_\alpha$ is the $\alpha$-quantile of standard normal distribution $N(0,1)$.

**Jeffreys interval.** This interval is based on the Bayesian approach. *Beta*-distribution is conjugate priors for binomial distribution. Let random variable $X \sim Bi(n,\pi)$ and $\pi \sim Beta(k_1,k_2)$. Then the posterior distribution of $\pi$ is $Beta(x + k_1, n - x + k_2)$. Thus $100\cdot(1-\alpha)$% Bayesian interval is

$$p_L = Beta_{\frac{\alpha}{2}}(x + k_1, n - x + k_2),$$
$$p_L = Beta_{1-\frac{\alpha}{2}}(x + k_1, n - x + k_2).$$

It is known that non-informative Jeffreys prior is $Beta\left(\frac{1}{2},\frac{1}{2}\right)$.

Then the lower and upper bounds of $100\cdot(1-\alpha)$% Jeffreys interval are

$$p_L = Beta_{\frac{\alpha}{2}}\left(x + k_1, n - x + \frac{1}{2}\right),$$
$$p_u = Beta_{1-\frac{\alpha}{2}}\left(x + \frac{1}{2}, n - x + \frac{1}{2}\right). \quad (8)$$

where $Beta(k_1,k_2)$ is the $\alpha$-quantile of *Beta*-distribution with $k_1$ and $k_2$ degrees of freedom.

For $X = 0$ is $p_L = 0$, for $X = n$ is $p_U = 1$.

## 3. Criteria for Comparing the Confidence Intervals

In this section we introduce the criteria that are used for comparing the confidence intervals.

**Coverage Probability.** For the fixed values $n$ and $\pi$ the coverage probability is the probability that the confidence interval $CI(X, n)$ contains the parameter $\pi$. The coverage probability is defined for the given $n$ and $\pi$ as

$$C(n,\pi) = P_\pi\left(\pi \in CI(X,n)\right) = \sum_{x=0}^{n} I(x,\pi)\binom{n}{x}\pi^x(1-\pi)^{n-x} \quad (9)$$

where $0 < \pi < 1$, $I(x,\pi) = \begin{cases} 1 & if\ \pi \in CI(X,n) \\ 0 & if\ \pi \notin CI(X,n) \end{cases}$.

The confidence interval is strictly conservative, if $C(n,\pi) \geq \geq 1 - \alpha$ for all $\pi$.

Due to the discrete nature of the binomial distribution the coverage probability can not be exactly equal to the nominal level $(1 - \alpha)$ at all possible values. Therefore, our goal is to construct such a confidence interval for the parameter $\pi$ that the coverage probability is near the nominal level $(1 - \alpha)$.

**Average Coverage Probability.** The average coverage probability (*AVEC*) for the uniform averaging of the parameter values is defined as

$$AVEC(n) = \int_0^1 C(n,\pi)d\pi\,. \quad (10)$$

The confidence interval is conservative on average, if $AVEC(n) \geq 1 - \alpha$.

**Expected Length.** The expected length of the confidence interval is defined as

$$EL(n,\pi) = \sum_{x=0}^{n}\left[p_U(x,n) - p_L(x,n)\right]\binom{n}{x}\pi^x(1-\pi)^{n-x} \quad (11)$$

where $p_L(x,n)$, $p_U(x,n)$ are lower and upper bounds of a particular confidence interval.

This criterion measures the confidence interval length. In addition to the coverage probability the interval length is important for evaluating the confidence interval. The confidence interval is better if it has a shorter expected interval length with the similar performance of the other criteria.

**Average Expected Length.** The average expected length (AVEL) of the confidence interval is defined as

$$AVEL(n) = \int_0^1 EL(n,\pi)d\pi\,. \quad (12)$$

**Root Mean Square Error.** The root mean square error (RMSE) is defined as

$$RMSE(n) = \sqrt{\int_0^1 \left|C(n,\pi) - (1-\alpha)\right|^2 d\pi} \quad (13)$$

This criterion is used to describe how far the coverage probability typically falls from the nominal level. This criterion measures the variability of coverage probability about the nominal level $(1 - \alpha)$.

## 4. Comparsion of Confidence Intervals

In this section we demonstrate the performance of the confidence intervals which are compared in terms of the criteria mentioned above. The coverage probability, conservatism and interval length are important for evaluating the confidence intervals. To evaluate and compare the performance of confidence intervals the coverage probability was computed in 2001 values equally spaced in the interval $\langle 0,1\rangle$ for $n = 1$ to 1 000 and for $\alpha = 0.05$. The calculations were performed in Matlab. As it is impossible to analyze a large number of plots we sumarize the results for the coverage probability in terms of the observed minimum coverage probability and AVEC. The confidence intervals were grouped into two classes of acceptable intervals:

1. **strictly conservative intervals** – intervals that have the minimum coverage probability at the least nominal level $(1 - \alpha)$, for all $n$ and for all $\pi : \min_{\pi} C(n,\pi) \geq 1 - \alpha$.

2. **not strictly conservative intervals, but conservative on** average – intervals whose average coverage probability is at the least nominal level $(1 - \alpha)$, for all $n : AVEC(n) \geq 1 - \alpha$.

In the first class such a confidence interval is ideal whose minimum coverage probability is equal or a little above the nominal level $(1 - \alpha)$. In the second class such a confidence interval is ideal whose AVEC is equal or a little above the nominal level $(1 - \alpha)$ and minimum coverage probability is little below the nominal level $(1 - \alpha)$. A shorter expected length and a smaller average expected length is preferred.

**Coverage probability.** Fig. 1 shows the coverage probabilities of 95% confidence intervals for the case $n = 50$. The figures for other values of $n$ are similar to this figure. It is evident why the Wald performs poorly and why the Clopper-Pearson is known as

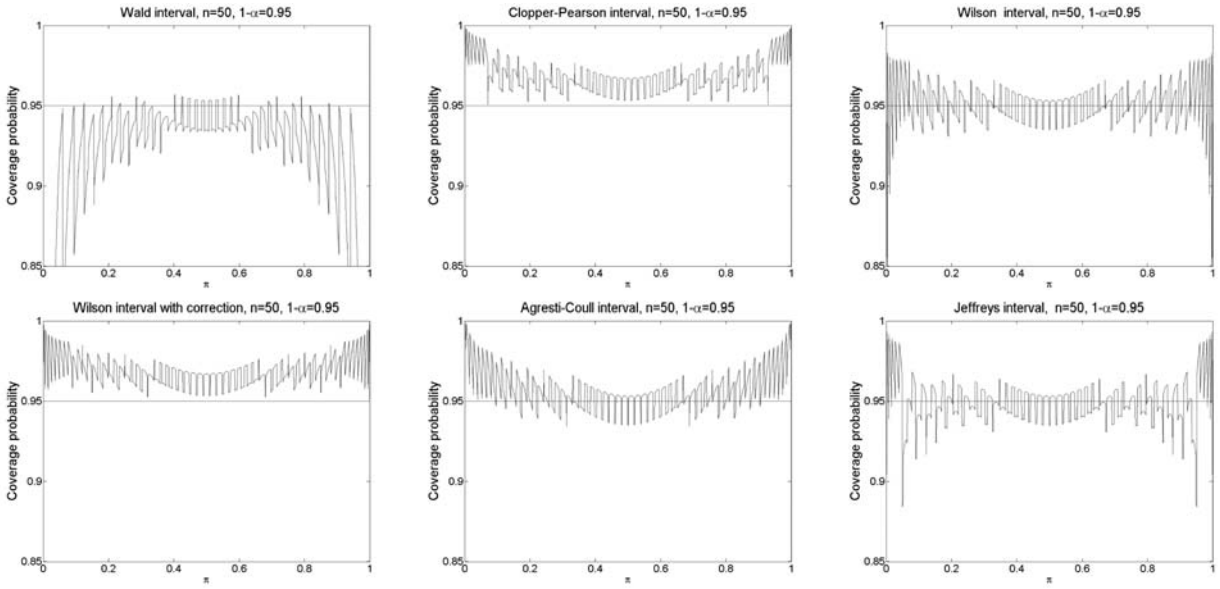



*Fig. 1 Coverage probability of 95% confidence intervals for n = 50*

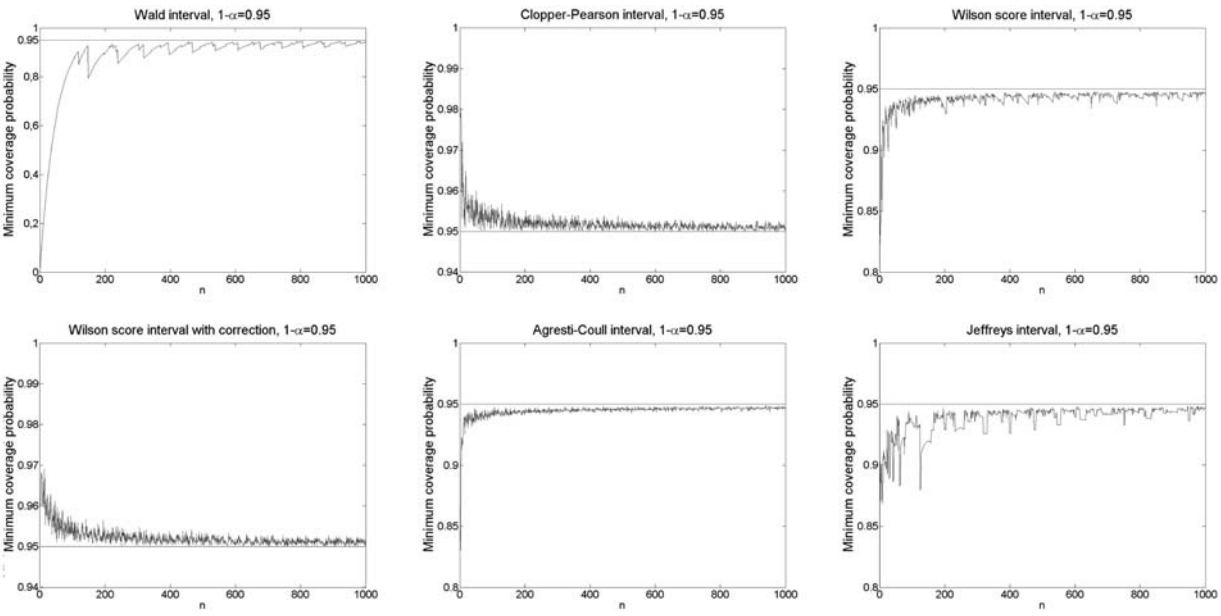



*Fig. 2 Minimum coverage probability of 95% confidence intervals for n = 1 to 1 000*

an overly conservative interval. The Clopper-Pearson guarantees that the coverage probability is always equal to or above the nominal level $(1 - \alpha)$. The coverage probability of the Wald is very poor for $\pi$ near boundaries 0 and 1. The problems with coverage probability exist even for $n$ large. This interval has more chaotic properties and can not be used (Brown, Cai, DasGupta, 2001). The Wilson has coverage probability which fluctuates near the nominal level $(1 - \alpha)$. As $n$ gets larger it comes to the significant improvement. The coverage probability near to boundaries 0 and 1 is problematic. The Wilson+CC falls into conservative intervals, with performance similar to the Clopper-Pearson. In comparsion to the Clopper-Pearson, the Clopper-Pearson is more conservative for $\pi$ near 0 and 1. The Agresti-Coull is even more conservative especially for $n$ small. In comparsion to the Wilson, the coverage probability is as good as the Wilson, but the Agresti-Coull is quite conservative for $\pi$ near the boundaries. The Jeffreys has a coverage probability qualitatively similar to the Wilson. Its coverage probability is reasonable, except for the very deep spikes near 0 and 1. As $n$ gets larger it comes to the improvement. Fig. 2 shows the minimum coverage probabilities of 95% confidence intervals for $n = 1$ to 1 000.

**Average coverage probability.** Fig. 3 shows the AVEC for $n = 1$ to 100. As it is showed in the figure the AVEC of the Wald tends to be under other intervals and under the nominal level $(1 - \alpha)$ for all values of $n$. The Clopper-Pearson and the Wilson+CC are comparable intervals and are too conservative on average. The Clopper-Pearson tends to be higher than other intervals and above the nominal level $(1 - \alpha)$ for all values of $n$. The Agresti-Coull is slightly conservative on average. The Wilson and the Jeffreys have the AVEC quite close the nominal level $(1 - \alpha)$ for most of values $n$, the Wilson even for $n$ small.

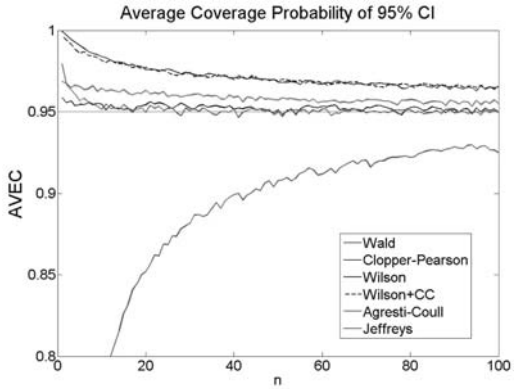



*Fig. 3 Average coverage probability of 95% confidence intervals for $n = 1$ to 100*

**Expected length.** Fig. 4 shows the expected lengths of 95% confidence intervals for case $n = 50$. The figures for other values of $n$ are similar to this figure. It is evident that the Clopper-Pearson and the Wilson+CC are wider than other intervals. For $\pi$ moderate the Wilson+CC is shorter. For $\pi$ small or large the Wald is the shortest. The Jeffreys and the Wilson are comparable intervals with a relatively small width. The Jeffreys performs better. The Agresti-Coull is quite wide. For $\pi$ moderate the Agresti-Coull, the Wilson and the Jeffreys are narrower than other intervals.

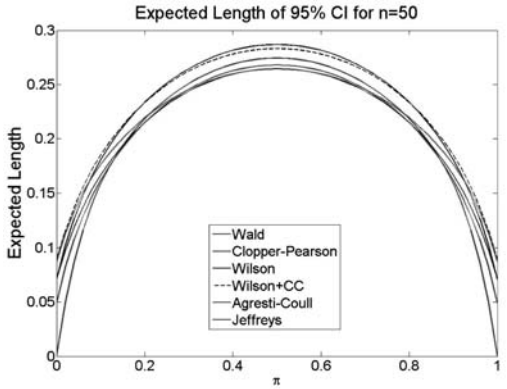



*Fig. 4 Expected length of 95% confidence intervals for $n = 50$*

**Average expected length.** Fig. 5 shows the AVEL of 95% confidence intervals for $n = 1$ to 100. As it is showed in the figure the Clopper-Pearson and the Wilson+CC are comparable intervals and their AVEL is the biggest of all the intervals. The Wilson and the Jeffreys are comparable intervals. In comparsion to them the Agresti-Coull is larger for $n$ small. From the given figure it is evident that as $n$ gets larger the difference between intervals starts to wear off.

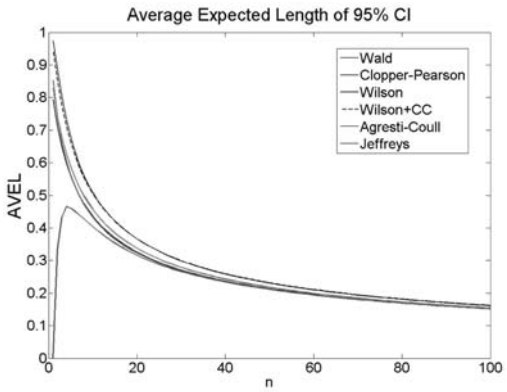



*Fig. 5 Average expected length of 95% confidence intervals for $n = 1$ to 100*

**Root mean absolute square error.** Fig. 6 shows the RMSE of 95% confidence intervals for $n = 1$ to 100. It is evident that the RMSE of the Wald is much larger than the other intervals. The Clopper-Pearson and the Wilson+CC are comparable intervals, the RMSE of the Clopper-Pearson is slightly larger than the Wilson+CC. The RMSE of the Jeffreys and the Agresti-Coull are comparable. The Wilson has the smalltest RMSE.

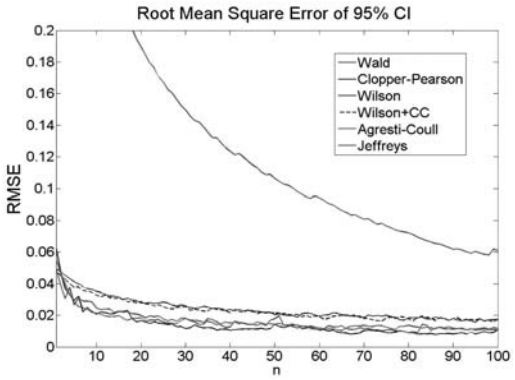



*Fig. 6 Root mean square error of 95% confidence intervals for $n = 1$ to 100*

Table 1. The comparsion of 95% confidence intervals in terms of minimum coverage probability (MCP), average coverage probability (AVEC), root mean square error (RMSE) and average expected length (AVEL) for $n = 10, 30, 50, 100, 500, 1\,000$.

| | Methods | MCP | AVEC | RMSE | AVEL |
|---|---|---|---|---|---|
| $n$=10 | Clopper-Pearson | 0.9610 | 0.9837 | 0.0349 | 0.5085 |
| | Wald | 0.0050 | 0.7696 | 0.2858 | 0.4035 |
| | Wilson | 0.8382 | 0.9540 | 0.0218 | 0.4354 |
| | Wilson+CC | 0.9511 | 0.9819 | 0.0330 | 0.5043 |
| | Agresti-Coull | 0.9168 | 0.9638 | 0.0217 | 0.4565 |
| | Jeffreys | 0.8681 | 0.9531 | 0.0244 | 0.4326 |
| $n$=30 | Clopper-Pearson | 0.9506 | 0.9734 | 0.0256 | 0.2990 |
| | Wald | 0.0149 | 0.8753 | 0.1674 | 0.2663 |
| | Wilson | 0.8475 | 0.9525 | 0.0141 | 0.2708 |
| | Wilson+CC | 0.9528 | 0.9730 | 0.0245 | 0.2998 |
| | Agresti-Coull | 0.9339 | 0.9599 | 0.0163 | 0.2789 |
| | Jeffreys | 0.8887 | 0.9505 | 0.0172 | 0.2687 |
| $n$=50 | Clopper-Pearson | 0.9509 | 0.9692 | 0.0215 | 0.2306 |
| | Wald | 0.0247 | 0.9009 | 0.1299 | 0.2113 |
| | Wilson | 0.8392 | 0.9516 | 0.0118 | 0.2129 |
| | Wilson+CC | 0.9512 | 0.9693 | 0.2008 | 0.2313 |
| | Agresti-Coull | 0.9345 | 0.9578 | 0.0141 | 0.2177 |
| | Jeffreys | 0.8842 | 0.9502 | 0.0135 | 0.2117 |
| $n$=100 | Clopper-Pearson | 0.9504 | 0.9646 | 0.0169 | 0.1614 |
| | Wald | 0.0488 | 0.9226 | 0.0915 | 0.1518 |
| | Wilson | 0.8606 | 0.9510 | 0.0087 | 0.1523 |
| | Wilson+CC | 0.9512 | 0.9647 | 0.0164 | 0.1619 |
| | Agresti-Coull | 0.9390 | 0.9664 | 0.0113 | 0.1543 |
| | Jeffreys | 0.8806 | 0.9499 | 0.0098 | 0.1517 |
| $n$=500 | Clopper-Pearson | 0.9503 | 0.9573 | 0.0091 | 0.0706 |
| | Wald | 0.2212 | 0.9434 | 0.0382 | 0.0687 |
| | Wilson | 0.9099 | 0.9504 | 0.0041 | 0.0687 |
| | Wilson+CC | 0.9502 | 0.9573 | 0.0087 | 0.0707 |
| | Agresti-Coull | 0.9448 | 0.9519 | 0.0059 | 0.0689 |
| | Jeffreys | 0.9145 | 0.9499 | 0.0046 | 0.0686 |
| $n$=1000 | Clopper-Pearson | 0.9501 | 0.9551 | 0.0068 | 0.0496 |
| | Wald | 0.3934 | 0.9464 | 0.0243 | 0.0486 |
| | Wilson | 0.9098 | 0.9501 | 0.0041 | 0.0486 |
| | Wilson+CC | 0.9501 | 0.9552 | 0.0070 | 0.0496 |
| | Agresti-Coull | 0.9453 | 0.9511 | 0.0045 | 0.0487 |
| | Jeffreys | 0.9172 | 0.9499 | 0.0040 | 0.0486 |

## 5. Concluding Remarks

In this section we summarize the classification and performance of the alternatives of confidence intervals.

The Wald interval should not be used. It performs poorly in terms of the coverage probability and the RMSE, though the expected length is short. In comparsion to the Wald all the intervals mentioned above outperform the Wald.

From the alternatives of confidence intervals mentioned above only the Clopper-Pearson is strictly conservative. The Clopper-Pearson guarantees the minimum coverage probability which is equal to or is above the nominal level. This interval is too conservative on average, too wide and has the larger RMSE.

The other confidence intervals the Wilson, the Wilson+CC, the Agresti-Coull and the Jeffreys are not strictly conservative, but are conservative on average.

The Wilson and the Jeffreys have similar properties such as a relatively small length, comparable AVEL and RMSE. The Wilson has excellent properties, the coverage probability near the nominal level, except for the problems with the coverage probability for values near 0 and 1 that makes a very low minimum coverage probability. Similar problems with the minimum coverage probability exist for the Jeffreys due to unlucky deep spikes near boundaries 0 and 1. But otherwise the Jeffreys has also good properties.

The Agresti-Coull has the minimum coverage probability better than others. To compare it to the intervals in this class, except for the Wilson+CC, the Agresti-Coull is slightly conservative and wider on average, but its advantages are easy calculation and presentation.

The Wilson+CC is similar to the Clopper-Pearson. It is too conservative on average, wide and has the larger RMSE. This interval is almost strictly conservative. The coverage probability for some values that are near boundaries 0 and 1 is slightly below the nominal level. (For example, for $n = 15$, $\pi = 0.003485$, $\alpha = 0.05$ is $C(n,\pi) = 0.9490$.)

Which method should be used in practical applications? The choice from the alternatives depends on the situation where they should be used and on preferences of users. The strictly conservative Clopper-Pearson is a choice for a situation when the coverage probability must be guaranteed to be equal to or above the nominal level. Otherwise if strict conservativeness is not a major criterion, the preference is to use the confidence intervals which are conservative on average, and their coverage probability is quite close the nominal level and are narrower. The almost strict Wilson+CC is also a valid choice. The Jeffreys is also an appropriate choice for practice but it is more complicated to compute. Considering properties of alternatives of the confidence intervals the Wilson and the Agresti-Coull are the best choice in this class. They perform very well and are simple to compute.

These recommended confidence intervals are much better to guarantee the estimation of a binomial proportion when compared with the standard and frequently used Wald interval.

**Acknowledgement**
This paper was supported by the Grant VEGA No. 1/0249/09.

## References

[1] AGRESTI, A., COULL, D. A.: *Approximate is Better than "Exact" for Interval Estimation of Binomial Proportion*. American Statistician 52, p. 119-126, 1998
[2] BLYTH, C. R., STILL, H. A.: *Binomial Confidence Intervals.* Journal of the American Statistical Association 78, pp. 108–116, 1983
[3] BROWN, D. L., CAI, T. T., DASGUPTA, A.: *Interval Estimation for a Binomial Proportion.* Statistical Science 16, pp. 101–133, 2001
[4] CLOPPER, C. AND PEARSON, S.: *The use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial.* Biometrika 26, pp. 404–413, 1934.
[5] FENG, X.: *Confidence Intervals for Proportions with Focus on the US National Health and Nutrition Examination Survey.* Master thesis, Simon Fraser University, 2006.
[6] LAPLACE, P. S.: *Theorie Analytique des Probabilities.* Paris, France, Courier, 1812
[7] NEWCOMBE, R. G.: *Two-sided Confidence Intervals for the Single Proportion; Comparison of Several Methods.* Statistics in Medicine 17, pp. 857–872, 1998
[8] PIRES, M. A., AMADO, C.: *Interval Estimators for a Binomial Proportion: Comparison of Twenty Methods.* REVSTAT – Statistical Journal, Volume 6, Number 2, p. 165–197, 2008
[9] POBOČÍKOVÁ, I.: *Confidence Intervals for a Binomial Proportion.* Proc. 8th International Conference APLIMAT, Bratislava, February 3.-6. , pp. 791–800, 2009
[10] REICZIGEL, J.: *Confidence Intervals for a Binomial Proportion: Some New Considerations.* Statistics in Medicine 22, pp. 611–621, 2003
[11] VOLLSET, S. E.: *Confidence Intervals for a Binomial Proportion.* Statistics in Medicine 12, pp. 809–824, 1993
[12] WILSON, E. B.: *Probable Inference, the Law of Succession, and Statistical Inference.* Journal of the American Statistical Association 22, pp. 209–212, 1927

Darina Stachova *

# MATHEMATICAL FOUNDATIONS OF CARTOGRAPHY

*The aim of this contribution is to acquaint the reader with a specific part of mathematical cartography that is interesting not only from the point of view of mathematics, but also from the point of view of constructive geometry.*

## 1. Introduction

The surface of the Earth is quite irregular and complex which makes it difficult to represent in models. Despite significant progress in geographic information systems which give us more detailed and more precise view of the landscape surrounding us, the map remains the most common form of representation. In many domains of human activity, maps are one of the key tools without which we cannot imagine work of experts, nor even our leisure time.

## 2. Mathematical cartography

The purpose of cartography is construction of maps of the Earth and its parts, the so-called *geodetic systems*, on which the surface of the Earth is depicted together with real objects and phenomena. Map construction is a process based on a geometrical transformation. Therefore, geometry plays an important role in cartography. It allows us to mathematically describe cartographic transformations enabling us to rigorously study properties of cartographic projections.

The methods of representing geodetic systems in the plane are the primary concern of mathematical cartography. Namely, mathematical cartography is that part of cartography which is concerned with mathematical and geometrical fundamentals of cartography in the general sense of the word. It studies processes of transformation of space coordinates of objects and phenomena on surfaces of reference to the surface of a map. It is concerned with calculations and constructions used for transforming the curved surface of the Earth to the plane through cartographical projections. Therefore, mathematics and geometry form the theoretical basis for cartography.

## 3. Classification of cartographic projections

We consider the *reference surface* of the Earth to be the geometrical surface that is a mathematical model of the Earth. In practice, this is usually chosen to be the surface of a sphere, or a rotational ellipsoid, three-axis ellipsoid, or the plane when a small portion of the surface is to be depicted neglecting the curvature. In this contribution, we choose as the reference surface of the Earth the surface of a sphere. On this surface, we shall use two types of coordinates.

- *cartesian coordinates* $(O, x, y, z)$ where $O$ is the origin of the reference surface and the z-axis corresponds to the Earth's axis.
- *spherical coordinates* $[r, \phi, \lambda]$, where $\phi$ is the *latitude* (northern, southern, calculated from the equator to the poles so that $\phi \in \left\langle -\frac{\pi}{2}, \frac{\pi}{2} \right\rangle$) and $\lambda$ is the *longitude* (eastern, western, calculated from the plane defined by the Earth's axis and a fixed point – Greenwich – so that $\lambda \in \langle -\pi, \pi \rangle$). The set of points on the reference surface with constant $\phi$ is called a *parallel*, and the set of points with constant $\lambda$ is a *meridian*. The system of meridians and parallels is called the *cartographic grid*.

If the prime meridian (with $\phi = 0$) lies in the plane $(x, z)$, then the relationship between spherical $[r, \phi, \lambda]$ and cartesian $[x, y, z]$ coordinates is expressed as follows:

$$x = r \cdot \cos\phi \cdot \cos\lambda,$$
$$y = r \cdot \cos\phi \cdot \sin\lambda,$$
$$z = r \cdot \sin\phi.$$

* **Darina Stachova**
Department of Mathematics, Faculty of Science, University of Zilina, Slovakia, E-mail: darina.stachova@fpv.uniza.sk

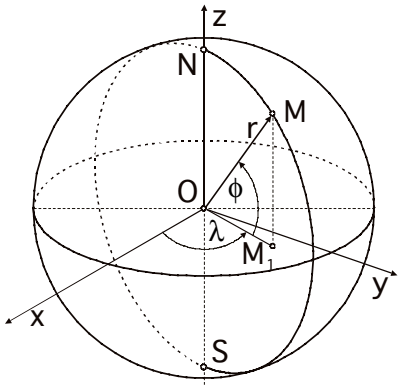



*Fig. 1 Spherical coordinates on the reference surface*

Cartographic projection is a system of methods and calculations transforming the cartographic grid to the plane of a map. Cartographic projection is defined by its transformation equations which in general are of the form $x = f(\phi, \lambda)$, $y = g(\phi, \lambda)$. If each equation depends only on one coordinate of the reference surface, we call such projections *simple projections* which in their general form can be expressed as follows:

$$x = n \cdot \lambda, \quad y = g(\phi).$$

Projections can be expressed in cartesian coordinates, as is shown in the above example, or alternatively, in *polar* coordinates $(\varepsilon, \rho)$.

A separate group of projections form the so-called *pseudoprojections*. In these projections, one of the two transformation equations is a function of both coordinates of the reference surface. These include also the so-called *polyconic* projections which are essentially simple conical projections with an infinite number of projecting cones.

In total there are around 300 known types of cartographic projections out of which around 50 are simple; yet in practice only few of them are used. Since there are so many of them, we categorize the cartographic projections according to their common features such as the surface of projection, the position of the Earth's axis, the distortion of various geometric shapes such as angles, lengths of curves, areas of surfaces, etc.

According to the type of the surface of projection we distinguish the following types:

- *azimuthal projection* – mapping the reference surface to a planar surface
- *cylindrical projection* – mapping the reference surface to a cylindrical surface
- *conical projection* – mapping the reference surface to a conical surface

Projections that can be derived geometrically from the central projection are called map projections. In the case of azimuthal

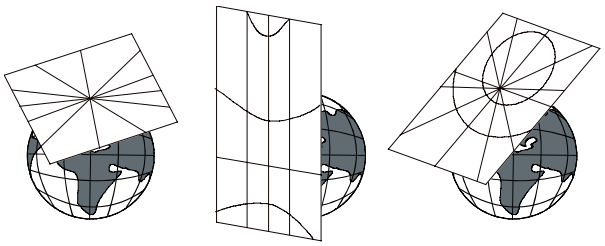


*Fig. 2 Azimuthal projection*

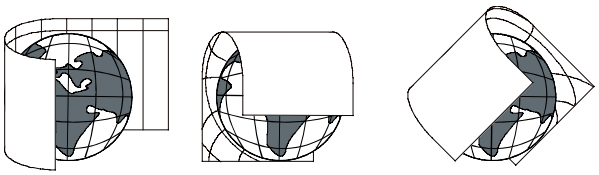


*Fig. 3: Cylindrical projection*

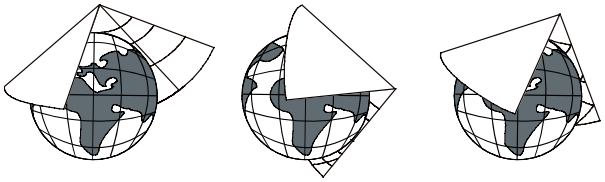


*Fig. 4 Conical projection*

map projections, depending on the position of the center, we further distinguish the following projections:

- *gnomonic projection* – the center of the projection is the center of the sphere; it is used for maps of polar regions and for naval maps. Its transformation equations are as follows:

$$\varepsilon = \lambda \text{ and } \rho = r \cdot tg\,\delta$$

where $\rho$ and $\varepsilon$ are polar coordinates on the map and $\delta = \frac{\pi}{2} - \phi$ is the zenith angle,

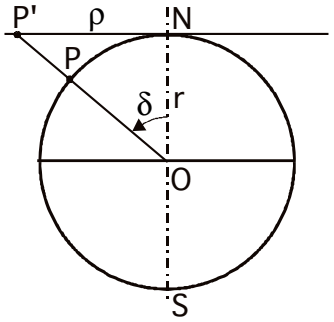



*Fig. 5 Gnomonic projection*

- *stereographic projection* – the center of the projection lies on the pole; it depicts larger part of the surface than gnomonic projection and has the following transformation equations:

$$\varepsilon = \lambda \text{ and } \rho = 2r \cdot \operatorname{tg}(\delta/2)$$

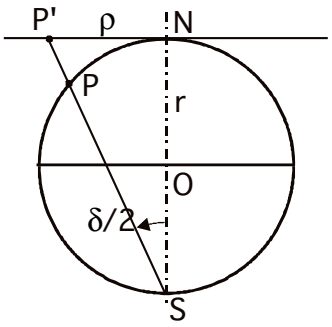



*Fig. 6 Stereographic projection*

- *orthographic projection* – the center of the projection is the point at infinity; the projection represents a view of the Earth from the outer space. Its transformation equations are as follows:

$$\varepsilon = \lambda \text{ and } \rho = r \cdot \sin\delta$$

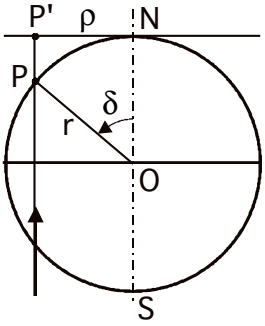



*Fig. 7 Orthographic projection*

- *satellite projection* – the center of the projection lies outside the Earth's surface; it is used for satellite imagery in weather forecast. Its transformation equations are as follows:

$$\varepsilon = \lambda \text{ and } \rho = \frac{(2r+s)\cdot s \cdot \sin\delta}{r + s\cdot\cos\delta + s}$$

- *double projection* – is a double stereographic projection; the point is first mapped to an auxiliary circle with radius 2*r* and then it is stereographically projected to the plane. This projection introduces less distortion and has the following transformation equations:

$$\varepsilon = \lambda \text{ and } \rho = 4r \cdot \operatorname{tg}(\delta/4)$$

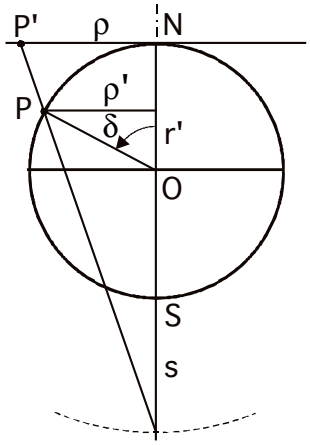



*Fig. 8 Satellite projection*

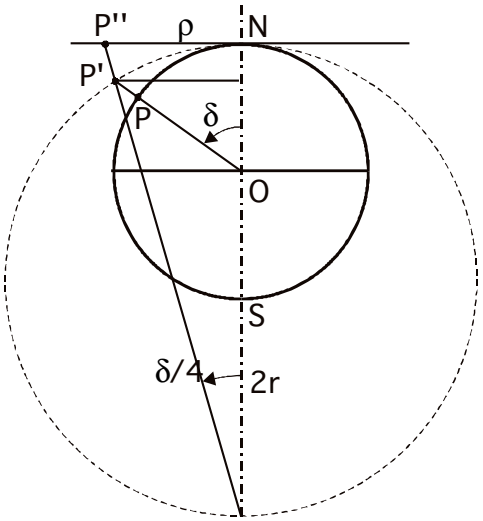



*Fig. 9 Double projection*

As we know from historical sources, the oldest known cartographic imagery is considered to be the geographic drawing of a settlement of hunter-gatherers from the region of Pavlov hills whose age is approximately 24 000 years. Scientific foundations of cartography were established in the ancient Greece over a period of several centuries. During this time, geometry, physics, and other scientific disciplines were also founded in connection to the names such as Strabo, Euclid, Pythagoras, Archimedes, Erastothenes, Ptolemaios, Thales, and many others. Indeed, the word cartography does not have its Greek root by accident. Next significant impulse for development of cartography was the age of sea exploration and colonization in the years 1492-1522. For these reasons many cartographic projections are named after their inventors. For instance:

- *azimuthal Lambert projection* – Johan Geinrich Lambert (1772) – it is used for about 15% of all maps and is area-preserving. If the center of the projection is on the equator or in a general

position, parallels and meridians are mapped to complicated curves. If the center is on the pole, the transformation equations are as follows:

$$\varepsilon = \lambda \text{ and } \rho = 2r \cdot \sin\delta.$$

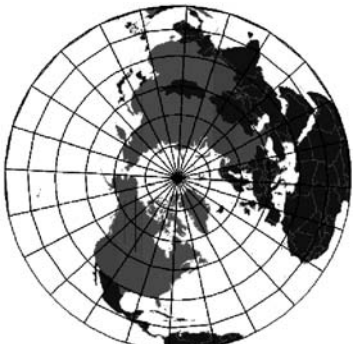


*Fig. 10 Azimuthal Lambert projection with center on the pole*

- *azimuthal orthographic projection* – Apollonius (third century B.C.) – the projection preserves lengths of parallels. If the center of the projection is on the equator, it maps meridians to elliptic arcs and parallels to parallel lines. In general position of the center, both meridians and parallels are represented as elliptic arcs. With the center on the pole, the transformation equations are of the form:

$$\varepsilon = \lambda \text{ and } \rho = r \cdot \sin\delta.$$

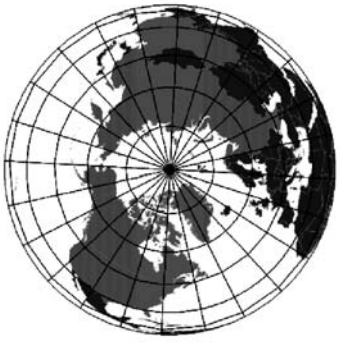


*Fig. 11 Azimuthal orthographic projection with center on the pole*

- *Ptolemaic conical projection* – Klaudios Ptolemaios (second century A.D.) – the projection preserves lengths of meridians and the length of the tangent parallel $\phi_0$. It is used for about 40% of maps in atlases using $\phi_0 = 45°$. Its transformation equations are:

$$\varepsilon = \lambda \cdot cos\,\delta_0 \text{ and } \rho = r \cdot [\mathrm{tg}\,\delta_0 - (\delta - \delta_0)].$$

- *Delisle conical projection* – Joseph Nicholaus de l'Isle (1745) – preserves the lengths of parallels $\phi_1$ and $\phi_2$ and lengths of all meridians. Areas and angles are distorted less than with Ptolemaic projection. The transformation equations are:

*Fig. 12 Ptolemaic conical projection*

$$\varepsilon = \lambda \cdot \frac{\sin\delta_1 - \sin\delta_2}{\delta_1 - \delta_2} \text{ and}$$

$$\rho = r \cdot \left( \frac{\delta_2 \cdot \sin\delta_1 - \delta_1 \cdot \sin\delta_2}{\sin\delta_2 - \sin\delta_1} + \delta \right)$$

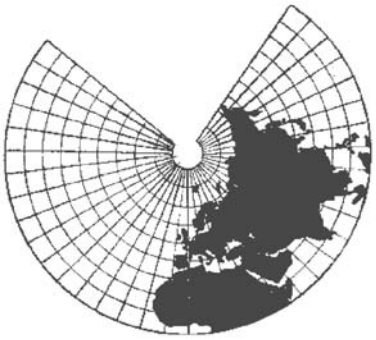


*Fig. 13 Delisle conical projection*

- *Marinos' cylindrical projection* – Marinos of Tyre (third century B.C.) – preserves lengths of meridians and parallels and introduces large distortion in the polar regions. Its transformation equations when expressed in spherical coordinates are as follows:

$$x = r \cdot \lambda \text{ and } y = r \cdot \varphi$$

*Fig. 14 Marinos' cylindrical projection*

- *Mercator's cylindrical projection* – Gerhard Mercator (1569) – preserves angles and is used for navigation charts; it does not allow depicting the poles and it introduces large aerial distortion. Its transformation equations have the following form:

$$x = r \cdot \lambda \quad \text{and} \quad y = r \cdot \ln(\operatorname{cotg}(\delta/2))$$

*Fig. 15 Mercator's projection*

- *Mercator-Sanson's projection* (also known as *sinusoidal* or *Sanson-Flamsteed projection*) – attributed to several authors – pseudocylindrical projection based on Marinos' projection; meridians are depicted as half sinusoids. It preserves lengths of parallels and of the prime meridian, both of which are mapped to line segments. Its transformation equations are of the form:

$$x = r \cdot \lambda \cos\varphi \quad \text{and} \quad y = r \cdot \varphi$$

*Fig. 16 Mercator-Sanson's projection*

- *Bonne's projection* – Rigobert Bonne (1752) – pseudoconic projection formed from the Ptolemaic projection; it preserves lengths of parallels and of the prime meridian. It has the following transformation equations:

$$\varepsilon = \frac{2\pi \cdot \sin\delta}{tg\delta_0 + \delta - \delta_0} \quad \text{and} \quad \rho = r \cdot (tg\delta_0 + \delta - \delta_0).$$

*Fig. 17 Bonne's projection*

- *Hassler's polyconic projection* – Ferdinand Rudolph Hassler (19th century) – the projection maps parallels to circular arcs of non-concentric circles and preserves lengths of parallels and of the prime meridian.

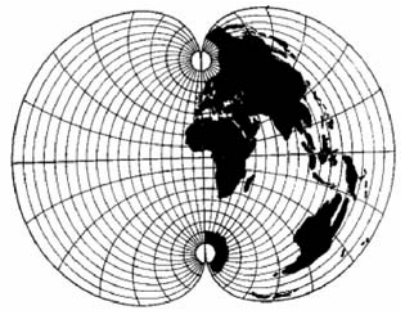


*Fig. 18 Hassler's polyconic projection*

## 4. Conclusion

The aim of this contribution was to acquaint the reader with cartographic projections, their categorization, and their precise mathematical description. Naturally, we do not present here a complete list of projections or aspects considered in their classification. We have focused mainly on those projections that we consider esthetically appealing and also those that are frequently used yet whose origin is not well known.

**Acknowledgement**

The author acknowledges support from the project KEGA SR 3/7090/09.

## References

[1] BARANOVA, M.: *Mathematical Cartography (in Czech)*, manuscript, 2009, *http://home.zcu.cz/~baranov/SPS_Kartografie/MatKart.pdf* (accessed 09/09)
[2] DRABEK, K., HARANT, F., SETZER, O.: *Descriptive Geometry - part II (in Czech)*, Praha, SNTL, 1979, ISBN 04-007-79
[3] KADERAVEK, F.: *Geometry and Art in the Past (in Czech*, Praha, Pudorys, 1997, ISBN 80-90791-5-6
[4] MEDEK, V., ZAMOZIK, J.: *Constructive Geometry for Engineers (in Slovak)*, Bratislava, Alfa, 1978, ISBN 63-552-76
[5] TALHOFER, V.: *Fundamentals of Mathematical Cartography (in Czech*, Brno, Univerzita obrany, 2007, *http://uzivatel.unob.cz/talhofer/Zaklady matematicke kartografie.pdf* (accessed 09/09) ISBN 978-80-7231-297-9
[6] VAJSABLOVA, M.: *Cartographic Projections from the Perspective of Constructive Geometry (in Slovak)*, Zbornik zo seminara o pocitacovej geometrii SCG'99, Kocovce, 1999
[7] *http://www.progonos.com/furuti/MapProj/CartIndex/cartIndex.html* (accessed 09/09).

Peter Hockicko – Peter Bury – Peter Sidor – Stanislav Jurecka – Igor Jamnicky *

# MATHEMATICAL MODELS FOR ACOUSTIC SPECTRA SIMULATION

*The experimentally obtained acoustic spectra of some investigated materials; particularly the ion conductive glasses and MOS (metal-oxide-semiconductor) structures are analyzed using suitable theoretical models and mathematical procedure to fit the experimental data. The acoustic spectra of ion conductive glasses can reflect the basic features of the relaxation and transport processes of the mobile ions. The results obtained form acoustic deep level transient spectroscopy (A-DLTS) spectra of MOS structures are used for the characterization of deep centers and determination of some of their physical parameters. Suitable theoretical model and mathematical description of acoustic spectra are necessary to understand both the principle of the ionic hopping motion and relaxation processes connected with the mobility of conductive ions including the role of their composition and the distribution of interface traps.*

## 1. Introduction

A considerable interest is given to experimental study of glassy materials with the fast ion transport because they play an important role in a number of modern electrochemical devices such as solid-state batteries (portable batteries for heart pacemakers, mobile telephones and laptop computers), solid-oxide fuel cells, electrochemical sensors, electrochronic displays and oxygen-separation membranes [1].

The acoustic attenuation measurement seems to be a useful technique for nondestructive investigation of transport mechanisms in conductive glasses and compared to the electrical ones it has even some advantages as high sensitivity, the absence of contact phenomena and so on [1, 2]. Acoustical measurements made over a wide range of frequencies and temperatures can characterize different relaxation processes according to corresponding transport mechanisms due to the strong acousto-ionic interaction. In glassy electrolytes, the mobile ions encounter different kinds of site and ionic hopping motion and relaxation processes connected with charge mobility so that the modified jump relaxation model [3] connected with the genetic algorithm [4] can be used for transport mechanisms description.

The interface states in MOS structures have been investigated for more than thirty years using many useful experimental techniques. One of the most important methods is the deep level transient spectroscopy (DLTS) originally developed in 1974 [5] with some modification, one of them is also an acoustic version (A-DLTS) [6, 7]. The A-DLTS version is the high-frequency ultrasonic method based on analysis of acoustoelectric response signal (ARS) of the MOS structure when longitudinal acoustic waves propagate through a structure after applied bias voltage steps to the structure. The ARS reflects relaxation processes in both semiconductor and interface layers. It was namely found that the acoustoelectric signal produced by MOS structures reflects any changes in the space charge distribution due to the external condition changes [7].

In this contribution we present theoretical mathematical models for description of experimental results obtained from acoustic spectra of a set of glasses prepared in the system CuI-CuBr-$Cu_2O$-$P_2O_5$ with the purpose to study ion transport mechanisms and to find the role of cuprous halides producing $Cu^+$ ions. Another theoretical model is presented for the characterization of the acoustic spectra of MOS structure in the case when the A-DLTS spectra are weak and we cannot analyze Arrhenius plots. Using the simulation of one spectrum is necessary to find the activation energies and some other parameters of traps at the insulator – semiconductor interface in MOS structures.

## 2. Theoretical models

### *a) Ion conductive glasses*

The formal theory of all relaxation processes looks similar. In dilute system containing a low concentration of mobile ions the acoustic attenuation spectrum may be described as a Debye-like, single relaxation time process in which the individual ion hops occur independently of each other. In such cases, the attenuation $\alpha$ for a wave of angular frequency $\omega$ takes the form [8]

$$\alpha = \Delta\left(\frac{\omega^2\tau}{1+\omega^2\tau^2}\right), \qquad (1)$$

* **Peter Hockicko[1], Peter Bury[1], Peter Sidor[1], Stanislav Jurecka[2], Igor Jamnicky[1]**
[1] Department of Physics, Faculty of Electrical Engineering, University of Zilina, Slovakia
[2] Department of Engineering Fundamentals, Faculty of Electrical Engineering, University of Zilina, Slovakia

where the parameter $\Delta$ is the relaxation strength and it determines the magnitude of the attenuation peak. It is related to the strain dependence of the mobile ion site energy, or deformation potential $B$ by

$$\Delta = NB^2/(4\pi\rho v^3 k_B T), \tag{2}$$

in which $N$ is the number of mobile ions, $v$ the velocity of the acoustic wave, $\rho$ the density of the solid, $T$ thermodynamic temperature and $k_B$ the Boltzmann constant.

The term in the equation (1) in the round brackets describes a Debye peak. The acoustic attenuation will exhibit a maximum when the condition $\omega\tau$ is equal to 1, where [9]

$$\tau = \tau_0 \exp(E_A/k_B T_{peak}) \tag{3}$$

is the most probable relaxation time, $\omega = 2\pi v$, $v$ the frequency of acoustic wave. The relaxation processes, described by the Arrhenius equation (3), are characterized by activation energy $E_A$ for jumps over the barrier between two potential minima and typical relaxation frequency of ion hopping $1/\tau_0 \approx 10^{13} - 10^{14}\ \mathrm{s}^{-1}$.

In fact all the investigated relaxation peaks are much broader than Debye peak. It can be interpreted as arising from the existence of a distribution of relaxation times due to random deviations in the local arrangement of the system. According to this hypothesis that the relaxation losses are not to large one can write for the acoustic attenuation, in case of a distribution $f(\tau)$ of relaxation times [9],

$$\alpha = \frac{\Delta}{2v}\int \frac{\omega^2\tau}{1+\omega^2\tau^2} f(\tau)d\tau\,. \tag{4}$$

As a consequence, the $\tau$ distribution can be connected with a distribution of activation energies $E$, representing the heights of the barriers that the ions must surmount to go into the near allowed positions. A useful form of equation (4) that takes into account only $E$ distribution can be derived by the microscopic theory of Jäckle et al. [9]

$$\alpha = \frac{B^2}{4\pi\rho v^3 k_B T}\int P(E)\frac{\omega^2\tau(E)}{1+\omega^2\tau^2(E)}dE\,, \tag{5}$$

where an average deformation potential $B$ expresses the coupling between the ultrasonic stress and the system and $P(E)$ represents the $E$ distribution function. We can assume for $P(E)$ a Gaussian distribution [10]

$$P(E) = \frac{N}{\sqrt{2\pi E_0^2}}\exp\left[-\frac{(E-E_m)^2}{2E_0^2}\right], \tag{6}$$

in which $N$ is the total number of jumping particles per unit volume, $E_m$ the most probable activation energy and $E_0$ the width of the distribution.

This approach depends on the assumption that ion migration may be treated in terms of a set of non-interacting Debye-like processes. However, in solid electrolytes the mobile ions concentrations are large and conduction mechanisms are thought to be cooperative. The relaxation phenomena observed in a wide variety of materials exhibit a power-law type of frequency dependence. The relationship to Debye behaviour is expressed in the form [8]

$$\alpha \approx \frac{1}{T}\left(\frac{(\omega\tau)^m}{1+(\omega\tau)^{1+m+n}}\right), \tag{7}$$

where $m$ and $n$ are power-law exponents, which take values between 0 and 1. When $m = 1$ and $n = 0$, equation (7) reduces to the equation for a single Debye-like process.

Two functions have mainly been used to fit mechanical loss data [11]. The first function is the Kohlrausch-Williams-Watts (KWW) function

$$\Phi(t,T) = \exp\left[-\left(\frac{t}{T}\right)^{\beta}\right] \tag{8}$$

with $0 < \beta \leq 1$. The acoustical attenuation is then given by

$$\alpha(\omega,T) \propto \int_0^{\infty}\left(-\frac{d\Phi(t)}{dt}\right)\sin(\omega t)dt\,. \tag{9}$$

The second function is the double power law (DPL)

$$\alpha(\omega,T) \propto \frac{1}{(\omega\tau)^{-n}+(\omega\tau)^m}\,. \tag{10}$$

Using these functions, we can fit also the acoustic attenuation spectrum of the cuprous halides glasses. All glasses we studied using acoustic spectroscopy exhibit an Arrhenius – type relaxation between the peak temperature and the applied frequency

$$\upsilon = \upsilon_0 \exp\left(-\frac{E_A}{k_B T_{peak}}\right), \tag{11}$$

where $v$ is the frequency, $v_0$ the preexponential factor, $T_{peak}$ the temperature of peak maximum which can be easily and directly determined from theoretical fits.

***b) MOS structures***

The basic principle of A-DLTS technique consists in utilization of the ARS produced by MOS structure interface when a high frequency longitudinal acoustic wave traverses the structure. After an injection pulse has been applied to the semiconductor structure the amplitude of ARS follows the accumulated charge behaviour over the capacitance, so that the ARS is proportional to the nonequilibrium carrier density $\Delta n$ [7]

$$\Delta n(t) = \Delta n_{t_0}\exp\left(-\frac{t}{\tau}\right), \tag{12}$$

where $\Delta n_{t_0}$ represents the variation in trap occupancy due to the acoustoelectric effect and $\tau$ is the time constant (relaxation time) associated with the release of the carrier from deep centers.

The measured acoustoelectric response signal amplitude for the discrete level can be then given by

$$U_{ac}^0(t) = U_0\exp\left(-\frac{t}{\tau}\right). \tag{13}$$

Including the regime of small-signal excitation, the A-DLTS output signal for the rate window ($RW$) $\Delta t = t_2 - t_1$ takes the form

$$\Delta U^0_{ac} = U^0_{ac}(t_1) - U^0_{ac}(t_2) = U_0\left(\exp\left(-\frac{t_1}{\tau}\right) - \exp\left(-\frac{t_2}{\tau}\right)\right), \quad (14)$$

where $\tau$ is the time constant of the acoustoelectric response delay. Note that Eq. (14) is accurate if the trap could be refilled to its equilibrium occupancy before each bias voltage step $\Delta U$. Evidently, with the rate window $\Delta t$ and corresponding $\tau_{max}$ (for the Lang's method $\tau_{max} = (t_2 - t_1)/\ln(t_2/t_1)$, the $\Delta U^0_{ac}(T)$ (14) peak should appear at the temperature $T = T_{max}$, when the relaxation time is equal to the time constant of emission process ($\tau = \tau_{max}$).

The reciprocal value of $\tau$ gives the emission rate which for electrons can be expressed by the relation

$$e = \tau^{-1} = \gamma_n \sigma_n T^2 \exp\left(-\frac{E_A}{k_B T}\right), \quad (15)$$

were $\gamma_n$ is a constant, $\sigma_n$ the capture cross section, $E_A = E_C - E_T$ the trap activation energy related to the bottom of conduction band.

Using the A-DLTS technique based on the computer-evaluated isothermal transients and correlation procedure with higher order on-line filters and rectangular weighting function [12], the activation energies $E_A$ and corresponding capture cross-section $\sigma_n$ of traps can be determined from Arrhenius type dependence $\ln(\tau_{max} T^2_{max})$ versus $1/T_{max}$.

Another method for determining the basic parameters of traps at the insulator – semiconductor interface is the modelling of measured spectra [13]. With several trapping centers being present, $U^0_{ac}(t)$ is composed of corresponding components

$$U^0_{ac}(t) = U_{i0} + \sum_{j=1}^{n} U_{j0} \exp\left(-\frac{t}{\tau_j}\right), \quad (16)$$

where $U_{i0}$ is the instantaneous acoustoelectric response of the device and the second term on the right hand side of Eq. (16) is the excess acoustoelectric response due to charging (discharging) of the $n$ traps which are captured (emitted) by the rate $e_j = \tau_j^{-1}$. We can consider that only

$$\tau_j = \tau_{j0} \exp\left(\frac{E_{Aj}}{k_B T}\right) \quad (17)$$

undergoes changes due to scanning the device temperature $T$.

Using the Eq. (16) and Eq. (17), respectively Eq. (14) for discrete levels we can fit the measured A-DLTS spectra for the various rate windows $\Delta t$. The activation energies $E_A$ determined from this modelled spectra expressed by the relation

$$\Delta U^0_{ac}(T) = \sum_{j=1}^{n} U_{j0}\left(\exp\left(-\frac{t_1}{\tau_j}\right) - \exp\left(-\frac{t_2}{\tau_j}\right)\right) \quad (18)$$

can be then compared with the activation energies calculated from Arrhenius plots.

## 3. Experimental results

### *a) Ion conductive glasses*

The preparation of glasses in the investigated system CuI-CuBr-$Cu_2O$-$P_2O_5$ and the measure equipment has been already described [14]. Using the theoretical models we can fit the temperature-dependent acoustic attenuation spectra at constant frequency. It is usually supposed [10] that the double power law (DPL) provides a slightly better fit than the KWW function or Debye-like process.

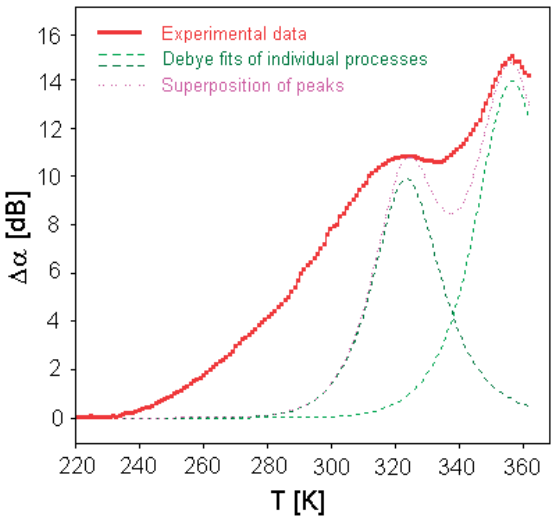



*Fig. 1. The acoustic spectrum of glass sample BIDP5 (starting glass composition in mol. %: 9.09CuI-9.09CuBr-54.55$Cu_2O$-27.27$P_2O_5$) (full line) and the Debye fits of the two relaxation processes (dashed line).*

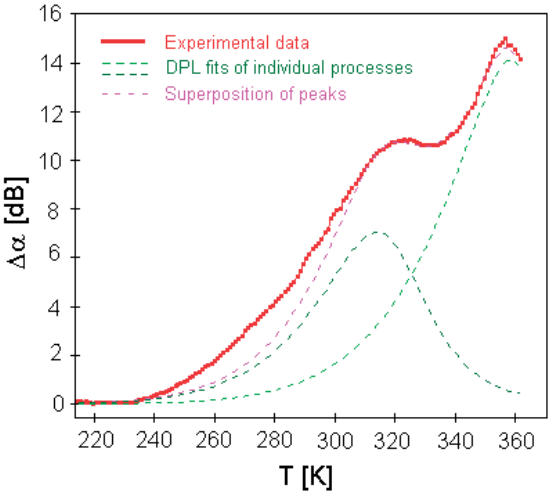



*Fig. 2. The acoustic spectrum of glass sample BIDP5 (full line) and the DPL fits of the two supposed relaxation processes (dashed line)*

Firstly, we tried to fit the measured temperature dependence of acoustic attenuation spectra at constant frequency $\upsilon = 18$ MHz using theoretical models (Eq. 1, 7, 8, 10).

The measurements of the temperature dependence of acoustic attenuation indicated in all investigated samples one broad attenuation peak at higher temperature, in which we initially supposed two separated peaks. Debye model (Fig. 1) isn't suitable for modelling relaxation processes in investigated glasses of the system CuI-CuBr-$Cu_2O$-$P_2O_5$. As we can see, the complete spectrum of sample BIDP5 illustrated in Fig. 2, cannot be fitted supposing only two relaxation processes. Using the theoretical double power law (DPL) model (Eq. 10) we tried to fit the broad attenuation peak as a superposition of three calculated lines represented by the cross-marked line in Fig. 3. The fit gave an excellent agreement with the measured spectrum in high and middle temperature range for the parameters $m = 0.27$ and $n = 0.42$. We found that the additional third relaxation process should be taken into account with maximum at the temperature around 280 K. The whole temperature dependence of acoustic attenuation was analyzed then assuming the existence of four thermally activated relaxation processes of ions in connection with different kinds of sites. The fourth peak, however, was detected at lower temperatures.

The distinctive peaks of acoustic attenuation spectra, which are caused by resonant interaction with the mobile ion hopping processes, enable us then to study relaxation and transport mechanisms in the ion conductive glasses. The attenuation spectra can be explained by the assumption that temperature peaks are caused by the relaxation processes of mobile $Cu^+$ ions in connection with different kinds of sites.

There are several categories of relaxation processes connected with temperature peaks of individual processes. High temperature activation energy presents the basic mechanism. The second mechanism observed by both acoustic and conductivity measurements [15] can play an important roll in ionic transport, too. The relaxation mechanisms at the temperature of about 280 K do not influence the ionic transport significantly. Comparing the activation energies obtained from acoustic and electrical measurements [15], it seems reasonable that essentially the same microscopic processes can be responsible for the acoustic and electrical relaxation processes. However, some differences can be caused by the different relaxation mechanisms connected with ion hopping transport at ac electric field and the hopping of mobile ions due to the interaction of acoustic wave with glass network.

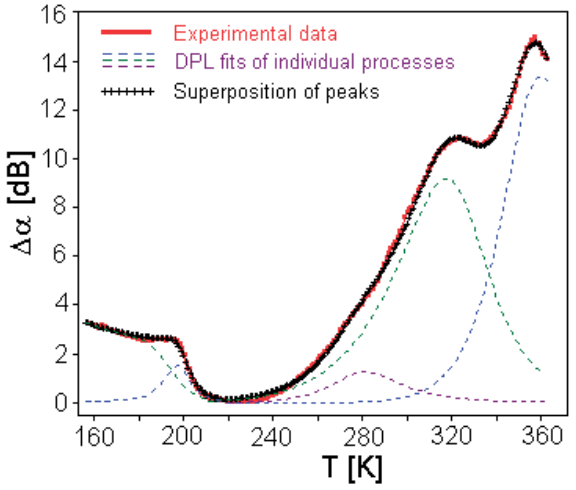



*Fig. 3. The acoustic spectrum of glass sample BIDP5 (full line). Cross-marked line represents the best fit of superposition at least four relaxation processes described by DPL model*

The results from IR spectra of the CuI-CuBr-$Cu_2O$-$P_2O_5$ glasses [16–18] indicate that the thermal activated processes of $Cu^+$ ions determined for all samples of investigated systems can be associated mainly with three different structural units – monomeric orthophosphate $PO_4^{3-}$, low-condensed dimeric diphosphate oxoanions $P_2O_7^{4-}$ and $P_3O_{10}^{5-}$ structure phosphate anions. The local electric field around the trivalent orthophosphate anion $PO_4^{3-}$ is stronger than the local electric field around dimeric diphosphate oxoanions $P_2O_7^{4-}$ and triphosphate $P_3O_{10}^{5-}$ anions because the diphosphate and triphosphate anions have smaller negative electric charge on non-bridging oxygen atoms. Moreover, expanded structure of chain groups creates advantageous conditions for ionic motion [18]. Because of this the electrostatic interactions between the mobile $Cu^+$ ions and orthophosphate anions are stronger than those between $Cu^+$ ions and diphosphate and triphosphate anions and we can suppose that the relaxation processes with biggest activation energies can be connected with monomeric orthophosphate anions and the processes with smaller energies can be connected with both dominant diphosphate oxoanions $P_2O_7^{4-}$ and low condensed triphosphate anions and maybe other polymeric structural units.

***b) MOS structures***

The investigated Al-$SiO_2$-Si MOS structure was fabricated on n-type Si substrates with (100) surface orientation and 1 – 20 Ωcm resistivity. The aluminium electrodes were deposited on oxide layer of the thickness of 4nm. All characteristics were obtained at the temperature range 160 – 360 K. The basic data were received from sampled acoustoelectric isothermal transients $U_{ac}(t)$ in response to a bias voltage step $\Delta U$ at point $t_1$ and $t_2 = 2\ t_1$ (or $t_3 = 4\ t_1$, $t_4 = 8\ t_1$, ..., $t_n = 2^{n-1}\ t_1$) respectively ($t_1 = 0.015$ s, $t_2 = 0.030$ s, $t_3 = 0.060$ s, $t_4 = 0.120$ s, $t_5 = 0.240$ s, $t_6 = 0.480$ s, $t_7 = 0.960$ s and $t_8 = 1.92$ s).

Fig. 4 represents the typical ARS transients measured at various temperatures and the same bias voltage $U_g = -0.2$ V ($\Delta U = 0.5$ V) for investigated MOS structure.

Fig. 5 shows the series of A-DLTS spectrum determined from the isothermal transients for various time constants. The obtained activation energy calculated from Arrhenius plot (Fig. 6) related to the bottom of conduction band $E_A = E_C - E_T = (0.37+/-0.01)$ $eV$ with the cross section $5.2\times10^{-17}$ cm$^2$ corresponding to the bias voltage $U_g = -0.2$ V.

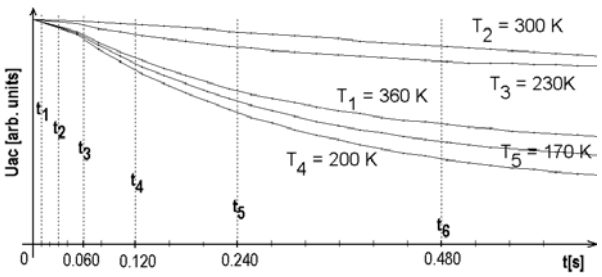



*Fig. 4. The ARS transients at various temperatures: $T_1$ = 360 K, $T_2$ = 300 K, $T_3$ = 230 K, $T_4$ = 200 K and $T_5$ = 170 K*

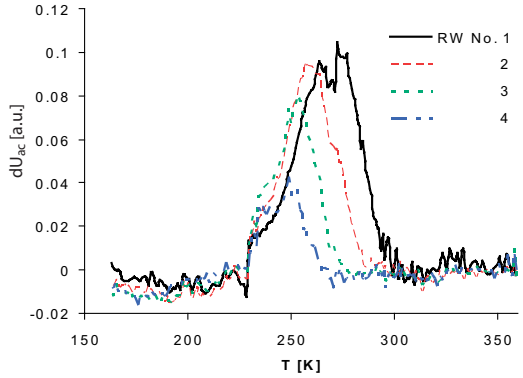



*Fig. 5. A set of A-DLTS spectra measured at $U_g = -0.2$ V, $\Delta U = 0.5$ V calculated from isothermal transients for the relaxation times 9.13 ms (rate window No. (1), 18.4 ms (2), 36.8 ms (3) and 73.6 ms (4)*

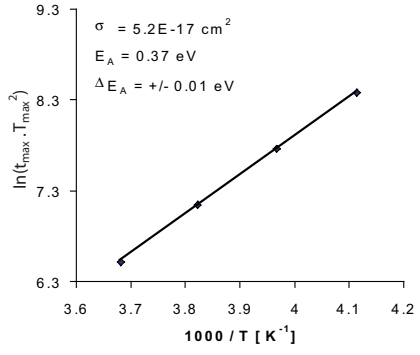



*Fig. 6. Arrhenius plot constructed from the positions of the peak maxima of the A-DLTS spectra*

From the A-DLTS spectra (Fig. 5) for various relaxation times we also tried to find some information about traps at the insulator-semiconductor interface applying the mathematical methods of simulation. For this reason it was important to correlate these measured results via exponential analysis. Using the Eq. (18) for the discrete level we simulated the main peak of measured spectra (Fig. 5) for the point $t_1$ = 15 ms and $t_2$ = 30 ms (Fig. 7). We found the energy level $E'_{A1}$ = 0.39 eV, which is very close to the activation energy calculated from Arrhenius plot. But using a modelling method of the whole measured spectra we could find next possible trap with energy level $E'_{A2}$ = 0.48 eV.

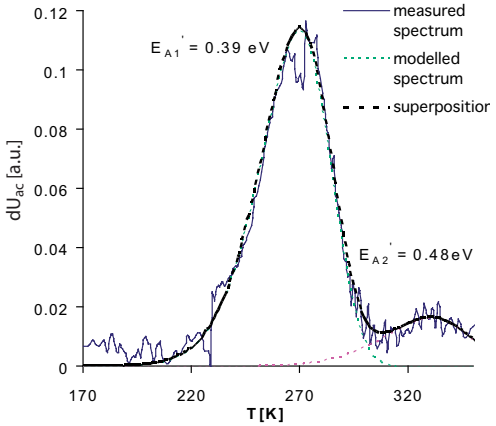



*Fig. 7. The simulation peaks of A-DLTS spectra obtained for the relaxation time 9.13 ms and their superposition*

## Conclusion

The experimental and theoretical investigation of ion conductive glasses in system CuI-CuBr-$Cu_2O$-$P_2O_5$ proved that acoustical spectroscopy connected with suitable mathematical modelling can be a very useful technique for the study of mechanisms in fast ion conductive glasses. Using the theoretical mathematical models of the relaxation processes and simulation of acoustic spectra we can better determine and describe the transport mechanism of mobile ions. Fitting the whole acoustic spectra as a superposition of individual relaxation processes helped us to find additional relaxation process, which could not be found directly from the position of peaks of temperature dependence of acoustic attenuation of ion conductive glasses or electrical measurements.

The A-DLTS measurement technique connected with simulation of acoustic spectra seems to be an effective method for the investigation of the trap at the insulator-semiconductor interface in MOS structures, too. Applying simulation of A-DLTS spectra we identified the similar activation energy of the deep centers at $SiO_2$-Si interface as we calculated from Arrhenius plot. But utilizing the method of simulation of acoustic spectra we could find also next possible traps.

Using other improved mathematical models connected with several mathematical correlation methods should help us to find other useful information about both relaxation or transport

processes in ion conductive glasses and interface states in MOS structures.

**Acknowledgements**

The authors would like to thank assoc. prof. M. Jamnicky from the Department of Ceramic, Glass and Cement, Slovak Technical University, Bratislava for preparation and provision of investigated ion conductive glasses, prof. H. Kobayashi, assoc. prof. M. Takahashi and Dr. K. Imamura from the Institute of Scientific and Industrial Research, Osaka University, Japan for preparation and provision of an investigated MOS sample and Mr. F. Cernobila for technical assistance. This work was supported by VEGA project No. 2/7120/07 and project APVV-0577-07 of the Ministry of Education of the Slovak Republic.

The authors wish to thank for the support to the R&D operational program "Centre of excellence of power electronics systems and materials" for their components. The project is funded by the European Community, ERDF – European regional development fund.

## References

[1] ROLING, B., HAPPE, A., INGRAM, M. D., FUNKE, K: *J. Phys. Chem.* B 103 (1999) 4122
[2] CHARNAZA, E. V., BORISOV, B. F., KULESHOV, A. A.: *Proc. World Congress on Ultrasonics,* Berlin (1995) 483
[3] FUNKE, K.: *Solid State Ionics 94* (1997) 27
[4] BURY, P., HOCKICKO, P., JURECKA, S., JAMNICKY, M.: *Physica Status Solidi* (c) 11 (2004) 2888
[5] LANG, D. V.: *Appl. Phys. 45* (1974) 3023
[6] TABIB-AZAR, M., HAJJAR, F.: *IEEE Trans. Electron Dev.* 36 (1989) 1189
[7] BURY, P., JAMNICKY, I., DURCEK, J.: *Physica Status Solidi* (a) 126 (1991) 151
[8] ALMOND, D. P., WEST, A. R.: *Solid State Ionics 26* (1988) 265
[9] CARINI, G., CUTRONI, M., FEDERICO, M., GALLI, G., TTIPODO, G.: *Physical Review B 30* (1984) 7219
[10] BORJESSON, L.: *Physical Review B 36* (1987) 4600
[11] ROLING, B., INGRAM, M. D.: *Physical Review B 57* (1998) 14192
[12] BURY, P., JAMNICKY, I., HOCKICKO, P.: *Communications No. 2,* (2003) 5
[13] PINCIK, E., KOBAYASHI, H., RUSNAK, J., TAKAHASHI, M., BRUNNER, R., JERGEL, M., MORALES-ACEVEDO, A., ORTEGA, L., KAKOS, J.: *Applied Surface Science 252* (2006) 7713
[14] HOCKICKO, P., BURY, P., JURECKA, S., JAMNICKY, M., JAMNICKY, I.: *Advances in Electrical and Electronic Engineering,* Vol. 3, No. 2 (2004) 243
[15] BURY, P., HOCKICKO, P., JAMNICKY, M.: *Advanced Materials Research 39-40* (2008) 111
[16] JAMNICKY, M., ZNASIK, P., TUNEGA, D., INGRAM, M. D.: *J. Non-Cryst. Solids 185* (1995) 151
[17] ZNASIK, P., JAMNICKY, M.: *Solid State Ionics 92* (1996) 145
[18] ZNASIK, P., JAMNICKY, M.: *Solid State Ionics 95* (1997) 207

Tomas Lengyelfalusy – Zdena Kralova *

# PERCEPTUAL ANALYSIS OF L2 PHONIC COMPETENCE

*The research is focused on the quality of English pronunciation of Slovak native speakers reflected in the score of perceptual evaluation by English native speakers. The primary objective is to detect the improvement of English phonic competence after the metaphonetic input focused on the contrastive analysis of the Slovak and English segmental systems.*

*Key words: L2 pronunciation – perceptual analysis – statistics*

## 1. Introduction

From a percipient`s point of view, the impression of "good" or "bad" foreign language (L2) pronunciation is dependent on many subsegmental, segmental, plurisegmental and suprasegmental phonic phenomena. Some studies [1] proved that many segmental sound substitutions significantly correlate with evaluating the pronunciation as non-idiomatic (non-native). However, the substitutions are not the only criterion. They are very easy to perceive, but the percipient's auditive impression represents a complex of phenomena and factors.

Besides the significant inter-individual variability of non-natives' phonic performance it is obvious that the majority of adult learners permanently speak L2 with a foreign accent without regard to the competence of other language levels. The ultimate level of a non-native speaker`s pronunciation often reflects significant inter-individual differences and intra-group tendencies. The primary objective of the research in this field should be the detection and verification of methods and techniques of L2 learning that help learners achieve high standards of L2 pronunciation effectively.

## 2. Methodology

The primary objective of our research was to detect the quality of the English phonic performance of adult Slovak speakers. The research was carried out before and after the speakers had received special metaphonetic input focused on the contrastive analysis of the English and Slovak segmental systems. The experimental and the control groups went through identical phonetic training focused on the segmental subsystem of the English language. The information the experimental group received during the training period was based on the comparative analysis of English and Slovak phonic systems [2]. On the other hand, the information provided to the control group included only the information about the English phonic (segmental) system without the Slovak-English contrastive aspects.

The first recordings of the respondents' spontaneous English monologues were made before the experiment was carried out (September). After 10 weeks of the theoretical input mentioned above and phonetic training in both groups more recordings were made (December). The audio-material was perceptually evaluated by four English native speakers (informants) who used a 5-point evaluation scale. The data obtained from the perceptual analysis were then evaluated through the basic descriptive statistical methods (mean, modus, median, standard deviation, variation coefficient).

We performed a longitudinal comparative quasi-experiment with two groups – the experimental group (E) and the control group (K). The intervention effect (metaphonetic input) was detected by the comparison of the pre-test ($E^1$ – $K^1$) and the post-test ($E^2$ – $K^2$) data. We applied the covariation analysis in the experimental plan and we measured the dependent variables before and after the experimental intervention.

*Respondents*

We worked with the group of 80 Year 1 students (60 female, 20 male) who studied English Language and Literature at the Faculty of Science, University of Zilina, Slovakia. The average age of the respondents was 19, all of them were Slovak, their mother tongue was Slovak and they all reported normal hearing dispositions. Their average English lexical and grammatical competence was at the B1 and B2 levels (CEFR)[1)]. Most of them started learning English at the first level of elementary school with a non-native English teacher. Most of them have never stayed in an English-speaking country for a long period (months/years). We randomized the respondents into two quasi-homogeneous groups of 40 people (30 female, 10 male).

* **Tomas Lengyelfalusy[1], Zdena Kralova[2]**
[1] Department of Mathematics, Faculty of Science, University of Zilina, Zilina, Slovakia, E-mail: tomas.lengyelfalusy@fpv.uniza.sk
[2] Department of English Language and Literature, Faculty of Science, University of Zilina, Slovakia

[1)] CEFR – Common European Framework of Reference for Languages.

*Informants*

Four British native speakers (from the southern part of England: Brighton, London, Oxford)[2] made the perceptual analysis of the respondents' English pronunciation. Concerning the gender structure of the group of respondents[3] we chose three female and one male informant to increase the compatibility of the experimental comparison.

The informants perceptually evaluated the respondents' pronunciation in a five-point interval scale (5 – very good pronunciation, 1 – poor pronunciation). The informants worked individually and were not informed about the evaluation of the other informants.

*Material*

The basic research material was the recording of the English monologues of each of the 80 respondents (the average length of 3.8 minutes). The recording was done in similar conditions, first in September (pre-test) and then, after ten weeks of specific metaphonetic input and phonetic training, in December (post-test). In order to maintain similar lexis and style of the respondents' utterances we chose an autobiographical topic. We considered the spontaneous monologue speech as more natural than reading isolated lexical units aloud. We tried to simulate a real language performance in which the speaker concentrates more on the contents than on the phonetic expression.

*Intervention*

In the experimental strategy we considered the equipment of relevant faculties in Slovakia. We did not want to apply the research in unrealistic laboratory conditions. The language departments frequently use available English textbooks with supplementary audiomaterials. The lectures for the experimental group were focused on the comparative analysis and practical presentation of the Slovak and English segmental systems with the emphasis on the potential phonic interference [2]. The lectures for the control group dealt with the English segmental system without mentioning the comparative aspect.

At the seminars both groups went through identical phonetic training focused on the segmental subsystem of the English language. The course followed the syllabus of the General and Comparative Phonetics and Phonology at the Department of English Language and Literature of the Faculty of Science, University of Zilina. Several techniques were applied during the course – the modification of the perceptual and articulatory base, the perception training, the articulation training in receptive and productive exercises combining imitative and analytical (cognitive) types of learning [3].

*Procedure*

Pretest (September)

1. We randomized 80 respondents into two groups – the experimental group and the control group, with respect to equal gender representation.
2. We recorded the spontaneous English monologues of each respondent using a condenser microphone.
3. The recordings were then evaluated by four English native speakers in a 5-point scale.

Intervention (10 weeks)

4. The respondents were provided the standard phonetic training in the seminars of the General and Comparative Phonetics and Phonology (50 minutes/once a week)[4].
5. The contrastive (experimental group) and non-contrastive (control group) metaphonetic input was provided in the lectures of the course General and Comparative Phonetics and Phonology (50 minutes/once a week).

Posttest (December)

6. After 10 weeks of phonetic training we recorded spontaneous English monologues of each respondent in similar conditions as those in the pre-test.
7. The recordings were then evaluated by the same English native speakers in a 5-point scale.
8. We made the conclusions for the language system and the didactic practice.

## 3. Discussion

We calculated an average evaluation mark for each respondent and for the whole group based on the perceptual analysis of the native speakers in the pre-test and in the post-test. The data are statistically characterized by the parameters of position and variability (Table 1, Table 2).

Position parameters:

$\bar{x}$ – arithmetic average of the data (mean);
$\tilde{x}$ – median (medium value, which divides the statistical set into two equal parts);
$\hat{x}$ – modus (the most frequent value of the statistical set).

Variability parameters:

$\sigma_x$ – standard deviation (absolute measure of variability);
$V_x$ – variation coefficient (relative measure of variability).

[2] The British pronunciation standard, RP English, is taught at most of the Slovak universities in the given study programmes.
[3] Male and female voices have different formant features which must be considered in the analysis.
[4] We used a textbook with CD for the phonetic training [4], specifically Units 2-9 focused on the English segmental subsystem. We also used complementary material in both groups when needed [5, 6, 7].

Perceptual analysis (experimental group) Tab. 1

| $E^1$ | 1 | 2 | 3 | 4 | $\bar{x}$ | $E^2$ | 1 | 2 | 3 | 4 | $\bar{x}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 3 | 3 | 2 | 3 | 1 | 4 | 3 | 4 | 4 | 3.75 |
| 2 | 2 | 1 | 2 | 2 | 1.75 | 2 | 3 | 3 | 4 | 3 | 3.25 |
| 3 | 2 | 2 | 2 | 1 | 1.75 | 3 | 4 | 3 | 2 | 3 | 3 |
| 4 | 2 | 2 | 3 | 2 | 2.25 | 4 | 4 | 3 | 3 | 3 | 3.25 |
| 5 | 2 | 2 | 1 | 1 | 1.5 | 5 | 3 | 3 | 3 | 3 | 3 |
| 6 | 3 | 3 | 3 | 4 | 3.25 | 6 | 4 | 3 | 3 | 4 | 3.5 |
| 7 | 2 | 3 | 3 | 2 | 2.5 | 7 | 3 | 3 | 3 | 2 | 2.75 |
| 8 | 3 | 2 | 3 | 3 | 2.75 | 8 | 3 | 3 | 3 | 3 | 3 |
| 9 | 2 | 2 | 2 | 2 | 2 | 9 | 3 | 3 | 4 | 3 | 3.25 |
| 10 | 1 | 2 | 1 | 2 | 1.5 | 10 | 3 | 2 | 2 | 2 | 2.25 |
| 11 | 3 | 2 | 3 | 2 | 2.5 | 11 | 4 | 3 | 1 | 3 | 2.75 |
| 12 | 2 | 2 | 2 | 1 | 1.75 | 12 | 2 | 3 | 3 | 2 | 2.5 |
| 13 | 2 | 2 | 2 | 1 | 1.75 | 13 | 2 | 2 | 1 | 2 | 1.75 |
| 14 | 3 | 2 | 3 | 3 | 2.75 | 14 | 4 | 4 | 4 | 3 | 3.75 |
| 15 | 2 | 2 | 3 | 2 | 2.25 | 15 | 3 | 2 | 3 | 3 | 2.75 |
| 16 | 3 | 3 | 3 | 2 | 2.75 | 16 | 3 | 3 | 3 | 3 | 3 |
| 17 | 4 | 3 | 5 | 3 | 3.8 | 17 | 3 | 4 | 5 | 4 | 4 |
| 18 | 4 | 3 | 4 | 2 | 3.25 | 18 | 4 | 3 | 3 | 3 | 3.25 |
| 19 | 3 | 3 | 4 | 3 | 3.25 | 19 | 3 | 3 | 3 | 3 | 3 |
| 20 | 5 | 4 | 5 | 4 | 4.5 | 20 | 5 | 5 | 5 | 5 | 5 |
| 21 | 1 | 2 | 1 | 1 | 1.25 | 21 | 3 | 2 | 2 | 2 | 2.25 |
| 22 | 4 | 3 | 4 | 3 | 3.5 | 22 | 3 | 3 | 3 | 3 | 3 |
| 23 | 3 | 3 | 4 | 2 | 3 | 23 | 4 | 3 | 4 | 3 | 3.5 |
| 24 | 3 | 2 | 3 | 2 | 2.5 | 24 | 4 | 2 | 3 | 3 | 3 |
| 25 | 2 | 2 | 3 | 2 | 2.25 | 25 | 3 | 3 | 4 | 3 | 3.25 |
| 26 | 2 | 2 | 3 | 2 | 2.25 | 26 | 2 | 3 | 3 | 3 | 2.75 |
| 27 | 3 | 2 | 3 | 3 | 2.75 | 27 | 3 | 4 | 3 | 3 | 3.25 |
| 28 | 3 | 2 | 2 | 3 | 2.5 | 28 | 3 | 4 | 4 | 3 | 3.5 |
| 29 | 3 | 2 | 4 | 3 | 3 | 29 | 3 | 3 | 4 | 4 | 3.5 |
| 30 | 2 | 2 | 3 | 2 | 2.25 | 30 | 3 | 2 | 3 | 3 | 2.75 |
| 31 | 3 | 3 | 4 | 3 | 3.25 | 31 | 4 | 4 | 5 | 4 | 4.25 |
| 32 | 3 | 2 | 4 | 3 | 3 | 32 | 4 | 3 | 3 | 4 | 3.5 |
| 33 | 2 | 2 | 2 | 1 | 1.75 | 33 | 3 | 3 | 3 | 3 | 3 |
| 34 | 4 | 3 | 4 | 3 | 3.5 | 34 | 4 | 4 | 4 | 4 | 4 |
| 35 | 2 | 2 | 2 | 3 | 2.25 | 35 | 3 | 3 | 2 | 4 | 3 |
| 36 | 3 | 2 | 2 | 1 | 2 | 36 | 3 | 2 | 3 | 3 | 2.75 |
| 37 | 3 | 2 | 2 | 1 | 2 | 37 | 3 | 3 | 2 | 3 | 2.75 |
| 38 | 1 | 2 | 1 | 1 | 1.25 | 38 | 2 | 2 | 1 | 2 | 1.75 |
| 39 | 4 | 1 | 3 | 2 | 2.5 | 39 | 4 | 3 | 3 | 3 | 3.25 |
| 40 | 3 | 1 | 3 | 2 | 2.25 | 40 | 4 | 3 | 3 | 3 | 3.25 |
| $\bar{x}$ | 2.7 | 2.25 | 2.85 | 2.18 | **2.495** | $\bar{x}$ | 3.3 | 3 | 3.1 | 3.1 | **3.125** |
| $\tilde{x}$ | 3 | 2 | 3 | 2 | **2.5** | $\tilde{x}$ | 3 | 3 | 3 | 3 | **3** |
| $\hat{x}$ | 3 | 2 | 3 | 2 | **2.25** | $\hat{x}$ | 3 | 3 | 3 | 3 | **3** |
| $\sigma_x$ | 0.91 | 0.63 | 1.03 | 0.84 | **0.72** | $\sigma_x$ | 0.69 | 0.68 | 0.98 | 0.67 | **0.61** |
| $V_x$ | 0.34 | 0.28 | 0.36 | 0.39 | **0.28** | $V_x$ | 0.21 | 0.23 | 0.32 | 0.22 | **0.20** |

Perceptual analysis (control group) Tab. 2

| $K^1$ | 1 | 2 | 3 | 4 | $\bar{x}$ | $K^2$ | 1 | 2 | 3 | 4 | $\bar{x}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 3 | 2.25 | 1 | 3 | 3 | 3 | 3 | 3 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 1 | 2 | 2 |
| 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 1 | 2 | 1 | 2 | 1.5 | 4 | 3 | 2 | 1 | 3 | 2.25 |
| 5 | 2 | 2 | 3 | 1 | 2 | 5 | 3 | 3 | 3 | 3 | 3 |
| 6 | 3 | 2 | 3 | 1 | 2.25 | 6 | 2 | 3 | 2 | 3 | 2.5 |
| 7 | 2 | 2 | 2 | 3 | 2.25 | 7 | 2 | 2 | 2 | 3 | 2.25 |
| 8 | 2 | 2 | 2 | 2 | 2 | 8 | 3 | 2 | 3 | 2 | 2.5 |
| 9 | 3 | 4 | 4 | 4 | 3.75 | 9 | 3 | 4 | 3 | 3 | 3.25 |
| 10 | 4 | 3 | 3 | 3 | 3.25 | 10 | 4 | 3 | 3 | 3 | 3.25 |
| 11 | 2 | 2 | 1 | 2 | 1.75 | 11 | 2 | 2 | 1 | 2 | 1.75 |
| 12 | 3 | 3 | 2 | 3 | 2.75 | 12 | 4 | 4 | 4 | 4 | 4 |
| 13 | 2 | 2 | 2 | 2 | 2 | 13 | 3 | 3 | 2 | 3 | 2.75 |
| 14 | 3 | 5 | 5 | 4 | 4.25 | 14 | 4 | 5 | 5 | 5 | 4.75 |
| 15 | 3 | 3 | 3 | 2 | 2.75 | 15 | 3 | 3 | 3 | 3 | 3 |
| 16 | 3 | 2 | 3 | 3 | 2.75 | 16 | 4 | 4 | 3 | 3 | 3.5 |
| 17 | 2 | 2 | 2 | 2 | 2 | 17 | 3 | 2 | 1 | 2 | 2 |
| 18 | 4 | 3 | 3 | 3 | 3.25 | 18 | 4 | 4 | 4 | 4 | 4 |
| 19 | 3 | 3 | 3 | 3 | 3 | 19 | 3 | 4 | 4 | 4 | 3.75 |
| 20 | 2 | 2 | 3 | 2 | 2.25 | 20 | 2 | 3 | 2 | 3 | 2.5 |
| 21 | 2 | 2 | 3 | 2 | 2.25 | 21 | 3 | 3 | 3 | 3 | 3 |
| 22 | 2 | 2 | 2 | 3 | 2.25 | 22 | 3 | 2 | 2 | 3 | 2.5 |
| 23 | 3 | 2 | 3 | 2 | 2.5 | 23 | 4 | 3 | 4 | 3 | 3.5 |
| 24 | 3 | 2 | 3 | 2 | 2.5 | 24 | 3 | 3 | 3 | 3 | 3 |
| 25 | 4 | 2 | 4 | 3 | 3.25 | 25 | 3 | 4 | 3 | 4 | 3.5 |
| 26 | 4 | 2 | 4 | 2 | 3 | 26 | 4 | 3 | 4 | 4 | 3.75 |
| 27 | 2 | 2 | 2 | 3 | 2.25 | 27 | 4 | 3 | 3 | 3 | 3.25 |
| 28 | 3 | 2 | 2 | 2 | 2.25 | 28 | 3 | 3 | 2 | 3 | 2.75 |
| 29 | 4 | 4 | 3 | 3 | 3.5 | 29 | 4 | 4 | 3 | 3 | 3.5 |
| 30 | 1 | 1 | 1 | 2 | 1.25 | 30 | 3 | 3 | 3 | 3 | 3 |
| 31 | 3 | 2 | 3 | 2 | 2.5 | 31 | 2 | 3 | 1 | 3 | 2.25 |
| 32 | 2 | 2 | 2 | 3 | 2.25 | 32 | 3 | 3 | 3 | 4 | 3.25 |
| 33 | 2 | 3 | 2 | 3 | 2.5 | 33 | 3 | 3 | 2 | 3 | 2.75 |
| 34 | 3 | 3 | 2 | 2 | 2.5 | 34 | 3 | 3 | 2 | 2 | 2.5 |
| 35 | 4 | 2 | 4 | 4 | 3.5 | 35 | 3 | 3 | 3 | 3 | 3 |
| 36 | 1 | 2 | 2 | 2 | 1.25 | 36 | 1 | 1 | 1 | 2 | 1.25 |
| 37 | 1 | 2 | 1 | 2 | 1.5 | 37 | 3 | 1 | 1 | 2 | 1.75 |
| 38 | 3 | 3 | 4 | 2 | 3 | 38 | 4 | 3 | 3 | 2 | 3 |
| 39 | 4 | 2 | 3 | 2 | 2.75 | 39 | 3 | 3 | 4 | 3 | 3.25 |
| 40 | 2 | 3 | 3 | 3 | 2.75 | 40 | 3 | 3 | 3 | 2 | 2.75 |
| $\bar{x}$ | 2.575 | 2.38 | 2.6 | 2.45 | **2.488** | $\bar{x}$ | 3.07 | 2.95 | 2.64 | 3 | **2.913** |
| $\tilde{x}$ | 2.5 | 2 | 3 | 2 | **2.375** | $\tilde{x}$ | 3 | 3 | 3 | 3 | **3** |
| $\hat{x}$ | 2 | 2 | 2;3 | 2 | **2.25** | $\hat{x}$ | 3 | 3 | 3 | 3 | 3 |
| $\sigma_x$ | 0.90 | 0.74 | 0.92 | 0.71 | **0.64** | $\sigma_x$ | 0.69 | 0.82 | 1.02 | 0.69 | **0.48** |
| $V_x$ | 0.35 | 0.31 | 0.35 | 0.29 | **0.26** | $V_x$ | 0.23 | 0.28 | 0.39 | 0.23 | **0.16** |

Results of perceptual analysis Tab. 3

| | E | K | *difference* | *%* |
|---|---|---|---|---|
| **pre-test** | 2.495 | 2.488 | ***0.007*** | ***0.28*** |
| **post-test** | 3.125 | 2.913 | ***0.212*** | ***6.78*** |
| ***difference*** | ***0.630*** | ***0.425*** | - | |
| *%* | ***25.25*** | ***17.08*** | - | |

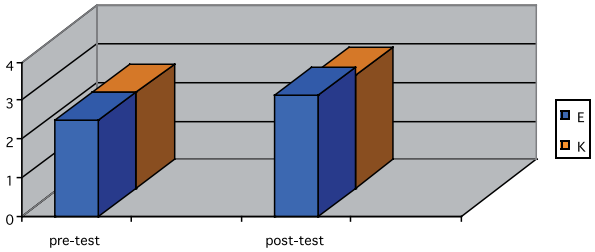



*Fig. 1 Results of perceptual analysis*

Limiting values of perceptual analysis Tab. 4

| | pre-test | | post-test | |
|---|---|---|---|---|
| | $E^1$ | $K^1$ | $E^2$ | $K^2$ |
| min | 1.25 | 1.25 | 1.75 | 1.25 |
| max | 4.50 | 4.25 | 5.00 | 4.75 |

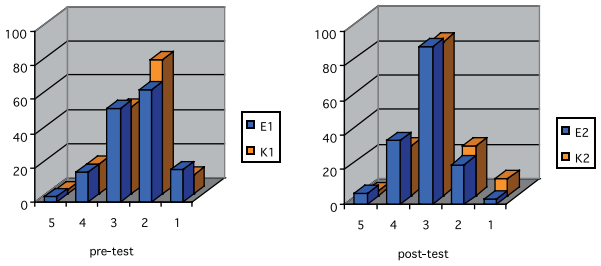



*Fig. 2 Frequency distribution of the values*

As can be seen in Table 3 and Figure 1 the average evaluation mark in the experimental group and in the control group was similar in the pre-test (difference 0.28%). The average evaluation mark of the experimental group in the post-test was higher than the average evaluation mark of the control group (difference 6.78%). While the improvement of the evaluation in the experimental group (pre-test vs. post-test) was 25.25%, in the control group it was lower (17.08%). The maximum and minimum pre-test values in both groups were similar (Table 4), but the limiting post-test values in the experimental group were higher than those in the control group.

The inter-individual variability of the evaluation shows relatively low variance of the values in a given phase of the experiment. An average pre-test and post-test variation coefficient in both groups was $V_x = 22.5\%$.[5)] The most frequent evaluation mark in the pre-test was 2 and in the post-test it was 3 (Figure 2). Modus in both groups in the pre-test was 2.25 and in the post-test 3.

## 4. Conclusions

We applied the synthesis of the theoretical analysis in the experiment and we detected the causal relations of the parameters by intentional manipulation of the independent variables and by the statistical analysis of the variables. We applied the covariance analysis in the experimental plan and we measured the dependent variable before and after the experimental intervention. The absolute randomization of the subjects was not possible in our experimental conditions so we applied the quasi-experimental method of the availability. We tried to interpret the results by the arguments resulting from the experiment which could approve or disprove the alternative theories.

The reliability of the measurement is approved by the agreement (measurement consistency) of the pre-test and post-test evaluation of an informant (test-retest reliability) with respect to the expected interaction of the maturation (the improvement of the respondents' pronunciation in a longitudinal design). The consistency of repeated measurements is approved by acceptable mean values $V_x$.

Besides the pre-test sensibilization (respondents' and informants') the interaction of maturation – the increase of experience – should be considered in the longitudinal type of experiment. In this case the interaction of maturation is one semester of the English language study which was partly eliminated by its homogeneous influence on all respondents.

The validity of measurements should be confirmed by the reliability of conclusions based on the measurements. To provide the internal validity we used statistical operations of the variance analysis. We tried to increase the external validity of measurements (possibility to generalize the results beyond an experiment) by work in fairly natural conditions – the common school environment and the research material reflecting natural communication.

One of the primary objectives of our research was to compare the effectiveness of the contrastive and non-contrastive metaphonetic input in teaching English phonetics and phonology at the university level in Slovakia. The abstract memory is intensively developing and the intentional type of memory is getting more effective with adult learners. The process of L2 pronunciation learning thus should be based on the theoretical information on the given activity. In case of adult learners the analytical (cognitive) type of phonetic training is considered to be more effective

[5)] Variation coefficient higher than 50% indicates high heterogeneity of the statistical set [8].

than the imitative type of training [3]. The awareness of differences, similarities and potential pronunciation mistakes (or deviations) resulting from the differences of the first and second language systems can contribute to the L2 phonic performance of an individual significantly.

Our research confirmed higher effectiveness of the contrastive metaphonetic input reflected in an overall quality of English pronunciation. Native speakers' perceptual analysis is subjective but the evaluation can be systematically interpreted and verified in an experiment. The level of an individual's phonic competence is often evaluated in terms of fluency and communicativeness of the speaker. However the interactive complexity of characteristics creating the final effect – L2 pronunciation – is undoubted. "Internal basis is an inseparable precondition of functionally adequate phonetic analysis of external speech factors" [9] and the performance is a point where the language – a potentially invariant and general phenomenon – meets with the variant and individual phenomenon – speech [10].

## References


[1] FLEGE, J. E.: *The Phonological Basis of Foreign Accent: A Hypothesis.* In: TESOL Quarterly, vol. 15, 1981, n. 4, p. 443–455.

[2] KRALOVA, Z.: *Slovak-English Phonic Interference [in Slovak].* Zilina: Faculty of Science, University of Zilina, 2005. 100 p. ISBN 80-89029-85-X.

[3] CHEBENOVA, V.: *Methods and Techniques in the German Pronunciation Training (I) [in Slovak].* In: Cizi jazyky, vol. 45, 2001-2002, n. 1, p. 6–8.

[4] ROACH, P.: *Eng1ish Phonetics and Phonology.* Cambridge: Cambridge University Press, 1989. 280 p. ISBN 9521786134.

[5] O´CONNOR, J. D.: *Better English Pronunciation.* Cambridge: Cambridge University Press, 1980. 149 p. ISBN 0-521-23152-3.

[6] BAKER, A.: *Ship or Sheep?* Cambridge: Cambridge University Press, 1981. 168 p. ISBN 052128354-X.

[7] BOWLER, B., CUNNINGHAM, S.: *Headway Pronunciation (Intermediate, Upper-Intermediate).* Oxford: Oxford University Press, 1999. 63 p. ISBN 0-194-36245-0.

[8] HENDL, J.: *An Overview of the Statistical Methods of Data Processing [in Czech].* Praha: Portal s. r. o., 2004. 584 p. ISBN 80-7178-820-1.

[9] KRAL, A.: *Speech Mechanism Model [in Slovak].* Bratislava: VEDA, 1974. 188 p.

[10] HLAVNOVA, A., HLAVNA, V.: *Is language the final barrier in a shrinking world?* In: Communications – Scientific Letters of the University of Zilina, vol. 5, 2003, n. 3, p. 96–97.

Ondrej Such *

# EXPERIMENTAL BEHAVIOR OF JURIK'S NEAREST POINT APPROACH ALGORITHM FOR LINEAR PROGRAMMING

*T. Jurik recently proposed a new algorithm for solving linear programming problems. The algorithm is iterative, finding points on the boundary of the problem's convex polyhedron. Global convergence of the algorithm is an open question. Jurik benchmarked the performance of the algorithm on standard sets of linear optimization problems, and the algorithm was on par with commonly used ones, and sometimes beating the simplex method by a wide margin. In this note we concentrate on testing local behavior of the algorithm near a vertex in the three dimensional Euclidean space. Our experiments indicate that the behavior of the algorithm is mostly very fast, but there appear to be cases where its behavior is worse than the simplex method.*

## 1. Linear programming problems

In linear programming problems [3] one tries to minimize a linear cost function on a set of higher dimensional vectors bounded by linear constraints. The simplex algorithm [4] to solve this problem was proposed by B. Dantzig in 1947. The algorithm is iterative, moving through vertices of the convex polyhedron bounded by the problem's constraints. It performs quite well in practice, and is still being used. However, it was shown in 1972 by Klee and Minty that in a worst case it takes exponentially many steps. Multiple polynomial time running algorithms were proposed such as L. Khachinyan's ellipsoid method in 1979 and N. Karmarkar's in 1984. The research in this topic of high practical importance is ongoing, with open problems even for the simplex method such as Hirsch's conjecture.

## 2. Jurik's algorithm

Jurik in [1] proposed a new algorithm for solving linear optimization problems. It starts with constructing a hyperplane $H$ orthogonal to the cost direction that lies sufficiently far from a convex polyhedron $P$. One then observes that to find the optimal $x$ is the same, as to find pairs of points $w$,$x$, where w lies in $H$ and $x$ lies in $P$ whose distance is minimal. The algorithm has an *outer* iterative cycle and embedded within it is an *inner* iterative cycle. Let us briefly describe what is computed at each iteration of the cycles.

The algorithm stipulates how to generate initial points $w^0$ and $x^0$ (the latter is generated first). Suppose in the $k$-th iteration of the outer cycle we generated points $w^k$ lying in the hyperplane H and $x^k$ in $P$ (necessarily on the boundary of $P$). A line $L^k$ is then defined as the projection of the line passing through $w^k$ and $x^k$ onto the hyperplane $H$. New points $w^{k+1}$ and $x^{k+1}$ are then generated in the inner cycle (which Jurik showed terminates in finite number of steps) as the points of minimal distance lying on $L^k$ and P respectively. Because the line $L^k$ contains the projection $r^k$ of point $x^k$ onto $H$, with increasing $k$ the distance of point $x^k$ to H (the cost function) cannot increase. In fact we have the following inequalities for distances

$d(x^{k+1}, H) <= d(x^{k+1}, L^k)$ because $L^k$ lies in $H$
$d(x^{k+1}, L^k) <= d(x^k, L^k)$ because $x^{k+1}$ is a point in $P$ minimizing distance to $L^k$
$d(x^k, L^k) = d(x^k, H)$ because Lk contains the projection of $x^k$ onto $H$

and thus

$$d(x^{k+1}, H) <= d(x^k, H)$$

At this time, it is not known whether the algorithm always converges. Jurik himself proved that the inner cycle stops after a finite number of iteration. The following conjecture was indicated by Jurik:

Conjecture A Jurik's algorithm converges to the optimum in a finite number of steps.

## 3. An example needing many line iterations

Jurik in his thesis [2] reports that in the vast majority of cases, he needed to use only two lines to find the optimum. In our experiments we found a 3-dimensional problem which takes many more iterations. Let us define the vectors

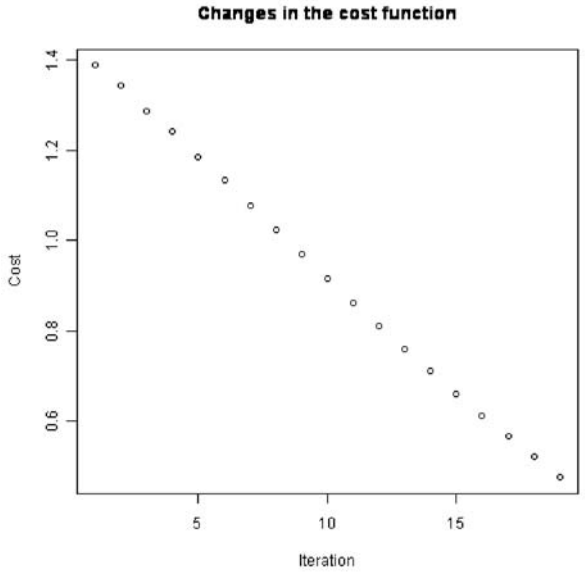
* **Ondrej Such**
Institute of Mathematics and Computer Science, Banska Bystrica, Slovakia, E-mail: ondrejs@savbb.sk

$V_1 = (0.18484\ \ 0.21435\ \ 0.95911)$
$V_2 = (-0.15919\ \ 0.97907\ \ 0.12678)$
$V_3 = (0.07727\ \ 0.0453\ \ 0.99598)$

and the line $L_1$ passing through the following two points

$P_1 = (3.09017\ \ 9.51057\ \ -1)$
$P_2 = (2.13911\ \ 9.81958\ \ -1)$

Consider the problem of minimizing the cost function $c(x,y,z) = z$ inside the convex hull of nonnegative multiples of vectors $V_1$, $V_2$, $V_3$. In this case one finds almost linear convergence of Jurik's algorithm towards optimum value of 0, as shown in the following figure.




This leads us to formulate the following conjecture, which indicates that in the worst case the running time of Jurik's algorithm is worse than the running time of simplex algorithm.

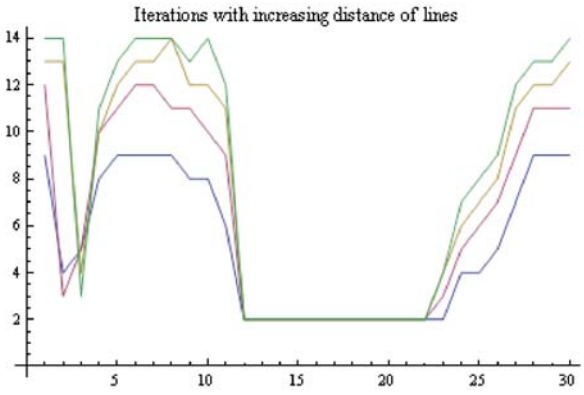
**Conjecture B** The number of steps in Jurik's algorithm is not bounded by any function of the problem's dimension and the number of constraints.

## 4. Typical behavior

In this section we present an example, that based on our experience, we believe is quite typical for running behavior of Jurik's algorithm on three dimensional problems. We chose three random vectors:

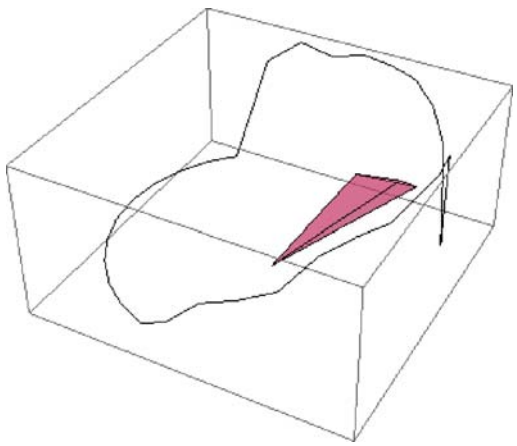
$V_1 = (0.524169\ \ -0{,}000129183,\ \ 0.851614)$
$V_2 = (0.674312\ \ 0.285704\ \ 0.680938)$
$V_3 = (0.631712{,}0.67865{,}0.374667)$

We looked at the cost function $c(x,y,z) = z$ and tried to minimize it using Jurik's algorithm on the convex set spanned by nonnegative multiples of $V_1$, $V_2$, $V_3$. For $H$ we chose the hyperplane $z = -1$, and there we chose 30 lines at distances 10, 20, 30 and 40 from the point $(0\ 0\ -1)$. The following figure shows that the number of iterations for those 30 lines increases slightly with the increasing distance.




We see that for quite a few starting line positions it took just two iterations to converge to optimum (the precise counts are 12, 13, 13, 13), confirming the behavior observed by Jurik.

The next figure shows how the position of the line relative to the optimizing set affected the number of iterations. The higher the point on the polygonal line the more iterations it took starting with the line (at distance 40 from $(0\ 0\ -1)$) passing through its projection to converge to the minimum.



## 5. Conclusion

Jurik presented in his works [1], [2] compelling evidence that his new algorithm is competitive with commonly used linear opti-

mization algorithms. Based on our, admittedly very limited, experimental experience we stated two conjectures about the behavior of the algorithm. In case the Conjecture B holds, a modification of the algorithm is needed to deal with some linear programming problems, perhaps by rescaling the problem, a clever choice of hyperplane H, or some other method.

## References

[1] JURIK, T.: *A new efficient two-path algorithm for linear programming problems,* Journal of Applied Mathematics, Statistics and Informatics, 4 (2008), pp. 129–138
[2] JURIK, T.: *Contributions to the Algorithms of Linear Programming,* Ph.D. thesis, Comenius University, 2009
[3] http://en.wikipedia.org/wiki/Linear_programming
[4] http://en.wikipedia.org/wiki/Simplex_algorithm

Maria Cernanska – Ondrej Skvarek *

# CLUSTERING OF SLOVAK SENTENCE MELODY – METHODS AND RESULTS

*Modern speech synthesis systems implement prosody features to achieve more naturally sounding voices, trying to produce sounds similar to the human speech. This article deals with obtaining one of prosody characteristics (sentence melody contour) from human speech recordings. Similarities of melody contours are studied using cluster analysis. Several clustering methods are evaluated for this purpose, estimation of proper number of clusters is described and typical melodies for different types of Slovak language sentences (declarative, interrogative, exclamatory, clauses with final "," etc.) are found and recommended for implementation in our text-to-speech system.*

*Keywords: speech synthesis, prosody, melody contour, cluster analysis, R-software*

## 1. Introduction

Progress in the area of speech processing enables new ways of communication between people and computers. Reading a text from the computer screen can be extended or replaced by a text to speech synthesis, writing a text on the keyboard or choosing commands using the mouse can be enriched by voice recognition systems. A text-to-speech synthesizer (TTS) converts a text from documents into audio voice files. One of such synthesizers (TTS-KIS) [10, 3] is developed in the Department of Information Networks at the Faculty of Management and Informatics of the University of Zilina.

Our TTS system implements a concatenative method of speech synthesis. The method is based on concatenation of speech elements "diphones" [3]. Diphones are selected from the TTS sound database at the time of speech synthesis.

Diphones in the database are stored in a normalised form with monotonous melody. Consequently a monotonous waveform corresponding to the input text is produced, and hence a proper melody modification is needed to achieve a naturally sounding sentence.

Our TTS system applies composed melody contour to the synthesized sentence waveform. The contour is composition of "sentence melody contour" (long-term melody trend spanning along the whole sentence or clause) and of "short-term melody contours" (related to shorter speech segments, words, bars, syllables). Our present work is focused on obtaining the long-term "sentence melody contour".

We use terms "word" and "bar" interchangeably to denote the word with neighbouring proclitics and enclitics. Terms "sentence" and "clause" are interchangeably used to denote smaller sentence parts of simple, complex or compound sentences separated by punctuation marks and conjunctions (see Table 2).

The TTS-KIS system analyses text of each input sentence and chooses melody contour corresponding to the characteristics found in the text: the sentence type (recognized by punctuation marks, interrogative words ...) and the number of words (bars) in the sentence. Our aim is to find typical "sentence (clause) melody contour" for each group of sentences characterised by the sentence type and by the number of words. Sentence types recognised by the TTS system are listed in Table 3.

Proposed method obtains original sentence melody contours using software Praat [11].

Then weighted-MA method [4] defined in formula (1) is used to remove short-term melody variations and to obtained sentence melody contour values. The smoothed values are computed in equidistant time events along the whole sentence (clause) waveform. The number of melody contour values is set proportionally to the number of words (bars) in the sentence. The number is equal for all sentences in the same sentence group, which makes comparison of different length waveforms possible.

Sentence melody contours are further centred to 0 Hz mean frequency value. Contour dynamics remains unchanged.

Similar melodies are grouped into separate clusters (the best clustering method for this purpose is found). Cluster with melodies best suited for analysed sentence type is selected and typical melody of this cluster is recommended for our TTS system.

* **Maria Cernanska, Ondrej Skvarek**
Department of Information Networks, Faculty of Management and Informatics, University of Zilina, Slovakia,
E-mail: maria.cernanska@fri.uniza.sk

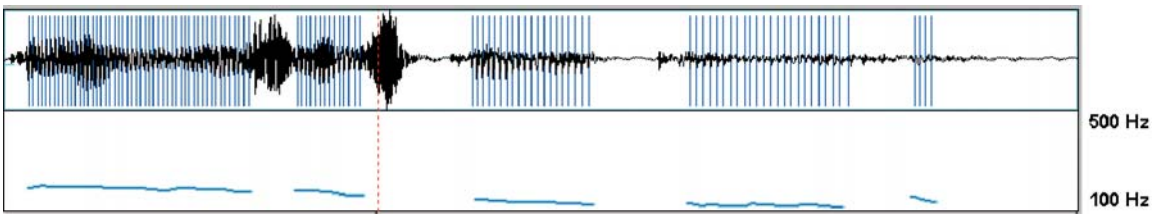



*Fig. 1 Waveform (upper part) and melody contour (lower part) of sentence „Lenze zostalo to v nej.“*

## 2. Speech material

We examined sound recordings of the novel [16]. The story was narrated by a female speaker in a faster speaking rate with an emotive accent (noticeable changes in the intonation, loudness and tempo).

Recordings were cut into smaller parts corresponding to the clauses. See Table 2 for punctuation marks and conjunctions used to determine clause boundaries. (Simple sentence and parts of complex or compound sentences were taken as separate clauses.)

Sound files were assigned names corresponding to the page number, sentence number, number of bars in the sentence and the type of the starting and ending punctuation marks. About 8000 sentence sounds (see Table 1) of the PCM format (22050 Hz, 16 bit/sample, mono) were stored in the 500 MB disk space.

## 3. Obtaining melody contours from voice recordings

We use the program Praat [11, 2] to obtain the glottal frequency contour F0. Praat implements a normalised autocorrelation function and best path selection algorithms. Proper settings of parameters are needed to obtain real F0 values [4]. To process the large number of voice records, the script of Praat commands was programmed.

## 4. Preparing melody contours for cluster analysis

Melody contours obtained by the program Praat are composed of equidistant values located inside detected voiced intervals (see Fig. 1).

To compare melodies of sentences with different time duration and with different distribution of voiced intervals, a smoothing method “weighted moving average” is used [4]. Weighted-MA values $y'(t)$ are computed according to the formulas (1) and (2).

$$y'(t) = \frac{\sum_{i=1}^{n} y_i * w_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

where

- $n$ denotes the number of original melody values taking into computation ($n/2$ values to the left and $n/2$ values to the right from the time event $t$),
- $y_i$ denotes the original melody values,
- $t_i$ denotes corresponding time events,
- $w_i$ denotes weight assigned to $y_i$ values, see the formula (2).

$$w_i = a * |t - t_i| + b \tag{2}$$

Weights decrease with raising distance from the time event $t$. Parameters were set to values $n = 12$, a $= -2$, b $= 1$.

Number of analyzed n-member clauses — Table 1

| Ending punctuation mark or conjunction | n | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | ∑ |
| a, i, aj, ani, či | 87 | 154 | 201 | 191 | 115 | 67 | 55 | 21 | 2 | 5 | 1 | 1 | 0 | 0 | 900 |
| „ . “ | 141 | 388 | 519 | 489 | 425 | 300 | 155 | 88 | 27 | 19 | 15 | 5 | 2 | 1 | 2574 |
| „ , “ | 502 | 607 | 582 | 540 | 373 | 201 | 112 | 52 | 21 | 12 | 5 | 1 | 2 | 0 | 3010 |
| „ : “ | 4 | 10 | 16 | 9 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 44 |
| „ ... “ | 145 | 118 | 124 | 144 | 108 | 85 | 40 | 20 | 9 | 7 | 2 | 0 | 0 | 1 | 803 |
| „ ? “ | 54 | 78 | 76 | 65 | 31 | 24 | 10 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 343 |
| „ ?! “ | 7 | 8 | 3 | 10 | 8 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 40 |
| „ - “ | 16 | 4 | 5 | 3 | 2 | 0 | 3 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 37 |
| „ ! “ | 37 | 30 | 33 | 28 | 18 | 11 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 164 |

The method removes short-term variations from the overall sentence melody and computes "smoothed sentence melody contour".

We compute melody values in equidistant time intervals. The number of contour points is set to 8-th multiply of number of words in the sentence. Each contour is further centred to 0 Hz mean frequency value. Dynamics of the contour remains unchanged.

## 5. Description of clustering methods

Once melody contours are represented by equal length vectors (values determined in equidistant time intervals) the cluster analysis of melody contour similarities is possible. Clustering methods found in the R-software's "hclust" method [12] were taken and fitness for melody contour clustering was examined. All the methods (ward, single, complete, average, mcquitty, median, centroid) perform hierarchical agglomerative clustering. Initially, each object (contour) is assigned to its own cluster, then (iteratively) distances between clusters are computed and the two closest clusters are joined.

Initial distances between clusters (dissimilarity matrix) are computed by the R-software's "dist" method as the Euclidean (geometric) distance.

The second and further iterations of a dissimilarity matrix are recomputed by the Lance–Williams dissimilarity update formula [1, 6]. The formula parameters are set to values [9] according to the method chosen (ward, single ...).

Each of the mentioned methods evaluate the distance between clusters in a different way, hence clusters of different properties are produced.

*Single Linkage method* In this method, the distance between two clusters is computed as the *smallest distance* of two objects, where the first object belongs to the first cluster and the second object belongs to the second cluster. Resulting clusters tend to represent long, straggly "chains" [13].

*Complete Linkage method* The distance between two clusters is computed as the *largest distance* of two objects, where the first object belongs to the first cluster and the second object belongs to the second cluster. This method tends to find extremely compact clusters [5]. The method is usually suitable when the objects form naturally distinct clusters. If the clusters tend to be somehow elongated or "chain" type nature then this method is inappropriate [13].

*Average Linkage method* The distance between two clusters is computed as the *average distance* of two objects, where the first object belongs to the first cluster and the second object belongs to the second cluster. The method tends to find globular clusters [5]. It is very efficient when the objects form natural distinct clusters and performs equally well with elongated "chain" type clusters [13]. This method is also called "the group average linkage algorithm" or "the unweighted pair group method average" (UPGMA) [14].

*McQuitty's method* In this method, when two clusters are merged into a new one, the distance from the new cluster to the old one is computed as an average of distances between two merged clusters and the old cluster [14]. Such rule corresponds to the weighted average computation, where *objects in small clusters have a larger weight* than those in large clusters. This method is also known as "the weighted average linkage algorithm" or "the weighted pair group method average" (WPGMA). This method (rather than the previous method) should be used when the cluster sizes are suspected to be greatly uneven [13].

*Centroid Linkage method* In this method, centroid (average of objects) of each cluster is computed, then the distance between two clusters is determined as the distance between centroids representing the two clusters.

*Median Linkage method* This method extends Centroid method by weights to consider different numbers of objects in clusters. This method is preferable to the Centroid method, when considerable differences in cluster size are expected [13].

*Ward's minimum variance method* In this method, the distance between two clusters is evaluated as the growth of total dispersion of objects around their respective cluster centroids. This method minimises clusters heterogeneity [8] and it tends to find globular clusters [5]. In general, it is regarded as very efficient [13].

## 6. Selection of the best method for melody contour clustering

We used two criteria to select the best method for melody contour clustering:

1. Correct separation of melody contours according to their similarities (similar melodies grouped in the same cluster, different melodies separated into different clusters).
2. Minimal number of clusters needed to accomplish the criterion 1.

Seven clustering methods (ward, single, complete, average, mcquitty, median, centroid) were evaluated. 4-member (4-word) declarative and interrogative sentences (clauses) were taken and clustering was computed by each of the methods (see corresponding dendrograms in Figs. 2 and 3). For each case criteria were evaluated and the method that best met the criteria was recommended to cluster all other sentence groups.

We evaluated Criterion 1 using:

- dendrograms obtained by clustering
- drawings of melody contour clusters
- hearing sounds grouped into clusters

Dendrograms depict inter-cluster distances and the size of clusters formed during computations. Figs. 2 and 3 shows that the

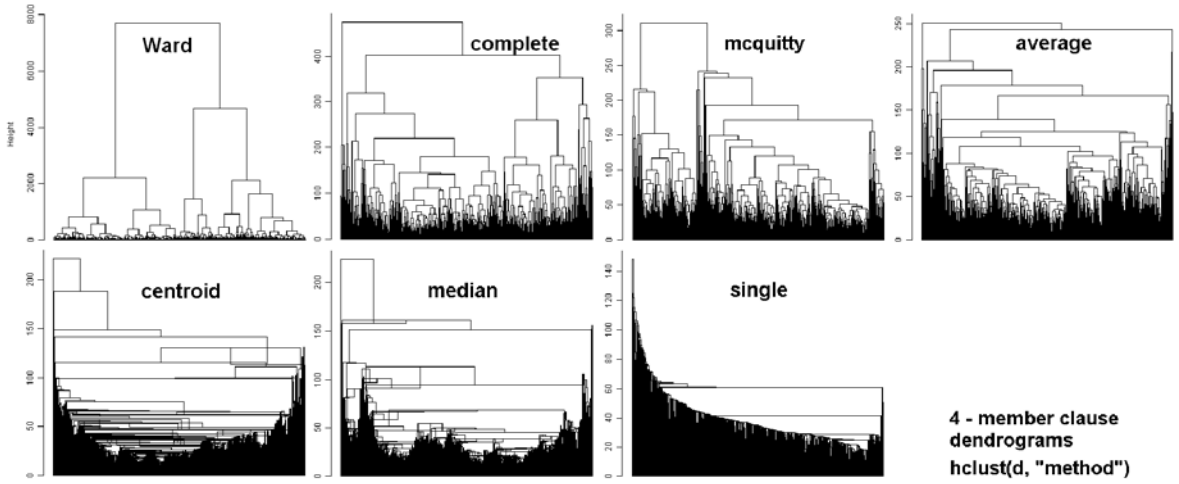



*Fig. 2 Declarative 4-member clause dendrograms (various clustering methods)*

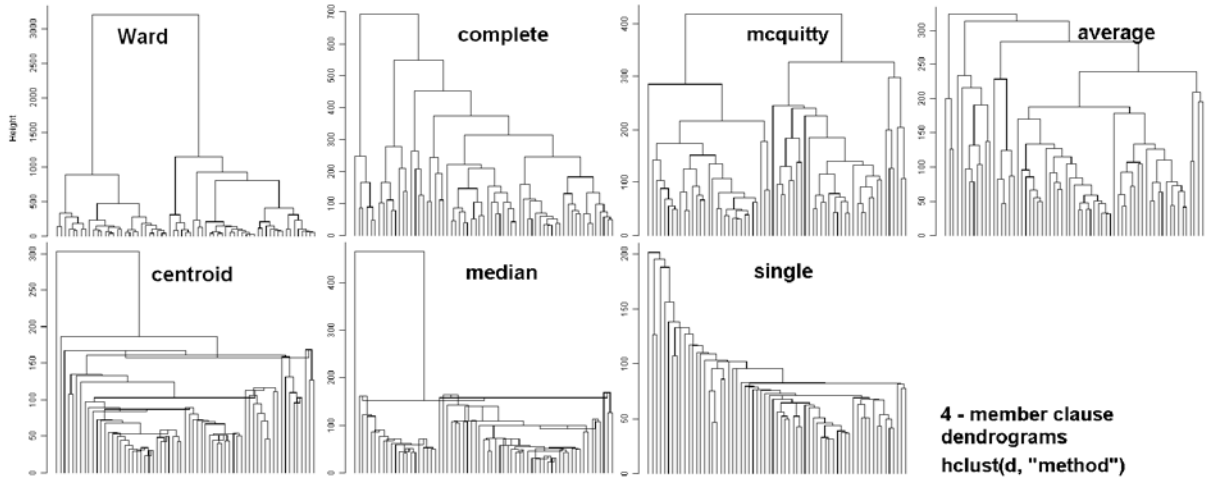



*Fig. 3 Interrogative 4-member clause dendrograms (various clustering methods)*

Ward's method tend to create clusters of similar size. Comparing these clusters to Fig. 4 we can see good separation of different melody shapes corresponding to these clusters. On the other hand the Single method separates very small clusters while retaining different melody shapes in one large cluster – not capable to meet the Criterion 1. The other methods exhibit properties between the Ward's and the Single method.

We investigated melody contour drawings of two, three and more clusters (starting from the top of the dendrogram). We stopped at the number of clusters when further cluster division would not give new clusters with significantly different melody shapes. Obtained number of clusters is taken as the Criterion 2. Proper separation of melody contours we verified by hearing of sounds from corresponding clusters.

*Ward's method* created clusters of similar size (see Figs. 2, 3 and 4) and postpone separation of single or small groups of contours to the later iterations. The proper melody separation was achieved at the number of clusters equal to five (Criterion 2 value). The *Complete method* put small groups of melodies into separate clusters in earlier iterations. In the case of interrogative clauses, at the moment of five clusters, different kinds of melodies (falling and rising) were still included in the same cluster. *McQuitty's method* kept two large clusters while separating small size clusters. The *Average method* separated even smaller clusters, still keeping one or two large clusters. The *Median method* separated objects into clusters of a very small size. The *Centroid method* separated melodies nearly one-by-one. The *Single method* also separated single objects; keeping dissimilar melodies in one large cluster (compare Figs. 2, 3 and 4).

The *Ward's method* was chosen as the most suitable for clustering melody contours. It performs *proper separation of melody contours* using the smaller number of clusters. All the methods

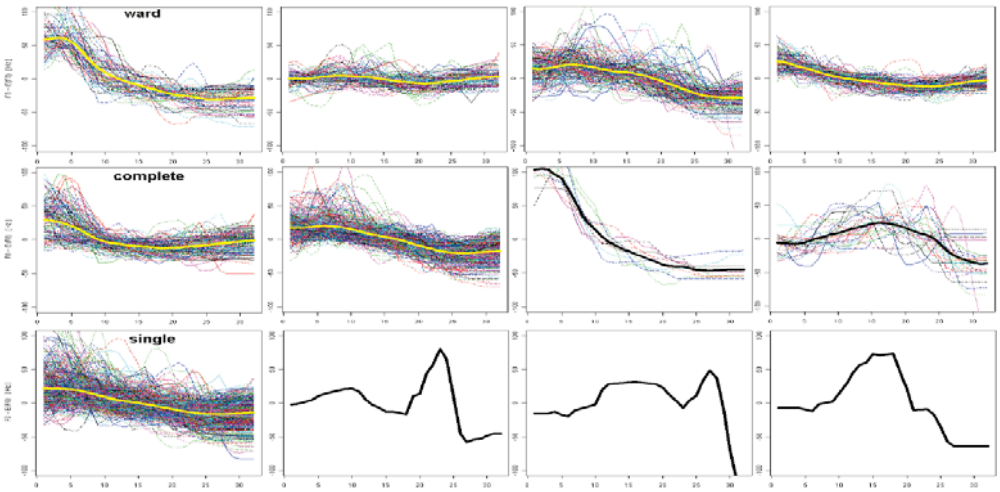



*Fig. 4 Clusters of 4-member declarative clauses computed by three clustering methods: Ward's, Complete and Single method.*

were ordered according to the suitability: Ward's, Complete, McQuitty, Average, Median, Centroid and Single.

## 7. Identifying melody contours for the TTS system

We used Ward's method (the best clustering method found in the previous paragraph) to compute clustering for all sentence types and different number of words. For each clustering we determined clusters with proper melody separation (also described in the previous paragraph). Then we determined the proper melody for particular groups of sentences using following criteria:

- melody features described by Slovak language scientists in [7]
- sentences uttered in a neutral way (without strong word accent)
- number of contours in the cluster

For example, analysing Slovak 4-clause determination sentences we stopped clustering at five clusters (see Fig. 5, clusters b1-b5). *Cluster b1* exhibits the rise at the beginning of the clause followed by steep melody fall. The hearing of sentences confirmed sentences with very strong accent on the first word in the sentence expressing emotional speech. *Cluster b2* contains flat melody contours, corresponding to repose even phlegmatic atmosphere, without conspicuous emotions. *Cluster b3* exhibits a rise at the middle part. Stress is heard on the third word or on the fourth word. *Cluster b4* has a tendency similar to the cluster b1, with slower decrease. The beginning of the sentence is less stressed, which adds moderate dynamics to the speech. Representative melody contour of this cluster was chosen as the "sentence melody contour" for 4-clause determination sentences in our TTS system (see Table 3). *Cluster b5* contours exhibit a slow rise in the melody at the end of the clause. Hearing it we found that the cluster contains: sentences with

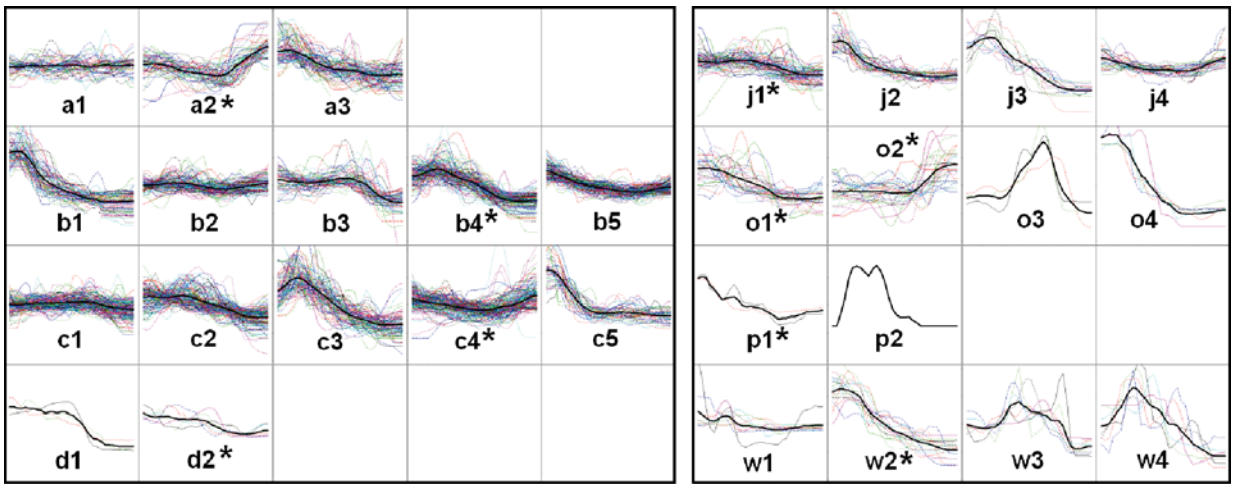



*Fig. 5 Results of cluster analyses (Ward's method) of 4-member clauses. Letters "a, b, c, d, j, o, p, w" denote ending of the sentence with different punctuation marks (see Table 2).*

Sentence ending labels and corresponding punctuation marks or conjunctions placed at the end of clauses

Table 2

| Label | *a* | *b* | *c* | *d* | *j* | *O* | *p* | *w* |
|---|---|---|---|---|---|---|---|---|
| Ending punctuation mark or conjuction | "a, i, aj, ani, či, alebo" not preceded by comma mark "," | " . " | " , " | " : " | " ... " | " ? " | " – " | " ! " |

melodies of non-ended sentences (similar to melody of comma "," ended sentences), sentences expressing theatricality of the story and sentences with a quiet ending (causing non-precise F0 calculation at the end of the sentence).

The clusters with melodies suitable for TTS systems are marked by asterisk "*" (see Fig. 5). Representative contours (arithmetic mean of contours) of these clusters are recommended for implementation in our TTS-KIS system. Alternation of chosen melodies with more expressive melodies (e.g. alternation of b4 and b1, Fig. 5) can be implemented to achieve more dynamic utterance production (see description of non-satisfying non-ending sentence melodies in [7]).

Another example – clustering of 4-member declarative clauses of different lengths is shown in Fig. 6.

## 8. Mapping of text characteristics into melodies recommended for TTS-KIS system

The cluster analysis showed the same results (shapes of melody contours) as language scientists described in [7] and we summarized in [15]. We recommend these melody contours for our TTS system and we have prepared mapping of the input text characteristics into recommended contours (e.g. see the 4-member clauses case in the Table 3). At the time of speech syntheses the contour is stretched to the length of the sentence waveform. The stretched contour and short-term melody contours are added to the synthesized sound.

## 9. Conclusion

In this article we described method of obtaining sentence melody contours for our TTS system. Cluster analysis is used to analyse 8000 melody sentence contours of Slovak language speech recordings. Melody contours were obtained from recordings by software Praat, prepared by the smoothing method and centred to the zero mean frequency. Then cluster analyses were performed. Seven clustering methods (implemented in "hclust" method of R-software) were compared. The Ward's method was chosen as the best one for the purpose of melody contour clustering. This method was used for clustering of different types of sentences with different numbers of words (bars). For each group of sentences clusters with correct separation of melody contours were found and melodies of individual clusters were analysed. The results of the analysis were formulated as recommended melody contours for individual sentence types. Text characteristics recognised by our TTS system were mapped to recommended sentence melodies.

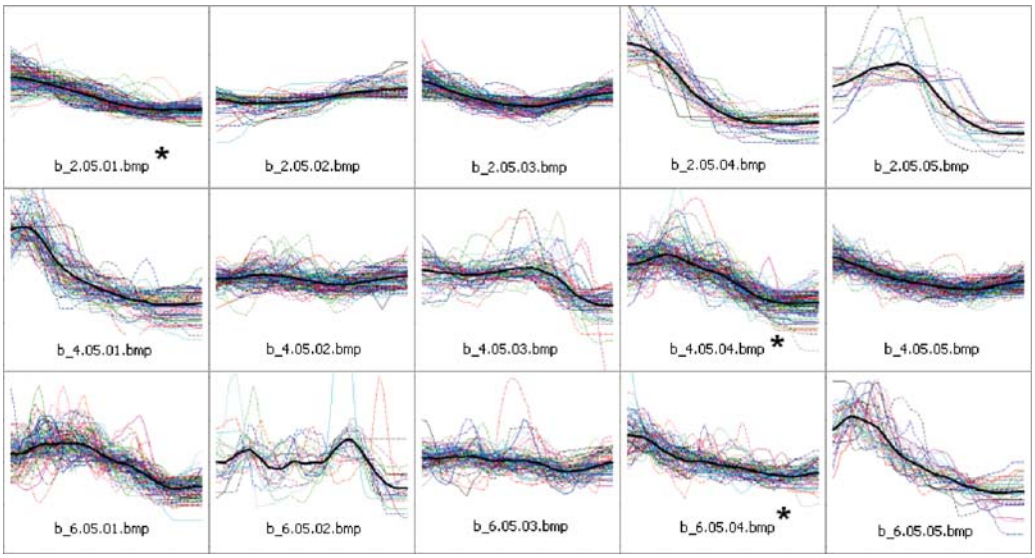
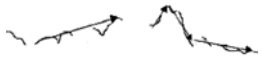
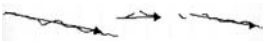



*Fig. 6 Ward's method clustering of 2-, 4- and 6-member declarative clauses. Number of analysed sentences: 2-member (607), 4-member (540) and 6-member clauses (201)*

Mapping of text characteristics into recommended sentence melodies and melodies expected by language scientists (4-member clauses case) Table 3

| Melody type | Sentence type | Characteristics recognised by the TTS system in the input text | 4-member clause melodies (Fig. 5) recommended for the TTS system | Examples of melody contours expected by language scientists (see [7, 15]) |
|---|---|---|---|---|
| **1. Satisfying ending (conclusive cadence)** | Declarative Exclamatory Wh-question ("doplnovacia otazka" in Slovak) | Final " . ", " ! "; " ? " when beginning with an interrogative word | b4, o1, w2 |  "Zavolal vsetkych priatelov." |
| **2. Non-satisfying ending (ant-icadence)** | Yes–no question (zistovacia otazka) Question to myself (rozvazovacia otazka) | " ? " when it does not begin with an interrogative word | o2 |  Ty? |
| | Alternative question (offers a choice of answer) (rozlucovacia otazka) | " ? " and the word "alebo" | a2 (& o1) |  "Priznavas chybu alebo ju popieras?" |
| **3. Non-satisfying non-ending (semi-cadence)** | Rising (stupava) – Flat (rovna) – Raised (zdvihnuta) – Not-raised (nezdvihnuta) – Falling (klesava) | " . " | c4 (& b4) |  "Opravnena je aj otazka, co budeme robit." |

We found slight differences between expectations (sentence melodies described by language scientists) and melodies obtained from real speech recordings of the book [16]. The speaker often uses "non-satisfying non-ending" melody for sentences ended with the period, where "satisfying ending" melody was assumed. This attracts attention of the listener by signalling the next action of the story.

The investigation of clusters also showed strong influence of word melody accent even after the smoothing of the overall melody contour. So the melody of shorter speech segments (words, bars, syllables) should be studied.

## References


[1] ANDERBERG, M. R.: *Cluster Analysis for Applications,* Academic Press, New York, 1973.

[2] BOERSMA, P.: *Accurate Short-term Analysis of the Fundamental Frequency and the Harmonics-to-noise Ratio of a Sampled Sound*, Inst. of Phonetic Sciences 17, 97–110. Univ. of Amsterdam, 1993.

[3] CAKY P., KLIMO, M., MIHALIK, I., MLADSIK, R.: *Text-to-speech for Slovak Language,* In proc. of the 7th international conference Text, speech and dialogue, TSD 2004, Brno - Berlin: Springer, 2004.

[4] CERNANSKA, M., SKVAREK, O.: *Sentence Melody Analysis for Speech Production in the TTS system.* In proc. of the TRANSCOM 2009: 8th European conference of young research and scientific workers, University of Zilina, 2009, Section 3: Information and communication technologies, p. 25–30.

[5] DOWNS, G. M., BARNARD, J. M.: *Hierarchical and non-Hierarchical Clustering.* Poster presented at Daylight EUROMUG meeting. GlaxoWellcome, Stevenage, 1995. BCI Barnard Chemical Information Ltd. Sheffield S6 6BX.

[6] KOWALSKI, G.: *Information Retrieval Systems,* Theory and Implementation, Kluver Academic Publishers, 1999.

[7] KRAL, A.: *The Pronunciation Rules of Slovak (in Slovak)*, Systematika a ortoepicky slovnik, Neografia, Martin 2005.

[8] MELOUN, M., MILITKY, J.: *Statistical Analysis of Experimental Data (in Czech)*, ACADEMIA, 2004

[9] MURTAGH, F.: *Correspondence Analysis and Data Coding with R and Java,* Chapman and Hall/CRC Press, 2005.

[10] TTS system version 1 demo, http://tts.kis.fri.utc.sk (taken: August 23 2009)

[11] Program Praat main web page, http://www.fon.hum.uva.nl/praat (taken: August-23 2009.)

[12] The R Project for Statistical Computing, http://www.r-project.org/

[13] Electronic Textbook StatSoft, http://statsoft.com/textbook (taken: August -18 2009)

[14] XU, R., WUNSCH, D. *Clustering.* Wiley-IEEE Press, 2009.

[15] CERNANSKA, M., SKVAREK, O.: *Clustering Methods used to Obtain Typical Sentence Melody Contours for Slovak Language TTS system.* Accepted at the conference: The Second International Scientific Conference on Applied Natural Sciences, 2009, Trnava,

[16] KELEOVA, VASILKOVA, T. *Cukor a sol.* Ikar, Bratislava, 2004. (novel written in Slovak language)

Milan Majernik – Jana Chovancova *

# IMPLEMENTATION OF ENVIRONMENTAL MANAGEMENT SYSTEMS IN THE SLOVAK REPUBLIC

*The human environment is always influenced by enterprise activity, which is mostly negative. The introduction of an Environmental Management System (EMS) can be an important step towards fulfilling the principles of sustainable development on a practical level. However, there are few ways of EMS implementation. Authors of the paper present and analyze possible approaches of EMS implementation in Slovak enterprises. According to this analysis possible development trends in the environmental management can be indicated.*

*Key words: Environmental management systems, ISO standard, pro-environmentally oriented voluntary tools, enterprise.*

## 1. Introduction

Historically, companies have been confronted with various regulations regarding their emissions. These problems have always been closely related to living and working conditions at and around production facilities.

Prior to the 20th century, there was a long history of local regulations. Many cities prohibited dumping of waste in streets, canals or rivers. For example in the Netherlands, the first national law in this respect was the Factory law of 1975, which contained provisions regarding nuisances such as noise and odour [8]. However, simple answers were no longer satisfactory once it was no longer just waste that was criticized but also excessive resource consumption and the need of system approach became inevitable. Environmental management gradually emerged into a sophisticated environmental management system (EMS).

## 2. History of environmental management systems

EMS is a method designed to systematically improve the environmental performance of an organization. [4] It addresses the immediate and long term impacts of products, services and processes on the environment and is embedded in the organization's overall management structure. Results are achieved through the allocation of resources, assignment of responsibility and ongoing evaluation of practices, procedures and processes.

In the 1980s, more and more organizations systematically paid attention to environmental matters with the aim of reducing pollution and minimizing energy and raw material consumption. Guidance [7] was published to help companies setting up their own EMS. Environmental management systems were further developed in the early 1990s when ISO published its ISO 14001 standards leading to certification for an EMS.

In December 1996 the Technical Normalization Committee (TNC no. 72) entitled "Environmental management" was established by the Slovak Institute for Technical Normalization. In 1997 the first five standards of ISO 14000 were processed [6].

After final approval by the Office for Standards, Metrology and Testing, which took place in early 1998, these standards were included to the STN system. EN ISO 14 001 which specifies requirements for environmental management is the basic technical standard for the implementation and certification of an EMS. Another document which allows voluntary participation by industrial companies is Council regulation EEC No. 1836/93 of 29 June 1993 "Eco-management and Audit Scheme" (EMAS). This regulation encourages organizations to voluntarily evaluate the environmental impact of their activities from the processing of raw materials to waste management.

In April 2001 a new review by the European Parliament and Council was released (761/2001/EEC), allowing voluntary participation in the program for environmental management and audit scheme (EMAS II).

The revised regulation was implemented to Slovak legislation by Law No. 468/2002 Z. z. "Environmental Management and Audit", as amended by Law No. 491/2005.

Historical development of the environmental management systems can be seen in Fig. 1.

For the most important changes brought by this review can be considered:

- the broadening of the EMAS scope all sectors and economic activities, including local authorities

* **Milan Majernik[1], Jana Chovancova[2]**
[1] Technical university of Kosice, Faculty of mechanical engineering, Kosice, Slovakia, E-mail: milan.majernik@tuke.sk
[2] University of Presov in Presov, Faculty of management, Presov, Slovakia, E-mail: jchovancova@unipo.sk

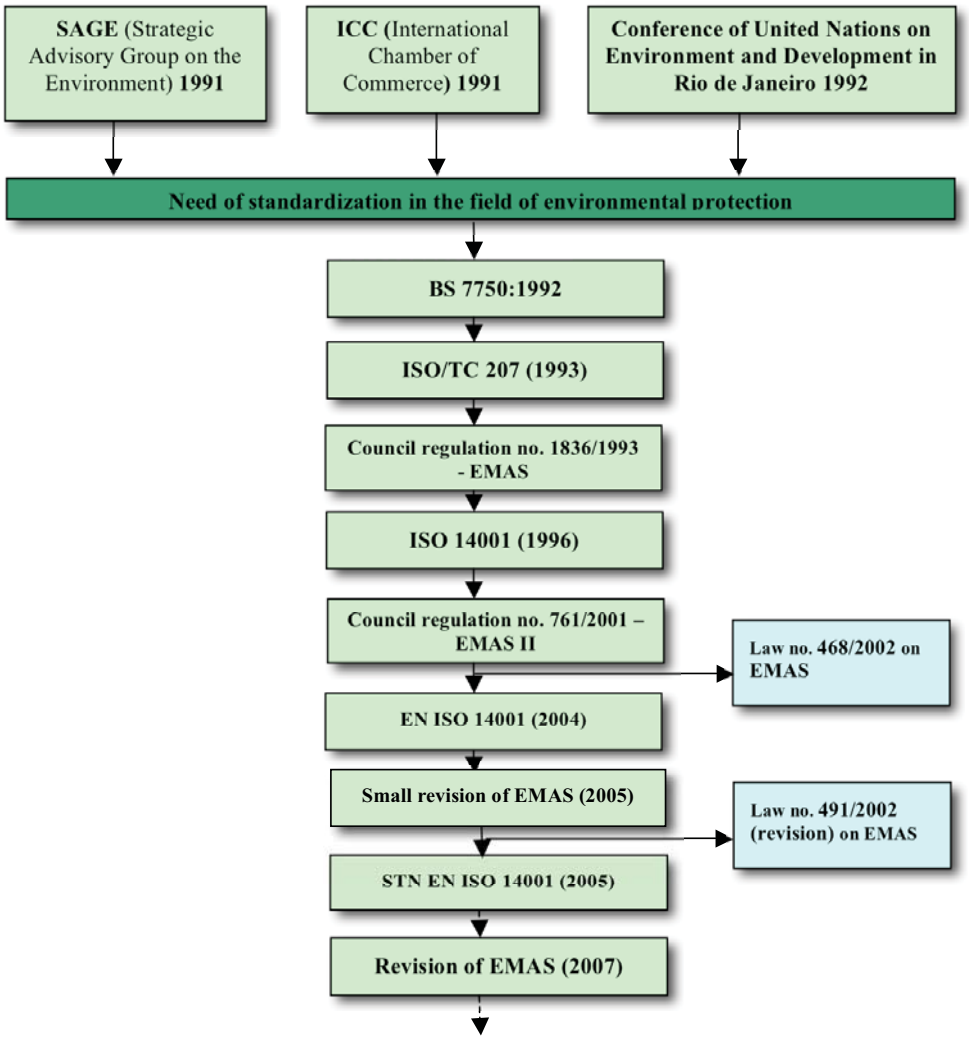



*Fig. 1 Development of environmental management standards and regulations*

- integration of standards such as ISO 14001, which facilitates the transition from ISO 14001 to EMAS and avoids duplication
- adoption of the EMAS logo, which allows registered organizations to promote their participation in the scheme more effectively
- employee participation in the implementation of EMAS strengthening the role of the environmental statement to improve the transparency of communication of environmental performance [1] between registered organizations and stakeholders and the public,
- more attention devoted to indirect environmental impacts such as capital investment, administrative decisions, planning, procurement [11].

## 2. State of Art in EMS Certification in Slovakia

The ISO 14001 Standard is the certification standard for EMS implementation and certification. As it is not an obligatory standard, it was designed in a way that can be applied to any size and type of organization and type of organization taking into consideration various geographical, cultural and social conditions. Efficiency of the implemented system is verified by certification auditing. Certification auditing results are used by certification organizations in the process of issuing official certificates.

The significance and relevance of the voluntarily adopted standards by traditional polluters is increasing rapidly in the SR. This is shown by growing numbers of established and certified environmental management systems according to ISO 14001. Up to January 2009, 669 enterprises in the Slovak Republic received ISO 14001 certificate.

Number of organizations with certified EMS according to ISO 14001 is illustrated in Fig. 2.

A number of sources identify organizations in the SR with a certified EMS according to international standard ISO 14001. The current list of organizations is published annually in the Envi-

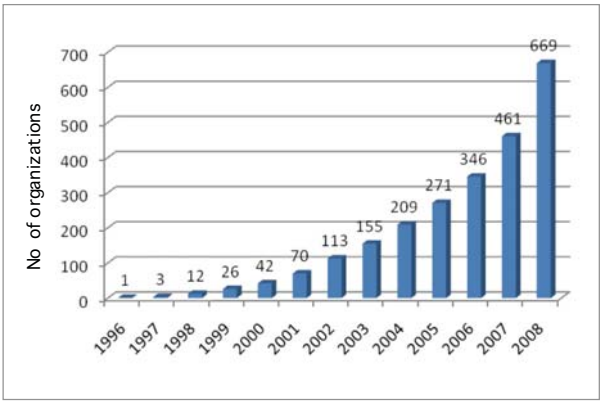



*Fig. 2 Number of certified EMS according to ISO standard 14001 in the SR [10]*

ronmental report of the Slovak Republic issued by the Ministry of Environment and the Slovak environmental agency (SAZP).

Fig. 3 shows EMS certifications in particular regions of the SR. The greatest number of companies with certified EMS is in Bratislava and the Kosice region. This situation is understandable because a disproportionately larger number of enterprises are in these industrialized areas where there is greater competition for customers and thus the likelihood that organizations will adopt EMS to gain competitive advantages.

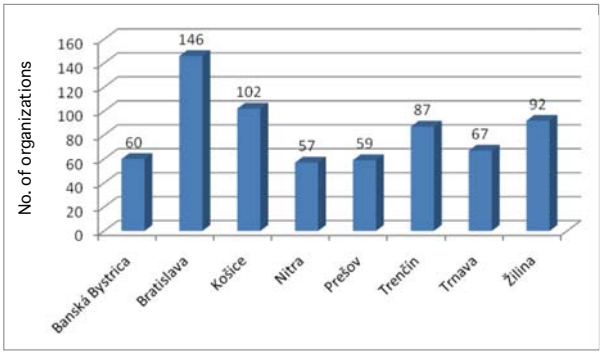



*Fig. 3 EMS certifications according to ISO standard 14001 in particular regions of the SR [10]*

Fig. 4 shows that the greatest share of all companies with certified EMS in Slovakia consists of large enterprises. Those are naturally the biggest financial, personnel and equipment options for the implementation of EMS [9].

## 3. EMAS Registration in Slovakia

EMAS is currently the most reliable and effective management tool on the market for organizations that want to improve their environmental performance. The scheme has been available for participation by companies since 1995 and was originally restricted

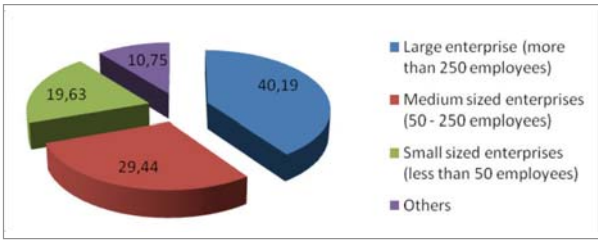



*Fig. 4 Percentage of EMS certified organizations by size*

to companies in industrial sectors. Since 2001 EMAS has been open to all economic sectors including public and private services.

In July 2008 the European Commission proposed to revise EMAS to increase the participation of companies and reduce the administrative burden and costs, particularly for small and medium sized enterprises (SMEs).

On November 25, 2009, the Council and the European Parliament adopted the revised EMAS Regulation. Publication in the Official Journal and entry into force of the Regulation is expected to happen at the end of 2009.

Revision in 2001 established ISO 14001 as a basis for EMAS implementation. Though EMAS goes beyond EN ISO 14001 in a number of ways, requiring the undertaking of an initial environmental review, the active involvement of employees in the implementation of EMAS, and the publication of relevant information to the public and other interested parties.

Notable differences include:

- Preliminary review: EMAS requires a verified initial environmental review – ISO does not.
- Public availability: EMAS requires that the policy, programme, environmental management system and details of the organisations performance are made publicly available as part of the environmental statement. ISO requires only that the policy be publicly available.
- Audits: EN ISO 14001 requires audits, although the frequency is not specified nor is the audit methodology set out in as much detail as in EMAS.
- Contractors and suppliers: EMAS is slightly more explicit in its control over contractors and suppliers, requiring that procurement issues are addressed and that the organisation endeavours to ensure that contractors and suppliers comply with the organisation's environmental policy. EN ISO14001 requires that relevant procedures are communicated to contractors and suppliers. In effect there should be no difference.
- Commitments and requirements: EN ISO14001 does not stipulate the extent to which performance must be improved. EMAS specifies that organisations must attempt to "reduce environmental impacts to levels not exceeding these corresponding to economically viable application of best available technology".

Up to November 2008 only 6 Slovak enterprises were registered in the EMAS scheme. The lack of interest in participation in the scheme is mostly due to the administrative complexity of the process. Also ISO 14001 has stronger position worldwive.

## 4. Solution for Small and Medium Sized Enterprises

Environmental management systems established according to ISO 14001 or EMAS are the best known, widely used voluntary tools, which provide a systematic approach to improving the environmental performance of an organization. But as showed above, these tools are mostly used in large enterprises – especially because of their financial, personal and administrative complexity.

But it is necessary also for small and medium sized enterprises (SMEs) to consider their impact on the environment. We are convinced of this due to the following reasons:

- New small and medium sized enterprises (SMEs) in Europe and Slovakia are the most common type of companies (with micro companies it is 99.8 %).
- SMEs provide more than 80 million jobs, out of total number of 122 million in the EU. In Slovakia SMEs has 70,8% share in overall employment.
  - SMEs are usually characterized by:
  - production of only one kind of product or service,
  - they are under great economic pressure,
  - they are limited to their local market,
  - established on family tradition,
  - low level of management,
  - limited possibilities of advertisement,
  - little possibilities of extended training and education
  - difficulties in information access.
- As a result of the heterogeneous focus of their activities, SMEs have different environmental problems and therefore require different types of solutions.

SMEs are active in very different actions, but they have some common features, for instance:

- they are a source of innovation,
- provide more jobs than large enterprises,
- they comprise a very heterogeneous group in terms of size and activity,
- they have considerable influence on environment. [3]

SMEs are also eminent polluters of environment; something they may not realize. In solving their environmental problems they have to get through a lot of difficulties related to lack of finances and know how [2], [5].

But there are a variety of voluntary pro-environmentally oriented tools for environmental performance improvement applied in different countries and branches which are less administratively, financially and personally demanding. The aim of their application is to improve the environmental performance of organizations, and to bring some other advantages in the form of risk elimination via reduction of accidents and breakdowns, environmental awareness improvement and increased economic effectiveness and cost saving in material and energy sources, etc.

These voluntary pro-environmentally oriented tools are more suitable for SMEs but they can also serve as a first step in EMS implementation according to ISO 14001 or EMAS or they can improve existing EMS.

The mostly used pro-environmentally oriented tools are:

- Environmental performance evaluation (EPE),
- Cleaner production (CP),
- Environmental benchmarking (EB),
- Environmental reporting (ER),
- Corporate social responsibility (CSR),
- Ecodesign (ED),
- Life-cycle assessment (LCA),
- Eco-labeling (EL),
- Environmental accounting (EA),
- Responsible care (RC) etc.

Through our research in collaboration with the Association of Industrial Ecology in Slovakia (ASPEK) we discovered that the most commonly used pro-environmentally oriented tools were: responsible care (10), environmental reporting (9), evaluation of environmental performance (8), corporate social responsibility (8) and cleaner production (7). Companies most frequently consider the introduction of environmental performance (6), environmental accounting (4), environmental reporting (4) and corporate social responsibility (4). Least preferred were ecodesign, environmental labeling, life cycle and extended producer responsibility, which may be due to specific focus of the tool, but also due to lack of awareness of other tools.

The above mentioned information is graphically illustrated in Fig. 5.

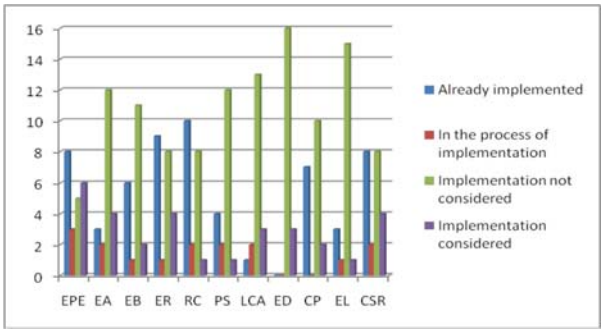



*Fig. 5 Pro-environmentally oriented tools most commonly used by Slovak enterprises*

## 5. Conclusion

It is difficult to give one specific answer to companies asking "what type of environmental management to choose" due to the factors mentioned above.

Our experience with certification audits of environmental management systems, however, revealed an interesting phenomenon. We found that companies were attempting to switch to European and world standards and certificate at any price, but often without the necessary environmental awareness of all enterprise workers.

The main feature of environmental management systems according to ISO 14001 as well as EMAS and pro-environmentally oriented voluntary tools is the management of environmental aspects, impacts and risks and the systems of continuous improvement of environmental performance.

Let us therefore build environmental performance in enterprises from the ground up – environmental awareness to environmental performance through the application of the system suitable for us (EMS, EMAS, pro environmentally oriented voluntary tools) and we will be able to manage – personally, economically and organizationally, the continuous improvement of environmental performance.

We are in the European Union so the focus should first be put on the EMAS. This, however, is problematic, as outlined, in Slovakia, where despite efforts on the legislation and methodology only 6 organizations/sites are registered.

It appears that the implementation of ISO prevails and Slovak enterprises will just focus in this direction, which in itself is no problem now and can serve as a basis for meeting the requests of EMAS in the future.

## References


[1] CLAUSEN, J., LOEW, T., KLAFFKE, K., RAUPACH, M., SCHOENHEIT, I.: *The INEM Sustainability Reporting Guide,* International Network for Environmental Management, 2001

[2] CHOVANCOVA, J., RUSKO, M.: *Spectrum of Voluntary Tools Used in Application of Environmental Policy,* In: CO-MAT-TECH 2007: 15th Intern. Scientific Conference: Quality assurance of products, safety of production and environment,Proc. of the contributions, Trnava, AlumniPress, 2007, p. 134–143. ISBN 978-80-8096-032-2.

[3] KOLLAR, V., BROKES, P.: *Environmental Management,* Sprint 2005. ISBN 80-89085-37-7

[4] LAIR, J., RISANEN, T., SARJA, A.: *Methods for Optimization and Decision Making in Lifetime Management of Structures,* Lifecon, 2004, Available on internet <http://life.vtt.fi>

[5] MAJERNIK, M. et al.: *Review of Voluntary Tools Focused on Environmental Performance Improvement,* Bratislava, ASPEK, 2007. 85, p. ISBN 978-80-88995-08-1.

[6] MAJERNIK, M., CHOVANCOVA, J.: *Environmental Management – Development and Trends, Environmental engineering and management,* Conf. proc. of 4th International conference EIaM, 2007, Herlany, Kosice, SjF TU, 2007, p. 22–29, ISBN 978-80-8073-894-5.

[7] MORRIS, Alan S.: *ISO 14000 Environmental Management Standards:* Engineering and Financial Aspects, John Wiley & Sons Ltd, 2004, ISBN 0-470-85128-7.

[8] MULDER, K.: *Sustainable Development for Engineers,* A handbook and Resource Guide, Delft university of Technology, The Netherlands, Greenleaf publishing, 2006, ISBN 978-1-874719-19-9

[9] RUSKO, M.: *Implementation of EMS in Slovak Industry,* In.: Experiences of enterprises using particular voluntary tools, Planeta, 2/2007, Ministry of Environment, Praha, ISSN 1801-6898

[10] *Slovak environmental agency,* Official Web page < http://www.sazp.sk>

[11] *Technical Workbook on Environmental Management Tools for Decision Analysis,* Available on internet: <www.unep.or.jp>.

Jana Chovancova – Milan Majernik – Jana Jurikova *

# INTEGRATED MANAGEMENT SYSTEMS

*Over the past few years an increasing number of companies certified according to the requirements of ISO 9001:2008 for Quality Management System (QMS) but also according to the requirements of other standards, for example environmental management system according to ISO 14001:2004 (EMS) and / or a management system for Occupational Health & Safety Management System, according to OHSAS 18001:2007, or other standards (ISO 22000, ISO 27000, etc.) can be seen in industrial sector and in services as well. Integration of management systems has become a key aspect of more sustainable management of a company. The aim of paper is to (1) better explain role of management systems implementation in enterprises, (2) point out necessity of their integration, (3) analyze common items of particular management systems, which is inevitable to their further integration, (4) point out some benefits and obstacles linked with integration.*

*Key words: Integration, management system, ISO standard, organization.*

## 1. Introduction

In the past, a clear trend of our organizations was the quality management system certification in accordance with the requirements of ISO 9001 and 9002:1994 and according to the requirements of ISO 9001/2000 [5]. Today, we are meeting the increasingly frequent cases of entry of foreign capital in the Slovak market or development of partnership relations within the EU as well as the U.S., Britain, Canada, and their requirements (e.g. automotive industry) establishment and implementation of management system of environmental management according to ISO 14001 and management system of occupational safety and health at work in terms of specification OHSAS 18001 or system HACCP for food safety management, or requirements for information safety etc. [8]

Many larger and smaller organizations with foreign capital, or with the business primarily in foreign markets have now implemented the Integrated Management System (IMS) in line with international requirements, which assist for managing their dominant market position at home and abroad and in the immediate extent manage their losses from the poor quality of products, from negative impacts on the environment, or threat of safety and health at work [3]. At present it is assumed that the company operating on the market can not produce high-quality, if it is not managing its losses and doesn't consider environmental protection and safety of its employees beyond the legislative requirements of our country [2].

Ensuring the success of the organization implementing the various management systems in today's difficult economic situation becomes more and more obvious. Among the best known and most frequently applied internationally recognized standards can be included:

- Quality Management System (QMS),
- Environmental Management System (EMS / EMAS),
- Occupational Health & Safety Management System (OHSAS) [4].

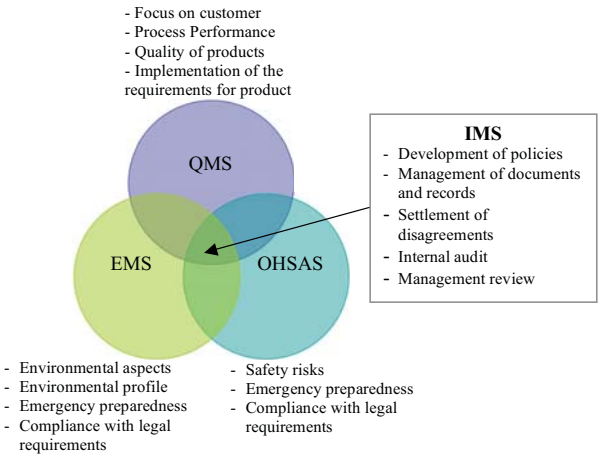



*Fig. 1 Integration of management systems*

## 2. Characteristic of the most Commonly used Management Systems

### Quality Management System

Quality of products and production affects the welfare, efficiency and competitiveness of the business entity. The aim of the implementing quality management system in company is to create a good relation of employees to quality as a fundamental category of the market economy and make them produce this quality. Quality management system determines the prior characteristics

* **Jana Chovancova[1], Milan Majernik[2], Jana Jurikova[2]**
[1] University of Presov in Presov, Faculty of Management, Presov, Slovakia, E-mail: jchovancova@unipo.sk
[2] Technical University of Kosice, Faculty of Mechanical Engineering, Kosice, Slovakia, E-mail: milan.majernik@tuke.sk

of the product, through which the company meets the identified needs of the market [2]. Setting the correct relationship between enterprise and market environment, is realized by using a quality management system.

Quality management system represents the organizational structure of enterprise, application of methods, tools and processes used for the implementation, maintenance and improvement of activities and their outcomes.

The current standard ISO 9001 monitors the company's processes and management system concerning customer requirements, product requirements and management of the risks related to quality of products and processes. Certification is a mark of maturity and credibility of a certain company and becomes a significant competitive advantage.

**Environmental Management System**

Environmental Management System according to standard ISO 14001 is a part of a complex management system that includes structure, planning activities, responsibilities, practices, procedures, processes and resources for the preparation, implementation, review and maintenance of environmental policy [4]. EMS is a part of the company's management system oriented to implementation of activities aiming to environmental protection. This means implementation (integration) of elements of environmental protection in decision-making processes and management of a company.

EMS is a management system related to environmental protection and is based on a series of ISO standards 14000. It is also the application of sustainable development specified in the 16 principles of Business charter adopted by the International Chamber of Commerce (ICC) for sustainable business development in 1991 and expanded on the principle of corporate social responsibility, which was adopted in 1997.

ISO 14000 series of standards represents a system of normative documents aimed mainly at:

- Environmental management systems with guide for documentation development.
- Audits as a form of verification of the status of environmental protection,
- Environmental performance evaluation of a company,
- Environmental labeling,
- Life-cycle assessment.

Though EMS implementation is not mandatory, concerning ever increasing legislative requirements it becomes inevitable especially for large enterprises.

**Health and Safety Management System**

Requirements of quality management system mostly in the area of working environment influence not only company's employees, but also public in close neighborhood. Therefore emphasis is put on meeting requirements of safety and health protection of employees regarding to new legislative requirements related to industrial accidents and protecting citizens living in the risk area [10].

The structure of OHSAS Standard is compatible with the essential elements of the legislation of most European countries and above mentioned standards ISO 14001 and ISO 9001 [11].

Similarly to the EMS the basis of the system is searching for threats and evaluation of resulting risks to employees, then underpin all possible risks and minimize their impact [8]. Today, regarding to our alignment with EU law, this principle is anchored in SR legislation.

Control system of work security is a vast system of organizational structures, procedures, processes and resources, which includes compliance with all legislation requirements [9]. Management of safe operations is part of top management of the company, and incorporates the implementation of these actions:

- Assign the same priority to the safety and health and environmental protection as economic aspects;
- Managing business the way which will increase the level of health protection of workers and environmental protection;
- Increasing awareness and responsibility of employees regarding to self health protection, with aim to motivate their cooperation in the field of work security.
- Integration of OHSAS requirements, as well as environmental protection to the designing and construction of all technological entities;
- Providing information to employees on safely managing activities within the system "man - machine – environment";
- Cooperation with national organizations, as well as with local authorities in preventing the occurrence of major accidents and to create conditions for health protection at work

## 3. Approaches to the Integration of Management Systems

Comparison and analysis of models and standards of management systems in terms of integration can be divided into three logical levels:

- Comparison or identifying common elements in all systems and their standards.
- Comparison or determine the common elements on the level of only two standards at the time.
- Comparison on the level of incompatibilities or specific elements of all standards.

Comparison or identifying common elements in all systems and their standards

It regards the determination of intersections of all three systems: quality ∩ environment ∩ safety.

About 80% of the elements of standards is common to all three aspects - quality, environment and security (see table 1).

ISO Guide 72:2001 Guidelines for the justification and development of management systems and standards arranges these common elements under the following main subjects:

- Policy,
- Planning,

Comparing the elements of safety, environment and quality in particular standards Table 1

| OHSAS 18001 | ISO 14001 | ISO 9001:2000 |
|---|---|---|
| **Health and Safety Management System** | **Environmental Management System** | **Quality Management System** |
| 4.1 General requirements | 4.1 General requirements | 4.1 General requirements |
| 4.2 OHSAS policy | 4.2 Environmental policy | 5.3 Policy of quality |
| 4.3 Planning | 4.3 Planning | 5.4 Planning |
| 4.3.1 Planning of identification of danger, risk assessment and risk management | 4.3.1 Environmental aspects | 5.2 Focus on customer |
| 4.3.2 Legal and other requirements | 4.3.2 Legal and other requirements | |
| 4.3.3 Targets | 4.3.3 Long-term and short- term targets | 5.4.1 Quality targets |
| 4.3.4 OHSAS management program | 4.3.4 Environmental management program | 5.4.2 Planning of QMS |
| 4.4 Implementation and operation | 4.4 Implementation and operation | 7. Realization of product |
| 4.4.1 Structure and responsibility | 4.4.1 Structure and responsibility | Responsibility of management<br>Resource management |
| 4.4.2 Training, awareness and competence | 4.4.2 Training, awareness and competence | 6.2.2 Competence, awareness and training of personnel |
| 4.4.3 Consulting and communication | 4.4.3 Communication | 5.5.3 Internal communication |
| 4.4.4 Documentation | 4.4.4 Documentation of EMS | 4.2 Documentation requirements |
| 4.4.5 Documentation and data management | 4.4.5 Documentation management | 4.2.3 Control of documents |
| 4.4.6 Operational management | 4.4.6 Operational management | 7.1 Planning of product realization<br>Processes related to customer |
| 4.4.7 Emergency preparedness and response | 4.4.7 Emergency preparedness and response | 7.1 Planning of product realization (partly)<br>8.3 Control of nonconforming product |

- Implementation and operation,
- Performance assessment,
- Improvement,
- Management review. [6], [9].

**Comparison or determination of the common elements on the level of only two standards at the time**

QUALITY ∩ ENVIRONMENT

Ecological friendliness of products and their production are increasingly assigned to one criteria of quality, which has to be considered by the organization, because of growing environmental awareness of customers. Therefore organizations today must take into account the society as a whole, not only their direct customers.

QUALITY ∩ SAFETY

An open relationship relies on the fact that the creation of the working environment as the role of health protection and safety at work has a direct impact on the quality of work and thus the quality of product, e.g. lighting conditions in the workplace etc.

ENVIRONMENT ∩ SAFETY

This intersection is particularly highlighted when the damage or potential hazards doesn't influence only employees in their workplace, but also to the surrounding environment, e.g. noise emissions, or hazardous substances in the air.

**Comparison on the level of incompatibilities or specific elements of all standards**

These elements may include the objectives of individual management systems, the complexity of their content, interest groups and the possible negative effects in the case of non regulation of these aspects.

## 4. Benefits of Management Systems Integration and Obstacles of Integration

Management system implemented according to ISO 9001, ISO 14001 and OHSAS 18001 is advantageous to implement and integrate into a single functioning management system, which will become a functional tool for managing and ensures the prevention of all the risks in the activities of the organization. This brings many advantages but also many barriers. Among the benefits of integration can be included:

- *Avoidance of duplication and conflicts:* The requirements of the various systems shall be arranged in particular activities of the organization and then common solutions are adopted.
- *The benefits of synergies:* Solutions that have already been proposed, may also be used for other systems.
- *A comprehensive overview of the organization:* the activities are analyzed and improved in the integration. If contradictory requirements are revealed, they are clarified and acceptable solutions are designed for all interested parties.

- *Optimizing cost management systems:* necessary structures are commonly used, for example common documentation in one manual for the integrated management system.

The main barriers to the integration of management systems can be divided as follows:

- *External reasons:*
- Lack of standards for integrated management systems,
- Different understanding of existing standards,
- Lack of tools (or their nonuse) in auditing/management systems evaluation.
- *Internal reasons:*
- A formal approach to implementing systems,
- Promote the interests of different groups (e. g. quality at the expense of environment or safety)
- Effort to satisfy certification/consulting companies (also leads to a formal approach to the implementation of IMS).

## 5. Integrated Management System Assessment

IMS assessment is not only a problem for companies establishing and using integrated management systems, but also for the certification companies. These problems arise from the absence of standards for integrated system, the requirements for qualification of auditors of particular systems, but also from approaches to auditing or evaluation.

Assessment of an integrated management system in terms of its functionality and completeness should be conducted so that the compliance with requirements of the particular management systems forming IMS should be assured. In this case it means to assess the compliance with quality standards according to ISO 9001:2000, environmental standards according to ISO 14001:2004 and safety standards OHSAS 18001. IMS assessment should be done by a project team through internal audit and its results should be provided to top management of organizations. In case of discrepancies, it is necessary to take corrective measures, implement them and re-assess the system. For the assessment of IMS, it is appropriate to plan and conduct assessment systematically. Helpful tool in planning and carrying out audits is standard ISO 19011, which establishes certain principles and assists in planning and conducting the internal audit [6].

When assessing, it is essential that the project team asks the competent questions to which there may be a direct answer either yes or no, and that are aimed precisely at individual elements of standards of the basic module of IMS. The project team should create a catalog of questions or check list with questions related to particular standards. Table 2 shows the passage of control list for the assessment of an integrated system of quality, environmental and safety management.

Another approach to the evaluation of IMS is the use of risk management, as a common denominator of all management systems.

Example of Control list — Table 2

| | Yes | No | Note |
|---|---|---|---|
| 5.3 Policy (quality + environment + safety) | | | Notes / Remarks: |
| Does top management ensure that policy of quality: | | | |
| 5.3-01 is compatible to purpose of the organization? | | | Unconformities: |
| 5.3-02 includes a commitment to meeting the requirements and to the constant improvement of the effectiveness of quality, environment and safety management system? | | | |
| 5.3-03 provides a framework for setting goals and evaluating the quality, environmental and safety? | | | Recommendations / warnings: |
| 5.3-04 was presented and understood within the organization? | | | |
| 5.3-04 was evaluated in terms of sustainability? | | | |
| HSE 4.2-01 top management of organization established and documented environmental policy and policy of health and safety? | | | |
| HSE 4.2-02 is policy appropriate in relation to the type, scope, environmental impacts and safety risks of operations, products and services of the organization? | | | |
| HSE 4.2-03 includes a commitment to continuous improvement safety and health of employees and pollution prevention of the environment? | | | |
| HSE 4.2-04 includes a commitment to complying with relevant environmental and safety laws and regulations and other requirements? | | | |
| HSE 4.2-05 forms a framework for identifying and evaluating environmental and safety goals as well es partial goals? | | | |
| HSE 4.2-06 is documented, implemented and maintained and is familiar to the staff? | | | |
| HSE 4.2-07 is accessible to public? | | | |

If the processes and their risks (e.g. risk to product quality, environmental risks and safety risks) are defined, each process can be monitored, measured and improved as a separate component measured by several performance indicators.

It is important to set parameters that are really important for management and for the operation.

An example of this approach is the **Risk Based Assessment** (RBA). This is a customer-oriented approach, where survey areas are determined on the basis of analysis of company management focused to the objectives, barriers to their achievement and their relation to the management areas (that covers all areas of management, including its relation to quality, environment, health and safety). This approach requires the involvement of management to the analysis and also to the results of an overall assessment.

RBA process consists of 5 steps:

- Identification and establishment of priorities of business objectives and barriers that may prevent their achievement.
- Definition of the importance of various areas of an ideal management system with respect to managing the obstacles defined in the first step.
- Evaluation of success of implementation in particular areas of the organization.
- Evaluation is scheduled in close cooperation with the senior leadership.
- Conduct the process of evaluation focused on several important areas under the resulting matrix of risks and report on any findings and areas for improvement.

## 6. Conclusion

Integrated management systems are now becoming an essential part of organizations that wish to manage their aspects effectively and efficiently and thereby increase their value on the market. Building an integrated management system, however, is left to the organizations themselves, which often lack the necessary information and support in the form of methodologies, guidelines and a variety of tools, which discourage many small and medium-sized organizations in the decision to implement IMS.

The IMS implementation in spite of many technical, economic and personnel problems, however, continues and is an important step on the way not only to the integration of enterprises in European and world trade area, but to the prosperity of business through more efficient management of risks associated with the quality of processes, environmental and safety risks.

## References

[1] BERANEK, Z.: *Experiences with Integrated Systems,* Planeta, 2/2007, Czech ministry of environment, Praha. ISSN 1801-6898
[2] DUFINEC, I.: *Virtual Process of Quality Management of Production in Integrated Management System of Company*, Intern. conference proc. QUALITY 2002, DT Ostrava, VSB-TU Ostrava, 2002, ISBN 80-02-01494
[3] MAJERNIK, M. et al.: *Conformity Assessment of Accreditation in the Field of Environmental,* Safety and Quality Management, Environmental management for education and edification, Vol. 2, 2/2005, p. 34–39, ISSN 1336-5762
[4] MAJERNIK, M., CHOVANCOVA, J.: *Environmental Management – Development and Trends, Environmental engineering and management,* Conference proc. of 4th international conference EIaM 2007, Herlany, Kosice, SjF TU, 2007, p. 22–29, ISBN 978-80-8073-894-5.
[5] MANKO, M., LIBERKOVA, L., HRICOVA, B.: *The Basic Access into the Integration Management Systems,* Engineering sciences 1. oral presentation: 5th Intern. conference of PhD students, University of Miskolc, Hungary, 2005, Miskolc University, 2005, p. 129–132, ISBN 963-661-673-6
[6] MATIAS, J. C. O., COELHO, D. A.: *The Integration of the Standards Systems of Quality Management,* Environmental Management and Occupational Health and Safety Management, Int. Journal of Production Research, Vol. 40, No.15, Taylor & Francis, 2004
[7] MORRIS, Alan S.: *ISO 14000 Environmental Management Standards: Engineering and Financial Aspects,* John Wiley & Sons Ltd., 2004, ISBN 0-470-85128-7.
[8] RUSKO, M., BALOG, K., TUREKOVA, I.: *Selected Chapters of Environmental and Safety Management,* Bratislava, VeV et Strix, Edicia EV-4, First Slovak edition, 2006, ISBN 80-969257-5-X,
[9] SCHWENDT, S., FUNCK, D.: *Integrierte Managementsysteme Konzepte,* Werkzeuge, Erfahrungen, Physica, Heidlberg 2002, ISBN 3-7908-1442-3
[10] SMITH, D.: *IMS. Implementing and Operating,* British Standard Institution, London 2002, ISBN 0 580 33328
[11] SMITH, D.: *IMS. The Framework,* British Standard Institution, London 2001, ISBN 0580332985.

## COMMUNICATIONS – Scientific Letters of the University of Zilina
## Writer's Guidelines

1. Submissions for publication must be unpublished and not be a multiple submission.
2. Manuscripts written **in English language** must include **abstract** also written in English. The submission should not exceed **10 pages** with figures and tables (format A4, Times Roman size 12). The **abstract** should not exceed 10 lines.
3. Submissions should be sent: **by e-mail** (as attachment in application MS WORD) to one of the following addresses: *komunikacie@uniza.sk* or *holesa@uniza.sk* or *vrablova@uniza.sk* or *polednak@fsi.uniza.sk* **with a hard copy** (to be assessed by the editorial board) **or on a CD** with a hard copy to the following address: Zilinska univerzita, OVaV, Univerzitná 1, SK-010 26 Zilina, Slovakia.
4. Abbreviations, which are not common, must be used in full when mentioned for the first time.
5. Figures, graphs and diagrams, if not processed by Microsoft WORD, must be sent in electronic form (as GIF, JPG, TIFF, BMP files) or drawn in contrast on white paper, one copy enclosed. Photographs for publication must be either contrastive or on a slide.
6. References are to be marked either in the text or as footnotes numbered respectively. Numbers must be in square brackets. The list of references should follow the paper (according to **ISO 690**).
7. The author's exact **mailing address of the organisation where the author works, full names, e-mail address or fax or telephone number,** must be enclosed.
8. The editorial board will assess the submission in its following session. In the case that the article is accepted for future volumes, the board submits the manuscript to the editors for review and language correction. After reviewing and incorporating the editor's remarks, the final draft (before printing) will be sent to authors for final review and adjustment.
9. The deadlines for submissions are as follows: September 30, December 31, March 31 and June 30.

COMMUNICATIONS

SCIENTIFIC LETTERS OF THE UNIVERSITY OF ZILINA
VOLUME 12



**Address of the editorial office:**
Zilinská univerzita
Office for Science and Research
(OVaV)
Univerzitna 1
SK 010 26 Zilina
Slovakia
E-mail: komunikacie@nic.uniza.sk,
polednak@fsi.uniza.sk

Each paper was reviewed by two reviewers.

Journal is excerpted in Compendex and Scopus

It is published by the University of Zilina in
EDIS – Publishing Institution of Zilina University
Registered No: EV 3672/09
ISSN 1335-4205

Published quarterly

Single issues of the journal can be found on:
http://www.uniza.sk/komunikacie