



Romas LEONAVIČIUS

MELISMA SYNTHESIS USING ARTIFICIAL NEURAL NETWORKS

Summary of Doctoral Dissertation

Technological Sciences, Electrical Engineering and Electronics (01T)

1330



Vilnius LEIDYKLA
TECHNIKA 2006

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Romas LEONAVIČIUS

**MELISMA SYNTHESIS USING
ARTIFICIAL NEURAL NETWORKS**

Summary of Doctoral Dissertation

Technological Sciences, Electrical Engineering and Electronics (01T)



Vilnius LEIDYKLA
TECHNIKA 2006

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2002–2006.

Scientific Supervisor:

Assoc Prof Dr Dalius NAVAKAUSKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronics Engineering – 01T)

The dissertation is being defended at the Council of Scientific Field of Electrical Engineering and Electronics at Vilnius Gediminas Technical University.

Chairman

Prof Dr Habil Romanas MARTAVIČIUS (Vilnius Gediminas Technical University, Technological Sciences, Electrical Engineering and Electronics – 01T)

Members:

Prof Dr Habil Julius SKUDUTIS (Vilnius Gediminas Technical University, Technological Sciences, Electrical Engineering and Electronics – 01T)

Prof Dr Habil Algimantas KAJACKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical Engineering and Electronics – 01T)

Dr Algimantas RUDŽIONIS (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07T)

Assoc Prof Dr Antanas Leonas LIPEIKA (Institute of Mathematics and Informatics, Sciences, Informatics – 09P)

Opponents:

Assoc Prof Dr Šarūnas PAULIKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical Electronics and Engineering – 01T)

Prof Dr Habil Romualdas APANAVIČIUS (Vytautas Magnus University, Humanities, Ethnology – 07H)

The dissertation will be defended at the public meeting of the Council of Scientific Field of Electrical Engineering and Electronics in the Senate Hall of Vilnius Gediminas Technical University at 10 a.m. on 21 December 2006.

Address: Saulėtekio al. 11, LT-10223 Vilnius-40, Lithuania.

Tel. +370 5 274 49 52, +370 5 274 49 56; fax +370 5 270 01 12;

e-mail doktor@adm.vtu.lt.

The summary of the doctoral dissertation was distributed on 21 November 2006.

A copy of the doctoral dissertation is available for review at the Library of Vilnius Gediminas Technical University (Saulėtekio av. 14, Vilnius, Lithuania).

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Romas LEONAVIČIUS

**MELIZMŲ SINTEZĖ
DIRBTINIŲ NEURONŲ TINKLAIS**

Daktaro disertacijos santrauka

Technologijos mokslai, elektros ir elektronikos inžinerija (01T)



Vilnius LEIDYKLA
TECHNIKA 2006

Disertacija rengta 2002–2006 metais Vilniaus Gedimino technikos universitete

Mokslinis vadovas

doc. dr. Dalius NAVAKAUSKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01).

Disertacija ginama Vilniaus Gedimino technikos universiteto Elektros ir elektronikos inžinerijos mokslo krypties taryboje:

Pirmininkas

prof. habil. dr. Romanas MARTAVIČIUS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Nariai:

prof. habil. dr. Julius SKUDUTIS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

prof. habil. dr. Algimantas KAJACKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

dr. Algimantas RUDŽIONIS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07T),

doc. dr. Antanas Leonas LIPEIKA (Matematikos ir informatikos institutas, fiziniai mokslai, informatika – 09P).

Oponentai:

doc. dr. Šarūnas PAULIKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

prof. habil. dr. Romualdas APANAVIČIUS (Vytauto Didžiojo universitetas, humanitariniai mokslai, etnologija – 07H).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties tarybos posėdyje 2006 m. gruodžio 21 d. 10 val. Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius-40, Lietuva.

Tel. +370 5 274 49 52, +370 5 274 49 56; faksas +370 5 270 01 12;

el. paštas doktor@adm.vtu.lt

Disertacijos santrauka išsiųsta 2006 m. lapkričio mėn. 20 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 11, Vilnius).

VG TU leidyklos „Technika“ 1330 mokslo literatūros knyga

NOTATION

Symbols

k	Time index.
m	Frame index.
g_s	Glottal signal extracted from song signal.
$\widehat{g_m}$	Glottal signal for melisma synthesis.
$f(\cdot)$	Neuron activation (in general nonlinear) function.
$s(k)$	Original signal value at time instance k .
$\hat{s}(k), \hat{s}_A(k), \hat{s}_{MLP}(k)$	Estimated signal in general or with Adaline, or with Multilayer Perceptron.
$s_m(k), s_s(k), s_v(k)$	Signal of melisma or song, or voiced part of song.
$w_{i,j}^{(L)}$	L -th layer neuron synapse weight value.
$\mathbf{a}$	Linear prediction coefficients.
$A(q), B(q), C(q), D(q), F(q)$	Polynomials of a general system model.
$I(m)$	Intensity of signal at frame m .
$\mathbf{H}$	Hessian matrix.
$\mathbf{J}$	Jacobian matrix.
N_w	Frame length for signal analysis.
N_Δ	Frame shift length for signal analysis.
$N_w^{(0)}$	Number of neurons in 0-th (input) layer.
$N_w^{F,G,M,T}$	Number of hidden neurons in the Fortis, Gruppert, Mordent and Trill models.
$T_0, \bar{I}$	Pitch of voiced signal.
$\bar{T}_0, \bar{I}$	Estimated (constant) values of the pitch and intensity.
$\widehat{T}_0, \widehat{I}$	Modelled pitch and intensity.
$\widehat{T}_0^{F,G,M,T}, \widehat{I}^{F,G,M,T}$	Modelled pitch and intensity of Fortis, Gruppert, Mordent and Trill.
$W_m^{\leftarrow}(k), W_m^{\rightarrow}(k)$	Weighting functions (from the right and from the left).
$\Delta^{(i)}$	Varying frame shift length for melisma analysis.
$\omega^{(i)}$	Varying frame length for melisma analysis.
$\hat{\theta}_k$	Estimated parameter values at time sample k .
$\Psi(\cdot)$	Generalized model of melisma.
$\Phi_T(\theta_T, k), \Phi_I(\theta_I, k)$	Pitch and intensity models of melisma.

Abbreviations

ANN	Artificial Neural Network.
ARX	Autoregressive model with external output.
LDS	Linear Dynamic System.
LP(C)	Linear Prediction (Coding).
MLP	Multilayer Perceptron.
MOS	Mean Opinion Score.
MSE	Mean Square Error.
NARX	Nonlinear Autoregressive model with external output.
NDS	Nonlinear Dynamic System.

INTRODUCTION

Topicality of the Research Work

Digital speech processing is already widespread and facilitates the communication among people both mutually and with intelligent systems in many fields. Though there is a number product of economic use, however, the majority of them are oriented at a monotonic speech synthesis. The classical methods of speech synthesis lack naturality, vitality and intonation.

Lithuanian ethnographic song records urgently require remastering. In order to apply during remastering signal restoration techniques, they must assure that no signal degradation will be introduced.

Melismas are inherited features of old songs and their preservation during remastering process needs special attention. When dealing with one of most complicated ornaments of songs – melismas – we seek here to elaborate qualitative speech synthesis aspects and to propose methods that will allow us to raise the quality of songs synthesis.

Statement of the Problem

Modern methods of speech synthesis take into account only considerable slow variations of pitch and locally use linear approximation techniques. Meanwhile the pitch in the song varies very quickly. Actually one of decorative ornaments of song – melisma – stands out by an abrupt change of the pitch and intensity.

Aspects of melisma synthesis were considered only in a few publications world wide. In Lithuania processing of melismas was considered two-fold.

Works of G. Raškinis concentrated in a transcription of melismas into music scores [16]. Procedures based on signal-extracted energy and fundamental frequency traces were developed. Application of Artificial Neural Networks for these purposes were not considered.

A pioneering work on the use of Artificial Neural Networks [7] for the restoration of melismas was done by D. Navakasas [11]. Special structures of Neural Networks called – Reduced Size Lattice-Ladder Multilayer Perceptrons – were developed and employed as predictors in order to synthesize melisma waveform. Vocal tract modelling or speech production mechanism were not considered here, speech waveforms were synthesized directly.

On the other hand in the speech synthesis area the method based on determination of Linear Prediction coefficients [10] is widespread. It decomposes speech into glottal signal and vocal tract models, thus rely on the speech production mechanism. In Lithuania, e.g., the use of vocal tract and residue signal Linear Prediction coefficients for speaker recognition was studied by A. Lipeika and colleagues [8].

By applying method of Linear Prediction and modelling speech characteristics one can attempt melisma synthesis. By the use of universal approximators – Artificial Neural Networks – one can model required melisma characteristics that are nonlinear or time varying. Thus, establishment of the model for melisma synthesis can be partly reduced to selection and training a special Artificial Neural Network that approximates melisma characteristics.

The aim of Work

The aim of this work is to synthesize melismas met in Lithuanian folk songs, by applying Artificial Neural Networks.

Tasks of the Work

1. After reviewing melismas and signal modelling methods, to select the basic features of melismas and to establish the possibilities of applying Artificial Neural Networks to synthesize melismas.
2. To propose the technique for synthesizing melismas after investigating the possibility to apply the Linear Prediction method.
3. To construct mathematical models of melismas suitable for the synthesis of the main kinds of melismas.
4. To compile a collection of melisma records ranging over different kinds of melismas and sung by several performers.
5. To test the developed models of melismas experimentally and to find the proper size of models as well as the values of their parameters.

Applied Methods

The methods used in this work are as follows: speech signal processing, approximation of nonlinear functions, search optimization as well as Artificial Neural Networks and digital simulation.

Scientific Novelty

- ♦ A new melisma synthesis technique allowing the application of generalized melisma model has been developed, based on bidirectional processing influenced by approximated characteristics of melismas.
- ♦ On the basis of the generalized melisma model and by the use of Artificial Neural Networks original mathematical models of all the four kinds of melismas have been created.
- ♦ After doing experiments with more than 500 melismas, the values of original melisma model parameters have been defined and suitability of applying these models for the synthesis of melismas of one performer have been confirmed.

Presented for Defence

- ♦ *The technique of synthesizing melismas* based on bidirectional processing applying: the pitch and intensity characteristics approximated by songs, an original fragment of the glottal signal and Linear Prediction Coefficients.
- ♦ *The generalized melisma model* grounded by two types of Artificial Neural Networks – a Multilayer Perceptron and Adaline – and the network learning algorithms: Levenberg-Marquardt and minimization of Least-Squares errors.
- ♦ *Original mathematical models of Fortis, Gruppett, Mordent and Trill*, their minimal size and the values of model parameters found during experiments.
- ♦ *The experimental results*, of synthesis of one performer melismas and that independent of a performer, using over 500 melisma records.

Links with Scientific Programmes

The author took part in researches pursued according to the international project “Nonlinear Dynamic Signal Processing” (2002–2004) supported by the Swedish Institute, as well as in two scientific works of Vilnius Gediminas Technical University: “Improvement of Nonlinear Digital Signal Processing Technologies” (2005–2006) and “Improvement of Digital Processing Technologies of Video and Audio Signals” (2002–2004).

Approval of the Work

The main results of the thesis were reported at the following scientific conferences:

- ♦ International Conference “Modern Information Technology”, 2004 and 2005, Minsk, Belarus.
- ♦ International Conference “Electronics”, 2004 and 2005, Vilnius.
- ♦ Young Scientist Conference “Lithuania without Science – Lithuania without Future”, 2001 and 2003, Vilnius.

Publication of Results

Five scientific papers have been published on the topic of this work: one paper in the prestigious national journal quoted in the international database Inspec, two papers in periodical reviewed journals published abroad, and two papers in the proceedings of national conference. Research results were used in the technical reports on three scientific works.

- [A1] Leonavičius, R.; Navakasas, D. Aspects of Melisma Synthesis. *Elektronika ir Elektrotechnika*, ISSN 1392-1215, 6(48), 2003, p. 18–21. Available from Internet: <<http://www.ktu.lt/lt/mokslas/zurnalai/elektr/z48>>.
- [A2] Leonavičius, R.; Navakasas, D. Improvement of the Restoration of Melisma by a Signal Synthesis. *Izvestija Beloruskoj Inzenernoj Akademii*, 1(19), No. 1, 2005, p. 110–113.
- [A3] Leonavičius, R.; Navakasas, D. Restoration of Melisma by Signal Synthesis. *Izvestija Beloruskoj Inzenernoj Akademii*, 1(17), No. 2, 2004, p. 64–67.
- [A4] Leonavičius, D. Review of Nonlinear Dynamic Systems Identification Methods. 6th Conference for Lithuanian Junior Researchers in Lithuania – “Science - Lithuania’s Future”, 2003, p. 33–40. ISBN 9986-05-695-0 (In Lithuanian).
- [A5] Leonavičius, R.; Navakasas, D. Reseach of Melisma Synthesis. 4th Conference for Lithuanian Junior Researchers in Lithuania – “Science - Lithuania’s Future”, 2001, p. 62–71. ISBN 9986-05-473-7 (In Lithuanian).

The Scope of the Scientific Work

The dissertation is written in Lithuanian. The explanatory part takes up 102 pages, its essential text – 94 pages. The work contains 90 mathematical expressions, 38 figures, 6 tables, 2 algorithms, 1 example, cites 95 references.

The thesis consists of 6 chapters, the first is Introduction and the last one – Conclusions, and two Appendices – Tools for Melisma Modelling and Vocabulary of Terms. The index of main concepts as well as lists of used notations and abbreviations are presented here as well.

DISSERTATION THESIS

The Place of Melisma in the Modelling of Nonlinear and Dynamic Systems

The aim of this chapter is to present an analytical survey of rather a widespread literature. First let us classify and comprehensively discuss melismas; next we shall briefly but essentially discuss the theory of dynamic systems which will make the basis for studying melismas, and finally we shall present the relationship for modelling a melisma with nonlinear and dynamic systems.

Melisma and its Characteristics

Both the musicological and technological definitions are of utmost importance for further studies.

Melisma is a small rhythm valued melodic ornamental figure [12], [3]. In terms of signal theory, melisma is a non-stationary signal characterized by an abruptly changing amplitude and periodicity. Melismas are usually classified into:

- Fortis (a complicated articulated and vividly expressed tensions of vocal chords),
- Gruppets (produced by two auxiliary sounds: one higher and one lower than the principal note),
- Mordents (produced by rapidly varying higher or lower by a single interval note on the principle note),
- Trills (generated by a rapidly alternating principal note with another note a full tone or semitone above it).

There are quite a number of characteristics by which a speech signal can be described. Two main characteristics that define the signal of melisma are intensity (Fig 1(b)) and pitch (Fig 1(a)). The work relies on these characteristics because they distinguish melismas out of other speech signals.

The most suitable method to determine intensity $I(m)$ is the integration method. In order to calculate it, one should choose the frame N_w long and shift it over the whole

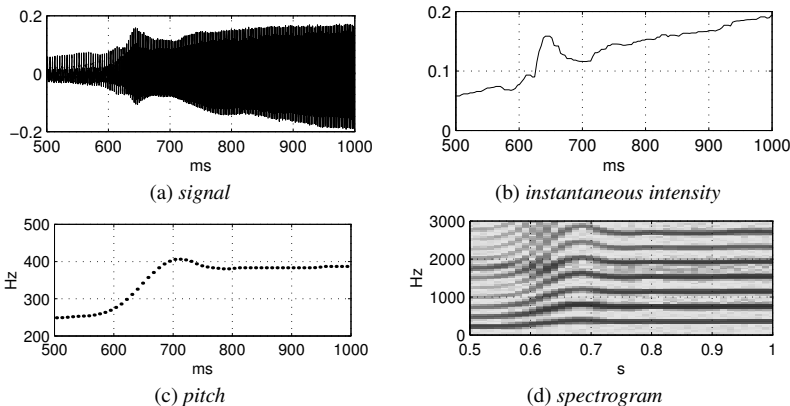


Fig 1. The main characteristics of Fortis

signal $s(k)$ to sum signal amplitudes in each m frame:

$$I(m) = \sum_{i=0}^{N_w} s(N_w \cdot m - i + 1). \quad (1)$$

The methods for calculating the pitch differ by the complexity of calculations used that in their turn stipulate the accuracy of pitch definitions. However, even the simplest methods enable us to obtain exact enough results. After trying several major methods [5, 17, 19], finally we have chosen a fast method based on autocorrelation calculations [5].

Speech signals are additionally described by the following derivative time characteristics: cepstrum, variation of MEL scale filter bank output signals, and variation of Linear Prediction Coding (LPC) coefficients. Variation of LPC coefficients is rather important for melismas, therefore let us discuss it in more detail. The LPC method is based on the assumption that the speech signal $s(k)$ can be approximated at the moment k by linear combination of N_m previous values of speech signals:

$$\hat{s}(k) = \sum_{i=1}^{N_m} -a_{i+1} \cdot \hat{s}(k - i). \quad (2)$$

Here a_i are i -th coefficients of the autoregressive process and to find them, the widespread autocorrelation method was applied [10].

In order to encompass melismas as widely as possible, we will investigate further the signals of all the four main kinds of melismas. We choose intensity and pitch as the essential characteristics of melismas and use a spectrogram to visualize them conveniently. These characteristics illustrate a rapid variation of melisma in time that can be modelled by dynamic systems. The characteristics also show the dependence of melisma on the pitch that can be modelled by a nonlinear system. So, let us view the theory of dynamic systems – the basis of future modelling of melismas.

Dynamic Systems, their Structures, and Identification

Identification of nonlinear dynamic systems (NDS) is usually decomposed into the following tasks [13]: a) selection of input signals; b) selection of excitation signals; c) choice of model architecture; d) choice of model dynamic; e) choice of model structure and complexity; f) selection of model order; g) calculation of model parameters; h) model verification.

For simplicity, at first let us consider a linear dynamic system (LDS) model [9]. Then the model can generally be written as follows:

$$A(q)s(k) = \frac{B(q)}{F(q)} g(k) + \frac{C(q)}{D(q)} v(k). \quad (3)$$

Here $s(k)$ is output of linear system at the moment k under the input signal $g(k)$ and noise $v(k)$, and $A(q)$, $B(q)$, $C(q)$, $D(q)$ and $F(q)$ are polynomials of required order, e.g., $B(q) = b_1 q^{-1} + \dots + b_{N_B} q^{-N_B}$. The generalized model itself is not directly used – the model is simplified dependent on the complexity of system under

consideration. Most frequently, the autoregressive model with external output (ARX):

$$A(q)s(k) = B(q)g(k) + v(k), \quad (4)$$

is employed in LDS identification. The merits of this model are truthful simulation of real processes and not complicated calculation of parameters, using the Least-Squares error correction methods. Recursion calculation methods are also applicable to this method.

Having written the latter formula by the N_m -th order difference equation, we obtain:

$$\hat{s}(k) = b_1 g(k-1) + \dots + b_{N_m} g(k-N_m) - a_1 \hat{s}(k-1) - \dots - a_{N_m} \hat{s}(k-N_m). \quad (5)$$

As mentioned above, this model is linear, while we need a nonlinear model for melisma modelling. So we can extend this expression into nonlinear ARX (NARX) by changing linear dependency (5) by an unknown function $F(\cdot)$ so, that:

$$\hat{s}(k) = F(g(k-1), \dots, g(k-N_m), \hat{s}(k-1), \dots, \hat{s}(k-N_m)). \quad (6)$$

Thus the task of searching for the parameters $\mathbf{b}$ and $\mathbf{a}$ of expression (5) is extended up to approximation of the unknown function F . This class of input-output models is called models with exterior dynamics. This method allows simulation of NDS by separating a nonlinear part of the model, then treating this problem as two but considerably simpler problems. Dependent on the place of the nonlinear function, NDS models are subdivided into Hammerstein models [1, 2], where nonlinearity preceeds LDS, and that of Wiener [15] – nonlinearity follows after LDS.

Identification means a parameter determination of a system of selected structure. To identify LDS, the methods of Least-Squares and Recursive Least-Squares have been used as well as a linear optimization with unequal constraints.

The Least-Squares method is the one most frequent by applied in the solution of linear problems. This fact is stipulated by the obtained parabolic Least-Squares function:

$$I(\boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{\theta}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{h}^T \boldsymbol{\theta} + h_0, \quad (7)$$

here $\boldsymbol{\theta}$ – denotes the parameters being optimized, $\mathbf{H}$ – quadratic term – Hessian (second derivation of loss the function), $\mathbf{h}$ – linear term and h_0 – constant term.

Solution of nonlinear problems is much more complicated, because there exist many local minima and there are no analytic solutions. The Newton [7] method is applicable to identify NDS. The matrix $\hat{R}_{k-1}$ of direction is found after evaluating the Hessian $\hat{H}_{k-1}$ inversion of the error function $\hat{\boldsymbol{\theta}}_{k-1}$:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} - \eta_{k-1} \hat{H}_{k-1}^{-1} \hat{\mathbf{G}}_{k-1}. \quad (8)$$

Since the Newton method is the most complex thereby, we use derivative one, that allow us to simplify calculations.

Generally speaking, we can state that modelling of NDS is a topical but in fact a complicated problem, one has to choose a proper NDS model, to identify it, and to verify the model obtained. Each step of NDS modelling is not single-valued, therefore

application of these systems is closely connected with nonlinear optimization theory. The modern way of NDS implementation is methods grounded by Artificial Neural Networks.

Relationship of Artificial Neural Networks with Nonlinear and Dynamic Systems

The Neural Network is an approximator of general purpose. As an example we can express the Linear Prediction model (2) by a single neuron – Adaline – equation. A combination of previous N_m speech samples will constraint the number $N_w^{(0)}$ of neuron inputs. Expressing the autoregression process coefficients a_{i+1} by the neuron synaptic weight $w_{i+1}^{(1)}$ and introducing the neuron activation function f , we obtain:

$$\hat{s}_A(k) = f \left(\sum_{i=1}^{N_w^{(0)}} -w_{i+1}^{(1)} \hat{s}_A(k-i) \right). \quad (9)$$

Here zero index at $N_w^{(0)}$ indicates neuron input layer conditionally denoted by zero, while the weights $w_{i+1}^{(1)}$ precede the first neuron, therefore the index is equal to one. Employed nonlinear activation functions enable us to use Adaline for implementation of simple nonlinear models.

Several neurons are rarely used, most frequently they are joined into neuron networks. Let us by Multilayer Perceptron [7] express the frequently used (6) NARX model:

$$\begin{aligned} \hat{s}_{MLP}(k) = \sum_{i=1}^{N_w^{(1)}} w_i^{(2)} f \left(w_{i,1}^{(1)} g(k-1) + \dots + w_{i,N_m}^{(1)} g(k-N_m) - \right. \\ \left. w_{i,N_m+1}^{(1)} \hat{s}_{MLP}(k-1) - \dots - w_{i,2N_m}^{(1)} \hat{s}_{MLP}(k-N_m) \right), \end{aligned} \quad (10)$$

here $w^{(1)}$, $w^{(2)}$ are weights of first and second layer neurons.

The ANN theory laid the foundation for rather formalized and convenient realization of complex systems including dynamic and nonlinear systems of general purpose. Actually any nonlinearity of a system can be realized by the ANN approach. There exist quite a few specialized ANN structures that are closely connected with the NDS theory and can be implemented using not NDS identification, but ANN training algorithms.

With a view to develop models fit for a melisma, we will explore specialized structures of Artificial Neural Networks, and in order to establish the parameters of these structures, we will seek an appropriate way to train ANN.

Substantiation and Composition of Melisma Synthesis Technique

With a view to form an original methodology, in this chapter we investigate the most widely used Linear Prediction Coding method (LPC) and possibilities of its improvement. Later on, when exploring possibility to supplement the LPC method with the functions of glottal signal and pitch modelling, we propose the modification of original Linear Prediction method based on dynamic LPC frame positioning. We present the new melisma synthesis technique, showing and grounding the use of melisma model.

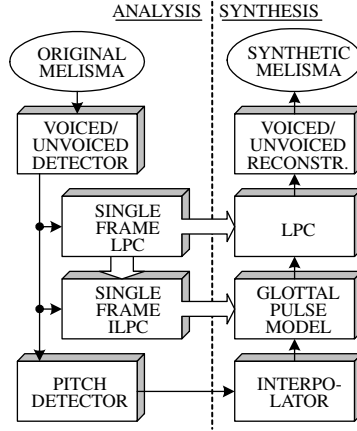


Fig 2. Synthesis using a single period of the original glottal signal

Investigation of Linear Prediction Method

Based on the traditional speech signal simulation method, it is possible to synthesize the melisma signal making use of the glottal signal and the vocal tract model. In the experiment we use LPC in the cause of which we define the coefficients of the N_m -th order linear autoregression process (1) as well as intensity.

The fitness of the LPC method for melisma synthesis was verified by an experiment during which the melisma was decomposed into glottal signal and vocal tract filter. First of all we find the coefficients by means of LPC, and during an inverse LPC process we establish the glottal signal. Using the LPC coefficients and the glottal signal, we restore the melisma during the synthesis. The complete analysis was made with 30 ms fragments by shifting the frame every 10 ms. The experiments with all the types of melismas have corroborated the suitability of LPC for the melisma analysis. However, redundancy of information is typical for this method.

In order to diminish redundant information, modern speech synthesis employ an artificial glottal signal. To this end, the glottal signal model in [4] is often applied:

$$\hat{g}(t) = \begin{cases} \frac{1}{2} \left[1 - \cos \left(\frac{\pi k}{P} \right) \right], & 0 \leq t \leq P; \\ \cos \left(\frac{\pi(k-P)}{2(K-P)} \right), & P \geq t \geq K; \\ 0, & t > K, \end{cases} \quad (11)$$

here P denotes the rate of glottal signal ascent, and K denotes descent.

By applying the latter expression during the synthesis experiments, an artificial glottal signal was generated, based on the pitch variation law (Fig 2). Using an unaltered vocal tract model, a melisma has been synthesized. Unfortunately, melismas obtained in this way were of low quality – high frequency components were missing. Though the moment amplitude of synthesized signal was rendered not so bad, however, the obtained sound was voiceless and mechanical [A1]. This is due the fact that glottal signal conveys

namely those speech properties pertain to that make it human-like. This leads to a conclusion that an artificial glottal signal is not good for synthesizing melismas.

The experimental results [A3] show that the glottal signal influences the quality of the sound restored a great deal. Even using a segment of the original glottal signal the timbre is imparted in melismas synthesized, but some signal vibrations is heard. It is caused due to discrepancy of phases. There are lot of methods to reduce this influence, e.g., by applying additional filtering or introducing 20 dB noise [6, 14, 18], however, the experimental tests have show that these methods cannot afford to reach the desired results with melismas.

Modification of the Linear Prediction Method

In the classical LPC method, a speech signal is decomposed into fragments each of which is N_w sample long (usually about 30–45 ms). Segments of the signal are shifted over N_Δ samples respect to one another (often $N_w/3$, i.e. about 10 ms). The modification of the LPC method proposed is based on the following two alterations:

$$N_\Delta \rightarrow \Delta^{(i)} \neq \text{const}; \quad N_w \rightarrow \omega^{(i)} \neq \text{const}. \quad (12)$$

Let us discuss that more in detail. Firs of all, let us write an expression of the voiced signal $s_v(k)$ of an idealized constant pitch as finite sum of T_0 duration signals $s(k)$:

$$s_v(k) = \sum_{m=1}^M s(k - mT_0). \quad (13)$$

With a view of expressing the melisma signal in a similar way, we have to evaluate the variation of the pitch T_0 and the signal amplitude $s(k)$ with the variation of the pitch index m :

$$s_m(k) = \sum_{m=1}^M s_m \left(k - \sum_{i=0}^{m-1} T_0(i) \right). \quad (14)$$

The latter expression describes the melisma signal at different periods of the pitch. We can now define a varying frame length ω of melisma analysis and a shift Δ of the analysis frame as follows:

$$\omega^{(i)} = \sum_{k=(i-1)\delta^i}^{(i-1)\delta^i + M - 1} T_0(k), \quad (15a)$$

$$\Delta^{(i)} = \sum_{k=0}^{(i-1)\delta^i} T_0(k). \quad (15b)$$

Modification of the LPC method allow to analyze every frame in the as same as possible conditions (Fig 3). This technique increase the efficiency of classical LPC calculation applied for non-stationary voiced signals. Experimental results [A2] confirmed feasibility of improved method for restoration of melisma.

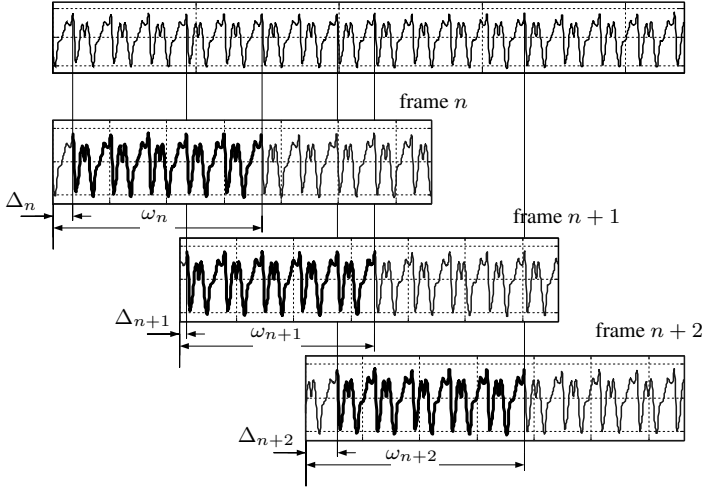


Fig 3. Illustration of modification of the LPC method

Composition of Melisma Synthesis Technique

In order to calculate parameters, a modified LPC method is employed, and the melisma is synthesized by filtering the modelled glottal signal through the LPC parameter filter. The glottal signal is known at the beginning of melisma and at the end. With a view to obtain a glottal signal of melisma, it is modelled. It is more convenient to express the melisma synthesis technique as Algorithm 1.

Technique of melisma synthesis is based on the glottal signal modelling performed by changing the glottal signal of song without a melisma according to the law of pitch and intensity variation. One period of the glottal signal is used which is resampled in accordance with the pitch variation and is amplified taking into consideration the melisma intensity. The aim of glottal signal synthesis is to obtain a continuous glottal signal of melisma, and pass it to the vocal tract model to obtain a synthesized melisma. Since the glottal signal of melisma alternates depending on the pitch value, we have used restoration in two directions which allowed us to reach a better quality of the melisma glottal signal and of the synthesized melisma thereby.

Algorithm 1 (Melisma Synthesis Technique)

A. Initial Data

- Song signal, $s_s(k)$;
- Melisma synthesis limits: start – n_s , finish – n_f ;
- Melisma model, Ψ .

B. Initialization

- Establishing of the pitch period and intensity: $\bar{T}_0(k)$, $\bar{I}(k)$;
- Calculation of Linear Prediction Coefficients, $\mathbf{a}(k)$.

C. Glottal Signal Synthesis using Forward and Backward Processing

- Calculation of the glottal signal in a segment without a melisma: before the melisma – $g_s^{\leftarrow}(k)$, and after the melisma – $g_s^{\rightarrow}(k)$;
- Application of the melisma model in restoring the glottal signal of melisma in forward direction:

$$\widehat{g_m^{\leftarrow}}(k) = \Psi(g_s^{\leftarrow}(k), \widehat{T_0}(k), \widehat{I}(k), \theta);$$

- Application of the melisma model in restoring the glottal signal of melisma in backward direction:

$$\widehat{g_m^{\rightarrow}}(k) = \Psi(g_s^{\rightarrow}(k), \widehat{T_0}(k), \widehat{I}(k), \theta);$$

- Computation of glottal signal of melisma by forward and backward results:

$$\widehat{g_m}(k) = \widehat{g_m^{\leftarrow}}(k)W_m^{\leftarrow}(k) + \widehat{g_m^{\rightarrow}}(k)W_m^{\rightarrow}(k).$$

D. Synthesis of Melisma Signal

$$\widehat{s_m}(k) = \frac{1}{\alpha(k)} * \widehat{g_m}(k).$$

Building a Melisma Models

The composed melisma synthesis technique makes it possible to apply any melisma model. Thus, its usability can be evaluated only in the case of application of particular model. The goal of this chapter is to present created original melisma models to be used later in experimental validation.

Development of a Generalized Melisma Model

Our earlier conclusion was that a melisma can be simulated by introducing the glottal signal modelling. Based on Algorithm 1, disregarding the processing direction, the glottal signal model is follows:

$$\widehat{g_m}(k) = \Psi(g_d(k), \widehat{T_0}(k), \widehat{I}(k)), \quad (16)$$

here $\widehat{T_0}(k)$ – pitch of melisma signal, $\widehat{I}(k)$ – intensity of melisma, $\widehat{g_d}(k)$ – glottal signal without melisma.

The glottal signal of song is known – it is isolated from a song without a melisma using LPC, after determining the vocal tract parameters of a song and applying direct filtering. The pitch is also known: both while restoring melisma and synthesizing it we know which type of melisma we are working with. The same is valid for intensity of melisma. Remains only to determine the way how to modify the pitch of a song at the time moment k to obtain the glottal signal of melisma.

The glottal signal of melisma depends on the pitch and intensity. The values of these two characteristics for a particular melisma are known, however information refers only to discrete samples, while we need to have a continuous variation of these characteristics. Also the pitch and intensity are varying in the melismas of the same performer. Since these characteristics of a particular melisma under synthesis are unknown, we only know what kind of melisma it is and the values of characteristics at the beginning and the end of it, it would be most reasonable to use the models of the pitch and the intensity to get generalized values of these characteristics.

The pitch and intensity of melisma are simulated accordingly to:

$$\widehat{T}_0(k) = \bar{T}_0 + \Phi_T(\theta_T, k); \quad (17a)$$

$$\widehat{I}(k) = \bar{I} + \Phi_I(\theta_I, k), \quad (17b)$$

here $\bar{T}_0$ and $\bar{I}$ are the pitch and intensity values of a song, θ_T and θ_I are parameters of the pitch and intensity models correspondingly; k is time. Having the pitch and intensity variation models, we shall be able to find the functional Ψ – solution how the glottal signal of the song without a melisma should change in order to obtain a glottal signal of melisma.

To build the pitch and intensity models we apply Artificial Neural Networks. Taking into consideration the variation of the characteristics being modelled, most suitable for the model there would be two layer Perceptron [7], the model of which is expressed as:

$$\Phi(\theta, k) = \sum_{i=1}^{N_w^{(1)}} \theta_i^{(2)} f(\theta_i^{(1)} k), \quad (18)$$

here $\theta = [\theta^{(1)}, \theta^{(2)}]$. For modelling of simpler functions, we propose to use Adaline:

$$\Phi(\theta, k) = \theta^{(2)} f(\theta^{(1)} k). \quad (19)$$

So, the glottal signal is chosen as the basis of the melisma model. In order to describe it, we suggest the use of the estimates of the pitch and intensity characteristics of melisma and one period of the glottal signal of the song. A generalized melisma model is proposed and several types of Artificial Neural Networks for implementation are considered. The generalized model of melisma in future will make it possible to build simple models of small size for a particular kind of melisma.

Identifying of Melisma Models

For training ANN we have selected the Levenberg-Marquardt [13] optimization method:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \eta_{k-1} \left(\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1} + \alpha_{k-1} \hat{\mathbf{I}} \right)^{-1} \hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{f}}_{k-1}. \quad (20)$$

There, just like in the Gauss-Newton method, we apply indirect inversion calculation to determine the vector:

$$\left(\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1} + \alpha_{k-1} \hat{\mathbf{I}} \right) \hat{\mathbf{p}}_{k-1} = \hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{f}}_{k-1}. \quad (21)$$

Additional $\alpha_{k-1} \hat{\mathbf{I}}$ member to the Hessian approximation $\hat{\mathbf{J}}_{k-1}^T \hat{\mathbf{J}}_{k-1}$ is nothing else then regularization applied to solving linear optimization problem by the Least-Squares method.

If α_{k-1} is small, the Levenberg-Marquardt algorithm approaches the Gauss-Newton method. And if α_{k-1} is large it approaches the fastest gradient descent method that ensures approaching towards a minimum which is typical of Gauss-Newton method. As usual, when using the Levenberg-Marquardt optimization algorithm firstly the initial positive value of α_{k-1} is set. Further, at each iteration α_{k-1} is decreased by a certain

step size. Thus, we approach the optimal parameter α_{k-1} value at which the Gauss-Newton method starts acting. However, if the decrease of α_{k-1} causes an increase of error (optimization is in the wrong direction), then α_{k-1} is increased again until a decrease of error is reached.

Composing the Models of Separate Kinds of Melisma

The pitch of Fortis modelled by Adaline is expressed as follows:

$$\widehat{T}_0^F(k) = \bar{T}_0^F + \theta_T^{(2)} f\left(\theta_T^{(1)} k\right). \quad (22)$$

Here $\bar{T}_0^F$ – pitch estimate for a song; $[\theta_1; \theta_2] = \theta_T$ – weights of adaline; $f(\cdot)$ – sigmoidal neuron activation function; k is the time index in the interval $[k_1, k_2]$.

Meanwhile the intensity of Fortis modelled by Multilayer Perceptron, is expressed as:

$$\widehat{I}^{F,G,M,T}(k) = \bar{I}^{F,G,M,T} + \sum_{j=1}^{N_w^{F,G,M,T}} \theta_{I_j}^{(2)} f\left(\theta_{I_j}^{(1)} k\right), \quad (23)$$

here $\bar{I}^{F,G,M,T}$ – intensity estimated for Fortis, Gruppert, Mordent and Trill, respectively; $[\theta_{I1}^{(1)}, \dots, \theta_{I1N}^{(1)}; \theta_{I1}^{(2)}, \dots, \theta_{I1N}^{(2)}] = \theta_I$ – weights of multilayer perceptron; $f(\cdot)$ – sigmoidal neuron activation function; k is the time index in the interval $[k_1, k_2]$; $N_w^{F,G,M,T}$ – number of neurons in the model.

The Gruppert, Mordent and Trill pitch of melisma modelled by a multilayer perceptron will be expressed as follows:

$$\widehat{T}_0^{G,M,T}(k) = \bar{T}_0^{G,M,T} + \sum_{j=1}^{N_w^{G,M,T}} \theta_{T_j}^{(2)} f\left(\theta_{T_j}^{(1)} k\right). \quad (24)$$

We have got minimal model sizes of all the four melisma kinds (see Table 1) where Adaline is marked as **A** and Multilayer Perceptron – as **MLP**). The Fortis model alone based on Adaline network for pitch modelling and Multilayer Perceptrons for intensity, while the models of other melismas – Gruppert, Mordent and Trill – are based on Multilayer Perceptrons.

Table 1. Initial description of melisma models

Melisma	Pitch model			Intensity model		
	Network	min N	min Θ	Network	min N	min Θ
Fortis	A	1	2	MLP	4	11
Gruppert	MLP	7	17	MLP	5	13
Mordent	MLP	3	9	MLP	6	15
Trill	MLP	3	9	MLP	6	15

Experimental Evaluation of Melisma Models in the Synthesis

In this chapter, collected database of melismas is discussed. Using the proposed melisma synthesis technique and created melisma models, melisma synthesis experiments with the sets of one performer and independent of performers are commented.

Table 2. *Description of the created sets of melismas*

Kind of Melisma	Type of Melisma	Count of melisma in set				Total
		V1	V2	M1	M2	
Fortis	F01	35	7	20	8	186
	F02	40	–	–	6	
	F06	35	–	35	–	
Gruppett	G01	35	8	40	10	139
	G02	17	9	20	–	
Mordent	M01	29	8	45	10	147
	M02	21	–	25	9	
Trill	T02	25	–	25	–	90
	T04	40	–	–	–	
Total		277	32	210	43	562

Formation of Melisma Data Sets

Four performers were asked to compile a set of melismas: two of them are men and two women. All the melismas were sung by people with musical education, two of them (one man and one women) are vocalists.

The melismas collected are of all four kinds (see Table 2): Fortis, Gruppett, Mordent and Trill. The set contains the most frequent types of melismas, but we tried to collect at least several types of each kind. With the view of simpler handling of melismas, the set was partitioned according to performers: a man vocalist V1, a women vocalist M1, non vocalist man V2 and non vocalist women M2. In total over 500 melismas have been collected, most of them were sung by vocalist.

Synthesis of Melismas of One Performer

In the synthesis of one performer melismas the characteristics (pitch and intensity) of the same performer's melismas and segment of the glottal signal obtained from performer in a signal without a melisma have been used. The synthesis of melismas consists of two parts: modelling of melisma characteristics and synthesis of melismas using the melisma characteristics and segment of glottal signal.

Experiments have been done with melismas of all the types of performers, however, for conciseness, the results are presented only of two of them – a man vocalist and women vocalist. The melismas of vocalists were chosen because there were at least several melismas of the same kind and type. The larger number of melismas will allow us to obtain a more generalized model and computations will more exactly reflect the general tendency of variations.

Two Artificial Neural Networks have been created during experiments: one was meant for modelling the pitch, the other – for intensity. In all the modelling experiments set of melismas was decomposed into parts: 50 % used to training, 30 % – for validation and remaining 20 % for testing. ANN has information on training and validation sequences, while the testing sequence is unknown. The network is trained by passing all melismas of training set, and error check was performed according to validation sequence. The testing sequence was used to calculate the error estimate. The Mean Square Error (MSE) method was chosen as training error criterion.

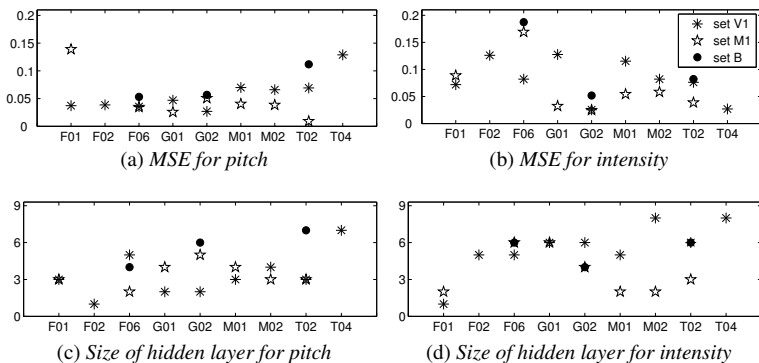


Fig 4. Pitch and intensity modelling results

The theoretical pitch model of Adaline for Fortis was rejected during experiments, because in reality the pitch variation of Fortis more complicated than in theory. Everywhere a Multilayer Perceptron with one input, one output, and a hidden neuron layer between them was used. Since the goal of experiment is to find the smallest network that would meet the model requirements, the experiment was carried out starting from the simplest network with one neuron in hidden layer and each time increasing the number of hidden neurons up to 12. Ten experiments have been done with each of MLP. Thus, for for each type of melisma, 120 experiments are done. From the results obtained, the best model is selected based on MSE and later is used in the synthesis of melisma.

As the criterion of good modelling we have experimentally selected MSE to not exceed 0.1, which makes up 10 % of the maximum value of the modelled signal. We can see from modelling results (Fig 4(a)) that in many cases the selected ANN modelled the pitch: for Fortis and Gruppert models are simplest, while for Mordent and Gruppert are more complicated. Due to great disagreement of signal we have got poor model for Fortis F01 of set M1 and Trill T04 of set M1. Meanwhile intensity modelling is much more complicated mainly because of characteristics disagreement: vocalists are able to keep well the pitch variation of melisma, however it much more difficult to keep intensity in different melisma of the same set.

The best melisma models have been used in synthesis. A segment of the original glottal signal of melisma is extracted out from the beginning and the end of melisma. The segment of melisma glottal signal is resampled according to the pitch model values and amplified according to the intensity model. The vocal tract model parameters are computed by means of a modified LPC from the beginning of melisma. A melisma is synthesized from the synthesized glottal signal and vocal tract model parameters. The synthesis is performed in two directions. Only the melismas of the testing sequence have been synthesized.

For qualitative evaluation of melisma synthesis we need criteria that allow to estimate the quality of the signal synthesized. To this end, acoustic estimation of the melisma synthesized has been used applying the mean opinion score (MOS) method. Since we are aimed at achieving a high quality synthesis, the MOS criteria was adapted taking into consideration the specificity of synthesis (see Table 3).

Table 3. MOS scale for melismas

Point	Description
5	The synthesized melisma does not differ from original one
4	The difference is weekly audible at places
3	An intervals the difference is distinctly audible
2	In major part the difference is distinctly audible
1	The synthesized melisma distinctly differs from original one

The synthesis of melisma sets according to the MOS criteria were evaluated by ten listeners. Depending on the size of the melisma set, from 3 to 9 synthesized melismas were presented for their evaluation. The results of one melisma synthesis are averaged according to the arithmetical mean. Later on the results of all the estimators are averaged.

The MOS estimate of melisma range from 3.7 to 4.4 points (see Fig 5). The synthesis of a women's melisma is generally of lower quality, because the variation of a female glottal signal is considerably greater. Through relatively the pitch varies to the same extent, in absolute value however the variations is 2–2.5 times greater. Therefore, the general mean MOS estimate of set V1 is 4.1 points, while of set M1 is only 3.9 points. It is also been observed, that deviation of the model intensity has a little influence on the acoustic quality of the melisma synthesis.

Synthesis of Melismas Independent of a Performer

Synthesis of melisma independent of a performer are analogous to that of one performers. Here, sets V1 and M1 were joined into one common set B1. The common set is used to compose a melisma model. The experiment was carried out with three types of melisma: F06, G02 and T02.

The results presented in Fig 4 show that the models obtained for Fortis and Gruppett are good, while the Trill model is poor. Decrease in quality of the Trill model was caused by different singing of a man's and woman's Trill – variation of the pitch of melisma in set V1 is at the end of the signal, while in M1 – on the beginning.

The quality of synthesis of melismas independent of a performer determined by MOS criterion is low. In this case the estimate of mean is 3.6 points. On the other hand a fall of quality was not so great. It was caused by the selecting synthesis method – a segment of the glottal signal of the synthesized melisma was used here.

Having done experiments the following conclusions have been drawn: a) Melisma modelling disregarding a performer is possible only if use the glottal signal of the same or a similar performer; b) Discrepancy of the signal model in amplitude does not effect the MOS estimate considerably.

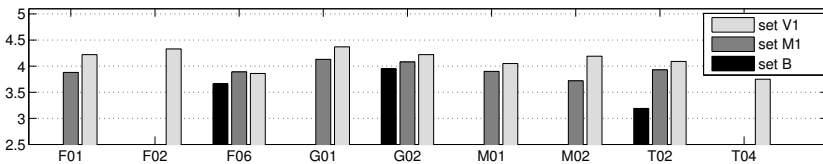


Fig 5. Averaged MOS values for melisma synthesis

GENERAL CONCLUSIONS

The problem of melisma synthesis by the use of Artificial Neural Networks was solved and following results were achieved:

1. A modification of Linear Prediction method has been proposed that assumes a quick alternation of pitch and is suitable for melisma analysis.
2. An original melisma synthesis technique has been composed that allows to apply the generalized melisma model. It is based on bidirectional processing using the pitch and intensity characteristics approximated by song, the original segment of the glottal signal as well as Linear Prediction Coefficients.
3. A flexible generalized melisma model has been developed, based on two Artificial Neural Networks – a Multilayer Perceptron and Adaline – as well as on two network training algorithms: that of Levenberg-Marquardt and the Least Squares Error minimization.
4. Original mathematical models of Fortis, Gruppert, Mordent and Trill have been created, fit for synthesizing melismas, and their minimal sizes have been proposed as well as the values of model parameters experimentally defined.
5. The melisma synthesis technique has been corroborated experimentally applying the new mathematical models to melisma synthesis of one performer. The MOS estimate of melisma range from 3.7 to 4.4 points. The general mean MOS estimate of male data set is 4.1 points and for female data set is 3.9 points.

Future research works are planned in connection using the improvement of the melisma methods proposed with a view to improve the melisma synthesis results independent of the performers, since MOS estimate of common data set is 3.6 points.

References

- [1] Akramizadeh, A.; Farjami, A. A.; Khaloozadeh, H. Nonlinear Hammerstein Model Identification Using Genetic Algorithm. // *ICAIS02*, 2002, p. 67–73.
- [2] Bauer, D.; Ninness, B. Asymptotic properties of identification of Hammerstein models with input saturation. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, June 2000.
- [3] Collins. English Dictionary, Updated Edition. 1994.
- [4] Deller, J. R.; Proakis, J. G.; Housen, J. H. L. Discrete-Time Processing of Speech Signals. New York: Macmillan, 1993.
- [5] Dubnowski, J. J.; Schafer, R. W.; Rabiner, L. R. Real-Time Digital Hardware Pitch Detector. *IEEE Trans. on ASSP*, 24, 1976, p. 2–8.
- [6] Hai, J.; Joo, E. M. Improved Linear Predictive Coding Method for Speech Recognition. // *ICICS-PCM 2003*, 2003.
- [7] Haykin, S. Neural Networks: A Comprehensive Foundation. Prentice-Hall, Inc., 2 edition, 1999.

- [8] Lipeika, A.; Lipeikienė, J. Speaker Recognition based on the use of Vocal Tract and Residue Signal LPC Parameters. *Informatica*, 10, No. 4, 1999, p. 377–388. ISSN 0868-4952.
- [9] Ljung, L. *System Identification - Theory for the User*. Prentice-Hall, 2 edition, 1999.
- [10] Makhoul, J. Linear prediction: A tutorial review. In *Proceedings of the IEEE*, 63, No. 4, 1975, p. 561–580.
- [11] Navakas, D. *Artificial Neural Network for the Restoration of Noise Distorted Songs Audio Records* Doctoral dissertation, Vilnius Gediminas Technical University, September 1999 No. 434.
- [12] Navickaitė-Martinonienė, E. *Elementarioji muzikos teorija*. Vilnius: Vaga, 1979.
- [13] Nelles, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer-Verlag, 2001.
- [14] Pereira, W. *Modifying LPC Parameter Dynamics to Improve Speech Coder Efficiency*. McGill University, Canada, 2001.
- [15] Pupekis, R.; Navakas, D.; Ljung, L. Identification of Wiener systems with hard and discontinuous nonlinearities. Technical report, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Mar. 2003.
- [16] Raškini, G. Preprocessing for Folk Song Acoustic Records for Transcription into Music Scores. *Informatica*, 9, No. 3, 1998, p. 1–9. ISSN 0868-4952.
- [17] Ross, M. J. et al. Average Magnitude Difference Function Pitch Extractor. *IEEE Trans. on ASSP*, 22, 1974, p. 353–362.
- [18] Rossi, P. S.; Palmier, F.; Cutugno, F. Inversion of F_0 Model for Natural-Sounding Speech Synthesis. // *ICASSP 2003*, 2003.
- [19] Seneff, S. Real-Time Harmonic Pitch Detector. *IEEE Trans. on ASSP*, 26, No. 4, 1978, p. 358–365.

About the Author

1995–1999: Studies at Electronics Faculty of Vilnius Gediminas Technical University and receives BSc in Electrical and Electronics Engineering.

1999–2001: Studies at Electronics Faculty of Vilnius Gediminas Technical University and receives MSc in Electrical and Electronics Engineering.

2002–2006: PhD studies at Electronic Systems Department of Vilnius Gediminas Technical University.

MELIZMŲ SINTEZĖ DIRBTINIŲ NEURONŲ TINKLAIS

Dauguma šiuolaikinių kalbos sintezės metodų pagrįsti nekintančiu pagrindiniu tonu ir orientuoti į monotoniškos kalbos sintezę. Šie metodai netinka tokiems uždaviniams kaip dainos restauravimas, nes sintezuotoje kalboje trūksta natūralumo. Dirbant su viena sudėtingesnių dainos puošmeninių figūrų – melizma – pasižyminčia staigiai kintančiu pagrindiniu tonu ir intensyvumu, siekiama išnagrinėti kokybinius kalbos sintezės aspektus ir pasiūlyti metodus leisiančius pakelti kalbos sintezės kokybę.

Šio darbo tikslas – taikant dirbtinių neuronų tinklus sintezuoti lietuviškose liaudies dainose sutinkamas melizmas. Darbe siekta: apžvelgus melizmas ir signalų modeliavimo metodus parinkti esmines melizmų charakteristikas bei atrasti ir parodyti dirbtinių neuronų tinklų taikymo melizmomis sintezuoti galimybes; pasiūlyti melizmų sintezės metodiką ištiriant tiesinės prognozės metodo taikymo galimybes; sukurti melizmų matematinius modelius tinkančius pagrindinių rūšių melizmų sintezei; sukurti melizmų įrašų rinkinį apimančią skirtingų rūšių melizmas bei įdainuotą keletos atlikėjų; atlikti sukurtų melizmų modelių eksperimentinę patikrą bei rasti tinkamus modelių dydžius ir jų parametrų reikšmes.

Disertaciją sudaro šeši skyriai iš kurių paskutinis – išvados. Pirmame įvado skyriuje suformuluojama problema, darbo tikslas ir uždaviniai, aptariami tyrimų metodai ir darbo mokslinis naujumas, pristatomas autoriaus indėlis bei disertacijos struktūra.

Antrame skyriuje nagrinėjama melizmos vieta netiesinių bei dinaminių sistemų modeliavime. Aprašoma melizma, jos tipai ir charakteristikos, parodoma kad melizma pasižymi netiesiškumu bei staigia dinamika. Apžvelgiamos dinaminės sistemos, jų struktūros. Aprašomi plačiausiai taikomi dinaminių sistemų identifikavimo ir patikros būdai. Paskutinėje antrojo skyriaus dalyje apžvelgiami dirbtinių neuronų tinklai ir parodomas jų ryšys su netiesinėmis ir dinaminėmis sistemomis – melizmų sintezės priemone.

Trečias skyrius skirtas apžvelgti melizmų sintezės metodiką. Nagrinėjamas tiesinės prognozės koeficientų metodas – pateikiami jo privalumai ir trūkumai. Atsižvelgiant į melizmos specifiką *siūlomas modifikuotas tiesinės prognozės metodas* ir jo pagrindu *sudaroma originali melizmos sintezės metodika*.

Ketvirtame skyriuje kuriami melizmų matematiniai modeliai: *pristatomas apibendrintas melizmos modelis*, parenkamas identifikavimo būdas ir *pristatomi originalūs pagrindinių melizmų rūšių – foršlago, grupeto, mordento ir trelės – modeliai*. Skyriaus pabaigoje pasiūlomi mažiausi atskirų rūšių melizmų modelių dydžiai.

Penktame skyriuje vykdoma sukurtų modelių taikymo melizmomis sintezuoti eksperimentinė patikra: aprašomas sudarytas melizmų rinkinys bei pateikiami atliktų vieno atlikėjo melizmų sintezės bei nuo atlikėjo nepriklausomų melizmų sintezės eksperimentų rezultatai. *Parodoma, jog sukurti melizmų modeliai yra tinkami vieno atlikėjo melizmų sintezei*.

Pagrindiniai disertacijos rezultatai paskelbti 6 mokslinėse konferencijose. Darbo tema paskelbti 5 moksliniai straipsniai: vienas straipsnis respublikiniame žurnale cituojamame Inspec duomenų bazėje, du straipsniai užsienio recenzuojamuose žurnaluose ir du straipsniai respublikinių konferencijų darbų medžiagoje.

Romas Leonavičius

MELISMA SYNTHESIS USING ARTIFICIAL NEURAL NETWORKS

Summary of Doctoral Dissertation

Technological Sciences, Electrical Engineering and Electronics (01T)

Romas Leonavičius

MELIZMŲ SINTEZĖ DIRBTINIŲ NEURONŲ TINKLAIS

Daktaro disertacijos santrauka

Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

2006 11 17. 1,5 apsk. leid. 1. Tiražas 100 egz.

Vilniaus Gedimino technikos universiteto leidykla „Technika“,

Saulėtekio al. 11, 10223 Vilnius

Spausdino UAB „Biznio mašinų kompanija“,

Jasinskio g. 16A, 01112 Vilnius