Comparative Analysis of the Applicability of Feed-Forward and Recurrent Neural Networks for Prediction of Surface Quality of Laser Polished Surfaces

Honghe Wu[1,2], Evgeni V. Bordatchev[*2,1], and O. Remus Tutunea Fatan[*1,2]

[1]*Mechanical and Materials Engineering Department, Western University, London, Ontario, Canada*
[2]*Automotive and Surface Transportation Research Center, National Research Council of Canada, London, Ontario, Canada*
[*]*Corresponding author's e-mail: evgueni.bordatchev@nrc-cnrc.gc.ca, rtutunea@eng.uwo.ca*

Laser polishing (LP) is an emerging manufacturing process capable to address some of the significant limitations of traditional surface quality improvement technologies such as abrasive or chemical based polishing processes. By reconfiguring the topography of the outer surface, surface characteristics such as quality, aesthetics, wettability, friction, bio-fouling resistance, and others can be enhanced at a relatively low cost. However, numerous LP process parameters have to be fine-tuned to achieve the intended surface quality. This makes the selection of optimal process parameters time consuming and often unrepeatable. This study suggests that while both feed-forward and recurrent neural networks can be used to predict LP surface quality with a reasonable accuracy, the latter is characterized by a superior performance.
DOI: 10.2961/jlmn.2022.02.2007
Keywords: laser polishing, surface quality prediction, artificial intelligence, feed-forward neural network, recurrent neural network

1. Introduction

Laser polishing (LP) is a common manufacturing process used in tooling, aerospace, and medical devices industries to achieve desired appearance or functionality. Abrasive and chemical polishing processes are also widely used for similar purposes, but these processes are slow, expensive and pollutant. Traditional surface polishing techniques are also unable to polish complex geometries with high aspect ratios or areas with a defined boundary. Traditional polishing techniques are also known to damage the surrounding geometry/structures/textures.

By contrast, LP is an emerging technology that offers solutions to many of the drawbacks of the traditional surface finishing techniques. During LP, a high-powered laser is used to transform the original topography of the surface, $h_{ini}(x, y)$, into a final topography, $h_{LP}(x, y)$. This happens by redistributing the molten material, a process that smoothens the micro asperities of the surface. The surface quality of $h_{LP}(x, y)$ is controlled by numerous inter-dependent process parameters such as laser power, P, surface scanning speed, v, laser beam diameter, d, and programed scanning trajectory. The combination of these process parameters controls numerous nonlinear thermodynamic phenomena such as rapid remelting and solidification, capillary and thermocapillary flows as well as various modes of heat transfer in the laser-material interaction zone. Furthermore, surface tension forces, temperature gradients, and continuously changing material properties with respect to temperature, material state, and process instabilities affect the formation of the LP surface topography [1]. These complex phenomena make the selection of optimal process parameters difficult due to inherent limitations of the present knowledge and/or capabilities associated with the nonlinear thermodynamics of laser-material interactions. Commonly, a grid search-based experiment design is needed to test every possible combination of process parameters in order to determine an optimal window of process parameters capable to yield the targeted and/or best surface quality. This surface characteristic is typically described by means of two metrics: areal arithmetical mean height S_a^{LP} (areal roughness) and W_a^{LP} (waviness). Nonetheless, the design of experiments (DoE)-based approach tends to be tedious, expensive and often unrepeatable, especially when it involves large matrices of experiments.

To address this, the current study aims to test whether artificial intelligence (AI) techniques can be used to establish a reliable model of the process capable to accurately predict the resultant surface quality. This AI-based approach has the potential of being faster and thereby more cost-effective. A survey of the available literature indicates that researchers have already implemented with success various AI based online control systems for laser powder bed fusion process [2]. The objective of this work is to compare the performance of feed-forward neural network (FFNN) with recurrent neural network (RNN) to evaluate the effect of RNN's ability to retain memory, and how RNN can be applicable in modeling the thermodynamic performance of the laser remelting process.

2. Feed-forward neural network-based model

Feed-forward neural networks are one of the most common network architectures. However, unlike some of their

newer successors – such as recurrent neural networks (RNN) – FFNNs lacks the feedback or cycle structure. Because of this, signals move through the network in a linear fashion such that previous results have no effect on the next network output. Recently, FFNN models have been widely used in process modeling of online control [3], medical disease prediction [4], handwriting recognition [5], and prediction of traffic conditions [6]. It has been demonstrated that FFNN is able to generalize and adapt to complex data patterns. At its core, the network structure is a universal function approximator according to which combinations of values of input and trained scalers produces the values at the output layer. Because of its internal structure, FFNN results are typically obtained with supervised learning techniques according to which datasets are divided into a training and a testing/validation set. The training set contains inputs and target outputs. The supervised learning algorithm relies on a predefined error function to evaluate the error between network predictions and the values targeted by the network. By tuning the network parameters to minimize the error function, the FFNN model can be built. The supervised FFNN learning algorithm used in this study relies on a mean square error (MSE) function. It is worth noting that MSE is a function involving all network parameters in a sense that the goal of training step is to find the global minimum of the error function surface.

2.1 Multilayer perceptron

Multilayer perceptron is the most common neural network structure that consists of a series of neurons, each representing a value ($\boldsymbol{a_n}$) interconnected with associated weights ($\boldsymbol{w_n}$). The weights and biases ($\boldsymbol{B_{bias}}$) are considered the network parameters. Taken together, they offer to the neural network the required degree of freedom to assimilate the dataset being modeled. The multilayer perceptron is composed of a preset number of layers. The first layer of neurons takes in values as inputs. The input values propagate through the hidden layers to the output layer (Fig. 1).

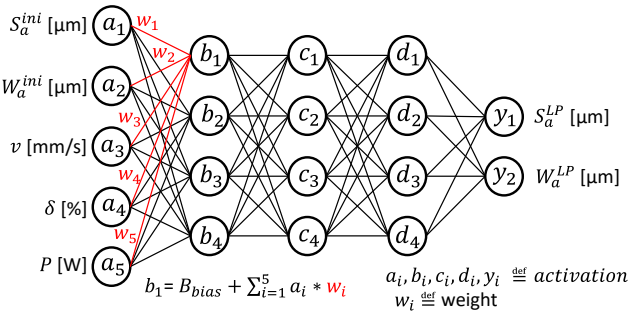

Fig. 1 Multilayer perceptron network structure.

The linear combinations of all neuron values in the input and the hidden layer proved to be a universal function approximator [7].

$$b_1 = B_{bias} + \sum_{i=1}^{5} a_i * w_i \quad (1)$$

The performance of multilayer perceptron networks is dependent on the number of neurons and the number of hidden layers. A shallow network will not contain enough parameters to approximate the nonlinear process with high accuracy. However, a deep network may overfit the dataset, in a sense that the network becomes overtrained on the input dataset and thereby it will fail to generalize the underlying patterns in the dataset. At this time, no rigorously proven methodology capable to determine the required network complexity exists. Typically, the optimal network structure is determined experimentally by finding – typically through heuristic searches – the boundary of under and over-fitting and by evaluating the performance of the trained neural network on the untrained testing dataset. Given its ability to model highly nonlinear processes, multilayer perceptron was used to model the LP process. This particular type of network requires five inputs: areal arithmetical mean height of the initial surface ($\boldsymbol{S_a^{ini}}$), areal arithmetical mean height of the initial surface ($\boldsymbol{W_a^{ini}}$), laser power ($\boldsymbol{P}$), scanning speed ($\boldsymbol{v}$), and step-over distance between the adjacent polished paths ($\boldsymbol{\delta}$). The network output was designed to predicts two important metrics of the polished surface: $\boldsymbol{S_a^{LP}}$ and $\boldsymbol{W_a^{LP}}$.

3. Recurrent neural network-based model

Building on the FFNN configuration, the recurrent neural network (RNN) structure encompasses a feedback portion, in which the model takes in previous numerical results into account when producing the next result. This type of network architecture is driven by real-life time domain dependencies in many applications. For example, in speech recognition, the meaning of the word being processed is inevitably dependent on the context, namely the meaning of previous words [8]. Along the same lines, most manufacturing processes are also time dependent. Since it is able to model time series with a recurrent nature makes, RNNs has a natural advantage [9].

The LP process encompasses two overlapping phenomena. First, the formation of the resolidified surface overlaps with a previously remelted surface where the amount of overlap is defined by the stepover percentage parameter (δ). This means that the surface quality of the previously polished laser track will be the initial starting condition of the current laser track and therefore, it will have a significant effect on the formation of the current laser track. Second, the remelting process constitutes the core of a thermodynamic system. Because of this, thermal gradients will affect the flow of molten material and thus impact the post-polished surface topography. In other words, the previously heated surface governed by conduction and convection heat transfer affects the remelting of the current melt pool.

The aforementioned natural time domain dependencies represent the primary motivation behind the utilization of RNNs to model the process. To verify its performance, the predictive output of the RNN will be compared with the traditional FFNN multilayer perceptron structure.

3.1 Long-Short Term Memory (LSTM) Cells

LSTM is a special type of RNN network structure, in which the network consists of numerous LSTM cells. Each LSTM cell contains four gates. These gates provide additional network parameters to control the rate at which the new information is retained whereas the old information is forgotten (Fig. 2). In case of RNN, the cell state $\boldsymbol{C_t}$ and cell output $\boldsymbol{h_t}$ are calculated according to the following equations:

$$f_t = \sigma * (W_f * X_t + R_f * h_{t-1} + b_u) \quad (2)$$
$$u_t = \sigma * (W_u * X_t + R_u * h_{t-1} + b_u) \quad (3)$$
$$i_t = \sigma * (W_i * X_t + R_i * h_{t-1} + b_i) \quad (4)$$
$$o_t = \sigma * (W_o * X_t + R_o * h_{t-1} + b_o) \quad (5)$$
$$C_t = f_t * C_{t-1} + i_t * u_t \quad (6)$$
$$h_t = \tanh(c_t) * o_t \quad (7)$$

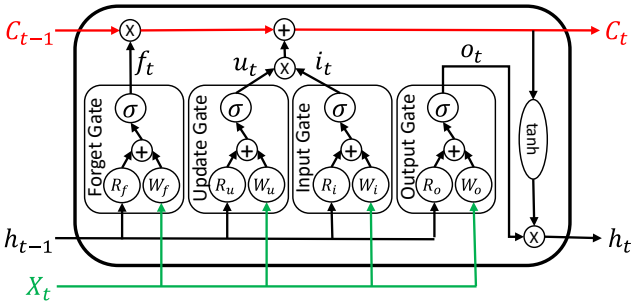

Fig. 2 LSTM cell structure.

These parameters are tuned during the training stage with a similar supervised learning technique. The overall RNN architecture was tuned to achieve the best accuracy of the prediction for the validation set. RNN consists of layers of interconnected LSTM cells (Fig. 3). The network inputs ($\boldsymbol{a_{1\sim5}}$) propagate through the hidden layers of LSTM cells to produce the resultant surface quality predictions ($\boldsymbol{y_{1,2}}$).

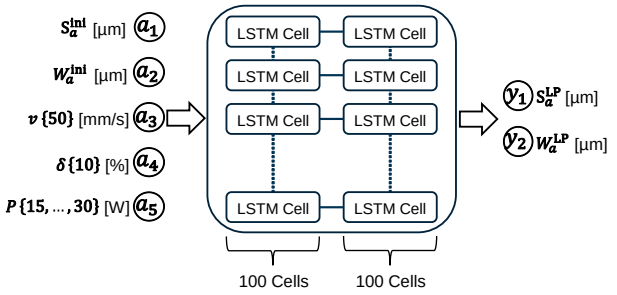

Fig. 3 RNN structure.

4. Experimental setup

The core of the LP system was a 60 W max continuous wave, 1070 nm laser (Fig. 4). The laser beam was initially passed to a collimator through fiber optic cable. The collimated beam was then passed through a camera beam splitter for online monitoring. A two-axis galvanometer scanner from SCALAB GmbH was used to precisely position the laser beam on the *XY* plane at high speeds. The *f*-theta lens ensured that the beam focus/defocus length remains constant when the beam was positioned at any angle from the *Z*-axis of the LP system.

The LP system also includes three-axis motion stages controlled through CNC software developed by Aerotech. Motion stages are meant to enable the positioning of larger workpieces whereas the scanner facilitates a rapid scanning of smaller polished patches. Scanner motions are controlled by means of laserDESK software that enables the tracing of a preprogramed trajectory. In this study, a zigzag scanning strategy was used to polish the workpiece areas depicted in Fig. 4. H13 tool steel was used for this research due to the wide use of this material in the tool and die industry. Experiments consist of power ranging from 15 W to 40 W, scanning speed from 50 mm/s to 750 mm/s, and stepover percentage of 5%, 10%, and 15%. For the context of the paper, the applicability of FFNN vs. RNN was evaluated for 50mm/s scanning speed, 10% track width stepover, and for 15 W, 20 W, 25 W, and 30 W power.

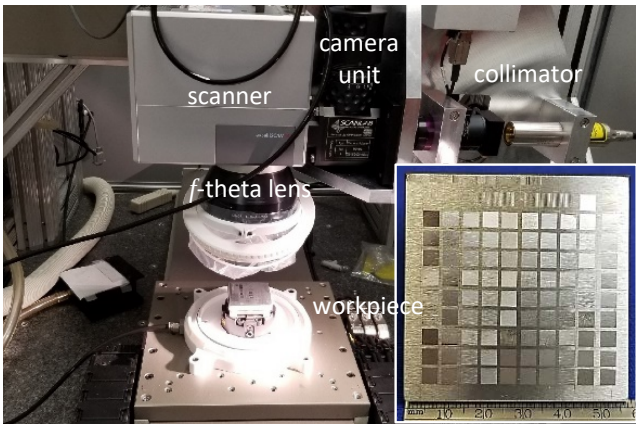

Fig. 4 Optical configuration of the LP system.

5. Data preparation methodology

To collect the data required for the training and testing of the neural network, a data extraction methodology was developed. The initial and polished surface was scanned by means of a Sensofar interferometry-based surface profiler (Fig. 5).

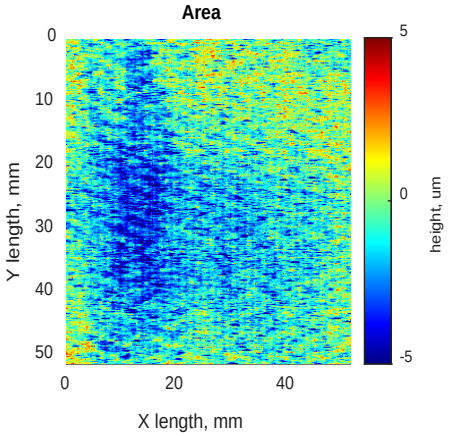

Fig. 5 Sample initial surface topography.

The surface topography was then passed through a high pass and a low pass gaussian 2D filter with a spatial cutoff frequency of 1.25 mm^{-1}. This filtering step splits the original topography into a low and a high frequency component that are associated with waviness (or surface error) and roughness, respectively. The results are presented in Figs. 6 and 7.

The two topographic images were then cropped to match the boundaries of the polishing areas. From each area, 49 subareas - each with a width of 0.825 mm – were then identified to generate more data for training and testing. After that, the last step was to calculate (according ISO 25178) the areal arithmetical mean height for waviness and roughness topography for each subarea. The last column of the subareas was intentionally left out of the training dataset, such that it can be used for the validation of NN's prediction accuracy

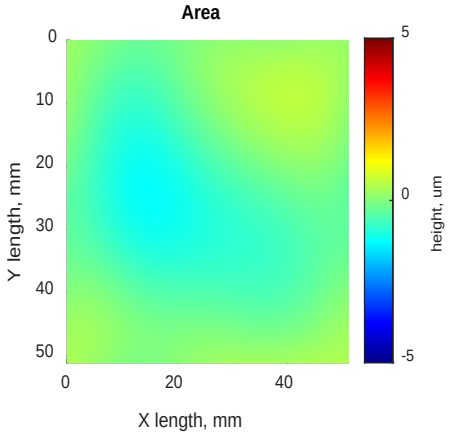

Fig. 6 Sample of roughness surface topography

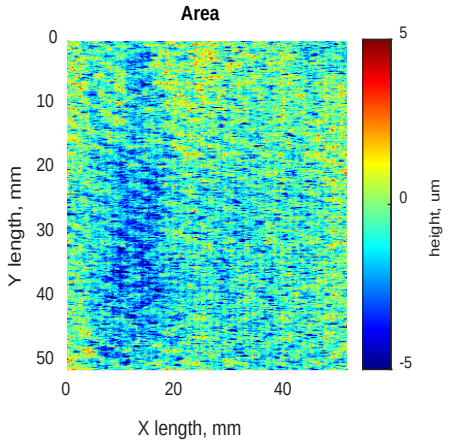

Fig. 7 Sample of waviness surface topography.

(Fig. 8). The training was done with 168 sets of datapoints, and training results were validated with 28 datasets consisting of process parameters mentioned above.

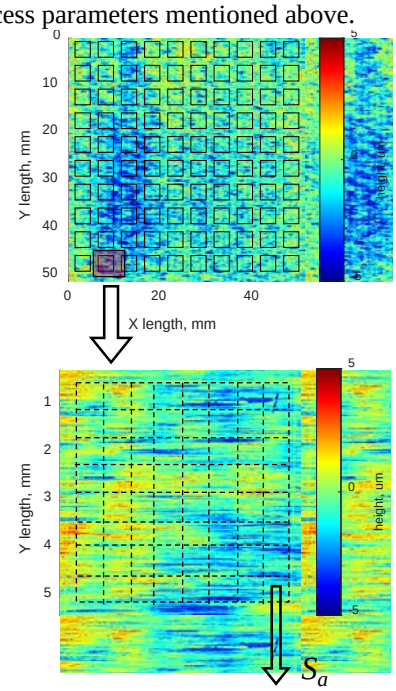

Fig. 8 Extracting surface characteristic parameters.

6. Results and discussion

Both FFNN and RNN were trained and tested with the same dataset and their performance was compared by means of correlation coefficients to determine the discrepancies between NN's predictions and the actual measured roughness/waviness values. The training and testing results are depicted in Fig. 9 whereas the validation results are shown in Figs. 10 and 11.

For the untrained testing dataset, the FFNN model achieved a Pearson correlation coefficient defined in equation 8 of 0.44 and 0.94 for roughness and waviness predictions, respectively. For the same dataset, RNN model achieved a correlation coefficient of 0.78 and 0.99 for roughness and waviness, respectively.

$$c(A,B) = \frac{1}{N-1} * \sum_{i=1}^{N} \frac{A_i - \mu_A}{\sigma_A} * \frac{B_i - \mu_B}{\sigma_B} \quad (8)$$

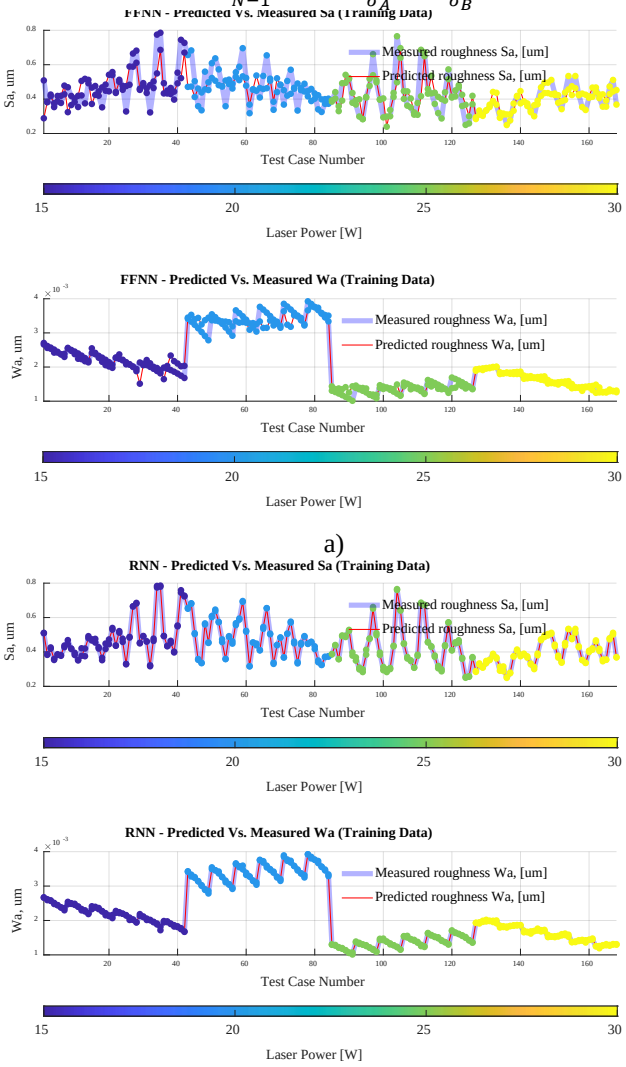

Fig. 9 Training results for: a) FFNN and b) RNN.

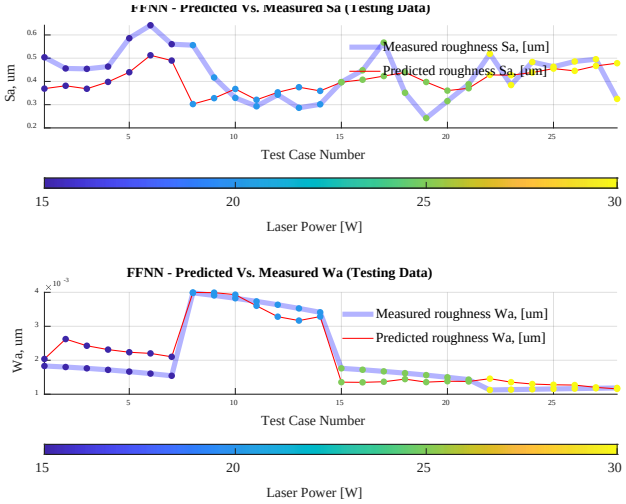

Fig. 10 FFNN validation results.

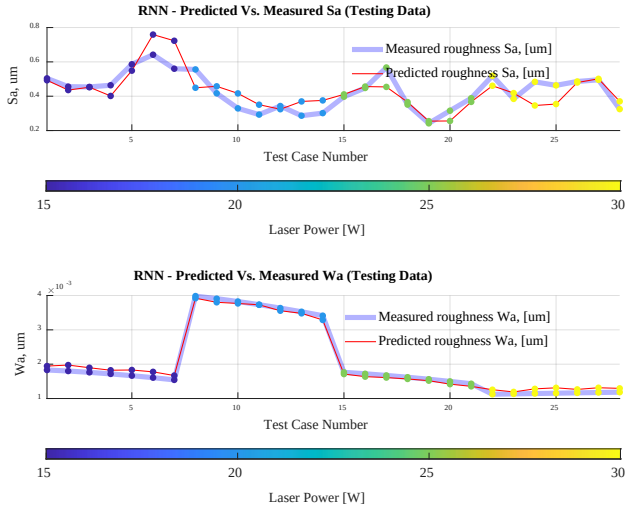

Fig. 11 RNN validation results.

7. Summary and conclusion

In this research, a multilayered perceptron-based FFNN and a LSTM-based RNN were developed for modelling of the post-polished surface quality. The input to both types of neural network models were process parameters and initial surface characteristics. While both FFNN and RNN structures showed good prediction results, RNN configuration performed slightly better in terms of prediction accuracy. Future work will attempt to integrate the developed NNs in online control and self-optimization of the LP process.

Acknowledgments

The work presented in this study is the result of the collaboration between the National Research Council of Canada and the University of Western Ontario (Canada) under the 2021-24 Canada-Germany 3+2 International Consortium project.

References

[1] E. Bordatchev, A. Hafiz, and R. O. Tutunea-Fatan: Int. J. Adv. Manuf. Technol., 73, (2014) 35.
[2] C. Knaak, L. Masseling, E. Duong, P. Abels, and A. Gillner: IEEE Access, 9, (2021) 55214.
[3] G. V. Bhagya Raj and K. K. Dash: Crit. Rev. Food. Sci. Nutr., 62, (2020) 2756.
[4] P. Siswanto and R. Rulaningtyas: Proc. 4th ICE on IMERI 2019, (2019) 17.
[5] S. Chaturvedi, R. N. Titre, and N. Sondhiya: Proc. ICESC 2014, (2014) 425.
[6] A. Derrow-Pinion, J. She, D. Wong, O. Lange, T. Hester, L. Perez, M. Nunkesser, S. Lee, X. Guo, B. Wiltshire, P. W. Battaglia, V. Gupta, A. Li, Z. Xu, A. Sanchez-Gonzalez, Y. Li, and P. Velickovic: Proc. 30th CIKM 2021, (2021) 3767.
[7] K. Hornick: Neural Netw., 2, (1989) 359.
[8] C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani: Proc. ICASSP 2018, (2018) 4774.
[9] E. Toyserkani, A. Khajepour, and S. Corbin: J. Laser. Appl., 14, (2002) 160610.

(Received: June 5, 2022, Accepted: October 8, 2022)