Radar based deep learning technology for loudspeaker faults detection and classification

1st A. Izzo
University of Strathclyde
Centre for Signal and Image Processing (CeSIP)
Glasgow, UK
alessio.izzo@strath.ac.uk

2nd C. Clemente
University of Strathclyde
Centre for Signal and Image Processing (CeSIP)
Glasgow, UK
carmine.clemente@strath.ac.uk

3rd L. Ausiello
Solent University
School of Media Arts and Technology
Southampton, UK
ludovico.ausiello@solent.ac.uk

4th J.J. Soraghan
University of Strathclyde
Centre for Signal and Image Processing (CeSIP)
Glasgow, UK
j.soraghan@strath.ac.uk

***Abstract*—Recently, radar based micro-Doppler signature analysis has been successfully applied in various sectors including both defence and civilian applications. A joint radar micro-Doppler and deep learning technology for End-Of-Line (EOL) test of loudspeakers is proposed in this paper. This approach offers the potential benefits of characterizing the mechanical motion of a loudspeaker in a noisy environment as a production line, in order to automatically identify and classify defects. Starting from real radar signal, the proposed Bidirectional Long Short-Term Memory (BiLSTM) classifier has been tested on training, validation and test dataset. The results show that the proposed approach produces a probability of correct classification above the 98%, outperforming the traditional k-NN classifier.**

***Index Terms*—Civilian Radar application, micro-Doppler analysis, loudspeaker analysis, classification, deep learning.**

I. Introduction

In a production chain of loudspeakers, condition monitoring cover an important topic as increasing quality checks at various stages of production (with limited costs) can provide substantial benefits to loudspeaker manufacturers. Nowadays, laser based analysis tools (i.e. Scanning Vibrometer System (SCN)) have been shown to yield significantly better results compared to traditional acoustic ones [1], [2]. In [3], the authors introduce for the first time a novel approach based on radar micro-Doppler to analyse and measure the return from loudspeakers. Furthermore, in [3] a model for the radar return from a loudspeaker based on the Thiele&Small parameters and a methodology to measure Mechanical Frequency Response (MFR) of loudspeakers with a single radar measurements are also presented. This approach is motivated by the potential cost effectiveness and operational advantages that a radar based approach could introduce over acoustic and laser based ones. With respect to the traditional acoustic measurement, a radar based approach is not affected by the acoustic environmental factors, allowing its use for End of Line (EoL) test. Unlike the SCN system, the micro-Doppler has the ability to cope with visual occlusion due to plastic parts and the capability of separation metallic components of a loudspeaker from non metallic ones through the use of the back-scattering intensity. Motivated by the recent advances in deep learning, aim of this paper is to show a potential benefit from a joint radar micro-Doppler and deep learning use for an industrial process application, such as EOL test for loudspeaker analysis. Following the recent trends of the manufacturing industry, a deep learning architecture will be introduced as classifier. While traditional statistical classification techniques rely on very restrictive assumptions and/or modelling, these are difficult to apply to all the possible defects of a speaker. For this reason, deep learning approaches that rely on observations represent good candidates to address this challenge. Moreover, deep learning approaches can offer a high degree of portability when the speaker model changes by selecting the appropriate training datasets. Although Convolutional Neural Networks (CNN) are mostly preferred in radar domain, here Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) is investigated. The reason behind the choice of such architecture lies on the nature of the data. Handling the MFRs of the DUT as time series or sequence data enables the use of LSTM-RNN architecture for classification purpose.

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. The key aspect of deep learning is that these layers of features are not designed by humans: they are learned from data using a general purpose learning procedure. In the radar domain, the use of deep learning methods lead to some benefits, solving a broad range of problems. The most straightforward application of deep learning to radar is within the area of Synthetic Aperture Radar (SAR) for Automatic Target Recognition (ATR). Deep learning algorithm is used to classify military target from SAR images [4], [5]. In the context of cognitive radar, deep learning algorithms have been used in order to detect the best sub-

978-1-7281-8942-0/20/$31.00

arrays of a phased array radar antenna, in order to increase the Direction of Arrival (DoA) estimation [6]. Due to the reducing cost of high frequency radar sensors, the use of radar-like sensors have become even more pervasive leading to joint radar and deep learning use for civilian applications. In [7], deep learning algorithms are used for human detection and activity classification based on radar micro-Doppler signature, while in [8] deep learning is used for hand gesture recognition. Following the recent trends in the manufacturing industry, in this paper a deep learning classifier based on MFRs of a loudspeaker is proposed in order to show the ability of the radar technology to automatically classify faulty speakers.

The remainder of the paper is organised as follows. In Section II, the model for the radar return from a loudspeaker based on the Thiele and Small parameters and the methodology to measure MFRs of loudspeakers in order to characterise the driver with a single radar measurement is summarized (a full description can be found in [3]). Belonging to the Recurrent Neural Network (RNN) class, Long Short-Term Memory (LSTM) is introduced in Section III. In Section IV the proposed Bidirection LSTM (BiLSTM) classifier is presented. Based on real measurements, results of the classification are reported in Section V and compared with the traditional $k-$NN classifier. Conclusions and future developments are presented in Section VI.

II. Radar signal model and MFR computation

In coherent radars, the range variations cause a phase change in the returned signal from a target. Thus, the Doppler frequency shift, representing the change of phase function over time, can be used to detect vibrations or rotations of structures in a target. The return signal from a target as a function of time is modelled as follows:

$$s_r(t) = \rho \exp\{j[2\pi f_0 t + \Phi(t)]\} \tag{1}$$

where ρ is the reflectivity of the vibrating point scatterer, f_0 is the carrier frequency of the transmitted signal and $\Phi(t)$ is the time varying phase change of the vibrating scatterer. Letting R_0 be the distance between the radar and the speaker's initial position, then the range function varies with time due to the speaker micro-motion:

$$R(t) = R_0 + D(t) \tag{2}$$

Assuming that the driver plays an exponential sine sweep signal $x(t)$ [9], then an arbitrary point of the cone located in P vibrates with instantaneous vibration frequency $f_v(t)$ equals to:

$$f_v(t) = f_1 k^t = f_1 \left(\frac{f_2}{f_1}\right)^{\frac{t}{T}} \tag{3}$$

with k the exponential chirp rate, and voice coil displacement $\tilde{\eta}_c(f_v(t))$ defined as [10]:

$$\tilde{\eta}_c(f_v(t)) = \frac{\tilde{e}_g}{2\pi f_r Bl Q_{es}} \mid \gamma_c(f_v) \mid \tag{4}$$

In (4), $\tilde{e}_g$ represents the voltage at the speaker's terminals, Bl is the force factor (magnetic flux density B multiplied by the length of the wire l), f_r is the resonance frequency of the speaker and $\gamma_c(f_v)$ is a dimensionless frequency response function given by [10]:

$$\gamma_c(f_v) = \frac{1}{1 - \frac{f_v^2}{f_r^2} + j\frac{f_v}{f_r Q_{ts}}} \tag{5}$$

where Q_{ts} is the total damping effect, composed of the electrical damping Q_{es} and the mechanical damping Q_{ms}, and j is the imaginary unit. Equation (4) describes the frequency dependent behaviour of the loudspeaker displacement. Depending on the amplitude of the displacement, the transducer will generate distorted signals that can be classified as linear (low displacement amplitude - high frequency) and non linear (high displacement amplitude - low frequency) distortion. Both of them are regarded as regular distortions because they are accepted within the design process and are results of optimization process giving the best compromise with other constraints (weight, cost, size). On the other hand, irregular distortions are non acceptable defects in a loudspeaker passing the EOL tests. Describing the voice coil displacement with (4) allows the second term in (2) to be modelled as follows:

$$D(t) = \tilde{\eta}_c(f_v(t))\, x(t) \tag{6}$$

while assuming the radar is in the line of the sight with the speaker. Then, the time varying phase can be written as:

$$\Phi(t) = \frac{4\pi}{\lambda} R(t) = \frac{4\pi}{\lambda}\left[R_0 + \tilde{\eta}_c(f_v(t))\, x(t)\right] \tag{7}$$

with λ the wavelength of the transmitted signal. Substituting (7) in (1) the received signal can be expressed as [3]:

$$\begin{aligned} s_r(t) =& \rho \exp\left\{j\frac{4\pi R_0}{\lambda}\right\} \\ & \exp\left\{j2\pi f_0 t + j\frac{4\pi\tilde{\eta}_c(f_v(t))}{\lambda} x(t)\right\} \end{aligned} \tag{8}$$

When a chirp stimulus is used in the simulated scenario, the voltage as the force factor is supposed to be constant with the frequency of the stimulus. In a real scenario this hypothesis is not justified due to the non linearities, introduced for example by stiffness ($K_{ms}(\tilde{\eta}_c)$), force factor ($Bl(\tilde{\eta}_c)$) and inductance ($L_e(\tilde{\eta}_c)$). To understand the influence of the non linearities on the speaker behaviour, in terms of deviation from the ideal piston mode behaviour, the Mechanical Frequency Responce (MFR) is computed [3]. If the model of the ideal received radar signal is found, matched filter technique can be applied to the radar received signal in order to characterize the mechanical behaviour of the speaker. Using (8) the received radar signal in baseband of an ideal loudspeaker, behaving as piston mode in the full frequency band, is described.

With an exponential sine sweep of $T = 60$ seconds as a test signal, and with the hypothesis of a linear system, the result of the matched filter would be a perfect peak centred in T, defined as linear impulse response. In a real scenario instead, where the Device Under Test (DUT) is never linear, along with the linear impulse responses, non linear impulse responses are also obtained, corresponding to the various harmonics of the

input signal. With the exponential sine sweep, these non linear products, do not contaminate the linear impulse response, as they are occurring at very precise anticipatory times Δt before the linear response, namely [11]:

$$\Delta t = T \frac{\ln(N)}{\ln\left(\frac{f_2}{f_1}\right)} \tag{9}$$

where N is the Nth distortion component. Thus, in the real scenario it is possible to define the time waveform at the output of the matched filter as the cross correlation function $r_{y,s_r}(\tau)$ between the measured signal $y(t)$ and the ideal one $s_r(t)$, where its Fourier Transform is referred as Cross Power Spectrum Density (CPSD) [12]. Depending on the status of the speaker, the matched filter output will show a different mechanical impulse response. Considering a *B&C 10CL51* LF driver with $f_r = 67.57$Hz, $Bl = 9.67N/A$, $Q_{ts} = 0.5425$ and $Q_{es} = 0.6075$[1], an example of matched filter outputs of the speaker with and without defects are shown in the Figure 5. As expected, different defects affect the impulse response of the system in different ways. In agreement with (9), the non linear products occur at very precise times before the linear response, namely at $\Delta t_{2nd} = 7.50$s, $\Delta t_{3rd} = 11.93$s and $\Delta t_{4th} = 15.04$s. The mechanical frequency response of the DUT can be evaluated by computing the Fourier Transform of the matched filter output. In case of no windowing, the CPSD of the signal is computed, where harmonics products and noise are incorporated into the frequency response. Applying a window around the main peak in $h(T)$, the linear frequency response can be assessed. Since the non linear products are a powerful indicator of possible manufacturing problems, they are analysed in the same way in order to compute harmonic frequency responses as well. These harmonic products affect the behaviour of the speaker mainly at low frequency, where the device is more susceptible to the non linear effects, in agreement with loudspeaker model theory.
In order to provide a classification of faulty speakers based on MFRs, as required by most of loudspeaker manufacturers, linear and non linear frequency responses will be used as input to the classifier.

III. Long Short Term Memory

LSTM are an improvement over the general RNNs, which is affected by vanishing gradient problem, making it suitable to learn long-term dependencies between time steps of sequence data. LSTM RNNs address the vanishing gradient problem commonly found in ordinary recurrent neural networks by incorporating gating functions into their state dynamics [14]. Figure 2 illustrates the flow of a time series $\mathbf{X}$ with D features of length S through an LSTM layer: h_t denotes the output (also known as the hidden state) and c_t denotes the cell state at the time step t. The first LSTM block takes the initial state of the network and the first time step of the sequence $\mathbf{X}_1$, and then computes the first output h_1 and the updated cell state c_1. At time step t, the block takes the current state of the network (c_{t-1}, h_{t-1}) and the next time step of the sequence $\mathbf{X}_t$, and then computes the output h_t and the updated cell state c_t. The hidden state at time step t contains the output of the LSTM layer for this time step, while the cell state contains information learned from the previous time steps. At each time step, the layer adds information to or removes information from the cell state, controlling updates and outputs state using gates. Figure 3 shows the flow of data at time step t and how the gates forget, update and output control the cell and hidden state. This is possible through the learnable weights. For an LSTM layer, they can be divided in input weights, recurrent weights and bias, which are concatenated as follow:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_i \\ \mathbf{W}_g \\ \mathbf{W}_f \\ \mathbf{W}_o \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} \mathbf{R}_i \\ \mathbf{R}_g \\ \mathbf{R}_f \\ \mathbf{R}_o \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{b}_g \\ \mathbf{b}_f \\ \mathbf{b}_o \end{bmatrix} \tag{10}$$

where the subscript *i*, *f*, *g*, and *o* denote the input gate, forget gate, cell candidate, and output gate, respectively. More concretely, the computation at time step t of each gates is defined as follow:

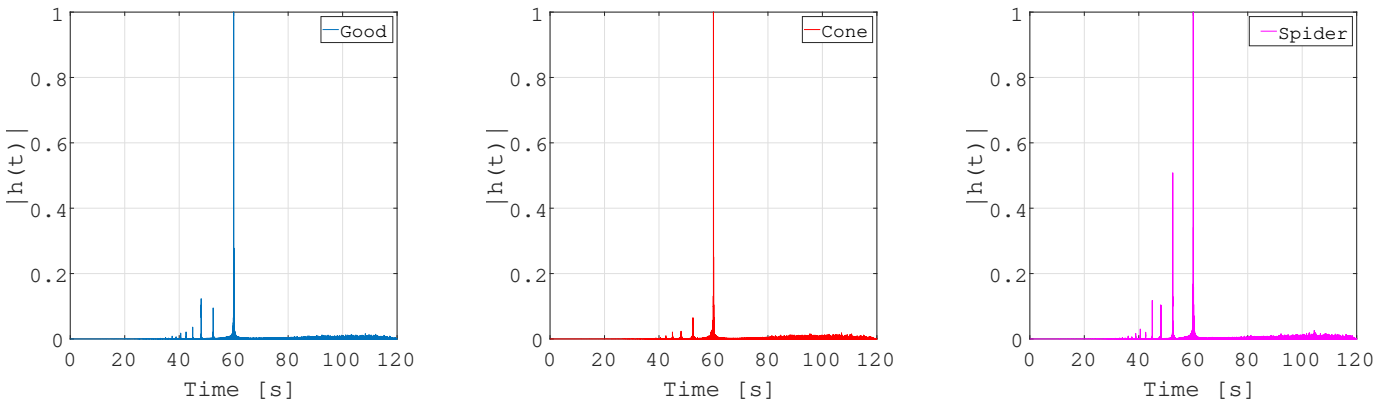
$$\begin{cases} \mathbf{i}_t = \sigma_g\left(\mathbf{W}_i \mathbf{X}_t + \mathbf{R}_i \mathbf{h}_{t-1} + \mathbf{b}_i\right) \\ \mathbf{f}_t = \sigma_g\left(\mathbf{W}_f \mathbf{X}_t + \mathbf{R}_f \mathbf{h}_{t-1} + \mathbf{b}_f\right) \\ \mathbf{g}_t = \sigma_c\left(\mathbf{W}_g \mathbf{X}_t + \mathbf{R}_g \mathbf{h}_{t-1} + \mathbf{b}_g\right) \\ \mathbf{o}_t = \sigma_g\left(\mathbf{W}_o \mathbf{X}_t + \mathbf{R}_o \mathbf{h}_{t-1} + \mathbf{b}_o\right) \\ \mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \\ \mathbf{h}_t = \mathbf{o}_t \odot \sigma_c\left(\mathbf{c}_t\right) \end{cases} \tag{11}$$

where $\odot$ denotes the Hadamard product. Usually, the activation function used for the input, forget and output gates is the sigmoid function $\sigma_g(x)$ while for the state activation function the hyperbolic tangent function (tanh) $\sigma_c(x)$ is used [14].
Although the problem of the vanishing gradient is solved, LSTM could still be affected by exploding gradient. If the gradients increase in magnitude exponentially, the training loss may assume an indeterministic form. As a consequence then, the training becomes unstable and can diverge within a few iterations. In order to avoid this problem, gradient clipping can be used to prevent gradient explosion by stabilizing the training at higher learning rates and in the presence of outliers [15].

IV. Proposed BiLSTM Classifier

While LSTMs possess the ability to learn temporal dependencies in sequences, they have difficulty with long term dependencies in long sequences. The solution proposed in [16] can help the LSTM RNN to learn these dependences. Combining a Bidirectional RNN (BRNN) with LSTM, it is possible to increase capacity of BRNNs by stacking hidden layers of LSTM cells in space, called deep Bidirectional LSTM (BiLSTM). In this way the output layer can get information from past (backwards) and future (forward) states simultaneously, making the BiLSTM networks more powerful than unidirectional LSTM networks, by involving all information of input sequences in the computation. In Figure 4 the architecture of the proposed BiLSTM classifier is shown,

[1]The T&S parameters are measured with Clio board [13]



(a) (b) (c)

Figure 1: Linear and non linear responses of *B&C 10CL51* LF driver with $f_r = 67.57$Hz, $Bl = 9.67N/A$, $Q_{ts} = 0.5425$ and $Q_{es} = 0.6075$. (a) Good speaker. (b) Cone manipulation. (c) Spider manipulation.

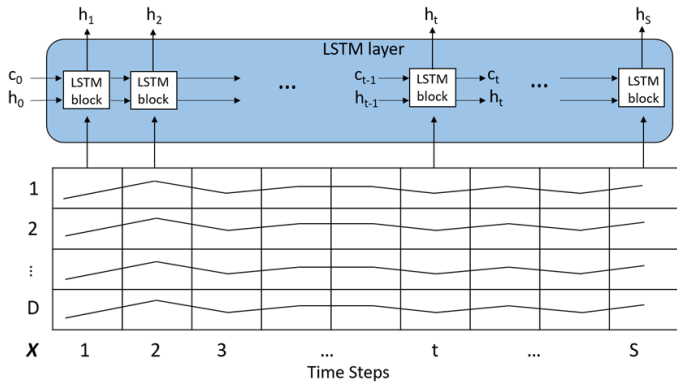


Figure 2: Long Short Term Memory (LSTM) network architecture.

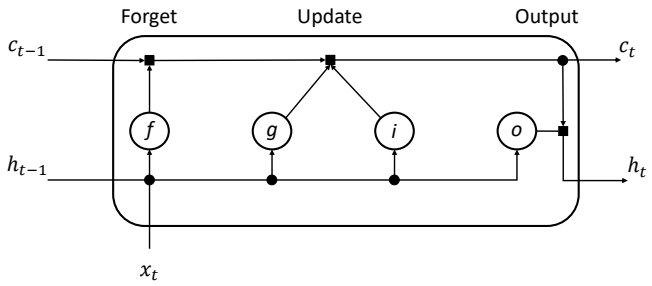


Figure 3: Diagram of a single LSTM block: flow of the data at the time step t.

where three BiLSTM layers are used with decreasing number of hidden units: 150, 125 and 100 for the first, second and third layer respectively. In particular, the selection of the number of layers as well as the size of the hidden units is based on empirical analysis. These parameters might depend on several aspects of the problem, such as the complexity of the dataset, the number of features and the number of data points. For instance, in case only CPSDs and linear frequency responses of the DUTs were considered, the input sequence of the classifier would have only two channels. In this case, the highest probability of correct classification would be achieved with only two BiLSTM layer, with an accuracy lower than 90%. Higher accuracy could be reached by adding another layer. In this case, due to the low number of training samples, adding a third layer would lead to over-fitting problem. To enable the use of three BiLSTM layers without the over-fitting problem, five channels of the input sequence are considered, taking into account the CPSD, linear frequency response and the first three harmonic frequency responses. To classify the class label, the

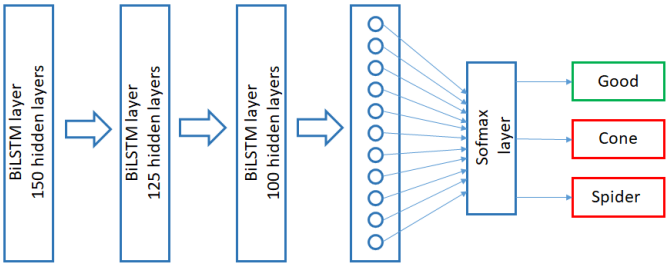


Figure 4: Architecture of the proposed BiLSTM network: input layer, three BiLSTM layers, full connected layer, sofmax layer and classification layer.

network ends with a fully connected layer, a softmax layer, and a classification output layer. As default, all weights are initialized from a Gaussian distribution with zero mean and standard deviation of 0.01, while for the initial bias the weights are set to be zero. Unlike the traditional Stochastic Gradient Descent algorithm with Momentum [17], where only single learning rate is used to upgrade all the network parameters, the ADAptive Moment (Adam) optimization is used here [18]. It keeps an element-wise moving average of both the parameters gradient and its squared values, such that:

$$\begin{cases} \mathbf{m}_l = \beta_1 \mathbf{m}_{l-1} + (1-\beta_1)\nabla L(\boldsymbol{\theta}_l) \\ \mathbf{v}_l = \beta_2 \mathbf{v}_{l-1} + (1-\beta_2)\left[\nabla L(\boldsymbol{\theta}_l)\right]^2 \end{cases} \tag{12}$$

where β_1 and β_2 are the gradient decay factor and the squared gradient decay factor of the moving averages $\mathbf{m}_l$ and $\mathbf{v}_l$, respectively. Thus, Adam algorithm is using the moving

averages to update the network parameters as:

$$\boldsymbol{\theta}_l = \boldsymbol{\theta}_{l-1} - \frac{\alpha \mathbf{m}_l}{\sqrt{\mathbf{v}_l} + \epsilon} \tag{13}$$

If gradients over many iterations are similar, then using a moving average of the gradient enables the parameter updates to pick up momentum in a certain direction. If the gradients contain mostly noise, then the moving average of the gradient becomes smaller, and so the parameter updates become smaller too [18]. In addition, the learning rate α in (13) is not meant to be a fixed learning rate. An alternative to using a fixed learning rate is to indeed vary the learning rate over the training process. The way in which the learning rate changes over time (training epochs) is referred to as the learning rate schedule or learning rate decay. The simplest learning rate schedule is to decrease the learning rate linearly from a large initial value to a small value. This allows large weight changes in the beginning of the learning process and small changes or fine-tuning towards the end of the learning process [19]. To avoid exploding gradient problem, a norm-based gradient clipping methods has been used, namely l2norm method. It rescales the gradient based on a threshold, avoiding a change of direction of the gradient [15]. Finally, the problem of overfitting is also tackled by adding a regularization term to the loss function. The training option and parameters of the proposed BiLSTM network are summarised in Table I.

Table I: Training option and parameters of the proposed BiLSTM network.

Parameter	**Value**
Epoch	500
Mini Batch	30
Solver	Adam
Initial Leaning Rate	$\alpha = 0.01$
Learning rate drop period	100 Epochs
Learning rate drop factor	0.1
Gradient Decay Factor	$\beta_1 = 0.9$
Squared Gradient Decay Factor	$\beta_2 = 0.99$
Gradient Threshold Method	l2norm
Gradient Threshold	5
Regularization coefficient	$\lambda = 0.0001$

V. Performance Analysis

The proposed BiLSTM classifier has been trained, validated and tested with real radar measurements. Since large amounts of balanced and diversified data are needed to properly train the network, a total amount of 620 measurements have been acquired from 24GHz CW radar sampling at 22kHz. For this purpose four different samples of the same speaker model (B&C 10CL51 LF driver of $10in$) have been considered. Due to the difficulty of acquiring samples with labelled defects, two different manipulations have been artificially applied first on the cone then on the spider. Note that, due to the irreversible damage, measurements related to the speaker with spider defect have been acquired always from the same driver. Consequently, as described in Section II, MFRs are obtained: with the number of frequency bins $S = 2048$, the dimension of the single sequence data is found to be equal to 5×2048. Thus, a balanced number of MFRs of the four speakers are taken per each class label: approximately 70% of the total measurements are used to populate the training dataset, and 15% for both validation and test datasets, with overall dimensions of $5 \times 2048 \times 450$, $5 \times 2048 \times 85$ and $5 \times 2048 \times 85$ respectively.

To gain a higher classification accuracy, preprocessing of the data is necessary. The average and the standard deviation of each channel are first calculated from the training dataset, then used to normalize each channel of training, validation and test dataset. In order to assess the performance of the proposed BiLSTM classifier, the loss function and accuracy have been considered and shown in the Figures 5a and 5b, respectively. In Figure 5b, the accuracy of the training dataset is compared with the accuracy of the validation dataset. After 500 epoch, a training dataset has been classified correctly with 100% accuracy, compared to the 98.82% on the validation dataset. As it can be noted from Figures 5a and 5b, the loss decreases rapidly in the first 100 epochs, with learning rate $\alpha = 0.01$ and an accuracy already above the 90%. Due to a large learning rate, the learning process is accelerated, though

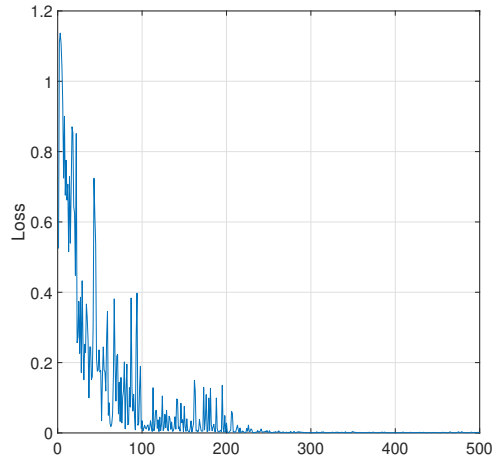


Figure 5: Performance analysis of the proposed classifier over 500 epochs. (a) Loss function on the training dataset. (b) Training and validation accuracy.

Output Class \ Target Class	Good	Cone	Spider
Good	100%	0	0
Cone	0	100%	0
Spider	0	4%	96%

Table II: Confusion matrix of the validation dataset with BiLSTM classifier.

Output Class \ Target Class	Good	Cone	Spider
Good	100%	0	0
Cone	0	100%	0
Spider	0	0	100%

Table III: Confusion matrix of the test dataset with BiLSTM classifier.

Output Class \ Target Class	Good	Cone	Spider
Good	88.33%	11.67%	0
Cone	10%	90%	0
Spider	2%	0	98%

Table IV: Confusion matrix of the cross-validation dataset with k-NN classifier.

a converging solution is not found yet. Result of this effect is the fluctuation of the loss, leading to the drop of the accuracy. Decreasing the leaning rate reduces this effect until almost zero loss is obtained at the end of training: with a learning rate of $\alpha = 10^{-4}$, a better tuning of the network parameters is found, leading to an accuracy of 100%. The comparison with a validation dataset is also a valid proof to show that the proposed network is not prone to overfitting problem.

To further validate the performance of the BiLSTM network, the confusion matrices of both validation dataset and test dataset are shown in the Tables II and III, respectively. Although the accuracy of the test dataset achieves the maximum percentage, an accuracy mismatch between training and validation dataset is registered. It comes from a single wrong fault classification, namely a spider defect that it has been classified as cone defect. As benchmark, the proposed BiLSTM based classifier has also been compared with the standard k-Nearest Neighbour classifier (k-NN). To empathise the ability of deep learning models to learn features directly from the data without the need for manual feature extraction, all the 5×2048 input samples are used for the k-NN. By setting the Euclidean distance as distance metric, and the number of nearest neighbours equal to the closest odd integer to the square root of the training dataset samples (as general practice), namely $k = 21$, the overall accuracy is 91.76% on the cross-validation dataset. In Table IV, the confusion matrix of the cross-validation dataset using the k-NN classifier is shown. Since multidimensional spaces sufficiently separated can not be found in the k-NN, it is less robust than the proposed BiLSTM classifier.

VI. Conclusion

In this paper the ability of the radar technology to automatically classify faulty speakers has been shown. The BiLSTM classifier has been used to classify loudspeakers affected by different damages. By using a 24GHz CW radar, signals scattered by the DUTs, with and without defects, have been acquired. The mechanical impulse responses of the DUTs have been computed through matched filter technique. To restrict the number of samples, Fourier Transform is applied on both linear and harmonic impulse responses. In this way, MFRs are obtained after a channel based normalization and used as input vector to the proposed BiLSTM architecture. The reliability of deep learning based classifier has been demonstrated by testing the network on training, validation and test dataset. The results showed that the proposed approach yields a classification accuracy above the 98%, outperforming the traditional k-NN classifier. In future, in the same loudspeaker testing method could be integrated in Linear Suspension Testing to assess quality control of moving parts.

References

[1] W. Klippel. "Scanning Vibrometer C5 Hardware and Software Module of the KLIPPEL R&D SYSTEM", *Klippel GmbH, Dresden, Germany.*

[2] W. Klippel. "Loudspeaker Nonlinearities: Causes, Parameters and Symptoms", *The Journal of Audio Engineer Socity, vol. 54 Issue 10 pp. 907-939; October 2006*, Klippel GmbH, Dresden, Germany.

[3] A. Izzo, and L. Ausiello, and C. Clemente and J. J. Soraghan,"Loudspeaker Analysis: A Radar Based Approach", *IEEE Sensors Journal*, pp. 1223-1237, February 2020, vol. 20, n. 3.

[4] S. Chen and H. Wang and F. Xu and Y. Jin, "Target Classification Using the Deep Convolutional Networks for SAR Images", *IEEE Transactions on Geoscience and Remote Sensing*, August 2016, vol. 54, n.8, pp. 4806-4817.

[5] J. Pei and Y. Huang and W. Huo and Y. Zhang and J. Yang and T. Yeo, "SAR Automatic Target Recognition Based on Multiview Deep Learning Framework", *IEEE Transactions on Geoscience and Remote Sensing*, April 2018, vol. 56, n. 4, pp. 2196-2210.

[6] A. M. Elbir and K. V. Mishra and Y. C. Eldar, "Cognitive radar antenna selection via deep learning", *IET Radar, Sonar and Navigation*, 2019, vol. 13, n. 6, pp. 871-880.

[7] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks", *IEEE Geoscience and Remote Sensing Letters*, 2016, vol. 13, n. 1, pp. 8-12.

[8] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures with Convolutional Neural Network", *IEEE Access*, 2016, vol. 4, pp. 7125-7130.

[9] M. C. Bellini, and L. Collini, and A. Farina, and D. Pinardi and K. Riabova, "Measurement of Loudspeakers with a Laser Doppler Vibrometer and the Exponential Sine Sweep Excitation Technique", *Journal of Audio Engineering Society*, vol. 65, n. 7/8, pp. 600-612, 2017.

[10] L. L. Beranek, and T. Mellow, "Chapter 6 - Electrodynamic loudspeakers" in "Acoustics: Sound Fields and Transducers", 2012, pp. 241-288, *Academic Press*.

[11] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique",*108th Audio Engineering Society Convention*, February 2000, Paris, FR.

[12] M. A. Richards, and J. A. Scheer and W. A. Holm, "Principles of Modern Radar", vol. 1, *SciTech Publishing, Incorporated*, 2010.

[13] Audiomatica S.r.l., CLIO poket 2.0, 2001.

[14] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, 1997, vol. 9, n. 8, pp. 1735-1780.

[15] R. Pascanu, T. Mikolov and Y. Bengio, "On the difficulty of training recurrent neural networks", *Proceedings of the 30th International Conference on Machine Learning*, 2013, vol. 28, n. 3, pp. 1310-1318.

[16] A. Graves and j. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures", *Neural Networks*, 2005, vol. 18, n. 5-6, pp. 602-610.

[17] I. Goodfellow and Y. Bengio and A. Courville, "Deep Learning", *MIT Press*, 2016. Available at https://www.deeplearningbook.org.

[18] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization", *International Conference on Learning Representations (ICLR) 2014*, April 2014.

[19] Y. Lecun and Y. Bengio and G. Hinton, "Deep learning", *Nature*, 2015, vol. 521, n. 7553, pp. 436-444.