

Noise Reduction in Industry Based on Virtual Instrumentation

Radek Martinek¹, Rene Jaros¹, Jan Baros¹, Lukas Danys¹, Aleksandra Kawala-Sterniuk²,
Jan Nedoma^{3,*}, Zdenek Machacek¹ and Jiri Koziorek¹

¹Faculty of Electrical Engineering and Computer Science, Department of Cybernetics and Biomedical Engineering, VSB–Technical University of Ostrava, 708 00, Ostrava-Poruba, Czechia

²Faculty of Electrical Engineering, Automatic Control and Informatics, Opole University of Technology, Opole, Poland

³Faculty of Electrical Engineering and Computer Science, Department of Telecommunications, VSB–Technical University of Ostrava, 708 00, Ostrava-Poruba, Czechia

*Corresponding Author: Jan Nedoma. Email: jan.nedoma@vsb.cz

Received: 03 February 2021; Accepted: 13 April 2021

Abstract: This paper discusses the reduction of background noise in an industrial environment to extend human-machine-interaction. In the Industry 4.0 era, the mass development of voice control (speech recognition) in various industrial applications is possible, especially as related to augmented reality (such as hands-free control via voice commands). As Industry 4.0 relies heavily on radiofrequency technologies, some brief insight into this problem is provided, including the Internet of things (IoT) and 5G deployment. This study was carried out in cooperation with the industrial partner Brose CZ spol. s.r.o., where sound recordings were made to produce a dataset. The experimental environment comprised three workplaces with background noise above 100 dB, consisting of a laser/magnetic welder and a press. A virtual device was developed from a given dataset in order to test selected commands from a commercial speech recognizer from Microsoft. We tested a hybrid algorithm for noise reduction and its impact on voice command recognition efficiency. Using virtual devices, the study was carried out on large speakers with 20 participants (10 men and 10 women). The experiments included a large number of repetitions (100 times for each command under different noise conditions). Statistical results confirmed the efficiency of the tested algorithms. Laser welding environment efficiency was 27% before applied filtering, 76% using the least mean square (LMS) algorithm, and 79% using LMS + independent component analysis (ICA). Magnetic welding environment efficiency was 24% before applied filtering, 70% with LMS, and 75% with LMS + ICA. Press workplace environment efficiency showed no success before applied filtering, was 52% with LMS, and was 54% with LMS + ICA.

Keywords: 5G; hybrid algorithms; signal processing; speech recognition



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Spoken word is still one of the most natural ways to directly transfer information between people [1,2]. Direct voice interaction with computers and machinery is slowly gaining in importance, as the industry is shifting towards the Industry 4.0 concept [3]. Voice communication systems are increasingly integrated into both industry and private life due to their significant benefits. Most applications are currently limited to a small set of tasks performed by specific machines using predetermined commands that they can recognize [4]. These systems are particularly useful when the operator must do several things at once.

In smart homes, control takes place between the device and the local gateway by power line communication (PLC), Transmission Control Protocol (TCP), or Message Queue Telemetry Transport (MQTT) protocol, which enables control through suitable clients (e.g., a smartphone or Amazon Alexa) [5].

A digital voice system was also designed for use with the internet of things (IoT) to control and simulate the process of an assisted robotic load as part of better human-robot interaction (HMI) [6].

He et al. [7] designed a well-known Arduino board with motion sensors and an audio receiver to control a robotic car using a cloud server and IoT technology, using preset voice commands integrated with Google Voice API.

Industrial applications often rely on radio-frequency technologies, such as Wi-Fi, IoT technologies (e.g., SigFox or LoRa), or broadband cellular networks (4G and 5G). In particular, 5G is often considered as a tool for artificial intelligence (AI), Industry 4.0, and IoT. New communication standards are resilient, and designed with smart sensors or devices and machine communication in mind. Its latency is much less than that of 4G, and it is considerably faster. Reliability is reaching levels of wired connections, but this is limited by the vast scope of manufacturing plants and Industry 4.0. However, 5G opens new fields of application in industry where conventional Wi-Fi fell short in the past. IoT and Industry 4.0 sensors are ever expanding—it is estimated that in 2022, almost 70 billion devices will be connected to IoT networks. Cellular networks often offer unparalleled coverage and scalability with robust and reliable connection. Manufacturers are expanding their plants by employing smart sensors to track cargo or employees, or to gather manufacturing data. The potential value can fuel the rise of automated factories. Automated vision quality checks, augmented reality construction, predictive maintenance, system-wide real-time processing control, and automated guided vehicles are the future of Industry 4.0. While low-power, wide-area solutions are sufficient for some simple connected devices, the opposite is true in manufacturing, where machines are data-intensive and in close proximity. The power of modern software-defined networks and the scalability supported in 5G networks offer a more agile and efficient model based on software rather than traditional hardware solutions. Virtual networks (or network slicing) and subnets adjusted to specific needs are possible with 5G. The 3rd Generation Partnership Project (3GPP) is working on a 17th release of mobile networks (expanding 4G and 5G), which was to be completed in 2022–2023, but may arrive later due to the ongoing COVID pandemic. Various teams are focusing on industrial deployment of 5G networks, including industrial IoT enhancement, IoT over non-terrestrial networks, dynamic power saving, Narrowband IoT enhancement, proximity-based services, and 5G-based location services. While the presented paper focuses on direct data gathering, it can be considered a work in progress.

A combination of sensors such as Raspberry Pi, cameras, and other inputs (voice, text, and visual) were used to facilitate laboratory communication [8]. While studying the implementation

of voice control of operational and technical functions in the virtualization of a production line, Kennedy et al. [9] investigated a passive attack called a fingerprinting attack and found that it may be possible to correctly derive up to 33.8% of voice commands just by eavesdropping on encrypted traffic.

Objectives of this work are as follows:

- Control assurance of operating and technical functions in the production line (production line on/off, arm activation, belt on, laser welder on/off, magnetic welder on/off, press on/off);
- Particular command recognition assurance for the control of operational and technical functions in the production line;
- Provision of a data connection between speech recognition technologies;
- Additive noise suppression in speech signals using the least mean square (LMS) algorithm and the independent component analysis (ICA).
- Ensuring the highest possible efficiency in voice command recognition in a real environment with additive noise.

2 Related Work

Speech signal processing is a promising research area. Automatic speech recognition, synthetic speech, and natural language processing will have a significant impact in business and industry [10–12].

The most important problems are related to automated or semi-automated equipment control (e.g., heating, cooling, lighting, ventilation, and air conditioning). Amrutha (MATLAB implementation) is one of the most successful tools used for spoken word identification in industrial environments, with success rates of up to 90% [13], and Kamdar and Kango are often used for smart home appliances [14–17].

Voice interaction is the most natural form of human interpersonal communication. Direct voice commands make it easy to control smart devices without time-consuming training. However, such a system controllable from multiple locations, requires a smart array of microphones and speakers connected to a centralized processor unit [18].

Automatic Speech Recognition (ASR) can be divided into three basic groups [19,20]:

- Isolated word recognition systems (voice commands are used separately, such as in banking or airport telephone services);
- Small systems for application commands and controls;
- Large systems for continuous speech applications.

As regards ASR, systems applied directly in industry are a mix of the second and third groups, which employ grammatically limited commands for administration and control purposes [19]. ASR systems can also be classified by voice interaction into two categories [18]:

- Specific control applications, which create the essence of smart homes (voice control of operational and technical functions and devices);
- General voice applications, which can be used in all ASR systems.

Obaid et al. [21] showed the system's broad applicability not only in industry but for personal use. Our proposed system consists of voice recognition and wireless systems, implemented with LabVIEW software and ZigBee modules, respectively. The system's greatest advantage is that it

must be trained only once. The required operations are performed based on the data received and stored in the wireless receiver, which is connected directly to the device.

A similar system, designed by Thakur et al. [22], can be used as a stand-alone portable unit to wirelessly control lights, fans, air conditioners, televisions, security cameras, electronic doors, computer systems, and audiovisual equipment [18].

Boeing is incorporating ASR in the new X-32 Strike Fighter aircraft, making it easier for the pilot to control the aircraft and focus on higher-priority aspects of a mission [10,23].

It is possible to regulate multiple factors of ASR systems, mainly speech variability, which is generally of limited use. The flexibility of the language can be limited by a suitable grammatical design. The ability to accurately recognize captured speech depends primarily on the size of the dictionary and the signal-to-noise ratio (SNR). Thus, recognition can be improved by reducing the vocabulary and by improving the SNR. Vocabulary restrictions in Voice Intelligence systems are based on the specific grammar. Reducing vocabulary, such as by shortening individual commands, can significantly improve recognition [24,25]. The quality of captured speech also affects recognition accuracy [26].

Real-time response is another requirement. Three aspects affect system performance [27]:

- Recognition speed;
- Memory requirements;
- Recognition accuracy.

It is challenging to combine all three aspects, as they tend to conflict with each other; e.g., it is relatively easy to improve recognition speed while reducing memory at the expense of accuracy of recognition.

ASR systems can also be found in industrial applications such as robotics, where today's powerful, inexpensive microprocessors and advanced algorithms control commercial applications in the areas of computer interaction, data entry, speech-to-text conversion, telephony, and voice authentication. Robust recognition systems for control and navigation are currently available in personal computers [22].

ASR systems are widely used in other fields, such as wheelchair management [28], defense and aviation [29], and telecommunications.

The IoT platform [30] within a cyber-physical system [31] can be understood as a combination of physical [32], network [33], and computational processes [34,35], and is important in simultaneous voice recognition.

Speech contains information usually obtained by processing a speech signal captured by a microphone through sampling, quantization, coding [36], parameterization, preprocessing, segmentation, centering, pre-emphasis, and window weighting [37,38].

Further steps include speech recognition, whose implementation may include the following:

- A statistical approach for continuous speech recognition using perceptual linear prediction (PLP) of speech [39–42], such as:
 - Audio-to-visual conversion in MPEG-4 [43];
 - Acoustic element modeling and extraction [44];
 - Speech detectors [45] or training of hybrid neural networks for acoustic modeling in automatic speech recognition [46].

- RASTA (RelAtive SpecTrAl) method [36]
- Mel-frequency cepstral analysis (MFCC), such as:
 - Reduction of the pathological system of voice quality evaluation dimensions [47];
 - Detection of clinical depression in adolescents [48];
 - Smart wheelchair speech recognition [49];
 - Speech recognition using spoken word signals [50].
- Hidden Markov models (HMMs) [51]
- Artificial neural networks (ANNs) [52], such as:
 - Automatic speech recognition (ASR) for speech therapy and phased patients [53];
 - Rapid adaptation of neural networks based on speech recognition codes [54];
 - A combination of the functions of the HMM/MLP hybrid system and HMM/GMM speech recognition system [55];
 - Hybrid systems for continuous speech recognition HMM, MLP, and SVM [56].
- Suppression of additive noise using single-or multi-channel methods [57], such as:
 - Speech enhancement using spectral subtraction algorithms [58];
 - Complex adaptive signal processing methods [59,60];
 - Model speech enhancement [61,62];
 - Improved removal of additive noise by spectral subtraction [63];
 - Reduction of speech signal noise by wavelet transform with adjusted universal threshold value [64] or suppression of speech signals by wavelet transform [65].
- Multi-channel methods, including:
 - Least medium quadrature (LMS) algorithm [66,67];
 - Recursive least squares (RLS) [68,69];
 - Independent component analysis (ICA) [70,71];
 - Principal component analysis (PCA) [72,73] or beamforming (BF) for speech acquisition in noisy environments [74] for linearly constrained adaptive beamforming [75] with a robust algorithm [76].

2.1 Classification of Speech Signal Processing Methods

Algorithms are applied to improve the quality of speech signals before processing them in speech recognition applications. These algorithms increase the intelligibility of speech signals and suppress interference while minimizing the loss of useful information. They can be categorized as adaptive or non-adaptive methods.

2.1.1 Adaptive Methods

Adaptive methods use a learning system that changes coefficients based on the working environment. They rely on continuous adjustments of control parameters influenced by fluctuations of environment or input and auxiliary signals. The basic element is feedback, which is used to adjust the parameters of the filter. These methods use a speech-to-noise signal as an input. The noise signal is used as a reference, which is subtracted to filter the speech signal. There are two categories of adaptive methods [66,77]:

- Linear filters are derived from a linear time-invariant system or one to which the principle of superposition applies. These include the Kalman filter, LMS, RLS, and the adaptive linear neuron (ADALINE).

- Nonlinear filters are not subject to the principle of superposition. They include the adaptive neuro-fuzzy inference system (ANFIS), multi-layer neural networks, and evolutionary algorithms.

2.1.2 Non-Adaptive Methods

Non-adaptive methods do not apply a learning system, and hence require no reference signal containing only noise. A speech signal with noise is sufficient. These methods can be categorized as follows:

- Multi-channel methods perform sensing using multiple microphones, where the primary one acquires the noisy speech signal and the others pick up only interference. Two or more channels may sense noisy speech signals in different places. Methods include ICA, PCA, singular value decomposition (SVD), and periodic component analysis (μ CA) [58,77].
- Single-channel methods require only one channel, with input consisting of a speech signal contaminated by interference. Interference suppression is based on the characteristics of the useful signal and the interference. These systems are simpler and less costly than multi-channel methods. They assume that the useful signal (speech) and background interference have different characteristics. They use the calculation of the frequency spectrum from sub-segments of the signal. Their effectiveness is usually limited because of non-stationary interference. Methods include frequency selective filters of the finite impulse response (FIR) and infinite impulse response (IIR) type, methods based on Wiener filtering theory, spectral subtraction using the fast Fourier transform (FFT), wavelet transform (WT), and empirical modal decomposition (EMD) [77].

2.2 Comparison of Speech Signal Processing Methods

Tab. 1 shows the advantages and disadvantages of basic speech signal processing methods.

In 2010, Borisagar et al. [78] tested the adaptive LMS and RLS algorithms for real-time speech signal processing. Both achieved much higher accuracy in MATLAB simulations than a fixed filter designed by conventional methods. In addition, LMS has a simple structure and is easy to implement. Its main disadvantage is slower convergence, but it requires much less memory than RLS.

Wang et al. [79] introduced a method in 2011 based on spectral reading using a multi-channel LMS algorithm. They performed recognition experiments on a distorted speech signal simulated by convolution of multi-channel impulse responses with pure speech. The method's error was 22.4% less than that of conventional cepstral mean normalization. When improved using beamforming, the error was 24.5% less than that of conventional cepstral mean normalization with beamforming. The test was focused on analysis of individual words with a duration of about 0.6 s.

In 2008, Cole et al. [63] applied different widths of the Hanning window and FFT signal conversion to the frequency domain to perform spectral readings using spectral subtraction, i.e., subtracting the noise spectrum from the spectrum of a speech signal contaminated with additive noise, assuming no correlation between the signals. The signal was further modified to form blocks called micro-segments. After conversion to the frequency domain, the interference component was removed by spectral subtraction and the signal was converted using an inverse FFT in the time domain. Testing used a speech signal with digitally added vacuum cleaner noise. Based on the SNR calculation, the best result was obtained using a Hanning window with a width of 256 points. However, the method can be considered inappropriate, as it is necessary to monitor the

amount of input information. The method's effectiveness depends on the determination of the noise spectrum, which is difficult in real conditions and unsuitable in a very noisy environment.

Table 1: Advantages and disadvantages of above methods for interference removal

Method	Advantages	Disadvantages
LMS	Simplicity	Lower speed
RLS	Optimization speed	Computational complexity
Spectral subtraction (FFT)	Simplicity	Necessity for noise spectrum separation; unsuitable for large interference
DWT	Information in time and frequency domain due to the division of components; easier for filtering non-stationary signals	Challenging to calculate stationary signals in the frequency domain
ICA	Noise detection accuracy	Computational complexity; Gaussian-type noises cannot be introduced
PCA	Sorting components according to their energy	Computational complexity; Gaussian-type noises cannot be introduced
μ CA	Robust for noise and filtering	Need for signal periodicity; Gaussian-type noises cannot be introduced

In 2009, Mihov et al. [65] employed the WT to reduce speech signals contaminated by interference. Files from a test database containing 720 male voice recordings were sampled at 25 kHz. Noise was added to the speech signal with SNRs of 0, 5, 10, and 15 dB. Due to its computational complexity, sym3 and higher (Symlet wavelet) were unusable for interference reduction in real-time. The best properties were shown by db3 and db5 (Daubechies wavelets), with a maximum SNR improvement of 14 dB.

Aggarwal et al. [64] used the DWT algorithm in 2011 to reduce interference, applying both soft and hard thresholding. Analysis was carried out on a speech signal contaminated with noise at SNR levels of 0, 5, 10, and 15 dB. The soft thresholding method provided better results at all measured levels of SNR input, and the maximum performance improvement was 35.16 dB. The hard threshold reached a maximum improvement of 21.71 dB.

In 2003, Visser et al. [70] analyzed the efficiency of the ICA method in automobiles. Driving at 40 km/h, the driver spoke a sequence of numbers while the passenger spoke on a mobile phone and the radio and heater were turned on. Stereo microphones on either side of the rearview mirror (15 cm apart) were used for recording, and recorded data were sampled at 8 kHz. The SNR of the mixture recorded by the microphone on the driver's side ranged from 2 to 5 dB. The recognition success rate was 46.9% before applying ICA separation. After using ICA, the success rate increased to 72.8%. The best results were achieved by a combination of the ICA and WT methods, where the recognition success rate was 79.6%. In the same year, Visser et al. [71]

examined the effectiveness of the ICA in a room ($3 \times 4 \times 6$ m) with two directional microphones placed 10 cm apart. Speakers placed in the four corners of the room generated spatially distributed noise. Two other speakers were placed 30 cm from the microphones. The first speaker transmitted a sequence of numbers, and the second transmitted interference consisting of prerecorded words. The SNRs of the mixtures recorded by the microphone were in the range of 5, 0, 5, and 10 dB. The recognition success rate was up to 49.34%, and this was improved to 84.89% through the ICA method.

In 2010, Kandpal et al. [80] used the PCA algorithm for both speech recognition and speech separation. A recording of seven voices, which was 2 s long, with a sampling frequency of 8 kHz, was used for analysis. Based on the correlation coefficient, they evaluated the output of the PCA method against these seven voices and concluded that the probability of a match between the PCA output and the voices was around 0.8.

In 2001, Saul et al. [81] developed the μ CA method for speech recognition. The algorithm had four phases. First, they used the eigenvalue method to combine and amplify weak periodic signals. They used a Hilbert transform to adjust the phase changes across the channels. They used effective sinus seizures to measure the periodicity. They performed a hierarchical analysis of the information through different frequency bands. The experiment was performed on synthetic data at a sampling frequency of 8 kHz. They showed that the μ CA method enabled extraction of the required signal segment from different parts of the frequency spectrum, and that the method is also effective on signals with an SNR input of 20 dB. They mentioned that the μ CA method is quite resistant to noise and filtering.

3 Applied Mathematical Methods

Based on the above studies that use advanced signal processing methods for speech filtering, the ICA method combined with an adaptive LMS algorithm was selected for interference suppression. A thorough study of the literature indicates that these methods provide promising results in various applications. We describe the selected methods below.

3.1 Independent Component Analysis

The independent component (ICA) method is a possible solution to the “cocktail-party problem,” as it can detect hidden factors that are the bases of groups of random variables, measurements, or signals. It is a multi-channel method, where two or more signals are converted to its input. The ICA is often used for analysis of a highly variable data from a large sample database. The variables are considered as linear mixtures of some unknown hidden variables, with no known mixing system. Hidden variables are considered to be non-Gaussian and independent, so they are called independent components of the observed data. Also called sources or factors, they can be found by ICA. Before applying this method, data preprocessing is necessary using centering (creating a vector with zero mean value) and whitening (creating uncorrelated data with unit variance). Eq. (1) represents the measured signals using microphones, where the matrix represents the mixing matrix \mathbf{A}_{mix} (e.g., environment and distance of the microphone from the source signals) and \vec{s} represents the contained source signals. The ICA method enables implementation of Eq. (2), for which it needs to estimate a matrix \mathbf{W} that is the inverse of \mathbf{A}_{mix} [70–72].

$$\vec{x} = \mathbf{A}_{mix} \cdot \vec{s}, \quad (1)$$

$$\vec{s} = \mathbf{W} \cdot \vec{x}. \quad (2)$$

An algorithm derived from ICA, called FastICA, is often used to solve such problems. It has four steps. A random vector \vec{x} is created, and kurtosis is calculated using Eq. (3), where \vec{w} is the vector of weights g and is the derivative of the non-quadratic function G . The data are standardized, and a scalar product is calculated between the new vector \vec{w} and its counterpart from the previous iteration. These steps are repeated until the scalar product reaches a value smaller than the selected convergence criterion, or the maximum number of iterations is reached. When working with FastICA, it is necessary to select the convergence criterion, maximum number of iterations, and number of output components, which is given by the number of source signals we are trying to estimate. In the case of speech processing, at least two components are used, where one should contain the speech itself, and the other only noise [70–72].

$$\vec{w}^+ = E \left\{ \vec{x}g \left(\vec{w}^T \vec{x} \right) \right\} - E \{ g'(\vec{w}^T \vec{x}) \} \vec{w}. \quad (3)$$

3.2 Least Mean Squares Filter

The LMS algorithm is currently one of the most widely used adaptive algorithms. Its main strength lies in its mathematical simplicity. Adaptive algorithms are in general used in unknown environments because they can adjust their coefficients based on varying circumstances. They are based on a gradient search algorithm, or maximum gradient method. The dependence of the standard deviation of the output error signal of the adaptive FIR filter on the filter coefficients is a quadratic curve with one global minimum. The basis of the adaptive algorithm is the calculation of the error function $\vec{e}(n)$ using Eq. (4), where $\vec{d}(n)$ is the required output and $\vec{y}(n)$ is the real output. The output of each iteration of the LMS algorithm is defined according to Eq. (5) and its modification, Eq. (6). Filter recursion (adjustment of filter weights) is given according to Eq. (7), where μ is the step size of the adaptive filter (which greatly affects the convergence rate), $\vec{w}(n)$ is the vector of filter coefficients, and $\vec{x}(n)$ is the input vector. These steps are repeated in each iteration until convergence is achieved. Another important parameter of the LMS algorithm is the order of the filter M , which has a significant effect on the computational complexity [67–69,78,82].

$$\vec{e}(n) = \vec{d}(n) - \vec{y}(n), \quad (4)$$

$$\vec{y}(n) = \vec{w}(n) \vec{x}(n), \quad (5)$$

$$\vec{y}(n) = \vec{w}^T(n) \vec{x}(n), \quad (6)$$

$$\vec{w}(n+1) = \vec{w}(n) + 2\mu \vec{e}(n) \vec{x}(n). \quad (7)$$

These algorithms require fewer demanding mathematical operations than RLS algorithms. Furthermore, they are one order less in complexity, and are therefore faster. The main disadvantage of LMS is its lower performance in time-varying environments and lower convergence speed [67,69,70,78].

4 Experiments

Five experiments were conducted in laboratory or real conditions to verify the above technologies. Five scenarios were evaluated by software-based simulations. The interference models were combined with audio recordings of individual commands to test speech processing methods in different conditions.

4.1 Applied Hardware

The measuring equipment consisted of a professional Steinberg UR44 sound card and four connected Rode NT5 microphones. The devices were controlled by a PC via virtual instrumentation-based software.

The Steinberg UR44 professional sound card [83] is primarily used for music and audio. It has four inputs for microphones or musical instruments. It supports various communication standards of the audio industry, e.g., ASIO, WDM, and Core Audio. Standardized values of sampling frequencies in the range of 44.1 to 192 kHz can be selected, with a resolution up to 24 bits. The sound card supports the supply of phantom power for connected microphones, from +24 to +48 VDC.

The Rode NT5 microphone [84] is a compact device that can be connected via an XLR connector. The 1/2" diaphragm consists of an externally deflected capacitor. The membrane is gold-plated, which improves its properties. The microphone has cardioid directional characteristics, and a frequency range of 20 Hz and 20 kHz (corresponding to the range of human hearing). The microphone must be connected to the input of a sound card supporting phantom power.

4.2 Software

Connectivity with the Steinberg UR44 sound card was maintained by the Audio Stream Input/Output (ASIO) audio communication standard. The LabVIEW graphical environment was chosen as the programming environment due to its high modularity and availability of usable libraries, including ASIO API, which is part of the WaveIO library [85].

The software was required to be as modular as possible, so that it could be used in various experiments and scenarios with minimal modification. The application was therefore designed in accordance with the queued message handler (QMH) design pattern [86]. The chosen architecture allows the consideration of each microphone as a separate measuring unit, eliminating the need to make large code changes when expanding the application.

A commercially available recognizer integrated into the Windows operating system was chosen as the speech signal recognizer. To enable its communication with LabVIEW required the installation of Speech SDK 5.1. The recognizer converted voice commands to text, which could then be used for either synthesis or further processing of the speech signal. A disadvantage is the limited database of languages (e.g., English, Chinese, French, and German), where local unexpanded languages (e.g., Czech and Slovak) are not supported. It was therefore necessary to set the Windows environment to a supported language. LabVIEW can be obtained through the freely available Speech Recognition Engine library. A conceptual diagram of the speech recognizer system can be seen in Fig. 1.

To have a speech signal modified by adaptive filtering before transmission to the recognizer, it is necessary to adjust the routing of the signal. Since the integrated Windows speech recognizer runs in the background of the operating system as a service, it is not possible for users to choose anything other than the input (such as LabVIEW or speech recording). The signal routing can be adjusted using SW (in our case, the VB-Cable [87]), which emulates both inputs and outputs of the sound card. The adjustment can be seen in the block diagram in Fig. 2.

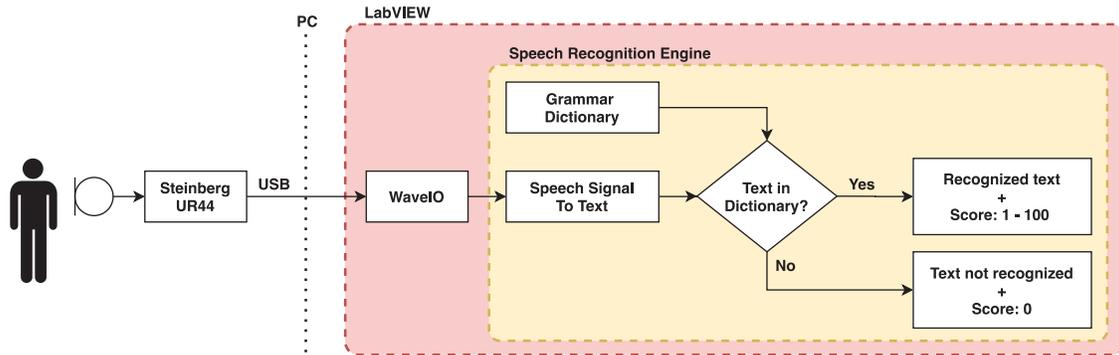


Figure 1: Concept diagram of speech recognizer system implemented in Windows

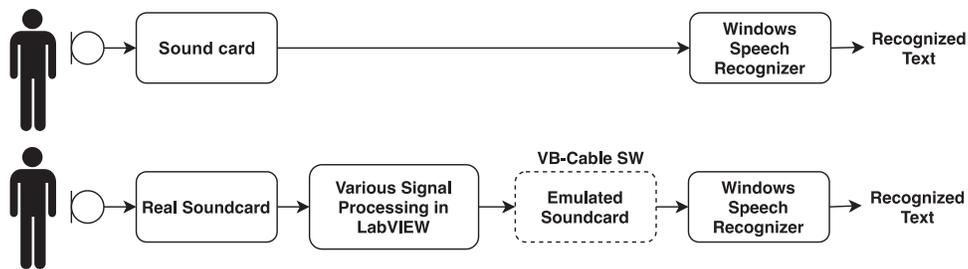


Figure 2: Principle of signal routing adjustments

4.3 Measurements of Interference Signals During Production Line Operation

The measurements were taken at Brose CZ spol. s.r.o. (Koprivnice, Czech Republic), which produces seat structures, electric motors, drives, and locks for rear and side doors of vehicles. Interference sources of the laser welder, magnetic welder, and press were measured. The primary microphone (index 0) was in the area where the operator directly operated the devices. The reference microphones (index 2 and 3) were on the sides of the device, and the reference microphone with index 1 was placed in back of the device (see Fig. 3).

As mentioned above, the ICA method and adaptive LMS method were chosen for noise suppression.

4.3.1 Setting up the LMS Algorithm

Offline identification was necessary since it was impossible to determine the exact command chains in advance. This was performed by gathering the ideal values for each voice command and interference selection in accordance with the global SNR. Based on the measured values, the best filter length M and convergence constant μ were selected. Filtering was carried out in two steps, as seen in Fig. 4. A suppressed speech signal and a reference noise were applied to the input filter, which was a bandpass filter with the interval set to 300–3400 Hz (frequency band of human speech). The filtered signals were input to an LMS algorithm, where $y(n)$ is the filtered signal and $e(n)$ is filtering error.

From Tab. 2 it is clearly seen that the higher the interference energy the greater the requirements for the adaptive filter, i.e., the greater the filter length M and convergence constant μ . We found that during the measurement, the higher the interference energy the longer the required

filter length, which led to distortion of the useful signal (filter length of 1000 or higher), which was partially filtered out. The situation was similar for the convergence constant, where the filter became unstable in the case of higher values (above 0.1). Another problem was the computational time, which was significantly longer with longer filter lengths and smaller convergence constants.

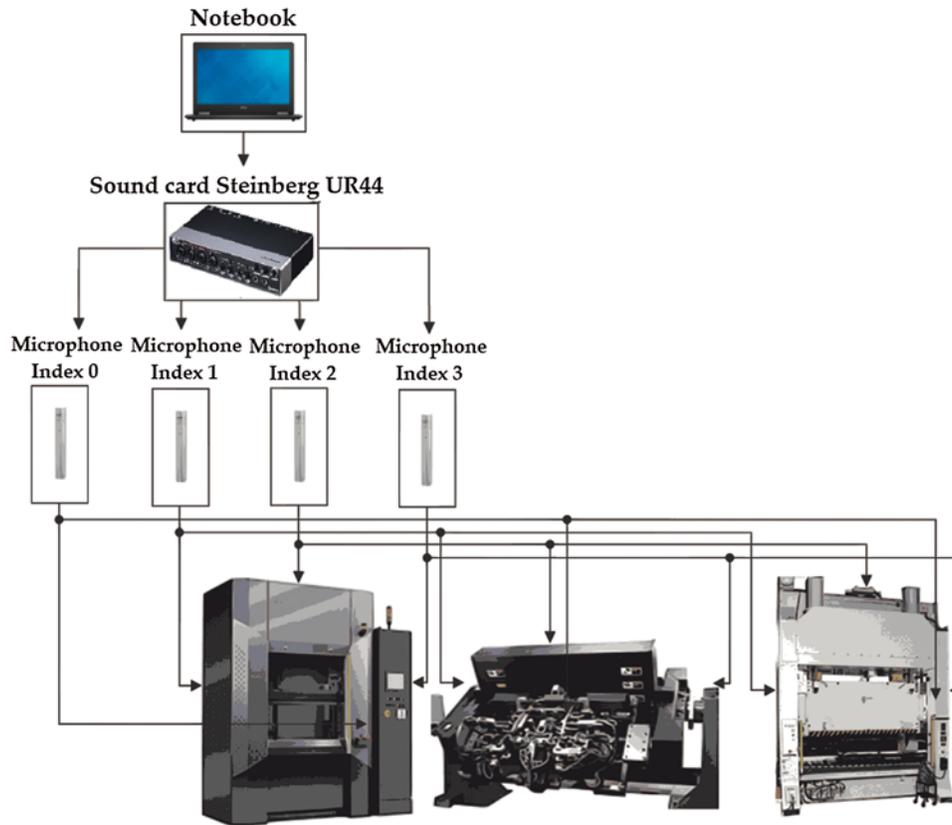


Figure 3: Implemented measurement system in production line

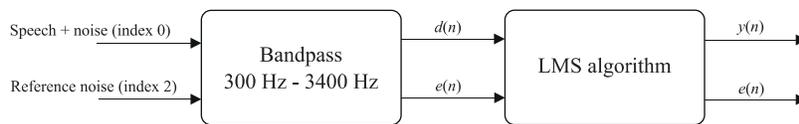


Figure 4: Measurement filtering chain using LMS algorithm

Table 2: Optimal parameter settings for LMS algorithm: production line visualization

Noise source	M	μ
Laser welder	180	0.001
Magnetic welder	910	0.001
Press machine	880	0.01

4.3.2 Independent Component Analysis

The presented study relied on hybrid filtering, and FastICA with two independent components on the output of the adaptive filter. The convergence constant was set to $\mu = 1 \cdot 10^{-6}$, and 1000 iterations were performed (Fig. 5). This specific constant was chosen, since it was impossible to use more than one microphone. However, it was impossible to solve the classic “cocktail party” scenario. The LMS algorithm noticeably suppressed the interference, but the sub-interference increased the filtering error. as partial speech filtering also occurred when suppressing this interference. This well-known feature of adaptive algorithms must be considered.

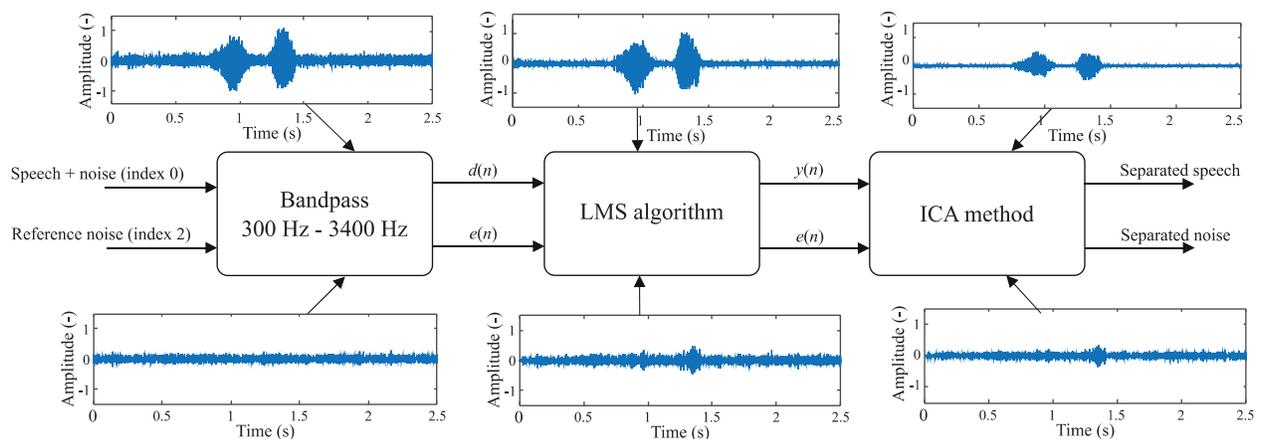


Figure 5: Measurement chain consisting of LMS and ICA hybrid algorithm system

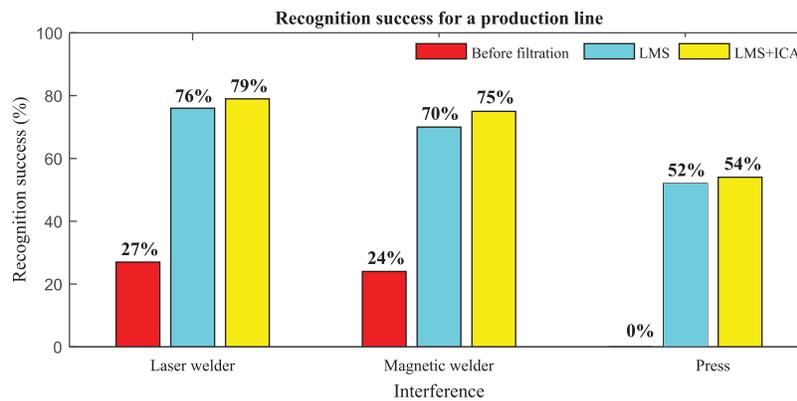
4.3.3 Recognition Success Rate

The recognition success rate was estimated based on the recognized/unrecognized status. One hundred repetitions were performed. During testing, the microphone had to be close to the mouth, mainly due to interference caused by the press. As a result, the sound card cut off the signal in some locations when the maximum resolution was exceeded.

From Tab. 3 and Fig. 6, it can be seen that when the laser welder was measured, the average success rate before filtering was 27%. Two commands, “turn on the middle arm” and “turn on the left arm,” were not recognized at all. After filtering, the average success rate of the LMS algorithm and ICA method were 76% and 79%, respectively. The average success rate of the magnetic welder was 24% before filtering, the least successful recognition being for the “turn on the left arm” command. After application of the filtering algorithms, the average success rates for the LMS method and ICA were 70% and 75%, respectively. While the LMS + ICA combination worked well in most scenarios, there were some exceptions. For example, the command “turn off the laser welder” achieved better results using only the LMS algorithm, and the signal deteriorated by up to 30% when the hybrid algorithm was used. The results from the press machine clearly show that the system was not able to distinguish the commands without LMS or hybrid algorithms. The interference energy was so high that it was necessary to place the microphone right next to the mouth to amplify the useful signal. Nevertheless, the average success rate after using the LMS algorithm was 52%, while the ICA method achieved 54%. This is an extreme scenario where the robustness of both methods was tested.

Table 3: Order recognition success for different types of production lines

Command	Production line								
	Laser welder			Magnetic welder			Press machine		
	No filtering (%)	LMS (%)	LMS + ICA (%)	No filtering (%)	LMS (%)	LMS + ICA (%)	No filtering (%)	LMS (%)	LMS + ICA (%)
Homing	57	90	100	30	72	100	0	56	63
Turn off the production line	51	70	81	54	74	69	0	48	51
Turn on the main arm	18	85	87	18	63	72	0	45	39
Turn on the right arm	9	66	84	12	57	74	0	42	45
Turn on the middle arm	0	63	72	10	66	70	0	38	36
Turn on the left arm	0	66	81	3	66	78	0	36	45
Turn off the laser welder	21	100	70	18	100	69	0	100	100
Turn on the belt	60	69	57	50	62	71	0	51	50

**Figure 6:** Average recognition success rates

From the spectrograms (Figs. 7–9) it can be noticed that the filtered signals from both welders have very similar waveforms. Both scenarios reached better success rates when employing the LMS and ICA combination. This is because the ICA method normalizes the signal to half the input signal, thus reducing the high values of interfering signals.

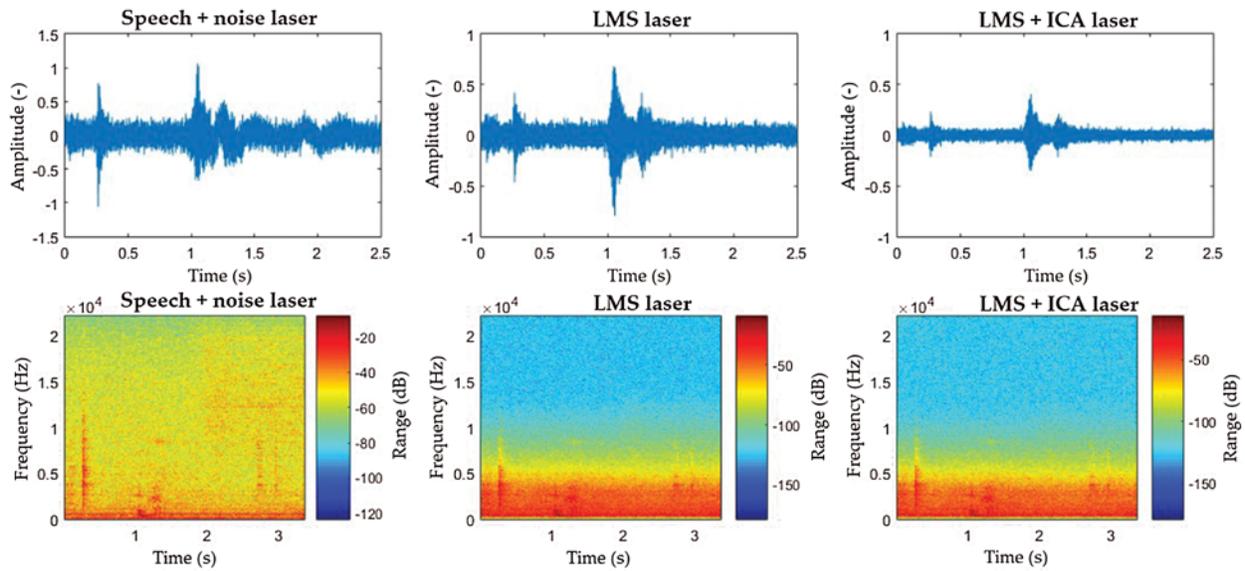


Figure 7: Laser welding machine interference spectrograms: “Homing” command

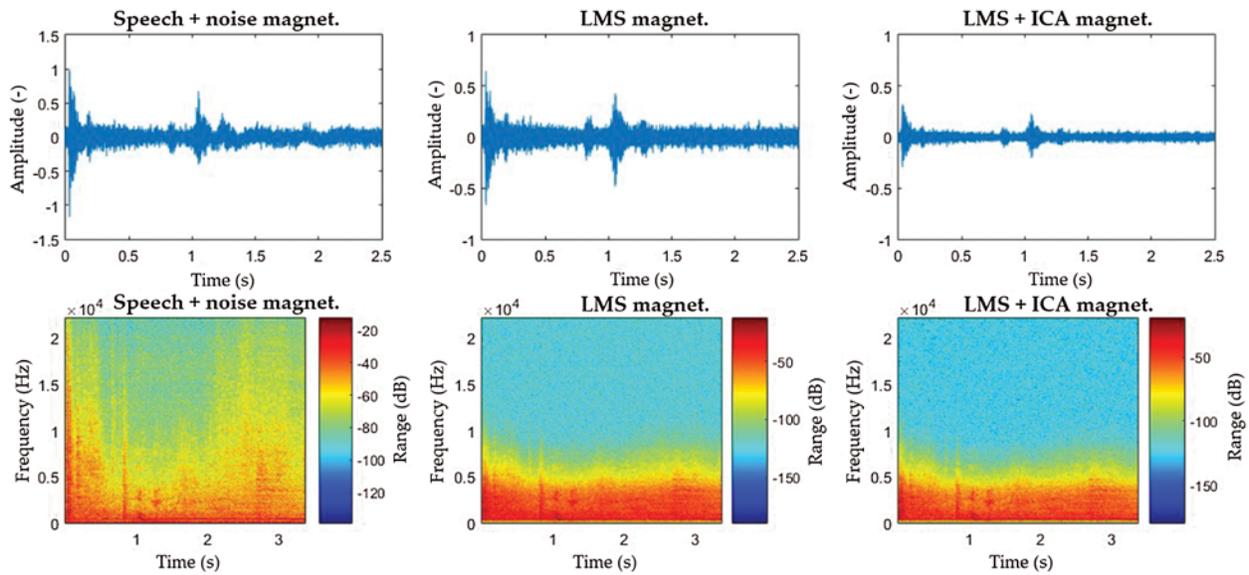


Figure 8: Magnetic welder interference spectrograms: “Homing” command

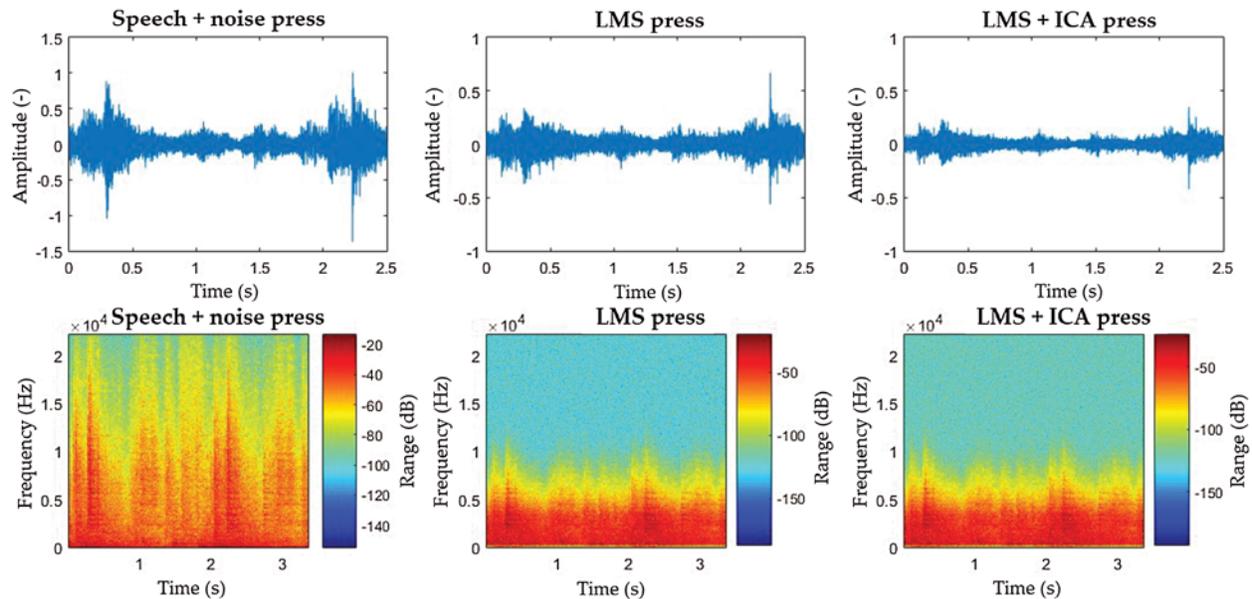


Figure 9: Press machine interference spectrograms: “Homing” command

5 Discussion

The proposed concept of interference reduction can be applied in other industrial areas. The so-called acoustic/mechanical analysis of production, which is part of the predictive maintenance concept, seems to be a promising approach. In cooperation with our industrial partner, Brose CZ spol. s.r.o., some pilot experiments were carried out on laser and magnetic welders, focusing on an acoustic analysis of welding quality (Fig. 10). The experiments focused on acoustic and mechanical analysis of a specific tool, as seen in Figs. 11 and 12. To carry out a reliable acoustic analysis requires the complete elimination of background noise, since it significantly influences the results. The presented early designs seem like an optimal tradeoff between costs and results, and will be the subject of further research.

The presented system can be used in other environments. Workers such as construction personnel, designers, artists, police, and fire fighters can leverage the power of direct voice commands in environments obstructed by noise. The presented algorithms are fully transferable and can be deployed for other uses. Noise reduction can be used in automobiles (to filter vehicular noise), construction (to filter background noise), or even in medicine (to filter life signs of a mother and/or fetus). The system has the advantage that it does not obstruct the worker in any way. Other systems require direct contact with the employee’s body. These solutions are often obtrusive and could potentially influence workers’ capabilities, or even their safety. A wireless and unobtrusive approach mitigates these problems and increases employees’ comfort.

However, some deployment areas might require effective advanced signal processing methods. Apart from the tested LMS and ICA combination, adaptive methods include normalized LMS (NLMS), RLS, QR-decomposition-based RLS (QR-RLS), and fast transversal filtering (FTF). The resulting signal can also be enhanced by post-processing techniques such as the wavelet transform (WT), empirical mode decomposition (EMD), and ensembled EMD (EEMD). Our

future research will seek the combination of the most suitable algorithms, and will also focus on advanced AI techniques.

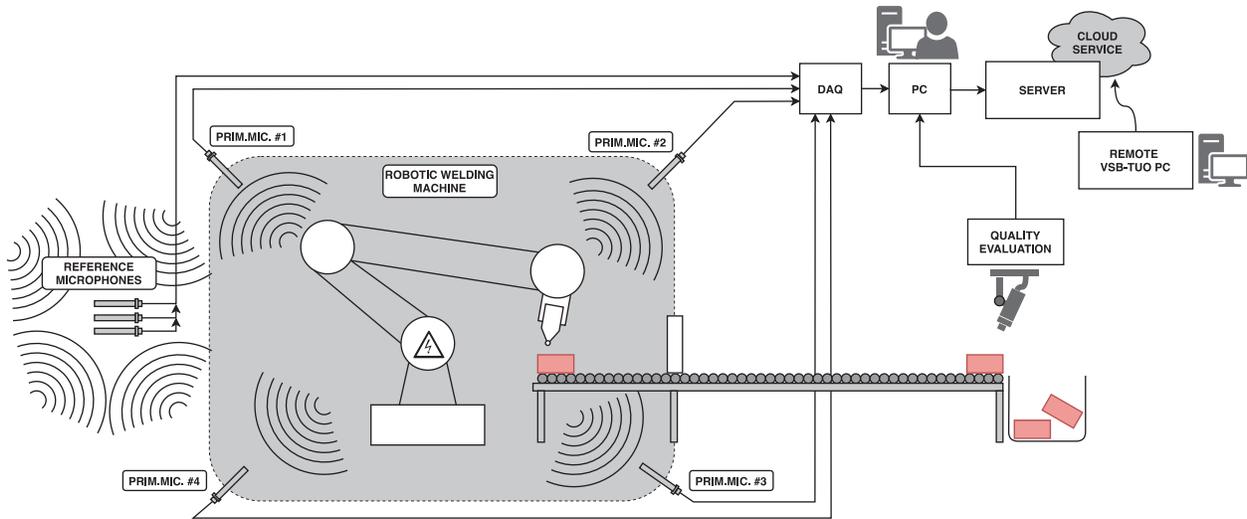


Figure 10: Acoustic analysis of welding process (detection of rejected products)

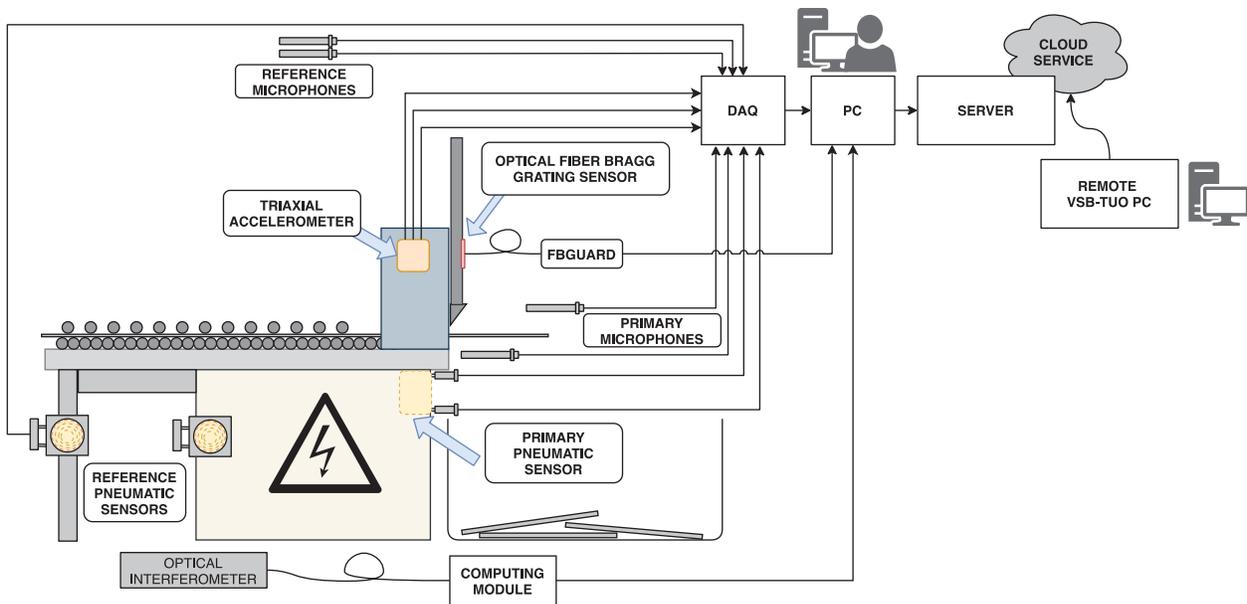


Figure 11: Long-term production line monitoring: cutting machine within press machine (use of artificial intelligence for predictive maintenance)



Figure 12: Collage from pilot measurements

Data in smart factories tend to differ greatly from regular IoT traffic, since they often contain higher amounts of data transferred over shorter periods of time. Manufacturing lines and machinery work at the best quality-time ratio and must maintain the highest effectivity possible. The machinery often contains a vast array of different commands, which report important data or influence precise manufacturing processes. It is therefore important to have a solution that offers higher transmit speeds [88] and the lowest possible latency [89], for which 5G is slowly surfacing as a candidate technology. The network is robust, operates in a licensed spectrum, has low latency, and is partially tailored for industrial deployment. Machine-to-machine (M2M) communication is a critical part of Industry 4.0, and is necessary to maintain coordination between various devices and components. Coordination can be maintained not only within a single plant but across plants. Synchronized machines offer a significant advantage in precise manufacturing processes, such as for automotive applications. Many teams are focused on requirements and challenges for wireless technologies in Industry 4.0. Most machines are currently wired. However, wires can obstruct machine movements, which can lead to malfunction or reject manufacturing. The industry is slowly shifting toward wireless technologies and there are certain necessary requirements needed for seamless transition. Varghese et al. [90] focused on these challenges, focusing on design criteria of latency, longevity, and reliability. The team benchmarked both WiFi and 5G in terms of latency and reliability parameters. Based on their information, a single wireless standard will not address all of the strict requirements of Industry 4.0; however, it too early to reject some technologies, since many are still undergoing revision. Ordonez-Lucena et al. [91] analyzed the newest 3GPP Release 16 specification of 5G and identified a number of deployment options relevant to non-public networks. Their work included a feasibility analysis covering technical, regulatory, and business aspects. They also discussed business models and regulatory aspects.

Based on this information, the presented speech recognition system could be integrated as a part of a 5G Industry 4.0 automated factory. The early concept is shown in Fig. 13. The system could employ an array of wireless microphones to assist in speech recognition. Workers would carry their own reference microphones as a source of voice commands, influenced by background noise. Arrays deployed on machinery could be used as a source of noise. The gathered data would be automatically evaluated on a local server based on various qualitative parameters. Due to the low latency of 5G networks, evaluation and interaction could be seamless. The machinery could therefore be operated by reliable and well-recognized voice commands, increasing workflow and safety. The local non-public network (NPN) can parse relevant data gathered from machines and evaluate advanced functions on auxiliary servers. The results could be distributed through public land mobile networks (PLMN) to other manufacturing plants.

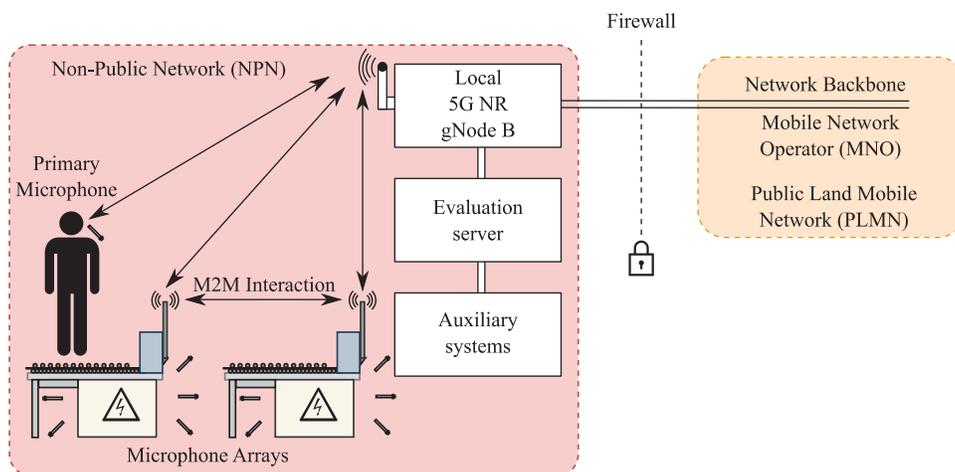


Figure 13: Concept of smart factory with 5G compatible voice evaluation platform

6 Conclusions

We presented innovative methods of speech signal processing for voice control of a production line in Industry 4.0. A commercially available Windows recognizer was used in order to recognize specific commands. The system was based on a commercially available sound card and LabVIEW programming environment. The analyzed data were gathered directly on the production line, making it possible to analyze a laser welder, magnetic welder, and press machine.

The linear adaptive filter LMS and the ICA method were chosen for environmental noise filtering. A dataset of 100 repetitions of each command was used to evaluate the designed system. A total of eight commands were tested in combination with three types of interference. The average recognition success before and after filtering was up to 49% higher in case of the LMS algorithm, and up to 52.3% for the hybrid filtering scenarios.

The overall results showed that the hybrid method had a 5% advantage over a conventional LMS algorithm. However, due to the computational complexity of the ICA method, it is significantly better to implement the LMS algorithm, which is much simpler and offers similar results. As the performance and price of available technology change rapidly, many more powerful algorithms might surface in the coming years.

Acknowledgement: This work was supported by the European Regional Development Fund in Research Platform focused on Industry 4.0 and Robotics in Ostrava project CZ.02.1.01/0.0/0.0/17_-049/0008425 within the Operational Programme Research, Development and Education, and in part by the Ministry of Education of the Czech Republic under Project SP2021/32 and SP2021/45.

Funding Statement: This work was supported by the European Regional Development Fund in Research Platform focused on Industry 4.0 and Robotics in Ostrava project CZ.02.1.01/0.0/0.0/17_-049/0008425 within the Operational Programme Research, Development and Education, Project Nos. SP2021/32 and SP2021/45.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. D. Goldinger, “Words and voices: Episodic traces in spoken word identification and recognition memory,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 22, no. 5, pp. 1166–1183, 1996.
- [2] K. Hendrickson, J. Spinelli and E. Walker, “Cognitive processes underlying spoken word recognition during soft speech,” *Cognition*, vol. 198, no. 4, pp. 1–15, 2020.
- [3] S. Elavarasi and G. Suseendran, “Automatic robot processing using speech recognition system,” in *Data Management, Analytics and Innovation*, 1st ed., vol. 2. Singapore: Springer, pp. 185–195, 2020.
- [4] T. Zoughi, M. M. Homayounpour and M. Deypir, “Adaptive windows multiple deep residual networks for speech recognition,” *Expert Systems with Applications*, vol. 139, no. 10, pp. 1–16, 2020.
- [5] R. Dotihal, A. Sopori, A. Muku, N. Deochake and D. Varpe, “Smart homes using alexa and power line communication in IoT,” in *Int. Conf. on Computer Networks and Communication Technologies*, 1st ed., vol. 15. Singapore: Springer, pp. 241–248, 2019.
- [6] A. P. Naik and P. Abraham, “Arduino based voice controlled robotic arm,” *Robotics*, vol. 1, pp. 1–8, 2020.
- [7] S. He, A. Zhang and M. Yan, “Voice and motion-based control system: Proof-of-concept implementation on robotics via internet-of-things technologies,” in *Proc. of the 2019 ACM Southeast Conf.*, Kennesaw, GA, USA, pp. 102–108, 2019.
- [8] N. J. Knight, S. Kanza, D. Cruickshank, W. S. Brocklesby and J. G. Frey, “Talk2Lab: The smart lab of the future,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8631–8640, 2020.
- [9] S. Kennedy, H. Li, C. Wang, H. Liu, B. Wang *et al.*, “I can hear your alexa: Voice command fingerprinting on smart home speakers,” in *IEEE Conf. on Communications and Network Security*, Washington, DC, USA, pp. 232–240, 2019.
- [10] J. Vajpai and A. Bora, “Industrial applications of automatic speech recognition systems,” *International Journal of Engineering Research and Applications*, vol. 6, no. 3, pp. 88–95, 2016.
- [11] K. Koumpis and K. Pavitt, “Corporate activities in speech recognition and natural language: Another “new science”-based technology,” *International Journal of Innovation Management*, vol. 3, no. 3, pp. 335–366, 1999.
- [12] J. Mocnej, T. Lojka and I. Zolotová, “Using information entropy in smart sensors for decentralized data acquisition architecture,” in *IEEE 14th Int. Symp. on Applied Machine Intelligence and Informatics*, Danvers, IEEE, pp. 47–50, 2016.
- [13] S. Amrutha, S. Aravind, A. Mathew, S. Sugathan, R. Rajasree *et al.*, “Voice controlled smart home,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 5, no. 1, pp. 272–275, 2015.
- [14] R. Martinek, J. Vanus, J. Nedoma, M. Fridrich, J. Frnda *et al.*, “Voice communication in noisy environments in a smart house using hybrid LMS + ICA algorithm,” *Sensors*, vol. 20, no. 21, pp. 1–24, 2020.

- [15] H. Kamdar, R. Karkera, A. Khanna, P. Kulkarni and S. Agrawal, "A review on home automation using voice recognition," *International Research Journal of Engineering and Technology*, vol. 4, no. 10, pp. 1795–1799, 2017.
- [16] R. Kango, P. Moore and J. Pu, "Networked smart home appliances-enabling real ubiquitous culture," in *Proc. 3rd IEEE Int. Workshop on System-on-Chip for Real-Time Applications*, Liverpool, UK, pp. 76–80, 2002.
- [17] I. Zolotová, P. Papcun, E. Kajáti, M. Miškuf and J. Mocnej, "Smart and cognitive solutions for operator 4.0: Laboratory H-CPPS case studies," *Computers & Industrial Engineering*, vol. 139, no. 105471, pp. 1–15, 2020.
- [18] I. V. McLoughlin and H. R. Sharifzadeh, "Speech recognition for smart homes, Speech Recognition," *Technologies and Applications*, vol. 2008, pp. 477–494, 2008.
- [19] L. R. Rabiner, "Applications of voice processing to telecommunications," *Proceedings of the IEEE*, vol. 82, no. 2, pp. 199–228, 1994.
- [20] D. Yu and L. Deng, *Automatic Speech Recognition*, 1st ed., London, UK: Springer, pp. 1–321, 2016.
- [21] T. Obaid, H. Rashed, A. A. El Nour, M. Rehan, M. M. Saleh *et al.*, "ZigBee based voice controlled wireless smart home system," *International Journal of Wireless & Mobile Networks*, vol. 6, no. 1, pp. 47–59, 2014.
- [22] D. S. Thakur and A. Sharma, "Voice recognition wireless home automation system based on Zigbee," *IOSR Journal of Electronics and Communication Engineering*, vol. 6, no. 1, pp. 65–75, 2013.
- [23] D. W. Beeks and R. Collins, *Speech Recognition and Synthesis*, 2nd ed., London, UK: CRC Press, pp. 1–316, 2001.
- [24] H. Chevalier, C. Ingold, C. Kunz, C. Moore, C. Roven *et al.*, "Large-vocabulary speech recognition in specialized domains," in *Int. Conf. on Acoustics, Speech, and Signal Processing*, Detroit, MI, USA, pp. 217–220, 1995.
- [25] C. A. Kamm, C. Shamieh and S. Singhal, "Speech recognition issues for directory assistance applications," *Speech Communication*, vol. 17, no. 3–4, pp. 303–311, 1995.
- [26] H. Sun, L. Shue and J. Chen, "Investigations into the relationship between measurable speech quality and speech recognition rate for telephony speech," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, pp. 865–868, 2004.
- [27] M. K. Ravishankar, "Efficient algorithms for speech recognition," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, 1996.
- [28] A. Rogowski, "Industrially oriented voice control system," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 3, pp. 303–315, 2012.
- [29] C. Spitzer, U. Ferrell and T. Ferrell, *Digital Avionics Handbook*, 3rd ed., London, UK: CRC press, pp. 1–848, 2017.
- [30] V. Marik, *Průmysl 4.0-Výzva Pro Českou Republiku*. Prague, Czechia: Management Press, 2016.
- [31] *Cyber-Physical Systems*. The Ptolemy project. 2021. [Online]. Available: <http://cyberphysicalsystems.org/>.
- [32] B. Mardiana, H. Hazura, S. Fauziyah, M. Zahariah, A. Hanim *et al.*, "Homes appliances controlled using speech recognition in wireless network environment," in *Int. Conf. on Computer Technology and Development*, Kota Kinabalu, Malaysia, pp. 285–288, 2009.
- [33] *Smart Device*. Techopedia. 2015. [Online]. Available: <https://www.techopedia.com/definition/31463/smart-device>.
- [34] M. Schiefer, "Smart home definition and security threats," in *Ninth Int. Conf. on IT Security Incident Management & IT Forensics*, Magdeburg, Germany, pp. 114–118, 2015.
- [35] O. Kyas, "How to smart home," *Tanggal Akses Terakhir*, vol. 3, pp. 1–208, 2013.
- [36] J. Psutka, L. Muller, J. Matousek and V. Radova, *Mluvíme s počítačem Česky*, 1st ed., vol. 111. Prague, Czechia: Academia, pp. 1–746, 2006.
- [37] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

- [38] R. E. Bellman and S. E. Dreyfus, *Applied Dynamic Programming*, 1st ed., Princeton, New Jersey, USA: Princeton University Press, pp. 1–390, 2015.
- [39] A. Kumar, M. Dua and T. Choudhary, “Continuous hindi speech recognition using monophone based acoustic modeling,” *International Journal of Computer Applications*, vol. 24, pp. 1–5, 2014.
- [40] S. J. Arora and R. P. Singh, “Automatic speech recognition: A review,” *International Journal of Computer Applications*, vol. 60, no. 9, pp. 34–44, 2012.
- [41] S. K. Saksamudre, P. Shrishrimal and R. Deshmukh, “A review on different approaches for speech recognition system,” *International Journal of Computer Applications*, vol. 115, no. 22, pp. 23–28, 2015.
- [42] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” *Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.
- [43] L. Xie and Z.-Q. Liu, “A comparative study of audio features for audio-to-visual conversion in mpeg-4 compliant facial animation,” in *Int. Conf. on Machine Learning and Cybernetics*, Dalian, China, pp. 4359–4364, 2006.
- [44] A. Garg and P. Sharma, “Survey on acoustic modeling and feature extraction for speech recognition,” in *3rd Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, pp. 2291–2295, 2016.
- [45] J. Rajnoha and P. Pollak, “Detektory řečové aktivity na bázi perceptivní keprální analýzy,” in *České Vysoké učení Technické v Praze, Fakulta Elektrotechnická Prague*, Czechia, pp. 1–9, 2008.
- [46] G. A. Saon and H. Soltau, “Method and system for joint training of hybrid neural networks for acoustic modeling in automatic speech recognition,” U.S. Patent No. 9,665,823, Washington, DC: U.S. Patent and Trademark Office, pp. 1–18, 2017.
- [47] J. I. Godino-Llorente, P. Gomez-Vilda and M. Blanco-Velasco, “Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters,” *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1943–1953, 2006.
- [48] L.-S. A. Low, N. C. Maddage, M. Lech, L. Sheeber and N. Allen, “Content based clinical depression detection in adolescents,” in *17th European Signal Processing Conf.*, Glasgow, UK, pp. 2362–2366, 2009.
- [49] N. T. Hai, N. Van Thuyen, T. T. Mai and V. Van Toi, “MFCC-DTW algorithm for speech recognition in an intelligent wheelchair,” in *5th Int. Conf. on Biomedical Engineering in Vietnam*, Ho Chi Minh City, Vietnam, pp. 417–421, 2015.
- [50] C. Ittichaichareon, S. Suksri and T. Yingthawornsuk, “Speech recognition using MFCC,” in *Int. Conf. on Computer Graphics, Simulation and Modeling*, Pattaya, Thailand, pp. 135–138, 2012.
- [51] J. Varak, “Možnosti hlasového ovládání bezpilotních dronů,” Bc. thesis, VSB-Technical University of Ostrava, Czechia, 2017.
- [52] M. Cutajar, E. Gatt, I. Grech, O. Casha and J. Micallef, “Comparative study of automatic speech recognition techniques,” *IET Signal Processing*, vol. 7, no. 1, pp. 25–46, 2013.
- [53] N. Jamal, S. Shanta, F. Mahmud and M. Sha’abani, “Automatic speech recognition (ASR) based approach for speech therapy of aphasic patients: A review,” in *AIP Conf. Proc.*, Bydgoszcz, Poland, pp. 1–8, 2017.
- [54] S. Xue, O. Abdel-Hamid, H. Jiang, L. Dai and Q. Liu, “Fast adaptation of deep neural network based on discriminant codes for speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1713–1725, 2014.
- [55] P. Pujol, S. Pol, C. Nadeu, A. Hagen and H. Bourlard, “Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 14–22, 2004.
- [56] E. Zarrouk, Y. B. Ayed and F. Gargouri, “Hybrid continuous speech recognition systems by HMM, MLP and SVM: A comparative study,” *International Journal of Speech Technology*, vol. 17, no. 3, pp. 223–233, 2014.
- [57] A. Chaudhari and S. Dhonde, “A review on speech enhancement techniques,” in *Int. Conf. on Pervasive Computing*, Pune, India, pp. 1–3, 2015.

- [58] N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study," *Procedia Computer Science*, vol. 54, no. 2, pp. 574–584, 2015.
- [59] R. Martinek, "The use of complex adaptive methods of signal processing for refining the diagnostic quality of the abdominal fetal electrocardiogram," Ph.D. dissertation, VSB-Technical University of Ostrava, Czechia, 2014.
- [60] J. Jan, *Číslicová Filtrace, Analýza a Restaurace Signálů*, 2nd ed., Brno, Czechia: Vutium, pp. 1–427, 2002.
- [61] P. Harding, "Model-based speech enhancement," Ph.D. dissertation, University of East Anglia, UK, 2013.
- [62] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed., London, UK: CRC Press, pp. 1–711, 2013.
- [63] C. Cole, M. Karam and H. Aglan, "Increasing additive noise removal in speech processing using spectral subtraction," in *Fifth Int. Conf. on Information Technology: New Generations*, Las Vegas, NV, USA, pp. 1146–1147, 2008.
- [64] R. Aggarwal, J. K. Singh, V. K. Gupta, S. Rathore, M. Tiwari *et al.*, "Noise reduction of speech signal using wavelet transform with modified universal threshold," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 14–19, 2011.
- [65] S. G. Mihov, R. M. Ivanov and A. N. Popov, "Denoising speech signals by wavelet transform," *Annual Journal of Electronics*, vol. 1, no. 6, pp. 2–5, 2009.
- [66] R. Martinek, J. Zidek, P. Bilik, J. Manas, J. Koziorek *et al.*, "The Use of LMS and RLS adaptive algorithms for an adaptive control method of active power filter," *Energy and Power Engineering*, vol. 5, no. 4, pp. 1126–1133, 2013.
- [67] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, 2nd ed., Hoboken, New Jersey, USA: John Wiley & Sons, pp. 1–800, 2013.
- [68] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 4th ed., Hoboken, New Jersey, USA: John Wiley & Sons, pp. 1–544, 2009.
- [69] R. Martinek and J. Zidek, "The real implementation of NLMS channel equalizer into the system of software defined radio," *Advances in Electrical and Electronic Engineering*, vol. 10, no. 5, pp. 330–336, 2012.
- [70] E. Visser, M. Otsuka and T.-W. Lee, "A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments," *Speech Communication*, vol. 41, no. 2, pp. 393–407, 2003.
- [71] E. Visser and T.-W. Lee, "Speech enhancement using blind source separation and two-channel energy based speaker detection," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Proc.*, Hong Kong, China, pp. 884–887, 2003.
- [72] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411–430, 2000.
- [73] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, 1st ed., Hoboken, New Jersey, USA: John Wiley & Sons, pp. 1–586, 2002.
- [74] S. Fischer and K. U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," *Speech Communication*, vol. 20, no. 3, pp. 215–227, 1996.
- [75] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.
- [76] Q. Zou, Z. L. Yu and Z. Lin, "A robust algorithm for linearly constrained adaptive beamforming," *IEEE Signal Processing Letters*, vol. 11, no. 1, pp. 26–29, 2004.
- [77] A. Rajani and S. Soundarya, "A review on various speech enhancement techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 8, pp. 296–301, 2016.
- [78] K. R. Borisagar and G. Kulkarni, "Simulation and comparative analysis of LMS and RLS algorithms using real time speech input signal," *Global Journal of Research in Engineering*, vol. 10, no. 5, pp. 44–47, 2010.

- [79] L. Wang, N. Kitaoka and S. Nakagawa, "Distant-talking speech recognition based on spectral subtraction by multi-channel LMS algorithm," *IEICE Transactions on Information and Systems*, vol. 94, no. 3, pp. 659–667, 2011.
- [80] N. Kandpal and B. M. Rao, "Implementation of PCA & ICA for voice recognition and separation of speech," in *IEEE Int. Conf. on Advanced Management Science*, Chengdu, China, pp. 536–538, 2010.
- [81] L. K. Saul and J. B. Allen, "Periodic component analysis: An eigenvalue method for representing periodic structure in speech," in *Advances in Neural Information Processing Systems*. Vancouver, British Columbia, Canada, pp. 807–813, 2001.
- [82] K. Cengiz, "Comprehensive analysis on least squares iteration for indoor positioning systems," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2842–2856, 2013.
- [83] *NT5 Compact 112" Cardioid Condenser Microphone*. RODE Microphones. [Online]. Available: <http://www.ode.com/microphones/nt5>.
- [84] M. A. Jamshed, M. Ur-Rehman, J. Frnda, A. A. Althuwayb, A. Nauman *et al.*, "Dual band and dual diversity four-element MIMO dipole for 5G handsets," *Sensors*, vol. 21, no. 3, pp. 1–13, 2021.
- [85] C. Zeitnitz, "WaveIO: A soundcard interface for labview," *Christian Zeitnitz*, 2017. [Online]. Available: <https://www.zeitnitz.eu/scms/waveio>.
- [86] J. Baros, R. Martinek, R. Jaros, L. Danys and L. Soustek, "Development of application for control of SMART parking lot," *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 19–26, 2019.
- [87] *VB-CABLE Virtual Audio Device*. VB-Audio Software, 2020. [Online]. Available: <https://vb-audio.com/Cable/>.
- [88] W. Tarneberg, O. Hamsis, J. Hedlund, K. Brunnström, E. Fitzgerald *et al.*, "Towards intelligent industry 4.0 5G networks: A first throughput and QoE measurement campaign," in *Int. Conf. on Software, Telecommunications and Computer Networks*, Split, Hvar, Croatia, pp. 1–6, 2020.
- [89] J. Garcia-Morales, M. C. Lucas-Estan and J. Gozalvez, "Latency-sensitive 5G RAN slicing for industry 4.0," *IEEE Access*, vol. 7, pp. 143139–143159, 2019.
- [90] A. Varghese and D. Tandur, "Wireless requirements and challenges in industry 4.0," in *Int. Conf. on Contemporary Computing and Informatics*, Mysore, India, pp. 634–638, 2014.
- [91] J. Ordonez-Lucena, J. Folgueira Chavarria, L. M. Contreras and A. Pastor, "The use of 5G non-public networks to support industry 4.0 scenarios," in *IEEE Conf. on Standards for Communications and Networking*, Granada, Spain, pp. 1–7, 2019.