Stowarzyszenie Elektryków Polskich
Oddział Wrocławski im. prof. Kazimierza Idaszewskiego

Stowarzyszenie Elektryków Polskich na Politechnice Wrocławskiej

Trendy i rozwiązania technologiczne w elektrotechnice



edycja 2022

Stowarzyszenie Elektryków Polskich na Politechnice Wrocławskiej 2022

Trendy i rozwiązania technologiczne
w elektrotechnice

Redakcja:
dr inż. Remigiusz Mydlikowski

Oficyna Wydawnicza Politechniki Wrocławskiej
Wrocław 2022

Komitet Naukowy
• dr hab. inż. Piotr Serkies
• dr hab. inż. Ryszard Zieliński
• dr inż. Dariusz Banasiak
• dr inż. Tomasz Grysiński
• dr inż. Maciej Gwoździewicz
• dr inż. Elżbieta Szul-Pietrzak
• dr inż. Remigiusz Mydlikowski
• dr inż. Damian Nowak
• dr inż. AdamWąż

Redakcja techniczna i korekta
Stanisław Gancarz

Skład i łamanie
Aleksandra Rzeszowska

Projekt okładki
Dominika Osadców-Będkowska

Wydrukowano na podstawie materiałów dostarczonych przez autorów.

Wszelkie prawa zastrzeżone. Niniejsza książka, zarówno w całości,
jak i we fragmentach, nie może być reprodukowana w sposób elektroniczny,
fotograficzny i inny bez zgody wydawcy i właścicieli praw autorskich.

© Copyright by Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2022
© Copyright by Stowarzyszenie Elektryków Polskich Oddział Wrocławski, Wrocław 2022

OFICYNA WYDAWNICZA POLITECHNIKI WROCŁAWSKIEJ
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
http://www.oficyna.pwr.edu.pl
e-mail: oficwyd@pwr.edu.pl
zamawianie.ksiazek@pwr.edu.pl

ISBN 978-83-7493-226-4

Druk i oprawa: beta-druk, www.betadruk.pl

Spis treści

Wydział Elektroniki, Fotoniki i Mikrosystemów **7**

1. Elektrodowy analizator rytmu serca **9**
inż. Bartłomiej Zieliński, opiekun: dr hab. inż. Adam Polak

2. Projekt i budowa robota typu SCARA **17**
inż. Adam Kubiak, opiekun: dr inż. Mateusz Cholewiński

3. Badanie hałasu urządzeń użytku domowego na przykładzie chłodziarek **27**
inż. Małgorzata Koziatek, opiekun: dr inż. Maurycy Kin

4. Design and control of agile robotic leg **37**
mgr inż. Jakub Tomaszewski, opiekun: dr inż. Mariusz Janiak

5. Silicon sensor characterisation and radiation hardness qualification for the CMS Endcap Calorimeter Upgrade for the High-Luminosity LHC **59**
mgr inż. Oliwia Kałuzińska, opiekun: dr hab. inż. Michał Mazur

6. Ultracienkie membrany dichalkogenków metali przejściowych **71**
mgr inż. Konrad Król, opiekun: dr hab. inż. Jarosław Serafińczuk

Wydział Elektryczny **83**

7. Eksperymentalny robot ramieniowy sterowany za pomocą mikrokontrolera ARM **85**
inż. Szymon Wąchała, opiekun: dr inż. Krzysztof Dyrcz

8. Programowalny model niewielkiego robota ramieniowego **95**
inż. Michał Zdobych, opiekun: dr inż. Marcin Pawlak

9. Zastosowanie platformy Omron Sysmac w układach sterowania wieloosiowych napędów pozycjonujących **103**
mgr inż. Marek Myśluk, opiekun: dr inż. Marcin Pawlak

10. Robot z funkcją przenoszenia paczek **113**
inż. Stanisław Piaskowski, opiekun: dr inż. Paweł Ewert

11. Budowa i sterowanie modelem magnetycznej lewitacji z wykorzystaniem sterownika X20 firmy B&R **121**
inż. Michał Szymecki, opiekun: dr hab. inż. Piotr Serkies

Wydział Informatyki i Telekomunikacji **129**

12. Design of a TinyML system used to analyze if the face mask is worn properly **131**
inż. Dominik Piątkowski, opiekun: prof. dr hab. inż. Krzysztof Walkowiak

13. Tożsamość cyfrowa w Polsce: ekosystem, stopień wdrożenia i zastosowanie dowodów osobistych z warstwą elektroniczną **139**
inż. Damian Rybak, opiekun: dr inż. Wojciech Wodo

14. Baza danych i warstwa dostępu do danych w aplikacji webowej dedykowanej sprzedaży dzieł sztuki **147**
inż. Jakub Smołka, opiekun: dr inż. Krzysztof Waśko

15. Zastosowanie filtrów morfologicznych w redukcji zakłóceń w obrazach cyfrowych **159**
mgr inż. Dawid Siuda, opiekun: prof. dr hab. inż. Olgierd Unold

16. Modelowanie interakcji społecznych na podstawie strumienia danych **173**
mgr inż. Michał Wiczk, opiekun: dr hab. inż. Radosław Michalski

17. Analiza aktywności podmiotów stanowiących cyberzagrożenie klasy State-Sponsored **185**
mgr inż. Mateusz Stafiniak, opiekun: dr inż. Wojciech Wodo

Wydział Podstawowych Problemów Techniki **193**

18. Konstrukcja urządzenia do monitorowania przebiegu treningu siłowego z funkcją pomiaru EMG **195**
inż. Joanna Zapała, opiekun: prof. dr hab. Krystian Kubica

19. Wpływ wybranych czynników zaburzających na dokładność pomiaru czujnika oksymetrycznego **205**
inż. Dominik Żyłka, opiekun: dr hab. inż. Agnieszka Ulatowska-Jarża

20. Zastosowanie uczenia maszynowego do klasyfikacji stadiów snu **213**
inż. Krzysztof Adamczyk, opiekun: dr hab. inż. Mirosław Łątka

21. Ocena wykorzystania sygnału EEG jako narzędzia do testowania funkcjonowania poznawczego **219**
inż. Kinga Kaczmarek, opiekun: dr inż. Wioletta Nowak

Przedmowa

Monografia przedstawia podsumowanie naukowej obecności Stowarzyszenia Elektryków Polskich na Politechnice Wrocławskiej w 2022 roku. Stowarzyszenie realizuje swoją działalność statutową organizując konkursy na najlepsze prace dyplomowe inżynieryjno-badawcze na czterech wydziałach Politechniki. Po raz pierwszy Stowarzyszenie pojawiło się także na nowym i jednocześnie największym wydziale Informatyki i Telekomunikacji.

Wszystkie wydziały łączą dyscypliny naukowe związane z szeroko rozumianą elektryką, tj. automatyka, elektronika i elektrotechnika, informatyka techniczna i telekomunikacja oraz inżynieria biomedyczna. Komisje Konkursowe powoływane na Wydziałach, oceniają zgłaszane prace pod kątem:

- sposobu rozwiązania problemu technicznego lub naukowego,
- powiązania pracy z aktualnymi problemami gospodarki narodowej lub regionu,
- stopnia samodzielności autora i oryginalności rozwiązania problemu.

Prace inżynieryjno-badawcze zgłaszane do Konkursu, charakteryzują się wysokim poziomem merytorycznym realizacji problemu technicznego. Poruszane są w nich aktualne problemy naukowo-badawcze odzwierciedlające obecne trendy stosowanych rozwiązań technologicznych.

W monografii zawarto streszczenia wszystkich nagrodzonych prac dyplomowych z edycji Konkursu 2022 roku.

dr inż. Remigiusz Mydlikowski
Przewodniczący Komisji Młodzieży i Studentów
SEP Oddział Wrocławski

Wydział Elektroniki, Fotoniki i Mikrosystemów

Elektrodowy analizator rytmu serca

inż. Bartłomiej Zieliński
Elektronika

dr hab. inż. Adam Polak
Katedra Metrologii Elektronicznej i Fotonicznej

Serce człowieka jest motorem napędowym krwiobiegu koniecznego dla wszystkich procesów życiowych i prawidłowego funkcjonowania całego organizmu, dlatego ważna jest kontrola jego pracy, w celu zapewnienia zdrowia człowieka. Celem pracy było zaprojektowanie, wykonanie, uruchomienie i przetestowanie mobilnego analizatora rytmu serca opartego na jednoodprowadzeniowej rejestracji sygnału EKG. Opracowanie analizatora obejmowało następujące etapy: przeanalizowanie literatury, sformułowanie założeń konstrukcyjnych, wyszukanie podzespołów i układów, opracowanie schematu elektronicznego, zaprojektowanie obwodu drukowanego, zamówienie elementów i PCB, montaż urządzenia, napisanie oprogramowania oraz przeprowadzenie szeregu testów. Ostatecznie uzyskano urządzenie niewielkich rozmiarów, które bardzo dobrze radzi sobie z pomiarem rytmu serca podczas różnych aktywności sportowych, zachowując przy tym niewielkie zużycie energii, a tym samym długi czas pracy na baterii.

1. Cel i założenia

Celem pracy jest zaprojektowanie, wykonanie, uruchomienie i przetestowanie mobilnego analizatora rytmu serca opartego na jednoodprowadzeniowej rejestracji sygnału EKG. Skonstruowanie urządzenia wymagało głębszego poznania tematu poprzez analizę literatury oraz dokumentacji technicznych. Na podstawie zdobytej wiedzy dotyczącej działania tego typu urządzeń możliwe było sformułowanie założeń konstrukcyjnych. W pracy skupiono się na zaprojektowaniu i wykonaniu rejestratora. Również zostało napisane oprogramowanie pozwalające na uruchomienie analizatora. Działający rejestrator przetestowano pod kątem dokładności pomiarowej oraz funkcjonowania w różnych warunkach pracy.

Główne założenia projektowe realizowanej pracy to:

- rozmiar analizatora nie powinien przekraczać wymiarów: 100 × 50 ×20 mm,
- kompatybilność z opaską piersiową z suchymi elektrodami,
- nieprzerwana praca przez minimum 12 godzin,
- możliwość ładowania poprzez złącze USB typu C,

- rytm serca wysyłany przy pomocy protokołu Bluetooth (kompatybilność z aplikacjami do rejestracji aktywności sportowych różnych producentów),
- informowanie użytkownika o trybie działania urządzenia.

2. Wstęp teoretyczny

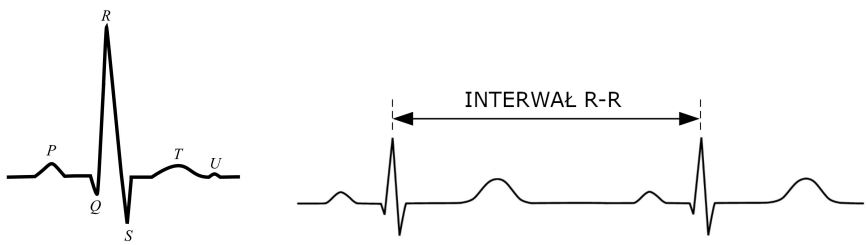
Sygnał jest wielkością fizyczną zmieniającą się w czasie, a w zachodzących zmianach zawarta jest jakaś informacja. Ludzki organizm wytwarza wiele rodzajów sygnałów określanych mianem sygnałów biomedycznych. Zaliczamy do nich sygnały biochemiczne, termiczne, mechaniczne, bioelektyrczne, dźwiękowe itd. Generowanie jest związane z występowaniem różnych zjawisk w organizmie, m.in. fizycznych i chemicznych [1]. Wszystkie rodzaje sygnałów biomedycznych charakteryzują się określoną periodycznością. Możliwe jest wyróżnienie trzech rodzajów okresowości: własna, dobowa oraz miesięczna [2].

2.1. Biopotencjały

Znaczącą liczbę sygnałów biomedycznych stanowią sygnały bioelektryczne. Każdy sygnał elektryczny charakteryzuje się określonymi parametrami, również w przypadku biopotencjałów. Typowymi cechami są amplituda rzędu od µV do mV oraz częstotliwość od 0 Hz do 20 kHz. Pomiar biopotencjałów może odbywać się na różnymi metodami. W opisywanym rozwiązaniu skupiono się na metodzie nieinwazyjnej, stosując elektrody suche. Metoda ta polega na przyłożeniu elektrod bezpośrednio do skóry. Suche elektrody charakteryzują się wysoką impedancją połączeń, której wartość sięga 10 MΩ.

2.2. Analiza rytmu serca

Dziedziną zajmującą się badaniem aktywności serca jest elektrokardiografia (EKG). Sygnał taki zawiera bardzo dużą ilość informacji diagnostycznych, co umożliwia określenie liczby uderzeń serca na minutę (BPM) oraz pozwala na wykrywanie różnego rodzaju chorób, takich jak arytmia. Mięsień sercowy jest największym źródłem prądowym w ludzkim organizmie. Podczas skurczów i rozkurczów uwalniana jest energia z mięśni, która jest źródłem sygnału EKG. Sygnał EKG jest okresowy, amplituda wynosi od 0,1 do 1 mV. Pasmo niosące informacje znajduje się w przedziale od 0,01 do 250 Hz [3]. W sygnale EKG można wyróżnić 6 różnych części związanych z załamkami PQRSTU (rysunek 1.1a). Najbardziej widoczny oraz często wykorzystywany do badań jest kompleks QRS.



(a) Normalny sygnał [3] **(b)** Interwał pomiędzy załamkami R-R [4]

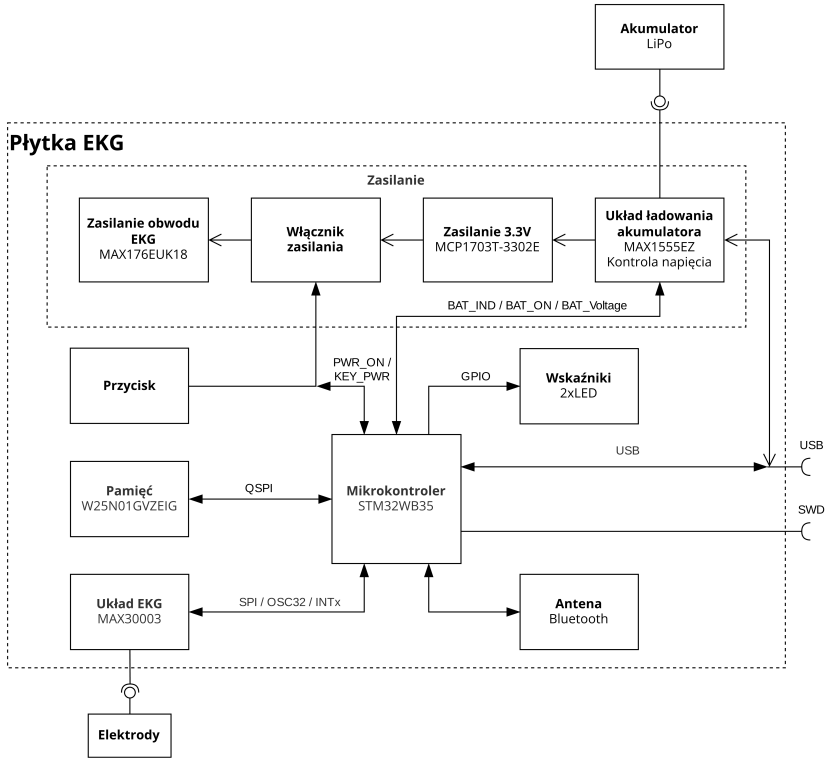
Rysunek 1.1. Sygnał EKG

W celu wyznaczenia tętna (HR) wymagany jest pomiar odcinka czasu między dwoma załamkami R (rysunek 1.1b). Następnie na podstawie wzoru (1.1) wyliczyć liczbę uderzeń na minutę (BPM), gdzie T_{R-R} jest czasem między dwoma załamkami R-R podawanym w sekundach [5]:

$$HR = \frac{60}{T_{R-R}} \tag{1.1}$$

3. Projekt i wykonanie

Elektrodowy analizator rytmu serca został podzielony na kilka elementów składowych, takich jak: zasilanie, mikrokontroler, układ EKG, peryferia, złącza, moduł oprogramowania. Wszystkie komponenty zostały przedstawione na rys. 1.2.



Rysunek 1.2. Diagram blokowy komponentów elektrodowego analizatora rytmu serca

Na podstawie założeń oraz diagramu blokowego przystąpiono do zaprojektowania schematu elektronicznego wraz z wykonaniem niezbędnych obliczeń. Podczas projektowania zwracano dużą uwagę na projektowanie komponentów w taki sposób, aby zużycie energii było jak najmniejsze.

W celu pomiaru biopotencjałów pojawiających się na skórze, a tym samym pomiaru sygnału EKG, zastosowano pasek z napylonymi elektrodami, obejmujący dookoła klatkę piersiową, wykonany z materiału rozciągliwego. W celu zapewnienia ochrony przed za-

kłóceniami elektromagnetycznymi (EMI) zastosowano typowe filtry EMI stosowane do wyeliminowania sygnałów wysokoczęstotliwościowych indukowanych przez szum radiowy.

Przestawiony analizator do pomiarów sygnałów związanych z aktywnością serca wykorzystuje układ MAX30003 firmy Analog Devices. Jest to wyspecjalizowany układ do jednoodprowadzeniowej rejestracji sygnału EKG stosowanej w systemach medycznych. Wybór urządzenia podyktowany był dostępnością na runku oraz kompleksowymi możliwościami. Układ zawiera m.in. filtr antyaliasingowy, programowalny wzmacniacz instrumentalny, przetwornik analogowo-cyfrowy (ADC) sigma-delta, filtr decymacyjny, układ detekcji załamków R oraz bufor FIFO na dane.

Sercem urządzenia jest mikrokontroler STM32WB35CEU7A wyposażony w moduł komunikacji bezprzewodowej Bluetooth. Charakteryzuje się niskim zużyciem energii za sprawą ośmiu trybów jej oszczędzania. Mikrokontroler posiada również przetwornicę obniżającą napięcie (SMPS) o wysokiej efektywności, dzięki której możliwe było obniżenie poboru energii przez mikrokontroler. Zastosowano dwie diody sygnalizujące w kolorze czerwonym i niebieskim do informowania użytkownika o stanie pracy urządzenia.

W urządzeniu została zaprojektowana antena PCB przeznaczona do pracy z częstotliwością 2,45 GHz. Jest to zakres pracy komunikacji bezprzewodowej Bluetooth. Ze względu na wykorzystanie elementu radiowego pracującego z wysokimi częstotliwościami sygnału cyfrowego, aby analizator nie zakłócał urządzeń pracujących w pobliżu oraz zapewniał kompatybilność elektromagnetyczną EMC, zastosowano metalową obudowę EMC, obejmującą mikrokontroler wraz częściowym torem radiowym.

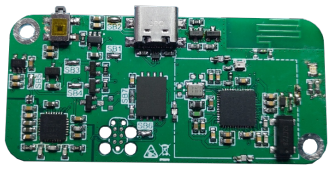
Oszacowano ilość zużywanej energii przez urządzenie oraz czas pracy na baterii, tym samym wybierając ogniwa litowo-polimerowe o pojemności 0,315 Wh. W celu zapewnienia możliwości ładowania akumulatora zastosowano układ MAX15555. Dostarczenie energii do ładowarki odbywa się przez złącze typu USB Typu-C w standardzie USB 2.0. Aby kontrolować poziom naładowania ogniwa LiPo, zastosowano układ pomiarowy złożony z klucza analogowego oraz dzielnika napięcia. Wykorzystano również dwa stabilizatory napięcia LDO 3,3 V oraz 1,8 V w celu zapewnienia zasilania mikrokontrolera oraz przetwornika MAX30003. Analizator został wyposażony w elektroniczny włącznik zasilania. W momencie wciśnięcia przycisku następuje zasilanie mikrokontrolera, który następnie podtrzymuje włącznik w stanie załączenia.

Obwód drukowany został wykonany w technologii 4-warstwowej. Podyktowane to było znacznie lepszymi parametrami obwodu radiowego oraz zwiększeniem odporności na zakłócenia elektromagnetyczne. Głównym materiałem dielektrycznym był popularny FR4 (laminat epoksydowy z włóknem szklanym). Schemat oraz obwód wykonano w programie Altium Designer 21.

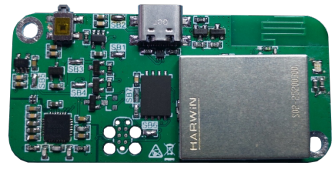
Następnie dokonano weryfikacji projektu oraz wprowadzono drobne poprawki, a po pozytywnym zakończeniu sprawdzania zamówiono płytkę PCB oraz wszystkie niezbędne elementy. Jednocześnie przygotowano dokumentację konieczną do montażu elementów. Na podstawie dokumentacji dokonano montażu urządzenia według wcześniej przygotowanej kolejności. Efekt montażu urządzenia przedstawiony jest na rysunku 1.3. Uzyskano płytkę o wymiarach 65 mm $\times$ 30 mm, a całkowita grubość nie przekracza 8,13 mm.

Następnie napisano oprogramowanie oraz przeprowadzono kolejną serię testów, sprawdzając przy tym poprawność konfiguracji zegarów wewnętrznych mikrokontrolera, stabil-

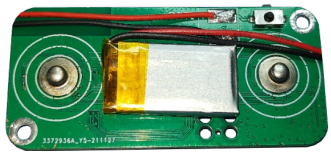
ności generowanej częstotliwości dla układu MAX30003, świecenia diodami LED, komunikacji z układem MAX30003, komunikacji USB, komunikacji Bluetooth.



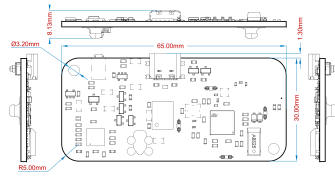
(a) Zmontowana płytka PCB, przód



(b) Zmontowana płytka PCB, przód z obudową EMC



(c) Zmontowana płytka PCB, tył z baterią



(d) Widok montażowy PCB analizatora

Rysunek 1.3. Widoki montażowy oraz zmontowanego urządzenia

4. Uzyskane wyniki

Zmontowane urządzenie przeszło przez następującą serię testów: zużycia energii przez urządzenia, poprawności detekcji rytmu serca przez analizator oraz stabilności pracy w trakcie wysiłku sportowego.

4.1. Zużycie energii

Analizator w trybie wyłączonego urządzenia zużywa 7 µW oraz 20 mW w trakcie działania. Na podstawie powyższych danych obliczono, że urządzenie pozwoli na pracę przez 16 dni przy założeniu, że trening nie trwa dłużej niż 1 godzinę na dobę.

4.2. Poprawność detekcji rytmu serca

Sprawdzenie poprawności detekcji tętna było jednym z podstawowych testów określającym poprawność działania urządzenia. Pomiar był wykonywany trzema urządzeniami: konstruowanym analizatorem, miernikiem ciśnienia oraz pulsoksymetrem. Wszystkie urządzenia mają możliwość pomiaru rytmu serca. Typowo dokładność takich pomiarów nie jest lepsza niż ±1 BPM, natomiast w przypadku pulsoksymetru zapewniona jest dokładność na poziomie ±2 BPM, miernik ciśnienia charakteryzuje się dokładnością pomiaru tętna na poziomie 5% wartości zmierzonej.

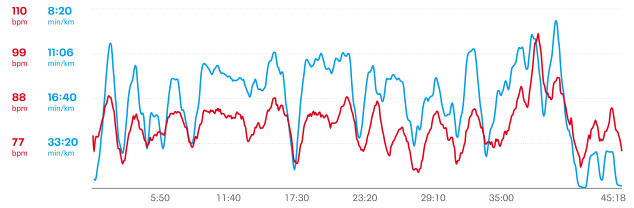
Pomiar wykonywany był wszystkimi urządzeniami w tym samym czasie, po zakończeniu różnych aktywności. Wartości które wskazały urządzenia przedstawiono w tabeli 1.1. Na podstawie znanych dokładności stwierdzono, że analizator poprawnie mierzy tętno.

Tabela 1.1. Pomiar tętna trzema urządzeniami przy różnych rodzajach aktywności, gdzie δHR_c to rozbieżność względna tętna między analizatorem i miernikiem cienienia, ΔHR_p to różnica tętna między analizatorem i pulsoksymetrem

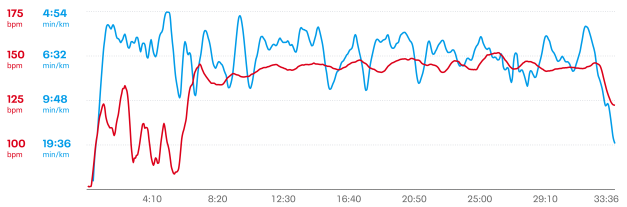
Rodzaj aktywności	**Analizator**	**Miernik ciśnienia**	**Pulsoksymetr**	$\delta\mathbf{HR_C}$	$\Delta\mathbf{HR_P}$
	[BPM]	[BPM]	[BPM]	[%]	[BPM]
Siedzenie w spoczynku	54	52	55	3,8	1
Po wykonaniu 5 przysiadów	72	73	72	−1,4	0
Po wykonaniu 15 przysiadów	83	83	80	0,0	3

4.3. Działanie urządzenia podczas aktywności fizycznej

Ostatnim z testów było sprawdzenie działania analizatora podczas rzeczywistych warunków pracy, dlatego wykonano dwa testy – w czasie spaceru oraz biegania (rysunek 1.4). Urządzenie przez cały czas trwania testów działało poprawnie. Analizując wyniki testów, zaobserwowano znane z literatury dwie składowe mechanizmu regulacji tętna: szybko- i wolnozmienną. Szybkozmienna regulacja decyduje o dobrej korelacji podczas niewielkiego wysiłku (rysunek 1.4a), wolnozmienna powoduje bardziej stabilne tętno po pewnym czasie większego wysiłku (rysunek 1.4b) [6].



(a) Podczas spaceru



(b) Podczas biegania

Rysunek 1.4. Wykres poziomu tętna oraz tempa względem czasu podczas spaceru i biegu (wykres z aplikacji MapMyRun). Kolorem niebieskim przedstawione zostało tempo w minutach na kilometr, natomiast kolorem czerwonym tętno

5. Podsumowanie

Celem pracy było zaprojektowanie, wykonanie, uruchomienie i przetestowanie mobilnego analizatora rytmu serca opartego na jednoodprowadzeniowej rejestracji sygnału EKG. Pracę nad urządzeniem rozpoczęto od wyszukania i przeanalizowania literatury. Pozwoliło to zdefiniować wstępne założenia projektowe: funkcjonalne, konstrukcyjne i progra-

mistyczne. Następnie przeprowadzono rozeznanie dotyczące dostępności podzespołów. Spowodowane to było trudną sytuacją na rynku podzespołów elektronicznych związaną z pandemią COVID-19. Przeanalizowano dokumentacje techniczne w celu zapoznania się z możliwościami poszczególnych elementów i układów, a następnie wybrano i zamówiono podzespoły spełniające wstępne wymagania. Kolejno przystąpiono do sformułowania szczegółowych założeń projektowych oraz opracowano diagram komponentów projektowanego analizatora. Na podstawie powyższych informacji zaprojektowano schemat elektroniczny oraz wykonano niezbędne obliczenia. Ukończenie projektu układu pozwoliło na przystąpienie do zaprojektowania obwodu drukowanego. Następnie dokonano weryfikacji schematu oraz obwodu, a po pozytywnym zakończeniu sprawdzania zamówiono płytkę PCB oraz wszystkie potrzebne elementy. Jednocześnie przygotowano dokumentację konieczną do montażu elementów. Na podstawie dokumentacji dokonano montażu urządzenia, na którym przeprowadzono kilka testów, weryfikując przy tym poprawność złożenia analizatora. Następnie napisano oprogramowanie oraz przeprowadzono kolejną serię testów. Ostatnim elementem było porównanie wyników badania rytmu serca z innych urządzeń oraz działania analizatora podczas różnego rodzaju aktywności sportowych. Ostatecznie uzyskano urządzenie o następujących wymiarach 65 mm $\times$ 30 mm, a całkowita grubość nie przekracza 8,13 mm. Analizator bardzo dobrze radzi sobie z pomiarem rytmu serca podczas różnych aktywności sportowych, którego dokładność detekcji nie przekracza ± 2 BPM. Również podczas długotrwałej rejestracji sygnału zachowana jest stabilność działania. Zastosowanie niskoenergetycznego standardu Bluetooth, wykorzystującego protokół GATT o profilu pomiaru tętna (HRP), umożliwiło współpracę urządzenia z szeroką gamą aplikacji mobilnych na smartfona. Rejestrator zużywa niewielką ilość energii, pozwalając tym samym na 1 godzinę treningu dziennie, przez 16 dni. Ostatecznie osiągnięto cel pracy.

Bibliografia

[1] Tadeusiewicz R., Augustyniak P., *Inżynieria biomedyczna: księga współczesnej wiedzy tajemnej w wersji przystępnej i przyjemnej*; Wydawnictwa AGH, Kraków, Polska, 2008, s. 116–120.

[2] Polak A., Materiały z wykładu „Elektroniczna aparatura medyczna" dla specjalności Aparatura Elektroniczna, semestr letni 2020/2021.

[3] Wannenburg J., Malekin R., Hancke G., *Wireless Capacitive-Based ECG Sensing for Feature Extraction and Mobile Health Monitoring.* IEEE Sensors Journal **2018**, 18, 14, s. 6023–6032.

[4] Ultra-Low Power, Single-Channel Integrated Biopotential (ECG, R-to-R Detection) AFE, MAX30003, Rev. 3, Analog Devices, **2021**.

[5] Wannenburg J., Malekin R., *Body Sensor Network for Mobile Health Monitoring, a Diagnosis and Anticipating System.* IEEE Sensors Journal **2015**, 15, 12, s. 6839–6852.

[6] van Brievingh R.P.W., Möller D.P.F., Shen X., *Biomedical modeling and simulation on a PC: a workbench for physiology and biomedical engineering*, wydanie pierwsze, Springer Science & Business Media, Nowy Jork, USA, 2013.

Projekt i budowa robota typu SCARA

inż. Adam Kubiak
Automatyka i Robotyka

dr inż. Mateusz Cholewiński
Katedra Cybernetyki i Robotyki

Robotyka jest młodą dziedziną nowoczesnej technologii, która łączy różne tradycyjne gałęzie nauk technicznych. Zrozumienie całej zawiłości robotów i ich zastosowań wymaga znajomości zagadnień elektrycznych, elektronicznych, mechanicznych, nauk komputerowych i matematyki. Obserwując ogromne możliwości robotów i rozwój dziedziny robotyki, można znaleźć duże pokłady inspiracji, aby być częścią tego skoku technologicznego ludzkości. Łatwy dostęp do mikrokontrolerów i niezbędnej elektroniki przyczynił się do tworzenia zaawansowanych projektów przez jednostki. Druk 3D jest kolejnym czynnikiem, który umożliwia tworzenie innowacji przez małe firmy i pojedyncze osoby. Dzięki zaawansowanym rozwiązaniom do druku 3D, oferowanym przez producentów, umożliwiającym szybki czas realizacji niestandardowych komponentów, inżynierowie mogą przyśpieszyć fazę badań i projektowania, wytwarzając roboty znacznie szybciej niż w przypadku konwencjonalnego prototypowania.

1. Cel i założenia

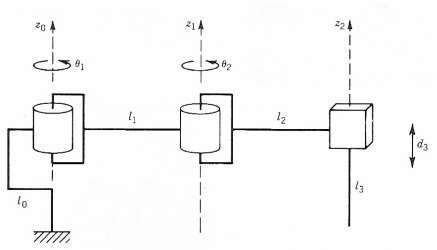
Celem pracy jest budowa robota typu SCARA o trzech stopniach swobody (RRT) oraz stworzenie aplikacji umożliwiającej jego sterowanie oraz obserwację położenia manipulatora w przestrzeni. Podczas projektu wykorzystano wiedzę z zagadnień tworzących dziedzinę naukową robotyki. Główne założenia projektowe realizowanej pracy to:

- stworzenie trójwymiarowego modelu robota,
- wykonanie modelu 3D manipulatora, gotowego pod druk 3D,
- skonstruowanie manipulatora,
- zaprojektowanie i wykonanie elektroniki niezbędnej do zasilania i sterowania podzespołów,
- wyprowadzenie modelu kinematyki prostej i odwrotnej dla robota typu SCARA,
- wykonanie programu sterującego na mikrokontroler STM32,
- stworzenie aplikacji okienkowej umożliwiającej sterowanie i obserwację położenia robota.

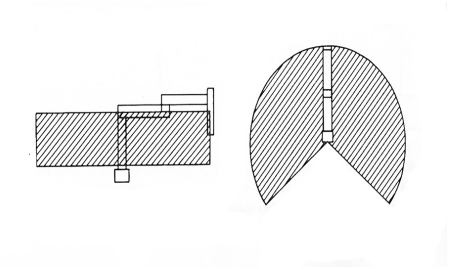
2. Wstęp teoretyczny

Manipulatory SCARA (ang. Selective Compliant Articulated Robot for Assembly) są idealnym przykładem robota przemysłowego. Ich konstrukcja ma na celu symulowanie złożonych ruchów, na które pozwala ludzkie ramię wraz z łokciem.

W typowych robotach nie istnieje bezpośredni pomiar położenia efektora. Dlatego też w celu opisania oddzielnych członów i ich połączeń wykorzystuje się kinematykę robota. W kinematyce manipulatorów i robotów elementarnymi operacjami są rotacje i translacje. Konfiguracja SCARA ma strukturę RRT (rys. 2.1), co oznacza że robot posiada dwa przeguby rotacyjne i jeden człon translacyjny. Cechą, która wyróżnia tego typu manipulatory, jest także ułożenie jego osi z_0, z_1, z_2 w przegubach, ponieważ są one względem siebie równoległe. Taka konstrukcja umożliwia dotarcie do każdego punktu znajdującego się w przestrzeni roboczej manipulatora dzięki wykonaniu jednego lub dwóch prostych ruchów napędów odpowiedzialnych za przemieszczanie efektora w tej przestrzeni (rys. 2.2).



Rysunek 2.1. Manipulator SCARA[2]



Rysunek 2.2. Przestrzeń robocza [2]

3. Projekt i wykonanie

3.1. Model kinematyki robota

Przez wykorzystane w projekcie czujniki realizowane jest sprzężenie zwrotne do bezpośredniego odczytania zmiennych przegubowych θ_1, θ_2 [2]. W manipulatorze typu SCARA dane te umożliwiają wyrażenie pozycji x_0 i y_0 w układzie bazowym $O_0x_0y_0$ za pomocą kinematyki prostej w zależności od tych zmiennych. Jednakże aby zadać robotowi ruch do pozycji (X, Y), która znajduje się w ograniczeniach przestrzeni roboczej (rys. 2.2), trzeba postąpić odwrotnie, tzn. znaleźć zmienne przegubowe θ_1, θ_2 w zależności od zadanych współrzędnych efektora. Jest to tzw. zadanie kinematyki odwrotnej [2].

Kinematyka prosta

Do obliczenia kinematyki prostej wykorzystano algorytm Denavita-Hartenberga (tabela 2.1). Po uwzględnieniu założonych wymiarów kluczowych elementów manipulatora otrzymano następujący wynik.

Tabela 2.1. Parametry kinematyczne manipulatora SCARA

Człon	θi	d_i	a_i	φ_i
1	q_1	l_1	a_1	0
2	q_2	0	a_2	π
3	0	$d_3 + l_2$	0	0

Oznaczenia parametrów:

q_i – kąt między osiami x_{i-1} i x_i mierzony wokół osi z_{i-1}; jeśli przegub jest obrotowy, to q_i jest zmienną przegubową [2],

d_i – przesunięcie przegubu po osi z_i; jeśli przegub jest pryzmatyczny, to d_i jest zmienną przegubową [2],

a_i – odległość wzdłuż osi x_i od początku O_i do przecięcia osi x_i i z_{i-1} [2],

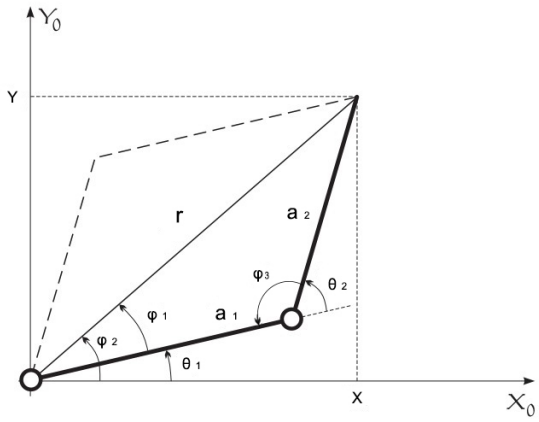
l_i – odległość wzdłuż osi z_{i-1} od początku O_i do przecięcia osi x_i i z_{i-1} [2].

Następnie na ich podstawie określono kinematykę manipulatora. Po wyznaczeniu macierzy transformacji między lokalnymi układami współrzędnych przemnożono je w celu otrzymania macierzy $\mathbf{A}_0^3$. Macierz ta opisuje orientację i położenie końcówki efektora w przestrzeni.

$$\mathbf{A}_0^3 = \mathbf{A}_0^1 \cdot \mathbf{A}_1^2 \cdot \mathbf{A}_2^3 \tag{2.1}$$

Kinematyka odwrotna

Obliczając kinematykę odwrotną, uproszczono konstrukcję manipulatora SCARA do podwójnego wahadła (rys. 2.3). Zastosowane rozwiązanie jest możliwe ze względu na to, iż przegub translacyjny manipulatora porusza się wzdłuż osi z_3, przez co nie wpływa na położenie efektora na płaszczyźnie $O_0x_0y_0$.



Rysunek 2.3. Zależności geometryczne w manipulatorze typu podwójne wahadło [1]

3.2. Konstrukcja mechaniczna

Projekt konstrukcji mechanicznej został w pełni wykonany w środowisku typu CAD, Autodesk Inventor. Wszystkie elementy drukowane zostały wykonane z materiału PCABS. W manipulatorze przeguby są głównymi punktami obciążenia, dlatego konstrukcja wymaga użycia wytrzymałych łożysk, pod które zostały wymierzone elementy. W założeniach projektowych przyjęto warunki, w których mowa jest o tym, że za ruch przegubów rotacyjnych są odpowiedzialne silniki DC, natomiast przegub translacyjny jest napędzany serwomechanizmem. Końcowy projekt przedstawiono na rysunku 2.4a, natomiast najważniejsze wymiary manipulatora przedstawiono w tabeli 2.2. W projekcie zastosowano szereg przekładni w celu polepszenia jakości wykonywanego ruchu przez ramię. Dzięki przekładni została zwiększona liczba impulsów enkodera zliczanych na jeden stopień obrotu przegubów. Mechanizm ten pozwolił także na zwiększenie momentu obrotowego oraz dokładności wykonywanego ruchu kosztem zmniejszonej prędkości.



(a) model **(b)** widok z boku

Rysunek 2.4. Wykonany robot

Tabela 2.2. Wartości parametrów manipulatora SCARA

Parametr	a_1	a_2	l_1	l_2
Wartość	94 mm	110 mm	100 mm	60 mm

3.3. Elektronika

W projekt elektroniki wchodzą następujące komponenty:

- STM32F411E Discovery,
- Konwerter USB-UART PL2303,
- Przetwornica impulsowa DC-DC step-down LM2596,
- Dwukanałowy sterownik silników L298N z wbudowanym stabilizatorem 5 V,
- Enkoder inkrementalny 22imp/obr zamontowany na wale silnika $\times 2$,
- Pakiet Li-Pol Dualsky 1800 mAh 25 C 7,4 V.

Na płytce prototypowej została umieszczona przetwornica z filtracją napięcia zasilającego układ, sterownik silników ze stabilizatorem 5 V, który jest wykorzystywany do zasilania części logicznej układu, oraz sekcja zasilania. Sercem układu jest mikrokontroler

STM32F11E Discovery. Jego zadaniem jest zapewnienie komunikacji z komputerem dzięki protokołowi UART, sterowanie silnikami przy pomocy regulatora PID i sterownika silników, sterowanie położeniem wału serwomechanizmu przy pomocy wygenerowanego sygnału PWM oraz odczyt impulsów wygenerowanych przez enkodery inkrementalne, dzięki którym jesteśmy w stanie odczytać zmienne q_1 i q_2 na poszczególnych przegubach manipulatora.

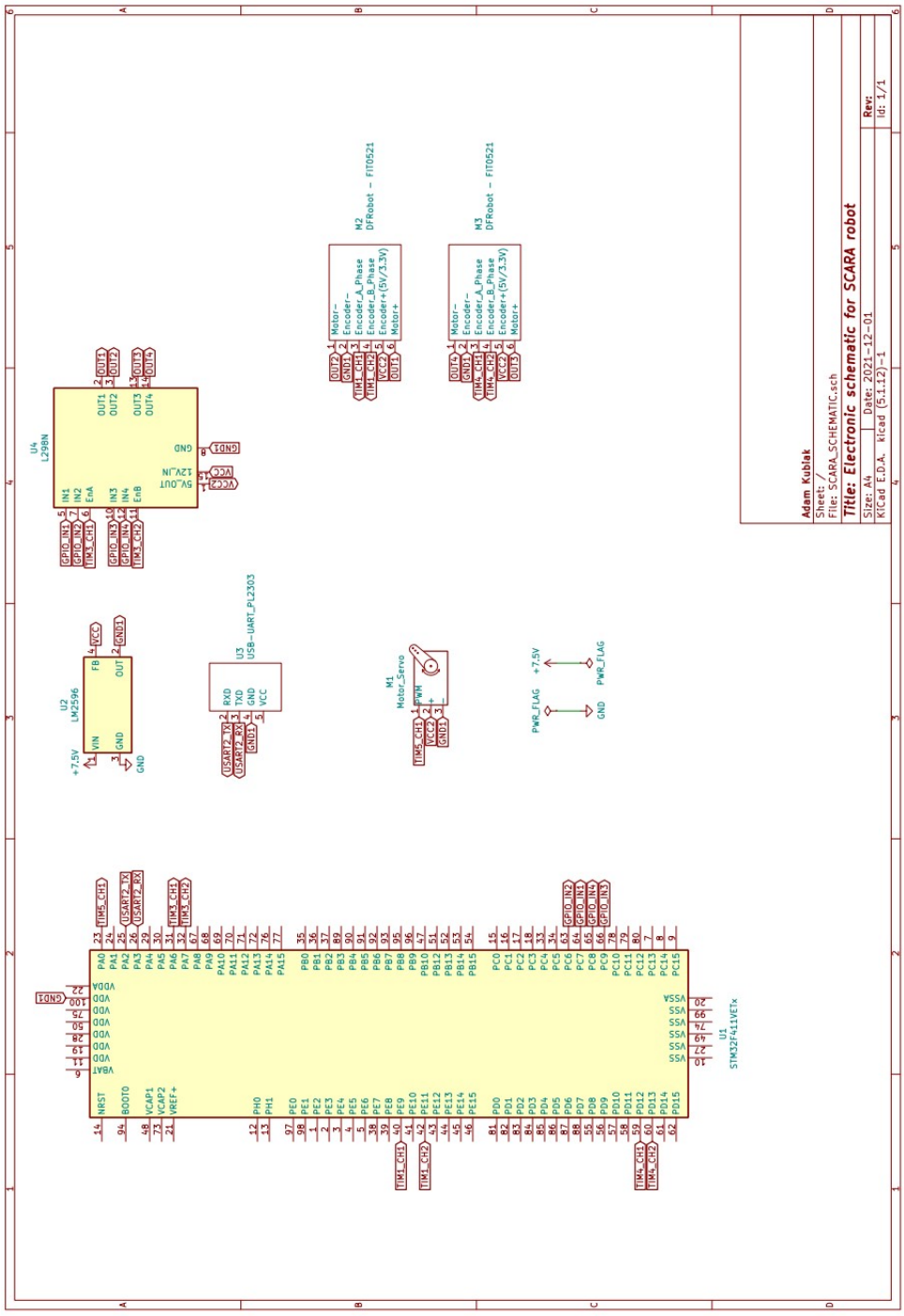
3.4. Aplikacja okienkowa

Aplikacja pełni rolę graficznego interfejsu użytkownika służącego do obsługi pełnej funkcjonalności wykonanego manipulatora SCARA. Oprócz tego aplikacja zapewnia wizualizację danych sensorycznych pobieranych z wchodzących w skład projektu czujników oraz wizualizację 3D modelu wykonanego w Autodesk Inventor, który reprezentuje konfigurację przegubów manipulatora 1:1. Interfejs w założeniach ma być prosty w obsłudze oraz wizualnie czytelny dla użytkownika. Aplikacja spełnia następujące funkcjonalności:

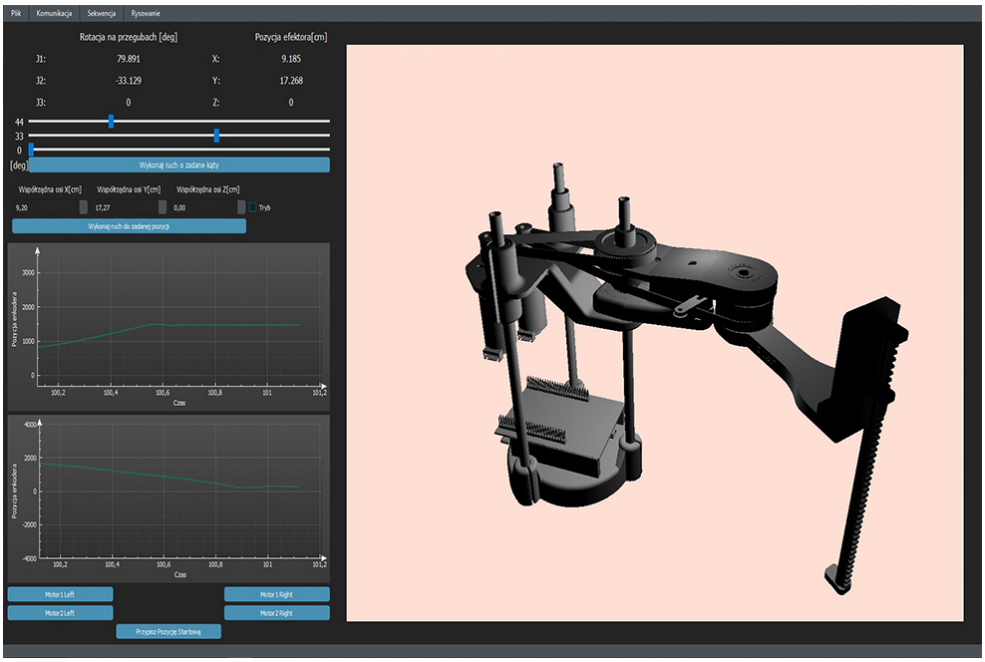
- Połączenie aplikacji z wybranym portem szeregowym w celu odbierania wysyłanych danych przez mikrokontroler z wykorzystaniem interfejsu komunikacyjnego UART,
- Informowanie użytkownika o aktualnej orientacji każdego z przegubów oraz aktualnej pozycji efektora wyliczonej z zaimplementowanej kinematyki prostej względem bazowego układu współrzędnych manipulatora za pomocą danych liczbowych,
- Kalibracja robota do pozycji zerowej za pomocą specjalnie przygotowanej sekcji w aplikacji,
- Prezentacja odczytów z magnetycznych enkoderów inkrementalnych za pomocą wykresów funkcji,
- Sterowanie manipulatorem za pomocą zadanych kątów na każdym z przegubów,
- Sterowanie manipulatorem z wykorzystaniem algorytmu odwrotnej kinematyki za pomocą zadanych współrzędnych w globalnym układzie współrzędnych znajdujących się w przestrzeni roboczej robota,
- Możliwość wyboru, w jakiej konfiguracji manipulator dotrze do zadanej pozycji przy użyciu zaimplementowanego algorytmu odwrotnej kinematyki,
- Specjalne menu do tworzenia i odtwarzania sekwencji ruchów, na którą składają się konfiguracje przegubów ustawione przez użytkownika,
- Wizualizacja trójwymiarowego modelu robota zgodnie z ruchem rzeczywistego ramienia,
- Przybliżanie i oddalanie modelu manipulatora, nie zmieniając przy tym jego położenia w układzie wewnętrznym programu.

Aplikacja została utworzona na podstawie zestawu bibliotek **Qt** dedykowanych dla języków C++, QML i Java. Podstawą tego środowiska są klasy służące do budowy graficznego interfejsu programów [3].

Podczas projektowania interfejsu graficznego zaczerpnięto inspirację z aplikacji sterującej wykonanej dla Cobota UR3 marki Universal Robots. Interfejs ten cechuje intuicyjne sterowanie i klarowny wygląd aplikacji, dzięki czemu użytkownik bez żadnego wcześniej zdobytego doświadczenia może wykonywać podstawowe operacje umożliwiające ruch manipulatora.



Rysunek 2.5. Schemat elektroniki wykonanej na potrzeby projektu



Rysunek 2.6. Finalny wygląd wykonanej aplikacji

4. Uzyskane wyniki

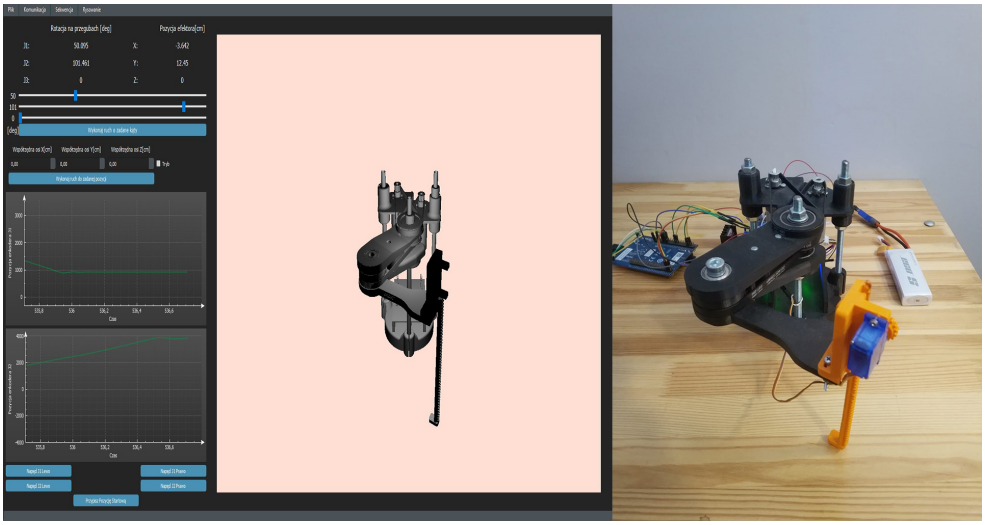
Początkowo sprawdzono poprawność wykonanego modelu matematycznego za pomocą oprogramowania Matlab. Po przeprowadzeniu badań mających na celu sprawdzenie dokładności, z jaką powtarzane są ruchy otrzymane, ich wyniki okazały się bardzo zadowalające, biorąc pod uwagę, że jest to pierwsza iteracja projektu manipulatora SCARA. Testy te pozwoliły na sprawdzenie jakości zaimplementowanych rozwiązań, w których skład wchodzą pozostałe elementy projektu takie jak regulator PID, aplikacja sterująca, jak i poprawność wykorzystania komponentów wchodzących w skład elektroniki. W tym celu wybrano losową konfigurację, a następnie powtarzano ruch do niej z różnych pozycji startowych i badano, z jaką dokładnością zostało wykonane przemieszczenie do zadanej konfiguracji końcowej. Wyniki przeprowadzonych badań można zobaczyć w tabeli 2.3.

Tabela 2.3. Wyniki przeprowadzonych badań

	$(q_1, q_2)[^\circ]$	(X, Y)[cm]
Zadana pozycja	$(166; -80)$	$(-8,353; 13,247)$
Ruch 1	$(165,916; -80,388)$	$(-8,272; 13,246)$
Ruch 2	$(166,024; -80,215)$	$(-8,318; 13,241)$
Ruch 3	$(166,024; -79,981)$	$(-8,363; 13,244)$
Ruch 4	$(166,187; -80,336)$	$(-8,332; 13,215)$
Ruch 5	$(165,916; -79,889)$	$(-8,355; 13,261)$

Analizując wyniki przeprowadzonych badań, można zauważyć, że niedokładność wykonanego ruchu nie przekracza 0,4° od zadanej zmiennej przegubowej. Jeśli chodzi o nie-

dokładność pozycji efektora manipulatora w bazowym układzie współrzędnych, to nie przekracza ona 0,4 cm.



Rysunek 2.7. Przemieszczenie robota do zadanej konfiguracji kątów na przegubach (nie jest to pozycja, w ramach której zostały wykonane testy 2.3)

5. Podsumowanie

Celem pracy było wykonanie projektu oraz skonstruowanie manipulatora typu SCARA z możliwością jego sterowania za pomocą wykonanej aplikacji okienkowej. Wykonano wszystkie zadania z dziedzin mechaniki, modelowania 3D, elektroniki oraz programowania. Projekt wymagał wykorzystania wiedzy z wielu dziedzin, które tworzą zagadnienie robotyki. Zadanie było bardzo wymagające i rozbudowane, jednak dobre rozplanowanie i rozłożenie projektu na poszczególne etapy pozwoliło na wykonanie wszystkich założeń projektowych. Zaprojektowany model z systemem przekładni okazał się być efektywny, dzięki czemu robot jest w stanie wykonywać ruchy przegubami nawet przy małym napięciu sterującym napędami. Zależność ta pozwala na płynne wykonywanie zadanych ruchów, co dodatkowo usprawnia zastosowany regulator PID, który został dostrojony na rzecz projektu. Prędkość protokołu UART zastosowanego w celu komunikacji płytki deweloperskiej z aplikacją okienkową spełniała oczekiwania projekowe. Za pomocą specjalnie przygotowanej ramki danych system jest w stanie dostarczać zarówno informację z czujników przekazujących pozycję wałów silników, jak i wartości sygnału PWM sterującego serwem bezpośrednio do aplikacji sterującej. Dzięki sprawnej komunikacji wizualizacje wykresów oraz modelu 3D robota działają w czasie rzeczywistym do ruchów fizycznego robota. W celu rozszerzenia możliwości wykonanego robota naturalnym następnym krokiem byłoby dołączenie chwytaka jako efektor. W tym celu należałoby przerobić aktualny mechanizm odpowiedzialny za ruch translacyjny w manipulatorze. Umieszczenie chwytaka umożliwiłoby wykorzystanie robota w wielu aplikacjach, takich jak przenoszenie detali czy też ich segregacja. Interesującym rozszerzeniem byłoby także dodanie systemu wizyjnego

oraz zintegrowanie go z aktualnym systemem sterowania manipulatora. W celu powiększenia przestrzeni roboczej manipulatora należałoby poprawić budowę niektórych elementów drukowanych konstrukcji mechanicznej, które zmniejszają zakresy ruchu przegubów.

Bibliografia

[1] Tchoń K., *Manipulatory i roboty mobilne: modele, planowanie ruchu, sterowanie*, Akademicka Oficyna Wydawnicza PLJ, Warszawa, Polska, 2000.

[2] Spongm M.W., Vidyasagar M., *Dynamika i sterowanie robotów*; Wydawnictwa Naukowo-Techniczne, Warszawa, Polska, 1997.

[3] Qt, Dostępny online: `https://doc.qt.io/` (dostęp: 04-04-2021).

ROZDZIAŁ 3

Badanie hałasu urządzeń użytku domowego na przykładzie chłodziarek

inż. Małgorzata Koziatek
Elektronika

dr inż. Maurycy Kin
Katedra Akustyki, Multimediów i Przetwarzania Sygnałów

Wraz z postępem technologicznym poprawił się standard życia oraz powstało zapotrzebowanie, na możliwość zamieszkania w cichszych i bardziej komfortowych przestrzeniach. Przemysł AGD aktywnie prowadzi badania w dziedzinie hałasu i wibracji w celu poprawy subiektywnej jakości dźwięku wraz ze zmniejszeniem poziomu hałasu chłodziarko-zamrażarek. W końcu produkty, które wytwarzają ten sam poziom w dB(A), mogą w rzeczywistości brzmieć zupełnie inaczej. W połowie lat osiemdziesiątych powstało pojęcie „jakości dźwięku" rozumianej jako wywołującej przyjemne wrażenia słuchowe. Praca obejmuje przeprowadzenie pomiarów poziomu mocy akustycznej zgodnie z międzynarodową normą mówiącą o procedurze pomiaru hałasu chłodziarko-zamrażarek IEC 60704-2-14, przeprowadzenie ankiety subiektywnej oceny jakości dźwięku wśród słuchaczy oraz skorelowanie wyników obydwu metod. Mimo że badane wyroby według nowych rozporządzeń i regulacji należą do tej samej klasy hałasu (klasa B), ocena subiektywna pokazała, że słuchacze odbierają dźwięki tych wyrobów lepiej lub gorzej w zależności od różnych czynników, związanych głównie z aspektami percepcyjnymi organu słuchu.

1. Cel i założenia

Celem pracy jest badanie jakości dźwięku rozumianego przez jego uciążliwość, pojmowanego w kategoriach psychoakustycznych. Badanie ma pokazać, że dźwięk lodówki jest zjawiskiem wielowymiarowym, a ocena subiektywna może różnić się od tego, co przedstawiają dane uzyskane wyłącznie z pomiarów poziomu mocy akustycznej, która dotychczas jest jedyną wartością obecną na etykietach produktów. Praca obejmuje przeprowadzenie pomiarów poziomu mocy akustycznej zgodnie z międzynarodową normą mówiącą o procedurze pomiaru hałasu chłodziarko-zamrażarek IEC 60704-2-14, przeprowadzenie ankiety mającej na celu określenie subiektywnej oceny jakości dźwięku wśród słuchaczy, szczegółową analizę wyników pomiarów z metody obiektywnej oraz skorelowanie wyników z subiektywną oceną słuchaczy. Przyjęto następujące założenia projektowe:

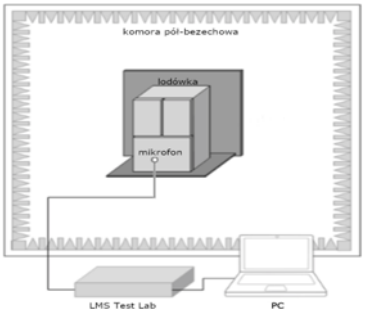
- badanie jakości dźwięku rozumianego przez jego uciążliwość, pojmowanego w kategoriach psychoakustycznych,
- dźwięk lodówki jest zjawiskiem wielowymiarowym, a ocena subiektywna może różnić się od tego, co przedstawiają dane uzyskane wyłącznie z pomiarów poziomu mocy akustycznej,
- aby uzyskać obiektywną ocenę jakości dźwięku należy uwzględnić więcej niż jeden parametr akustyczny.

2. Metoda badań

W pracy zbadano parametry akustyczne wpływające na wrażenie uciążliwości emitowanego dźwięku podczas pracy chłodziarek pochodzących od czterech producentów (ze względu na aspekty marketingowe będą one oznaczone numerami od 1 do 4).

2.1. Badanie średniego poziomu mocy akustycznej SWL

Pomiary przeprowadzono zgodnie z wymaganiami norm ISO 3743-1, ISO 3743-2 i ISO 3744. Do wyrażania hałasu urządzeń do użytku domowego wybrano średni poziom mocy akustycznej. W tym celu wykorzystano układ pomiarowy jak na rysunku 3.1.



Rysunek 3.1. Schemat układu pomiarowego

W celu zmierzenia hałasu lodówki w stałych warunkach, testowana lodówka została zainstalowana w tej samej pozycji i kierunku wewnątrz komory bezechowej, gdzie temperatura wewnętrzna jest utrzymywana na stałym poziomie 23 $\pm 2^{\circ}$, a hałas tła utrzymuje się na poziomie 23 ± 1 dB(A). W pomiarach wykorzystano 2 pomiarowe mikrofony pojemnościowe B&K 4189-A-021. Przed rozpoczęciem pomiarów lodówka była uruchomiona, natomiast pomiary rozpoczynano wtedy, gdy działała w normalnym stanie. Z tyłu lodówki znajdowała się ścianka, która symulowała warunki odbicia spowodowane przez wewnętrzną ścianę rzeczywistego środowiska mieszkalnego. Dodatkowo zainstalowano mikrofon, jak pokazano na rysunku, w odległości 1 m zgodnie z normą IEC 60704-2-1. Pomiar hałasu przeprowadzano, gdy temperatura wewnątrz chłodziarki ustabilizowała się, a praca różnych elementów powtarzała pewien cykl. Pomiary przeprowadzono dla pełnych trzech cykli pracy. Końcowa wartość poziomu wyrażona jest w decybelach ważonych krzywą A, względem poziomu 1 pW.

Analizę przetworzonego sygnału wykonano za pomocą FFT, a dane zebrano w czasie trwania trzech cykli regularnych. Sygnały analizowano w pasmach tercjowych w taki sposób, że widmo dźwięku było podzielone na 21 pasm tercjowych od 100 Hz do 10 kHz. Podziału pasm dokonano w następujący sposób: częstotliwość środkową 11. pasma tercjowego zdefiniowano jako $f_{11}^{ctr}1000$ Hz, co powoduje, że wszystkie niższe częstotliwości środkowe pasm 1/3-oktawowych mogą być zdefiniowane zgodnie ze wzorem (3.1). I odwrotnie, wszystkie wyższe częstotliwości środkowe dla pasm 1/3-oktawowych można określić na podstawie wzoru (3.2). Procentowa szerokość pasma na pasmo 1/3-oktawowe jest stała, opisana wzorem (3.3).

$$f_{n-1} = f_n/2^{1/3} \tag{3.1}$$

$$f_{n+1} = 2^{1/3} f_n \tag{3.2}$$

$$BW = 100\left(\frac{f_n^{high} - f_n^{low}}{f_n^{ctr}}\right) = 23,2\ \% \tag{3.3}$$

2.2. Testowanie BSR (Buzz, Squeak, Rattle)

W celu zdiagnozowania rzeczywistych odgłosów pracy urządzeń AGD wykonuje się tzw. Test BSR, za pomocą którego określa się uciążliwość zdarzeń dźwiękowych towarzyszących normalnej pracy urządzenia. Na te zdarzenia w przypadku urządzeń domowego użytku składają się: brzęczenie (ang. Buzz) – wzbudzane zwykle przez rezonans systemu, skrzypienie (ang. Squeak) – tarcie pomiędzy dwoma ciałami, które generuje hałas o różnej częstotliwości, oraz grzechotanie (ang. Rattle) – dźwięk generowany przez dwa stykające się ze sobą obiekty, który może powtarzać się w czasie i nie jest napędzany przez rezonans, często zaczyna się i kończy wraz ze wzrostem amplitudy drgań, zwykle niezależne od częstotliwości wzbudzenia. Test BSR przeprowadza się podczas regularnych cykli lodówki. Zadaniem prowadzącego test jest zaobserwowanie i zarejestrowanie wspomnianych odgłosów brzęczenia, grzechotania czy skrzypienia. W następnym kroku ocenia się w skali 5-stopniowej prawdopodobieństwo wystąpienia zjawiska oraz jego uciążliwości dla odbiorcy. Mnożąc prawdopodobieństwo wystąpienia zjawiska przez jego uciążliwość otrzymuje się wynik testu, gdzie maksymalna wartość to 25.

2.3. Metryka mR

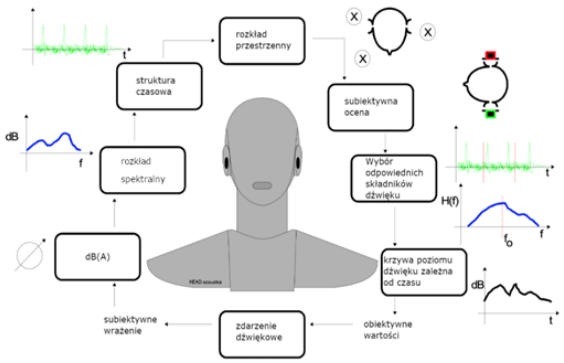
Jest to miara fluktuacji poziomu dźwięku, która wywołuje negatywne subiektywne wrażenie dźwiękowe. Moving Range czyli „zakres ruchomy” jest różnicą między dwoma kolejnymi punktami danych. Dla zbioru danych ten zakres jest listą wartości pokazującą stabilność danych i jest często przedstawiany w postaci wykresu. Im większa jest wartość współczynnika mR, tym dane urządzenie pracuje mniej stabilnie pod względem wytwarzanego hałasu.

2.4. Psychoakustyczne metryki oceny jakości dźwięku

Na rysunku 3.2 przedstawiono sposób, w jaki zarejestrowane informacje za pomocą fantomu sztucznej głowy mogą być oceniane z wykorzystaniem przetwarzania sygnału w celu identyfikacji i eliminacji uciążliwości. Obuuszne przetwarzanie sygnału w ramach sterowanego komputerowo systemu pomiarowego umożliwia wprowadzanie, przechowywanie i przetwarzanie zdarzenia oraz odsłuchiwanie wybranych segmentów, które mogą być stale powtarzane bez słyszalnych efektów. System pomiarowy wyświetla sygnały zarówno dla prawego, jak i lewego ucha w dziedzinie czasu i częstotliwości. Próbka segmentu dźwiękowego może być następnie bezpośrednio manipulowana, a wyniki są wyświetlane i odtwarzane przez słuchawki w „czasie rzeczywistym". Aspekty brzmieniowe są oceniane jako część analizy, tzn. obliczane są parametry decydujące o wrażeniach szorstkości, ostrości, wysokości i barwy dźwięku, jak również głośności w danym fragmencie, które uwzględniają właściwości słuchu [1]. Powyższe dane wyznaczające metryki psychoakustyczne są obliczane w programie ArtemiS firmy Head Acoustics [2] zgodnie z następującymi wytycznymi:

- poziom ciśnienia akustycznego (FFT, Hanning, ważony zgodnie z krzywą A),
- głośność (ISO 532-1),
- fluktuacja w funkcji czasu (wg algorytmu Sottek),
- tonalność (DIN 45681).

Dane są rejestrowane przez mikrofon w komorze półbezechowej.



Rysunek 3.2. Subiektywna i obiektywna identyfikacja, analiza i klasyfikacja dźwięku

Głośność

Głośność (wyrażona w sonach) wyraża wartość tego atrybutu, która występuje przez więcej niż 5% czasu trwania całego sygnału; z definicji 1 son to głośność wywołana przez ton o częstotliwości 1 kHz i poziomie ciśnienia akustycznego 40 dB. Wartościowanie wrażenia głośności oblicza się albo za pomocą określonych wzorców głośności w oparciu o koncepcję krytycznych szerokości pasma podobnych do pasm tercjowych (metoda Zwickera) [3], albo, według Glasberga i Moore'a [4], za pomocą skali ERB wykorzystującej koncepcję równoważnych prostokątnych szerokości pasma. Norma ISO 532-1 określa dwie metody szacowania głośności i poziomu głośności dźwięków odbieranych przez osoby bez ubytków słuchu, w określonych warunkach akustycznych. Pierwsza metoda jest przeznaczona dla dźwięków stacjonarnych, a druga dla dowolnych dźwięków niestacjonarnych (zmiennych

w czasie), w tym dźwięków stacjonarnych jako szczególnego przypadku. Metody te mogą być stosowane do każdego dźwięku zarejestrowanego jako pomiar jednokanałowy przy użyciu mikrofonu lub jako pomiar wielokanałowy, na przykład przy użyciu sztucznej głowy (HATS – ang. Head and Torso Simulator). Ponieważ większość dźwięków przemysłowych jest zmienna w czasie, preferowany jest model zmiennej w czasie głośności.

Ostrość

Ostrość (wyrażana w acumach) jest podcechą barwy dźwięku: opisuje proporcję między głośnością składowych o wysokich i niskich częstotliwościach w widmie. Im więcej energii w wyższych pasmach częstotliwości, tym większa wartość ostrości. Kształt obwiedni widma dźwięku jest odpowiedzialny za odczucie ostrości, natomiast drobna zmiana struktury spektralnej jest stosunkowo mało istotna w wartościowaniu odczucia ostrości [5]. Istnieją modele obliczania ostrości, które nie wykazują zależności od głośności lub wykazują pewną zależność od głośności wpływającą na wrażenie ostrości [6]. Jednostka ostrości, jaką jest acum, została zdefiniowana dla szumu o szerokości jednej wstęgi krytycznej, o częstotliwości środkowej 1 kHz i poziomie ciśnienia akustycznego 60 dB.

Poziom modulacji

Poziom modulacji stosuje się do opisania struktury czasowej sygnału. Jest zdeterminowana przez modulację o różnej głębokości, bierze się jednak pod uwagę inne częstotliwości modulacji – z zakresu od 1 do 20 Hz. Jednostką siły fluktuacji jest vacil – to siła fluktuacji wywołana przez ton o częstotliwości 1 kHz i poziomie ciśnienia akustycznego 60 dB, zmodulowany stuprocentowo sygnałem o częstotliwości modulacji równej 4 Hz. Dobrym przykładem jest para położonych blisko siebie tonów, które powodują dudnienie. Na siłę fluktuacji mogą wpływać: szybkość fluktuacji, poziom ciśnienia akustycznego oraz głębokość modulacji.

Tonalność 1–4 kHz DIN 45681

Tonalność jest miarą zawartości tonów i składowych harmonicznych w sygnale. Im więcej w nim szumów szerokopasmowych, tym niższa wartość tej cechy. Jednostka tonalności, tu, została zdefiniowana dla tonu o częstotliwości 1 kHz i poziomie ciśnienia akustycznego 60 dB. Wysokość dźwięku jest najczęściej kojarzona z częstotliwością. Jednakże wielotonowe dźwięki mogą wywoływać wrażenie pojedynczej wysokości, a szybko powtarzające się impulsy mogą mieć wysokość dźwięku związaną z częstotliwością powtarzania [7].

2.5. Ankieta

W celu przeprowadzenia badań subiektywnych przygotowano ankietę, w której słuchacze byli proszeni o ocenę uciążliwości odsłuchiwanej próbki zgodnie z podaną skalą (tabela 3.1). Nagrania przeznaczone do oceny zostały pozyskane z rejestracji trzech cykli pracy czterech badanych chłodziarek. Próbki dźwiękowe zawierają fragmenty uchwyconych zdarzeń słuchowych oraz zachowania lodówek podczas cykli regularnych. W badaniach wzięło udział 20 słuchaczy, a bodźce dźwiękowe prezentowano w porządku losowym przy użyciu

słuchawek. Odpowiedzi słuchaczy były zapisywane na specjalnie przygotowanym arkuszu. Należy zaznaczyć, że w badaniach wzięły udział osoby niezwiązane z tematem pracy, niektóre z nich miały kontakt z akustyką, niemniej żadna z badanych osób nie miała większego doświadczenia w badaniach odsłuchowych, co wiązało się z koniecznością wykonania ciągu uczącego w celu usystematyzowania odpowiedzi słuchaczy. Słuchacze byli także proszeni o ocenę prezentowanych dźwięków w kategoriach opisowych w przypadku gdy, ich zdaniem, dany dźwięk okazywał się irytujący.

Tabela 3.1. Pięciostopniowa skala ocen zdarzeń słuchowych

1	2	3	4	5
Neutralny	Lekko irytujący	Średnio irytujący	Irytujący	Uciążliwy

3. Uzyskane wyniki

Po zakończeniu badań subiektywnych otrzymane oceny poddano analizie i skonfrontowano z wynikami pomiarów parametrów obiektywnych dźwięku.

3.1. Pomiary obiektywne

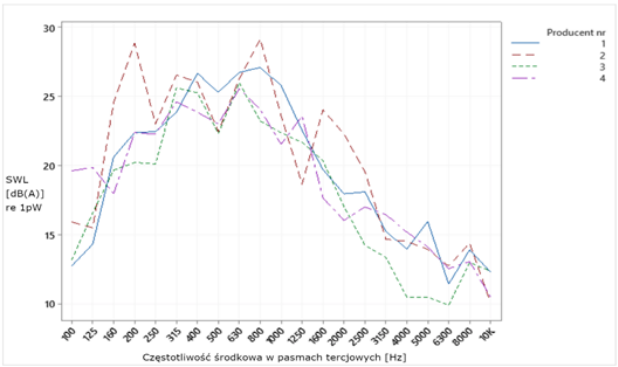
W tabeli 3.2 przedstawiono zmierzone wartości średniego poziomu mocy akustycznej dla każdego cyklu regularnego (SWL), średniego poziomu mocy akustycznej zmierzonego w okresie 3 cykli pracy oraz wartości metryki mR dla wszystkich badanych chłodziarek.

Tabela 3.2. Zmierzone wartości średniego poziomu mocy akustycznej oraz parametru mR

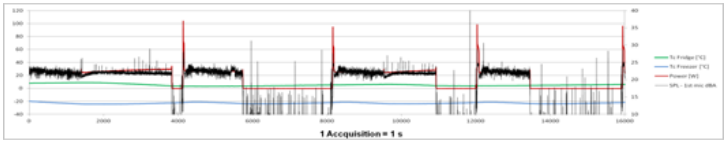
Producent [Hz]	SWL zmierzone [dBA]	Średnie SWL z trzech cykli [dBA]	mR [dBA]
1	35,35	35,52	0,57
1	36,00	35,52	0,45
1	35,20	35,52	0,57
2	36,89	36,62	0,34
2	36,52	36,62	0,28
2	36,47	36,62	0,32
3	34,15	33,89	0,28
3	33,77	33,89	0,26
3	33,74	33,89	0,25
4	33,81	34,30	0,27
4	34,23	34,30	0,30
4	34,86	34,30	0,44

Można zauważyć, że najwyższe wartości zmierzonego poziomu mocy akustycznej uzyskał produkt producenta nr 2, a najniższe produkt producenta nr 3. Wyniki są dość spójne, nie ma dużych różnic między kolejnymi cyklami. Zauważono także, że najwyższe zmierzone poziomy mocy akustycznej uzyskał wyrób producenta nr 2, a najniższe poziomy uzyskały wyroby producentów nr 3 i 4. Istnieje pewna zmienność między cyklami dla każdego z producentów, ale dla producenta nr 4 jest ona największa, a dla producenta nr 3 najmniejsza.

Na rysunku 3.3 przedstawiono uśredniony w trzech cyklach poziom mocy akustycznej w pasmach tercjowych, dzięki czemu można obserwować tonalne zachowanie badanych produktów. Z wykresu można wywnioskować, że każdy z produktów zmienia generowany poziom mocy w przestrzeni pasm tercjowych, w innych miejscach występują piki i nie każdy z hałasów ma charakter szerokopasmowy. Na rysunku 3.4 przedstawiono z kolei przebiegi poziomu ciśnienia akustycznego w funkcji czasu. Można zauważyć, że w połowie długiego cyklu lekko podniosła się temperatura wewnątrz chłodziarki, stąd też wiadomo, że wzrost poziomu mocy akustycznej bezpośrednio wiąże się z włączeniem wentylatora w celu schłodzenia i utrzymania odpowiedniej temperatury wewnątrz komory.



Rysunek 3.3. Średni zmierzony poziom mocy akustycznej w pasmach tercjowych



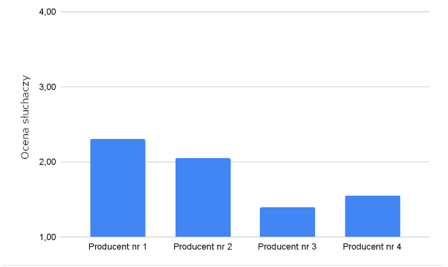
Rysunek 3.4. Zmierzony poziom ciśnienia akustycznego w funkcji czasu, uśredniony w czasie akwizycji trwającej 1 s

W drugiej połowie cyklu poziom ciśnienia akustycznego stabilizuje się, ale czasem występują większe skoki poziomu ciśnienia. Czarna linia wykresu wskazuje pośrednio na zwiększenie wartości mR oraz poziomu modulacji. Widoczny jest znaczny skok ciśnienia akustycznego na samym początku cyklu, a wraz z nim skok poboru mocy, co świadczy o włączeniu się sprężarki, która zużywa najwięcej energii elektrycznej jako element architektury lodówki.

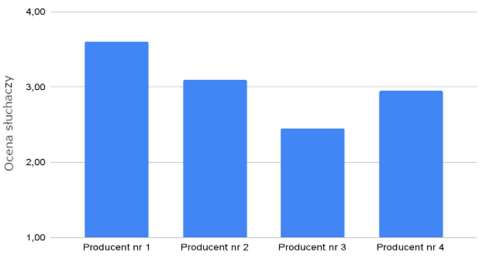
3.2. Badania subiektywne

Na rysunkach 3.5 oraz 3.6 przedstawiono wyniki oceny subiektywnej hałasu chłodziarki podczas całego cyklu regularnego oraz dla odcinka czasu zawierającego tzw. dźwięki charakterystyczne. Można zauważyć, że najlepiej oceniono dźwięk charakterystyczny wyrobu producenta nr 3, którego hałas oceniono jako lekko irytujący. Wyroby producenta nr 2 i 4 oceniono jako średnio irytujący, a producenta nr 1 jako irytujący. Nagrania te prezentują początek cyklu, kiedy sprężarka włącza się. Z otrzymanych przebiegów poziomów ciśnienia akustycznego w funkcji czasu (przykładowo przedstawiony na rysunku 3.4)

widać, że pojawia się w tym czasie wiele zdarzeń słuchowych. Zazwyczaj włączeniu sprężarki towarzyszy krótki, mocny dźwięk przypominający strzał, a także mogą występować miejscowe rezonanse struktury oraz niekontrolowane trzaski. Analizując wyniki ankiety, w której słuchacze wymieniali swoje wrażenia, zauważono, że są one przeważnie związane z testem BSR. Niektóre z komentarzy dotyczą włączenia się dodatkowego źródła hałasu, co naturalnie zwiększa generowany poziom mocy akustycznej. Dźwięki określone jako brzęczenie, buczenie czy stukanie najczęściej są spowodowane budową i pracą sprężarki. W tym badaniu każdy produkt miał zamontowaną produkcyjnie inną sprężarkę. Najwięcej komentarzy negatywnych otrzymała chłodziarka producenta nr 1, a najmniej producenta nr 4. Oceny uzyskane w teście BSR pokrywają się z komentarzami słuchaczy.



Rysunek 3.5. Wyniki uśrednionych ocen subiektywnych dla nagrania z cyklu regularnego



Rysunek 3.6. Wyniki uśrednionych ocen subiektywnych dla nagrania dźwięku charakterystycznego

4. Podsumowanie

Mimo że badane wyroby według nowych rozporządzeń i regulacji należą do tej samej klasy hałasu (klasa B), ocena subiektywna pokazała, że słuchacze odbierają dźwięki tych wyrobów lepiej lub gorzej w zależności od różnych czynników, związanych głównie z aspektami percepcyjnymi organu słuchu. Przykładem może być chłodziarka producenta nr 2. Dla tego wyrobu głośność została obliczona jako najwyższa, podobnie jak zmierzony poziom mocy akustycznej. Jednak słuchacze nie ocenili tego wyrobu jako najgorszy pod względem uciążliwości generowanego dźwięku, co potwierdza hipotezę, że na jakość hałasu (w skali uciążliwości) mają wpływ inne czynniki niż tylko mierzony poziom SWL. Z kolei najgorzej oceniono wyrób producenta nr 1, co może być związane z wysokim wynikiem mR (tabela 3.1) oraz dość wysokimi wynikami głośności, poziomu modulacji i wynikiem testu BSR.

W przypadku wyrobu producenta nr 1 słuchacze nie wykryli wyraźnie „słyszalnego tonu" w przeciwieństwie do pozostałych wyrobów. To zjawisko jest dobrze widoczne na rysunku 3.3, gdzie widać, że hałas produktu producenta nr 1 jest zdecydowanie szerokopasmowy, a żadne pasmo nie jest dominujące nad pozostałymi. W przypadku pozostałych wyrobów słuchacze podkreślali pojawianie się składowych tonalnych, co ma swoje odzwierciedlenie w widmach dźwięku w postaci wyraźnych maksimów (rysunek 3.3). Można więc wywnioskować, że dla tego atrybutu wrażenia słuchowego uzyskano dobrą zgodność pomiarów obiektywnym z subiektywnymi odczuciami oceniających.

Odnotowany przez słuchaczy fakt włączania się dodatkowego źródła hałasu naturalnie zwiększa generowany poziom mocy akustycznej, niemniej jednak jego wartość pozostaje w zakresie charakterystycznym dla tej klasy hałasu. Można przypuszczać, że autorzy norm obowiązujących dla poszczególnych klas hałasu urządzeń AGD uwzględnili możliwość niepercypowania tego typu zjawisk dźwiękowych w normalnych warunkach gospodarstwa domowego, które różnią się od scenerii sesji odsłuchowych choćby poziomem tła akustycznego.

Ze względu na charakter zmierzonych widm wrażenia chropowatości oraz fluktuacji nie zostały odnotowane przez słuchaczy jako istotne, ponieważ maksimum siły fluktuacji uzyskuje się przy częstotliwości modulacji 4 Hz, natomiast dla częstotliwości modulacji 70 Hz odnotowuje się maksimum wrażenia szorstkości dźwięku [8]. Porównanie zmienności danych z rysunku 3.5 i 3.6 pozwala na postawienie tezy o jednorodności ocen subiektywnych uciążliwości dźwięku badanych chłodziarek, dokonanych podczas regularnego cyklu pracy oraz uciążliwości powodowanych pojawieniem się dźwięku charakterystycznego – charakter zmian jest podobny, natomiast zmieniają się wartości ocenianej uciążliwości. Pozwala to na uproszczenie i skrócenie czasu pomiarów subiektywnych do odsłuchu wrażeniowego wyłącznie dźwięków charakterystycznych, a następnie przeskalowanie uzyskanych wyników w celu określenia uciążliwości dźwięków za cały cykl pracy. Wyznaczenie wartości współczynnika skalowania wymagałoby jednak przeprowadzenia badań na większej, bardziej reprezentatywnej liczbie produktów.

Wiele z dostępnych modeli „sound quality" to przedmioty badań i rozwoju. Proces ten ma na celu rozwinięcie zrozumienia percepcyjnego wpływu dodanej złożoności w prostych (regularnych) metodologiach. W związku z tym wiele z omawianych metryk nie jest jeszcze ustandaryzowanych, dlatego należy zachować szczególną ostrożność w ich analizie i stosowaniu, czego przykładem mogą być rezultaty uzyskane w niniejszej pracy.

W wyniku badań odsłuchowych, pomiarów i analizy hałasu urządzeń poprzez zainstalowanie w komorze bezechowej różnych modeli lodówek dostępnych na rynku w ostatnim czasie stwierdzono, że charakterystyka hałasu lodówki jest zależna od rodzaju i stanu pracy wentylatorów i sprężarek stosowanych w lodówce. W skali ocen wyniki poszczególnych produktów plasują się analogicznie, z tym, że wyniki oceny otrzymane dla dźwięku charakterystycznego są bardziej krytyczne, tzn. słuchacze ocenili ww. dźwięki jako bardziej uciążliwe.

Szumy z elementami tonalnymi, wyjące dźwięki i sygnały zmodulowane są często powodem reklamacji składanych przez użytkowników. Percepcja, a więc i ocena zdarzeń dźwiękowych zawierających elementy tonalne, staje się coraz bardziej istotna w przypadku urządzeń powszechnego użytku, np. w dziedzinie akustyki pojazdów, do oceny tonalności spowodowanej alternatywnymi napędami. Ponadto urządzenia i produkty technologii informacyjnej mogą także emitować dźwięki tonalne, które mimo bardzo niskich poziomów ciśnienia akustycznego są niepożądane i należy ich unikać lub je maskować nawet kosztem zwiększenia ogólnego poziomu szerokopasmowego hałasu urządzenia [9, 10] jako mniej dokuczliwego w odbiorze.

Wydaje się, że najlepszą metodą oceny uciążliwości dźwięków wytwarzanych przez urządzenia powszechnego użytku jest „słuchanie wszystkiego" i wykorzystanie własnych wrażeń na temat dźwięku oraz porównanie, czy dana metryka oddaje (lub nie) to, co

słyszymy. Analiza głównych składowych może zatem znaleźć wzorce metodologiczne, które sprawdzają się w badaniach jakości dźwięku, potwierdzając optymalizację ustalonych standardów czy też modeli „sound quality".

Bibliografia

[1] Ozimek E., *Dźwięk i jego percepcja. Aspekty fizyczne i psychoakustyczne*, wydanie pierwsze, Wydawnictwo Naukowe PWN, Warszawa, Polska, 2018.

[2] Head Acoustics. Psychoacoustics II: Calculating psychoacoustic parameters in ArtemiS SUITE, Application Note 03/18. Dostępny online: `https://cdn.head-acoustics.com/fileadmin/data/global/Application-Notes/SVP/Psychoacoustic-Analyses-II-01-2018.pdf` (dostęp: 16-12-2020).

[3] Fastl H., Zwicker E., *Psychoacoustics: Facts and Models*, Springer, Berlin-Heidelberg, Berlin, Niemcy, 2007.

[4] Glasberg B., Moore B., *A model of loudness applicable to Time-Varying Sounds*, Journal of the Audio Engineering Society **2002**, 50, s. 331–342.

[5] Pulkki V., Karjalainen M., *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*, John Wiley & Sons, Hoboken, Stany Zjednoczone, 2015.

[6] Aures W.,*Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale*, Acustica **1985**, 59, s. 130–141.

[7] Sottek R., *Modelle zur Signalverarbeitung im menschlichen Gehör*. Rozprawa doktorska, Rheinisch-Westfälische Technische Hochschule, Aachen, Niemcy, 1993.

[8] Bienvenue G.R., Nobile M.A., *Prominence ratio for noise spectra with discrete tones: a procedure based on Zwicker's critical band research*, Proceedings of Inter-Noise **2020**, 1991,53–5, s. 14.

[9] Johnston J.D., *Transform coding of audio signals using perceptual noise criteria*, IEEE J. Selected Areas in Communication **1985**, 6(2), s. 314–323.

[10] Kyung-Soo K., Weui-Bong J., Tae-Hoon K., Dae-Sik S., Se-jin A., *Psychoacoustical Study for Sound Quality Index of Steady State Noise of Refrigerators*, Transactions of the Korean Society for Noise and Vibration Engineering **2016**, 26, s. 536–545.

Design and control of agile robotic leg

mgr inż. Jakub Tomaszewski
Automatyka i Robotyka

dr inż. Mariusz Janiak
Katedra Cybernetyki i Robotyki

Performing jump has always been challenging research due to many factors that influence its proper execution. This thesis presents a method for designing dynamic robotic legs that are capable of performing a vertical jump. To achieve this, a simulation model of the jump motion was developed and evaluated. Based on the simulation, the requirements for the leg parameter were determined. On this basis, a model of the robotic leg was designed, and then using 3D printing technology, all elements were manufactured and the entire leg was assembled. The purpose of the leg is to perform the vertical jump trajectory obtained from the simulation. The launch phase in the vertical jump has a short duration. To make it possible to apply the trajectory on the real model, an appropriate control system was designed. The system provides high-frequency real-time communication that ensures smooth movement and reliable comparison with the results obtained from simulation. To summarize the results, a series of experiments were performed. In order to compare the simulation results with the real ones, the height of the jumps and various parameters such as position, velocities, and torques of the joints were measured. During the experiments, the leg reached a maximum jump height of 0.415 m.

1. Introduction

In nature, animals jump for many different reasons [1]. Predators, like lions, jump on their prey to knock them down. Squirrels and monkeys jump between the branches of trees. Frogs jump in directions that are difficult to predict to escape danger.

Jumping is associated with a large ground reaction force (GRF) and a short duration of the launch phase. The properties of humans and animals, as well as the results of robotic research [2–4] show that the launch phase is approximately 0.2 seconds. Therefore, large accelerations are needed to achieve high speed in the take-off moment (when the body loses contact with the ground). However, in nature, it is observed that animals with extremely different scales can achieve the same magnitude of jump height because they are geometrically similar. Animals prepare for a jump by bending their legs. To take off,

they rapidly straighten their leg, accelerating over a distance that varies depending on the size of the animal. Therefore, smaller animals need to accelerate much faster to reach the same take-off velocity [5] because the distance from start to take-off is shorter.

The analogy for robots is the same as for animals. The phenomena described above became a motivation to check how changes in the dimensions of the robotic leg can affect the requirements of the speed and torque generated by joint actuators to perform a vertical jump.

2. The purpose and assumptions

The purpose of this thesis is to investigate the influence of leg dimension, kinematic structure, and actuator parameters on a generated trajectory and jump height in a launch phase of a vertical jump. To examine this, the following assumptions were made:

- developing and evaluating a simulation model that will generate an optimal trajectory of the launch phase of the vertical jump for a predefined leg parameters,
- determining the requirements for the leg parameters based on the result from simulation,
- designing and manufacturing a leg prototype based on the determined parameters,
- designing a high performance, real-time control system that will be used to apply the obtained trajectory to the leg prototype,
- determining flight and landing phase of the vertical jump,
- performing experiments to examine the compatibility of the simulation model with the real one.

This thesis includes issues in the fields of robotics, optimization, control theory, mechanics, and electronics. Each of them has a significant impact on the final result of the research. Therefore, to guarantee reliable results, it is crucial to ensure that each stage is performed correctly.

3. Simulation model

The highly dynamic legged jumping motion is a challenging research topic due to the lack of established control schemes that handle the constrained control objective in the launch phase, which are coupled and affect each other [6]. There are many parameters of the leg and actuators that influence each other and the final movement possibilities. The most important are the following:

- mass of the leg [7],
- length of the links,
- kinematic structure of the leg [8, 9],
- inertia of the leg,
- selected motor,
- gear ratio of actuators [10, 11].

In order to investigate the influence of the leg dimensions, kinematic structure, and actuator parameters on the performed movement, a nonlinear oprimization was performed. The result of the optimization is an optimal leg trajectory in the launch phase of the vertical jump. Based on the results, the requirements for the geometry, kinematics

structure, and actuator parameters of the leg were determined. The optimization task also aimed to understand the forces, the relationship between the parameters, and physical phenomena associated with the execution of the jump.

During optimization, a two degrees of freedom leg model was used. The model consist of two joints, the hip (corresponding to an angle q_1) and the knee (corresponding to an angle q_2). The model was used to estimate the torque, speed, and leg dimensions requirements for the jump motion. It was assumed that both links in the leg model have the same length "l". Motors and gearboxes of both joints are placed in the base, which is the center of mass of the leg (CoM). Therefore, most of the mass is at the base of the leg, and the links have relatively small inertia. For this reason, it was assumed that both links are massless, so that all of the mass is lumped together at the base, which is mounted on a vertical rail.

The dynamics of the leg model was defined as a single point mass accelerated due to the ground reaction force (GRF), which is the only external force that could change the mechanical energy of the system. In [12] the researcher assumed that the GRF profile should finish and start at 0 to ensure smooth movement. However, the analysis of vertical jumps in [13] showed that from the beginning of movement there is a positive GRF, which is caused by body weight. The result of the analysis is represented in Figure 4.1.

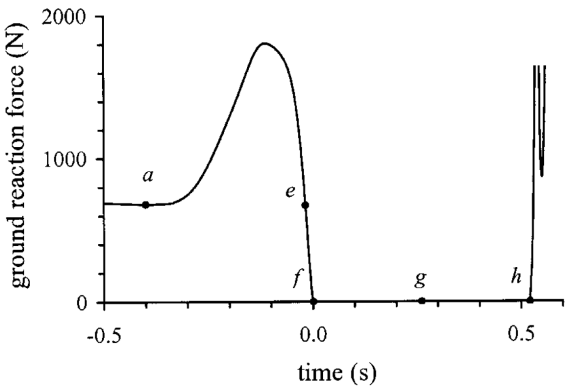


Figure 4.1. Ground reaction force in vertical jump [13]

In order to generate a trajectory of the leg in the launch phase of the vertical jump, it was crucial to introduce constraints to ensure that the leg remains in contact with the ground throughout the motion. They also take into account the limitation of the physical model, which ensures that the obtained trajectory is feasible. Therefore, the following constraints were introduced:

1. **Constraints of joint positions** – ensure that the position of the foot is consistent with the position of the leg resulting from the ground reaction force. The position was obtained by doubly integrating the vertical acceleration caused by the ground reaction force.
2. **Constraints of joint velocities** – ensure that the velocity of the foot is consistent with the velocity of the leg resulting from the ground reaction force. The velocity was obtained by integrating the vertical acceleration caused by the ground reaction force.

3. **Limits of the joint torques** – ensure that the joint torques are within limits caused by the parameters of the selected motor and the assumed gear ratio.
4. **Limits of the joint angles and velocities** – ensure that the joint angles and velocities are within feasible limits. The limits of the joint angles determine the range of leg movement, whereas the velocities of the joints depend on the selected motor, the gear ratio and the selected power supply, as well as the weight and inertia of the robot.
5. **Speed-torque curve constraint** – In motors, the torque is related to the speed. Therefore, the speed and torque of the joints must be below the spped-torque curve of the joint actuators.
6. **Initial and final posture of the robot** – because the developed optimization task generates a trajectory between an initial and final posture of the leg, it is crucial to add a constraint that ensures this.

Objective function

In a brushless motor, the torque is proportional to the current of the motor and the power is proportional to the square of the current. When performing a movement, it is important to minimize the energy required for its execution. Due to the physical limitation of the impetuous changes in joint acceleration, it is prudent to minimize acceleration during movement. That is why it is intuitive to minimize torque and acceleration [6]. To reduce sudden changes that can cause mechanical damage and high energy consumption, changes in torque and jerk are also minimized. The cost function of the optimization task in the launch phase is formulated in Equation (4.1).

$$J = \sum_{k=1}^{N} \left[|\tau_k| + |\ddot{q}_k| + |\tau_k - \tau_{k-1}| + |\ddot{q}_k - \ddot{q}_{k-1}| \right], (k = 1, 2, ..., N) \tag{4.1}$$

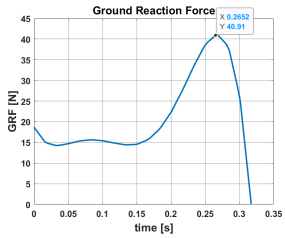
where τ_k and $\ddot{q}_k$ are the joint torque and acceleration at each grid point k, respectively.

3.1. Results

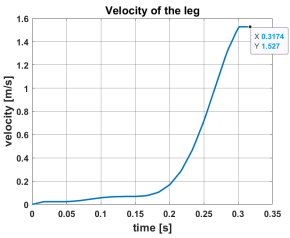
Having implemented the optimization task and selected motors, the next step was to check how leg parameters affect the generated trajectory. This allowed to determine the requirements of the leg in terms of:

- leg kinematics structure,
- length of the links,
- gear ratio of the actuators.

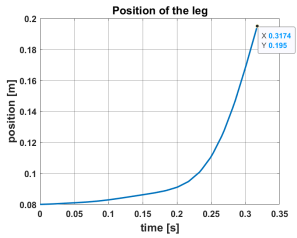
The results of the optimization also determined the velocities, accelerations and torques of the joints necessarry to perform a vertical jump for the selected parameters and given constraints.



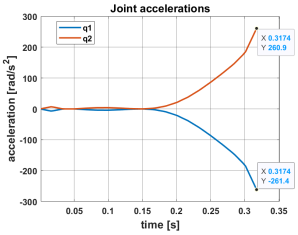
(a) Ground Reaction Force



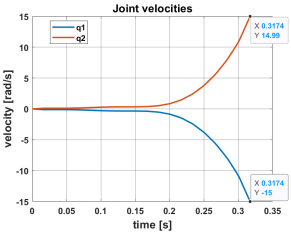
(b) Velocity of the leg base



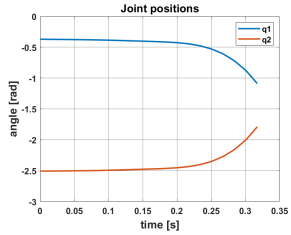
(c) Position of the leg base



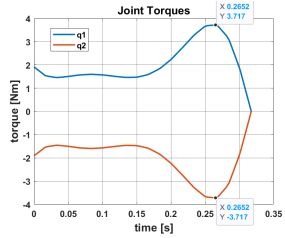
(d) Accelerations of the joints



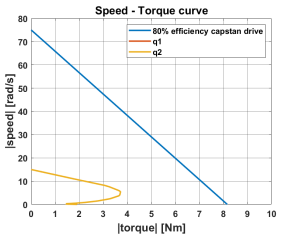
(e) Velocities of the joints



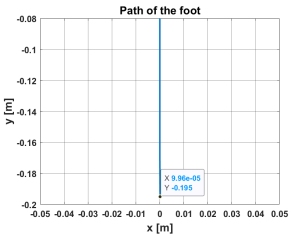
(f) Positions of the joints



(g) Torques of the joints



(h) Torque-speed curve of the leg and each joint



(i) Path of the foot

Figure 4.2. Result of optimization for take-off joint velocity equal to 15 rad/s; gear ratio 6:1, leg mas 1.467 kg, length of links = 0.11 m, start y coordinate = −0.08 m, end y coordinate = −0.195 m, offset in x coordinate = 0 m. The theoretical height of the jump is equal to 0.313 m.

There are many dependencies between the individual parameters of the leg. It is important to note that for inadequately selected parameters, the optimization may not find a solution that meets the given constraints. To make optimization a finite-dimensional problem, during all simulations, the time intervals were sampled at 20 grid points. Hundreds of simulations were performed during the experiment, from which an example was selected and shown in Figure 4.2.

Figures 4.2a, 4.2b and 4.2c show the ground reaction force, velocity, and position of the base of the leg, respectively. In Figures from 4.2d to 4.2g acceleration, velocities, position, and torques of the joints are presented (q_1 denotes the hip and q_2 denotes the knee). Figure 4.2h shows the torque-speed curve of the capstan drive and the relationship between the torque and the speed of each joint. The path of the foot is presented in

Figure 4.2i, which was calculated using inverse kinematics and the joint positions obtained at each grid point. For the obtained results the height of the jump is equal to 0.313 m. On the basis of the results, it can be concluded that the optimization works properly, and for the selected parameters it is possible to generate a trajectory in the launch phase. The GRF presented in Figure 4.2a is consistent with that shown in Figure 4.1. Therefore, it can be concluded that all necessary constraints are included in the optimization task.

Based on all performed experiments, the obtained results and dependencies between them, the requirements of the leg parameters were determined. The most important assumptions about the parameters of the leg are as follows:

- light weight: less than 1.5 kg,
- length of links: 0.11 m,
- special parallel kinematics structure,
- gear ratio: 6:1,
- most of the mass located at the base of the leg,
- small inertia of links,
- actuators that provide small backlash, low inertia, and backdrivability,
- possible to 3D print.

4. Mechanical structure

On the basis of the selected parameters, a 3D model of the leg was designed. Then, a prototype of the leg was manufactured. The purpose of the leg is to perform the vertical jump trajectory obtained from the simulation.

There are many different approaches to designing legged robots [14]. All aim to reduce weight and provide high torque and speed in the joints, which are necessary in legged locomotion. One of the most important parameters is the gear ratio of the actuators because it affects the maximum speed and torque of the joints. However, most walking robots use small gear ratios [11, 15–17]. The reason is that the feet of the robot need to detect contact with the ground. In addition, high impact forces with a large reduction ratio can risk damaging the actuator and the mechanical structure of the leg. Moreover, gearboxes should have minimal backlash to ensure accurate control and small mass to reduce inertia. Therefore, the mechanism should provide small backlash, low inertia, and backdrivability, which are all required in quadruped robots.

In robotics, there are many possible solutions. The best known and commonly used in quadruped robots are planetary gears [12, 15, 17, 18]. Another possible solution are cycloidal drives [19, 20]. There is also a mechanism that is rarely used in robotics. It is a capstan drive, which is a cable driven actuator. Based on the characteristic of each mechanism and the fact that the construction of the leg will be manufactured using 3D printing technology, it was decided to use a capstan drives. It is a highly efficient mechanism that can be used successfully in robotic applications [21, 22]. It allows for precise control, ensuring almost zero backlash [23]. In addition, there are many cable-driven mechanisms which have found applications in robotics [24–26].

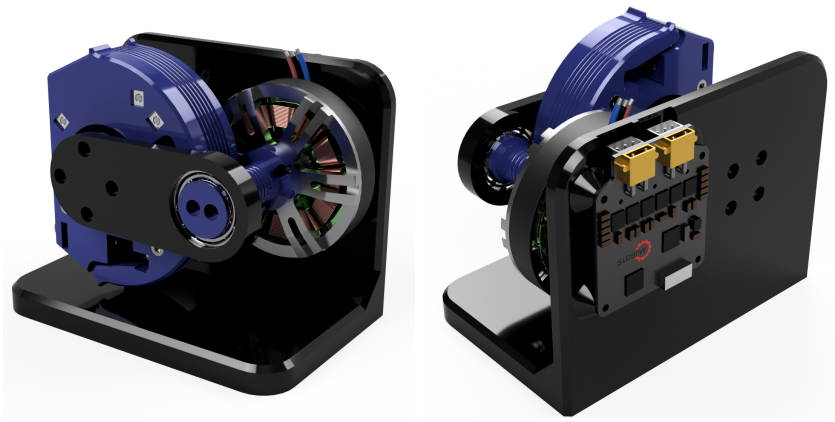
4.1. Gearbox design

The capstan drive consists of the input and output capstan, cable, and tensioning system. An example design of a capstan drive is shown in Figure 4.3. The capstan drive mechanism requires a high-quality tendon. The most popular cable material is stainless steel since it does not creep. However, they are usually inapplicable to legged robots as they have a high ratio of the maximum bending diameter to the cable diameter. The alternative solution is to use a fiber cable material. Based on the characteristics [27] and performed analysis [22, 28] it was decided to choose a Dyneema DM20 which is a new polyethylene fiber material. It has a high strength, small bending diameter, and resistance to creep.



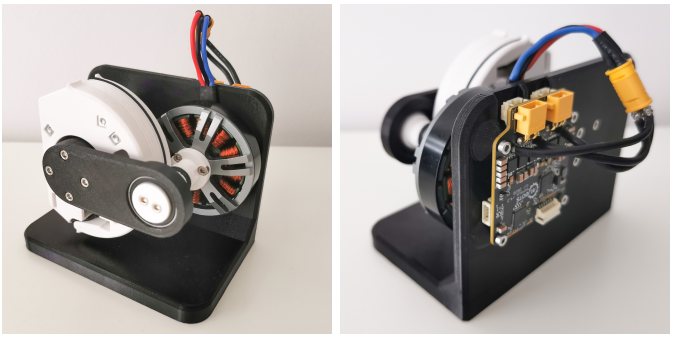
Figure 4.3. Example design of capstan drive [21]

In order to perform tests of the capstan drive, a simple test stand was designed. The designed model of the test stand is presented in Figure 4.4. Then, a prototype of the designed test stand was manufactured and assembled. All necessary parts were 3D printed. The assembled capstan drive test stand is presented in Figure 4.5.



(a) front **(b)** back

Figure 4.4. 3D model of the designed test stand



(a) front **(b)** back

Figure 4.5. Assembled test stand of capstan drive with gear ratio 6:1

After performed tests, the DM20 rope was still taut and none of the components was damaged. The mechanism provides zero backlash, low inertia, and backdrivability. The tensioning system is small, durable, and works as expected.

4.2. Leg design

Based on the determined parameters, the entire leg was designed. The 3D model of the leg is presented in Figure 4.6. It consists of:

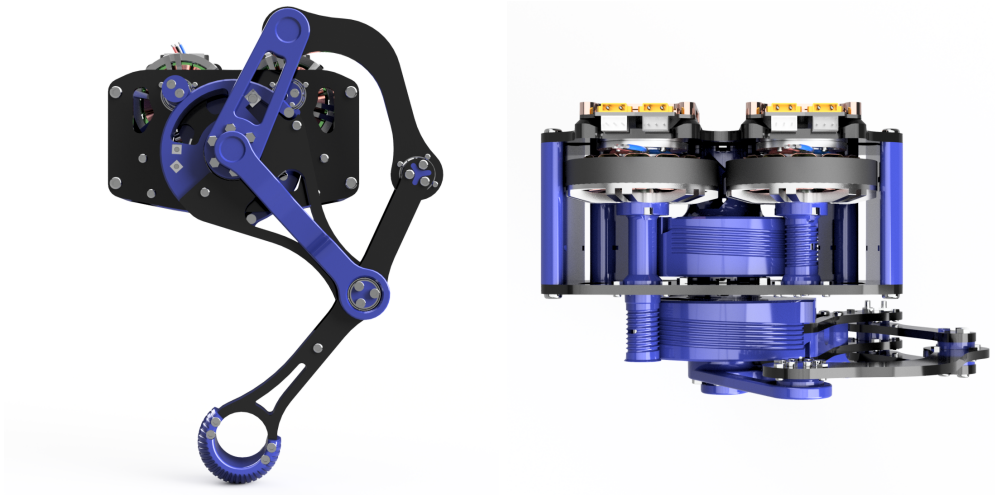
- hip joint with gear ratio 6:1,
- knee joint with gear ratio 6:1,
- base to which both joints are attached.

During modeling, the weight of individual elements was reduced to obtain approximately 1.2 kg of entire leg.



Figure 4.6. 3D model of the designed leg

The front and top views of the leg are shown in Figure 4.7. To demonstrate the connection between individual elements of the leg, a section analysis was performed, which is presented in Figure 4.8.



(a) Front view of the leg **(b)** Top view of the leg

Figure 4.7. Front and top view of the designed leg

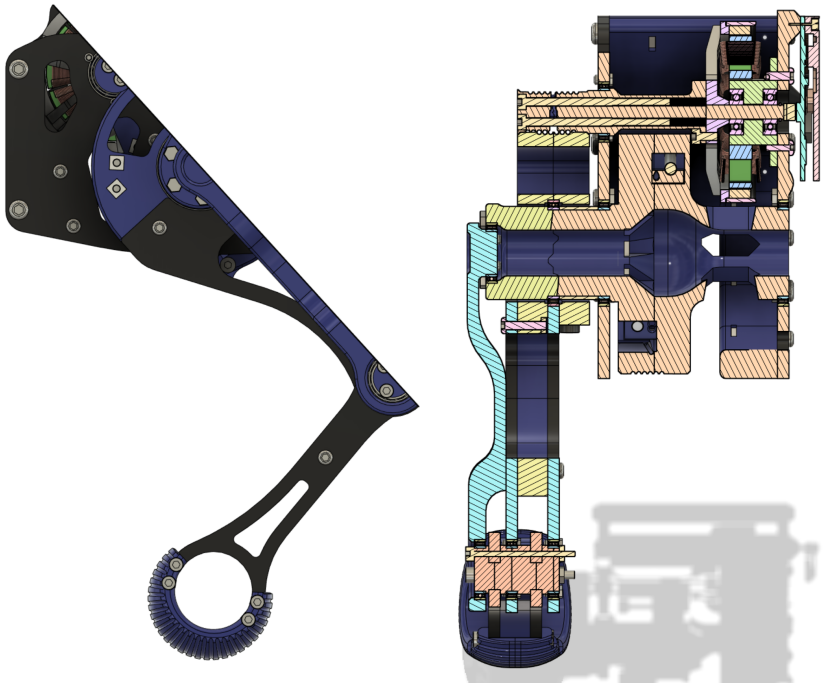
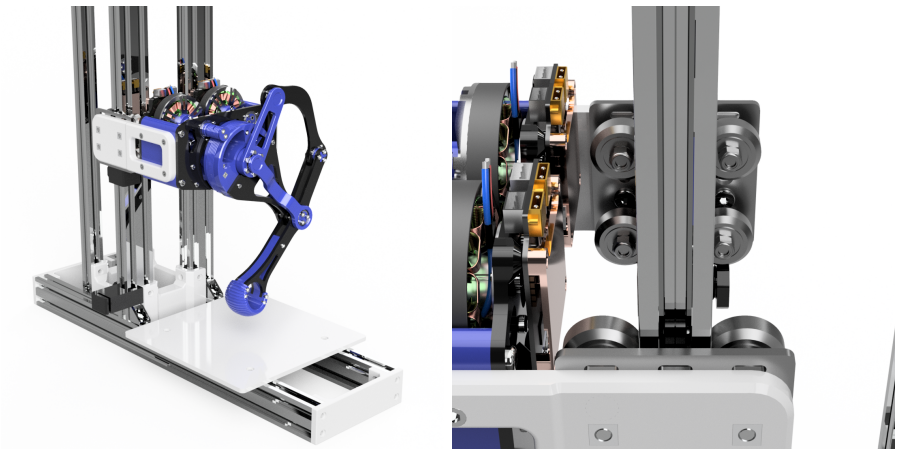


Figure 4.8. Section analysis of the leg showing the connection between the individual elements

To perform a vertical jump, a test stand is required, which ensures smooth vertical movement of the leg. For this purpose, a 3D model was designed, which is shown in Figure 4.10. It consists of 3D printed parts and aluminum profiles. It allows the leg to move vertically at a height of 1.25 m.



(a) Leg mounted to the test stand **(b)** Roller carriages that attached the leg to the aluminium profiles

Figure 4.9. Attachment of the leg to the test stand



Figure 4.10. 3D model of the test stand with attached leg

4.3. Manufacture of designed elements

After designing all the necessary elements, the next step was to manufacture them and assembly the leg. The main goal of the leg is to perform a jump that is associated with a large ground reaction force during the launch and landing phase. After comparing the materials available on the market, it was decided to use PCTG. It is a relatively new material that is an alternative to commonly used PETG. The main difference between them is that PCTG has up to 20 times greater impact strength. All 3D printed parts of the leg are shown in Figure 4.11.



Figure 4.11. All 3D printed part of the leg

After completing all necessary parts, the leg and the test stand were assembled. The manufactured prototype of the leg mounted on the test stand is presented in Figure 4.12. The front and top views are shown in Figure 4.13. The leg has the following specification: theoretical peak torque of each joint: 10.2 Nm; length of the links: 0.11 m; leg weight: 1187 g; number of 3D printed parts: 34; number of bearings: 11; number of screws and nuts: 226.

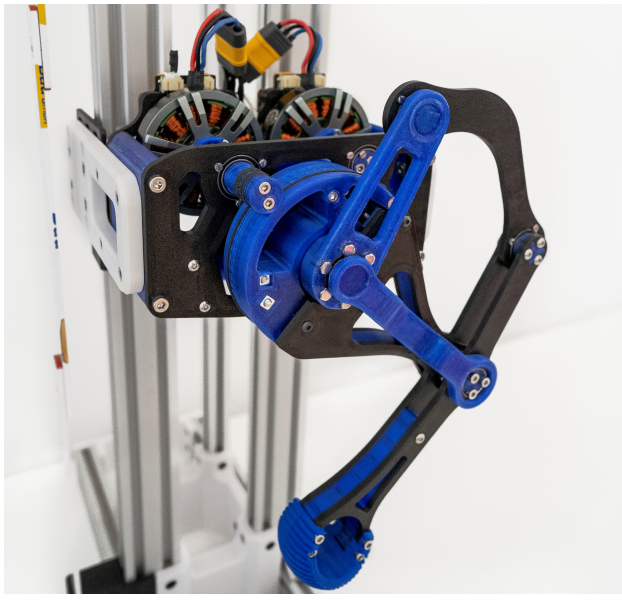
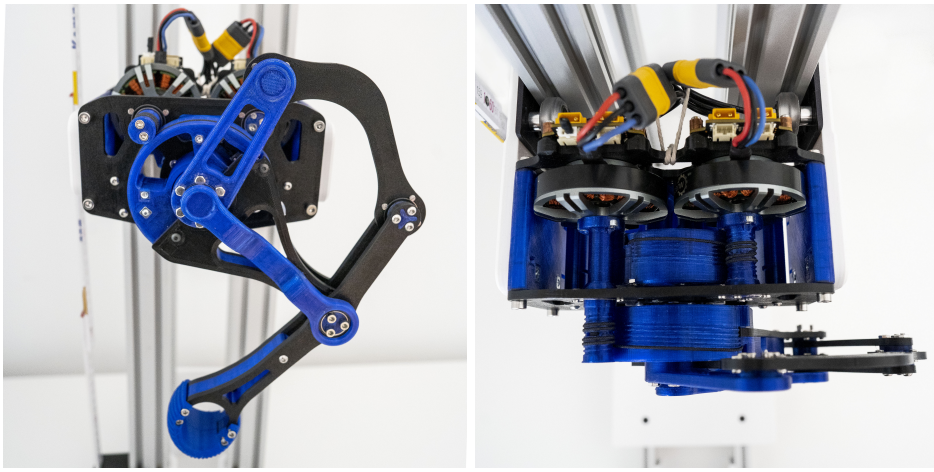


Figure 4.12. Manufactured prototype of the leg mounted on the test stand



(a) Front view of the leg **(b)** Top view of the leg

Figure 4.13. Front and top view of the manufactured leg

5. Control system design

In order to apply the trajectory obtained as a result of optimization to the physical model, an appropriate control system is required. First, all necessary electronics components that provide communication of the system, power management, and execution of a program logic were selected. To ensure a reliable comparison of the simulation model with the real one, a real-time system is necessary. The system was properly configured to ensure that commands are sent and received in real time. The configuration allows the control system to use the data obtained from the optimization properly and to save enough measurement for reliable comparison. The main purpose of the control system is to apply jump trajectories to the physical model. Therefore, an appropriate jump controller was developed. It is responsible for reading the trajectories obtained as a result of optimization and applying them to the physical model.

5.1. Hardware

The connection of all electronic components of the control system is presented in Figure 4.14, where the numbers from 1 to 9 indicate the following elements:

1. mj5208 brushless motors,
2. *Moteus* r4.8 brushless motor controller,
3. *Raspberry Pi* 4B with attached *mjbots pi3hat* r4.4b,
4. *Mjbots power dist* r4.3b,
5. two connected in series 5000 mAh 20 C, 4 cells 14.8 V, 74.0 Wh Li-Pol batteries,
6. power switch,
7. wires supplying power to the entire system,
8. CAN-FD wires providing communication between the *Raspberry Pi* and the motor controllers.

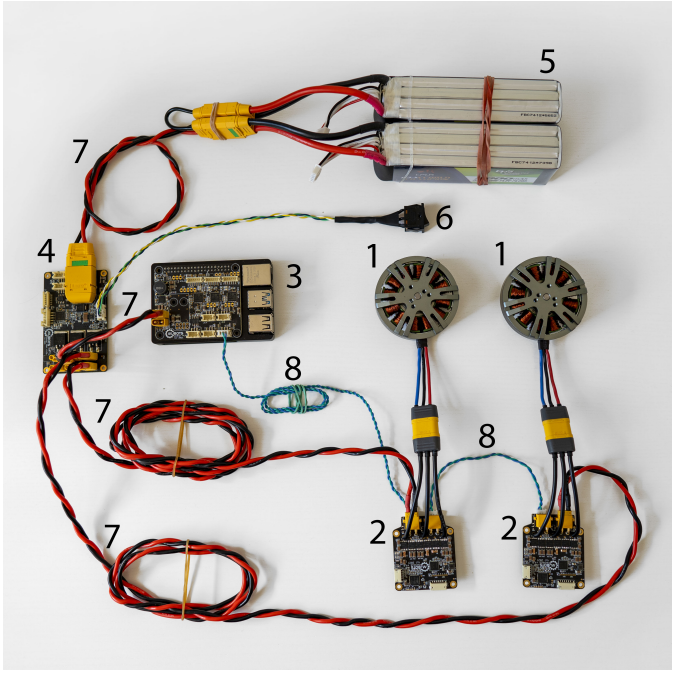


Figure 4.14. Connection of all electronic components of the control system

In the electronic connection, appropriate power and communication wires were used to ensure proper operation. The *mjbots power dist* board distributes power to the *Raspberry Pi* and two *motues* brushless motor controller. In addition, a power switch was used that allows the user to turn on and off the entire system. In the CAN-FD each node of the network must be connected in parallel. In this case, *Raspberry Pi* manages the communication by sending and receiving commands from the *moteus* controllers. The two motor controllers of the hip and knee joints are connected to each other by two CAN-FD interface wires. Therefore, the *Raspberry Pi* must be connected to one of the controllers to provide communication over CAN-FD of the entire system.

5.2. Software

In order to perform a vertical jump, an appropriate controller was developed, which flow is shown in Figure 4.15. At the beginning, depending on whether the system was turned off, the leg is calibrated or offset joint coefficients are read from the file. Then, the trajectories of a launch phase obtained as a result of optimization are read from the file. In the next step, a current time of a program is calculated, and based on them, a proper phase of jump is executed. It is necessary to set the appropriate parameters, i.e., kp and kd scales and the maximum torque in each phase of the jump. In the launch phase, it ensure that the position controller properly influences the motion and improves the results. In the landing phase, the appropriate kp and kd scales together with the maximum torque value allow the leg to avoid colliding and distribute the impact force over time as the leg bends. On the basis of the set parameters and the desired leg configuration, an appropriate command is created. In the end, the program waits enough time to ensure the desired

constant frequency of the loop. After that, the current time of the program is computed. Based on the value, a proper phase of jump is executed, or if its greater than the duration of the jump, the program ends.

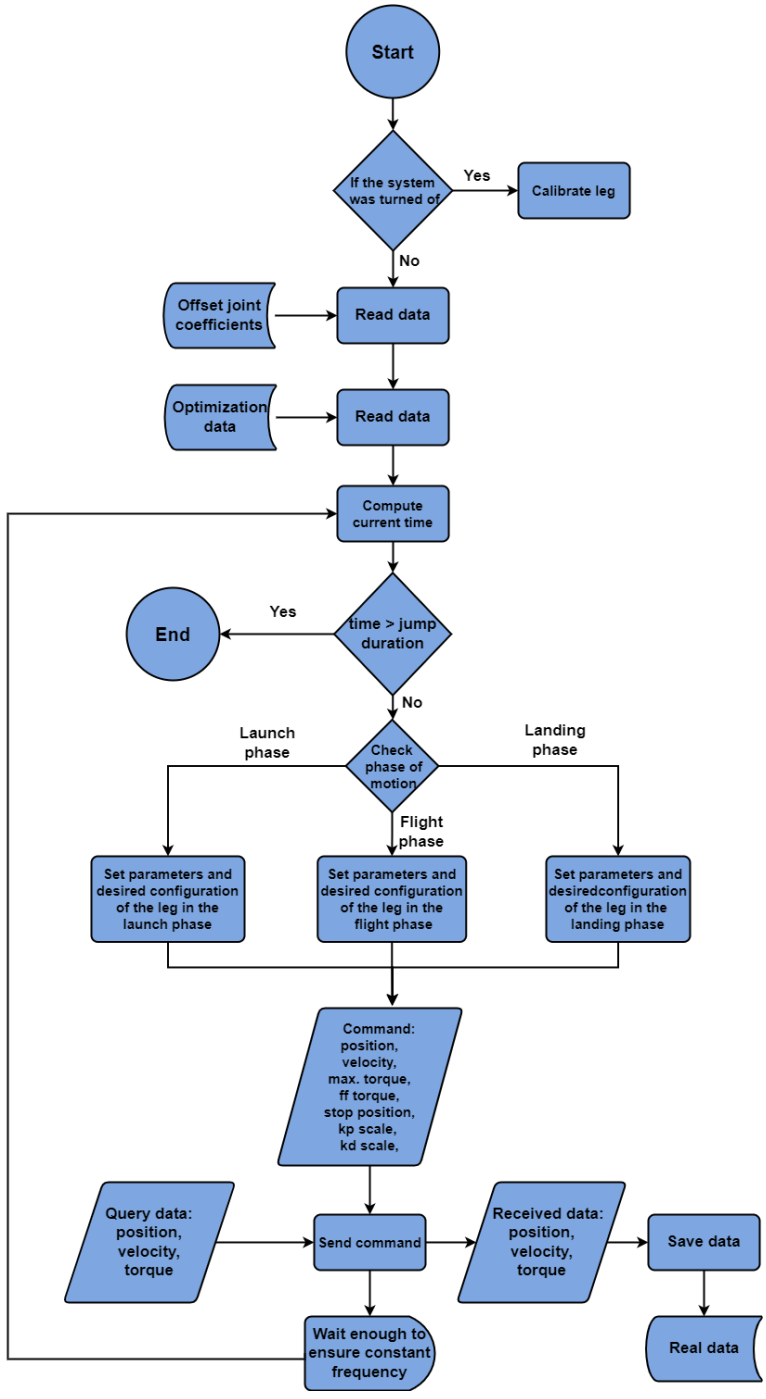


Figure 4.15. Flow of the jump controller

6. Experiments

The purpose of the experiments was to validate whether the trajectories obtained from optimization when applied to the robotic leg would result in a jump. In order to ensure that both models are compatible, appropriate changes were made to the simulation model based on the assembled leg prototype (i.e., mass of the leg and speed limits of the joints).

The optimization task generates an optimal trajectory in the launch phase. However, the vertical jump also consists of a flight and landing phase. During the landing phase, there occurs a large instantaneous ground reaction force. Therefore, an appropriate landing phase is necessary to ensure that the mechanical structure is not damaged.

To compare the simulation model with the real leg, a trajectories for different take-off joint velocities were generated. This affects the acceleration and torque required to perform the desired motion, which provides information on how accurate the simulation model is for different requirements.

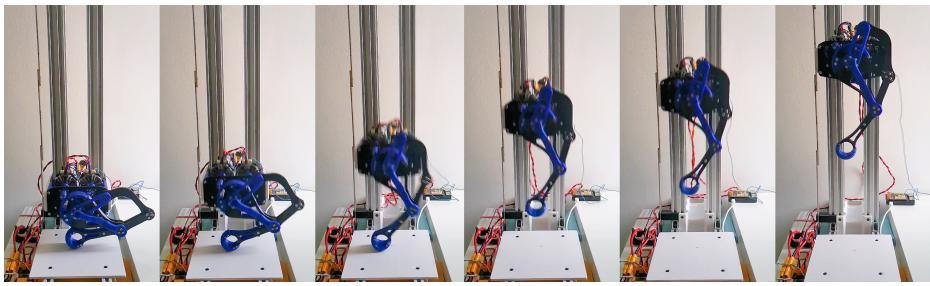


Figure 4.16. Sequential screenshots of one vertical jumping experiment in the launch phase

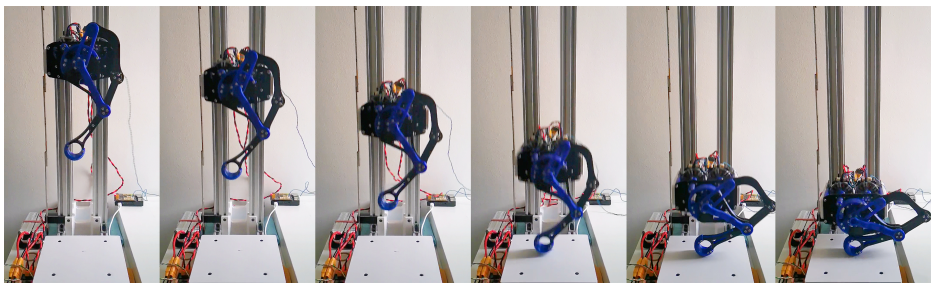


Figure 4.17. Sequential screenshots of one vertical jumping experiment in the landing phase

6.1. Custom Leg trajectory

The custom vertical jump trajectory consists of:

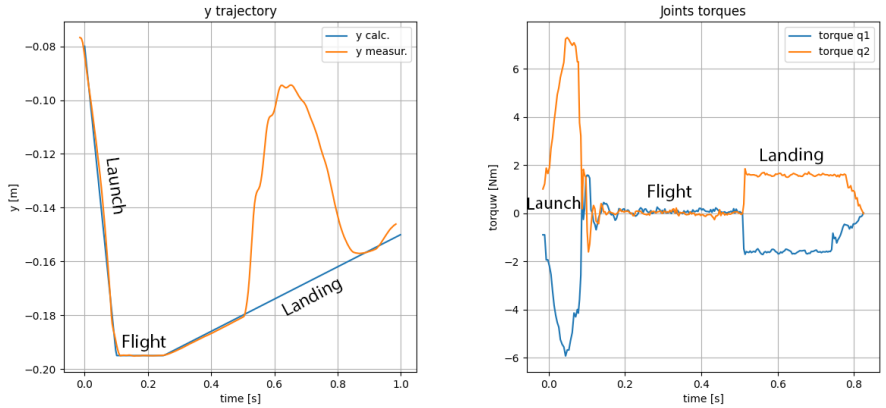
1. Launch phase;
2. Flight phase;
3. Landing phase.

The aim of the launch phase is to straighten the leg, while the landing phase is to bend the leg again. Therefore, the vertical position of the foot was implemented as a function of time. Similarly, as in the optimization task, it is assumed that the foot is vertically

aligned with the base. As a result, the foot moves along a vertical coordinate (denoted y), depending on the current time of each phase.

In order to follow the desired trajectory in the launch phase and achieve a smooth landing, the duration of the entire movement was gradually reduced. This allowed one to adjust the kp and kd scales in each phase and the maximum torque in the landing phase. Finally, the duration of the entire movement was set to 1 second including: launch phase of duration 0.1 s, flight phase of duration 0.15 s, and landing phase of duration 0.75 s.

The maximum torque of the landing phase was adjusted to avoid colliding the base of the leg with the foot. However, it must be small enough to distribute the impact force over time as the leg bends. Figure 4.18b shows the joint velocities and torques during the jump for a maximum torque in the landing phase equal to 1.8 Nm. Comparison of the y coordinate trajectory of the foot in the landing phase is shown in Figure 4.18a. It can be seen that during the jump, the leg follows the desired trajectory in the launch and flight phases. During the landing phase, the leg prepares for contact with the ground by slowly bending. At around 0.5 seconds of movement, there is a sudden change in position due to contact with the ground. The appropriate selected parameters allow the leg to avoid collision and distribute the impact force over time, which is the desired behavior.



(a) Trajectory of the foot – y coordinate for maximum torque in landing phase equal to 1.8 Nm

(b) Torques of the leg joints during vertical jump

Figure 4.18. Trajectory of the foot and torques of the leg joint during vertical jump for maximum torques in landing phase equal to 1.8 Nm

6.2. Optimized leg trajectory

To summarize the results a series of experiments was performed to examine the compatibility of the simulation model with the real one. For this reason trajectories for different take-off joint velocity were generated. Then the results of the optimization for different take-off joint velocity were compared. The increase in the take-off velocity significantly increases the torque and acceleration requirements of the joints. Additionally, as the take-off joint velocity increases, the duration of the launch phase decreases.

Since the optimization task generates only the trajectory of the launch phase, the appropriate flight and landing phases are necessary. The experiments performed in Section 6.1. provided the appropriate landing and flight phase.

To examine how accurate the simulation model is, for all results obtained from optimization, initially only the torque was used to control the motion of the leg. For this purpose, the torque of the joints obtained from optimization was applied as a feedforward torque. To ensure that the position controller does not affect the motion, the kd and kp scales in each sent command were set to 0.

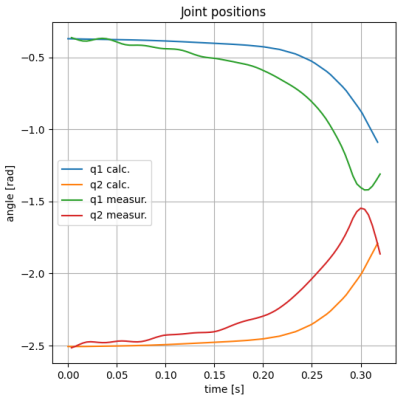
Then, to improve the obtained results and compensate for the inaccuracy of the simulation model to control the leg, the position and velocity trajectory of each joint was also applied. It was important to select appropriate kp and kd scales to ensure the proper influence of the position controller on the motion. Based on the performed experiments, it was decided to set the kp and kd scales during the launch phase at 1 and 0.2, respectively. The torque of each joint obtained from the simulation was applied as a feedforward torque, while the maximum torque of each joint was set to NaN. This ensured that the generated torque was limited only by the power limit of the *moteus* controller.

In the performed experiments, the jump controller described in Section 5.2. was used. The commands in the launch phase consist of:

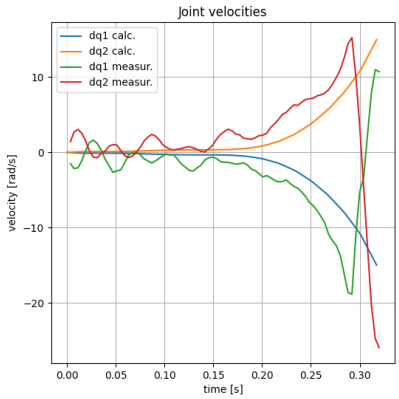
1. Position – set to NaN to ensure that the motion start from the current position of the joint;
2. Velocity – set to the velocity obtained as a result of optimization;
3. Maximum torque – set to NaN;
4. Feedforward torque – set to the torque obtained as a result of optimization;
5. Stop position – set to position obtained as a result of optimization;
6. kp scale – equal to 1 for position controller influence or 0 for torque control;
7. kd scale – equal to 0.2 for position controller influence or 0 for torque control.

The performed experiments were aimed to verify the compatibility of the simulation model with the real one. For this purpose, three trajectories for different take-off joint velocity were generated. The example results for the take-off joint velocity equal to 15 rad/s were presented in Figure 4.2 in Section 3.1. In order to compare the results, the optimal trajectories were applied to the physical model.

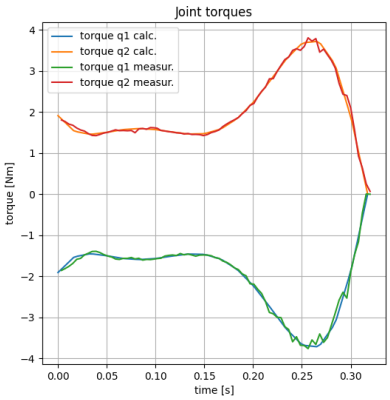
The comparison between the calculated (obtained from optimization) and measured data in the launch phase of the vertical jump when only the torque was used to control is shown in Figure 4.19. Joint positions and velocities are presented in Figures 4.19a and 4.19b, respectively. It can be seen that each joint reaches the final position and velocity, but does not accurately track the desired trajectories. However, the profile of the measured torque is almost identical to the calculated one, which can be observed in Figure 4.19c. The differences are due to measurement noise. The path of the foot during the launch phase is presented in Figure 4.19d. Despite the fact that the measured trajectory of the position differs from the desired one, the maximum errors in the measured path is around 3 mm. For the obtained results, the leg reached a jump height of 0.32 m.



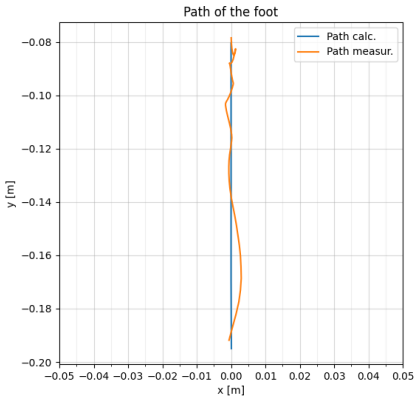
(a) Positions of the leg joints



(b) Velocities of the leg joints



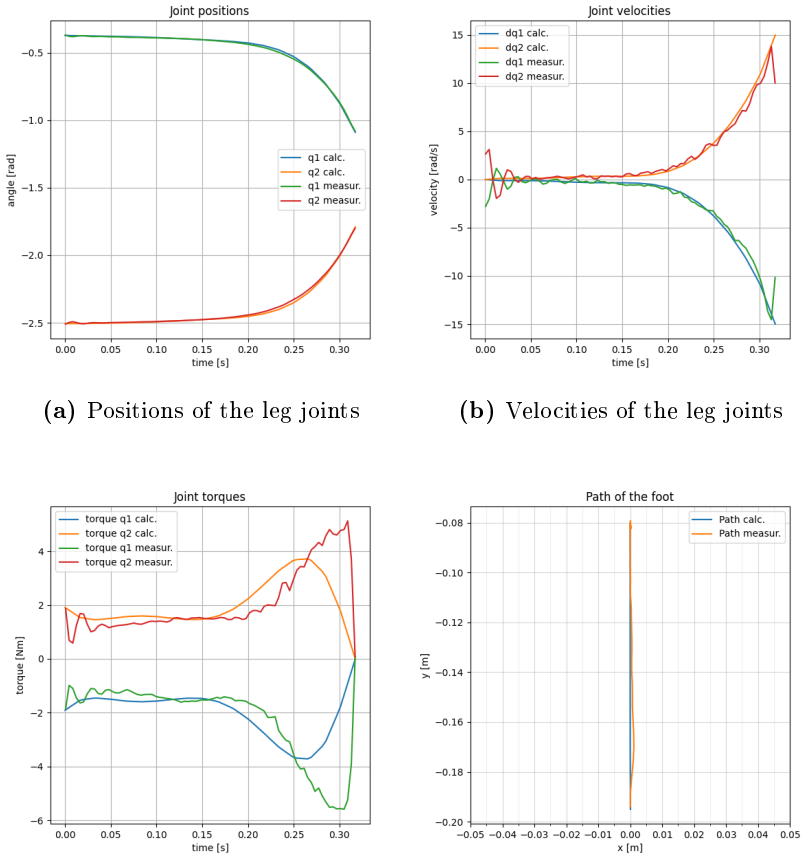
(c) Torques of the leg joints



(d) Path of the foot

Figure 4.19. Comparison of the calculated and measured data during the launch phase for take-off joint velocity equal to 15 rad/s when only the torque was used to control. The leg reached a maximum height of 0.32 m

In order to improve the obtained results, the position and velocity obtained from simulation were also used to control. The comparison between the calculated and measured data is shown in Figure 4.20. It can be seen that the position controller improved the result and that the joint position and velocity trajectories are accurately tracked, which is presented in Figures 4.20a and 4.20b. The differences in velocity are affected by the inaccuracy of the measurements, the fact that there is no set limit of joint acceleration, and the stiffness of the material used to manufacture the links. In Figure 4.20d, it can also be seen that the calculated and measured path are almost identical. The position controller influences the torque generated needed to track the position and velocity trajectory, which is shown in Figure 4.20c. In this experiment, the leg reached a jump height of 0.355 m.



(a) Positions of the leg joints

(b) Velocities of the leg joints

(c) Torques of the leg joints

(d) Path of the foot

Figure 4.20. Comparison of the calculated and measured data during the launch phase for take-off joint velocity equal to 15 rad/s when the joint position, velocity, and torque were used to control. The leg reached a maximum height of 0.355 m

7. Summary

This thesis presents a method for designing dynamic robotic legs. The research shows that using nonlinear optimization it is possible to choose design parameters that can be successfully used to manufacture a high-dynamic robotic leg. Moreover, the designed leg prototype proves that 3D printing and cable-driven actuators are a promising solution for dynamic legged robots. The purpose of this thesis was to investigate the influence of leg dimension, kinematic structure, and actuator parameters on the generated trajectory and jump height in the launch phase of a vertical jump.

The main purpose of the leg was to perform a vertical jump. It was decided to use a two DoF leg since the third one introduces additional mass and redundancy, which is unnecessary when performing planar vertical jump. There are many different structures

of a leg with two DoF. They significantly influence the torque and speed requirements to perform the desired motion and the final mass and inertia of the leg.

To examine the influence of the leg parameter on the generated trajectory and height of jump, a simulation model was developed. Since there are many parameters that influence each other, a nonliear optimization task was defined. The dynamics of the leg was defined as a single point mass accelerated due to the ground reaction force. It was assumed that the foot of the leg remains in contact with the ground during the launch phase. To ensure that foot position and velocity are consistent with those resulting from the ground reaction force, necessary constraints were introduced to the optimization. In addition, to ensure that the obtained trajectory is feasible, the constraints of position velocity and torque of the leg actuators were established. The objective of the optimization is to minimize the energy required to perform the motion. The profile of ground reaction force obtained as a result of optimization is consistent with the presented analysis of a vertical jump. This shows that all necessary constraints are included in the optimization task and the results correspond to reality.

To determine the requirements of the mechanical structure of the leg, several simulations were performed. It was decided to use 3D printing technology to manufacture the entire leg. By comparing different types of gearbox, a capstan drive was chosen which is a cable-driven mechanism. Then an appropriate actuator was manufactured and tested. Based on the test, it was concluded that the mechanism provides zero backlash, low inertia, and backdrivability, which are all required in legged robots. Having determined the requirements and the leg parameters, a two degrees of freedom leg was designed and manufactured. All elements of the leg, except the foot, were 3D printed using PCTG, which is characterized by high impact strength.

To apply the trajectory obtained as a result of optimization to the physical model, an appropriate control system was designed. First, the necessary hardware components were selected that ensure the power supply and communication of the entire system. The selected motor controllers were then calibrated to ensure precise control of the leg.

The main purpose of the control system is to apply jump trajectories to the physical model. This was achieved by implementing an appropriate jump controller, which was responsible for controlling the leg and saving measurement.

To ensure a reliable comparison of the simulation model with the real one, the control system was properly configured. This ensured real-time communication with a constant frequency of 400 Hz.

The optimization task generates an optimal trajectory for the launch phase. However, the vertical jump also consists of the flight and landing phases. The purpose of the flight phase is to ensure that the leg does not begin to bend before losing contact with the ground, while the goal of the landing phase is to prepare the leg for landing and distribute the impact force over time. To determine the trajectories and parameters of these phases, a custom vertical jump was implemented, during which the duration of the entire motion was gradually reduced. As a result, the flight and landing phases were determined which were used during the execution of the optimal launch phase trajectory from the simulation.

Having implemented the optimization task, leg prototype, and configured control system, the compatibility of the simulation model with the real one was examined. For this

purpose, the optimal launch phase trajectories for different take-off velocities were generated. Then the obtained trajectories that differ significantly in the acceleration, torque, and velocity requirements of the joints were applied to the physical model.

First, to determine how accurate the simulation model is, only torque was used to control the physical model. This ensured that the profile of the measured torque was almost identical to the calculated one. The simplified simulation model caused that, the joints did not accurately follow the desired trajectories of position and velocity. However, in each experiment, both joints reached the final position.

Then in order to improve the obtained results and compensate for the inaccuracy of the simulation model, the position and velocity trajectory were additionally used to control the leg. The position controller influenced the torque generated by each joint. This translated into better trajectory tracking in each experiment. For the take-off velocity equal to 15 rad/s, it ensured an almost ideal tracking of the position and velocity trajectory, as well as the path of the foot.

During performed experiments, the leg reached a maximum jump height of 0.415 m. On the basis of the results, it was observed that the position controller increased the height of jump compared to the case when only torque was used to control. The reason is the simplified simulation model, which does not include the dynamics of the leg.

Taking into account that in all experiments, the duration of the launch phase was shorter than 0.32 s and the maximum values of the joint acceleration reached the value of 880 rad/s^2, the differences between the calculated and measured trajectories can be considered small.

A full comparison of the results, the influence of the leg parameters, and kinematics structure on the generated trajectories, as well as the mathematical equation of the developed nonlinear optimization are presented in the full version of this thesis.

Further development

The main goal for further development is to add a third degree of freedom (an abduction/adduction joint) and then use the leg to design and manufacture an entire quadruped robot. The approach used in this thesis to determine optimal leg parameters will allow the quadruped robot to achieve high dynamic locomotion. This, in turn, will provide the opportunity to develop a variety of gait algorithms.

References

[1] McNeill A., *Exploring biomechanics. Animals in Motion*, Scientific American Library, New York 1992.

[2] Hudson P.E., Corr S.A., Wilson A.M., *High speed galloping in the cheetah (Acinonyx jubatus) and the racing greyhound (Canis familiaris): spatio-temporal and kinetic characteristics*, Journal of Experimental Biology **2012**, 14, pp. 2425–2434.

[3] Park H.W., Wensing P.M., Kim S., *High-speed bounding with the MIT Cheetah 2: Control design and experiments*, The International Journal of Robotics Research **2017**, 2, pp. 167–192.

[4] Haberland M., Karssen D., Kim S., Wisse M., *The effect of swing leg retraction on running energy efficiency*, International Conference on Intelligent Robots and Systems **2011**, pp. 3957–3962.

[5] Schmidt-Nielsen K., *Scaling: why is animal size so important?*, Press Syndicate of the University of Cambridge, 2011.

[6] Tian D., Gao J., Shi X., Lu Y., Liu C., *Vertical Jumping for Legged Robot Based on Quadratic Programming*, Sensors **2021**, 21, pp. 1–19.

[7] Seok S., Wang A., Otten D., Kim S., *Actuator design for high force proprioceptive control in fast legged locomotion*, International Conference on Intelligent Robots and Systems **2012**, pp. 1970–1975.

[8] Kang R., Meng F., Chen X., Yu Z., Fan X., Ming A., Huang Q., *Structural Design and Crawling Pattern Generator of a Planar Quadruped Robot for High-Payload Locomotion*, Sensors **2020**, 20, pp. 1–21.

[9] Wang L., Meng F., Kang R., Sato R., Chen X., Yu Z., Ming A., Huang Q., *Design and Implementation of Symmetric Legged Robot for Highly Dynamic Jumping and Impact Mitigation*, Sensors **2021**, 21, pp. 1–16.

[10] Asada H., Youcef-Toumi K., *Direct-drive Robots: Theory and Practice*, Cambridge, MA, USA: MIT Press, **1987**.

[11] Katz B., Carlo J.D., Kim S., *Mini Cheetah: A Platform for Pushing the Limits of Dynamic Quadruped Control*, International Conference on Robotics and Automation (ICRA) **2019**, pp. 6295–6301.

[12] Ding Y., Park H.-W., *Design and experimental implementation of a quasi-direct-drive leg for optimized jumping*, International Conference on Intelligent Robots and Systems (IROS) **2017**.

[13] Linthorne N.P., *Analysis of standing vertical jumps using a force platform*, American Journal of Physics **200**, 69, pp. 1198–1204.

[14] Zhong Y., Wang R., Feng H., Chen Y., *Analysis and research of quadruped robot's legs: A comprehensive review*, International Journal of Advanced Robotic Systems **2019**, 16, pp. 1–15.

[15] Katz B.G., *A Low Cost Modular Actuator for Dynamic Robots*, Massachusetts Institute of Technology **2016**.

[16] Levine G., Boyd N., *Blackbird: Design and Control of a Low-Cost Compliant Bipedal Robot*, 2019.

[17] Zeng X., Zhang S., Zhang H., Li X., Zhou H., Fu Y., *Leg Trajectory Planning for Quadruped Robots with High-Speed Trot Gait*, Applied Sciences **2019**, 9(7), pp. 1–21.

[18] Arm P., Zenkl R., Barton P., Beglinger L., Dietsche A., Ferrazzini L., Hampp E., Hinder J., Huber C., Schaufelberger D., Schmitt F., Sun B., Stolz B., Kolvenbach H., Hutter M., *SpaceBok: A Dynamic Legged Robot for Space Exploration*, International Conference on Robotics and Automation (ICRA) **2019**.

[19] Arm P., Zenkl R., Barton P., Beglinger L., Dietsche A., Ferrazzini L., Hampp E., Hinder J., Huber C., Schaufelberger D., Schmitt F., Sun B., Stolz B., Kolvenbach H., Hutter M., *A New Design of a Two-Stage Cycloidal Speed Reducer*, Journal of Mechanical Design **2011**, 133.

[20] Wang K., Marsh D., Saputra R. P., Chappell D., Jiang Z., Raut A., Kon B., Kormushev P., *Design and Control of SLIDER: An Ultra-lightweight, Knee-less, Low-cost Bipedal Walking Robot*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) **2020**.

[21] Mazumdar A., Spencer S.J., Hobart C., Dabling J., Blada T., Dullea K., Kuehl M., Buerger S.P., *Synthetic Fiber Capstan Drives for Highly Efficient, Torque Controlled, Robotic Applications*, IEEE Robotics and Automation Letters **2017**, 2.

[22] Hwangbo J., Tsounis V., Kolvenbach H., Hutter M., *Cable-Driven Actuation for Highly Dynamic Robotic Systems*, IEEE Robotics and Automation Letters **2018**.

[23] Lu Y., Fan D., *Transmission backlash of precise cable drive system*, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science **2013**, 227.

[24] Kim Y.J., *Anthropomorphic Low-Inertia High-Stiffness Manipulator for High-Speed Safe Interaction*, IEEE Transactions on Robotics **2017**, 33.

[25] Song H., Kim Y.S., Yoon J., Yun S.H., Seo J., Kim Y.J., *Development of Low-Inertia High-Stiffness Manipulator LIMS2 for High-Speed Manipulation of Foldable Objects*, International Conference on Intelligent Robots and Systems (IROS) **2018**.

[26] Mustafa S.K., Yang G., Yeo S.H., Lin W., Chen I.M., *Self-Calibration of a Biologically Inspired 7 DOF Cable-Driven Robotic Arm*, Mechatronics, IEEE/ASME Transactions on **2008**, 13.

[27] Peijs T., *High Performance Polyethylene Fibers*, Reference Module in Materials Science and Materials Engineering **2017**.

[28] Martin V., Bill F., Jorn B., Sergio L., Peter D., *Development of HMPE fiber for deep water permanent mooring applications*, Brazil, 2012.

Silicon sensor characterisation and radiation hardness qualification for the CMS Endcap Calorimeter Upgrade for the High-Luminosity LHC

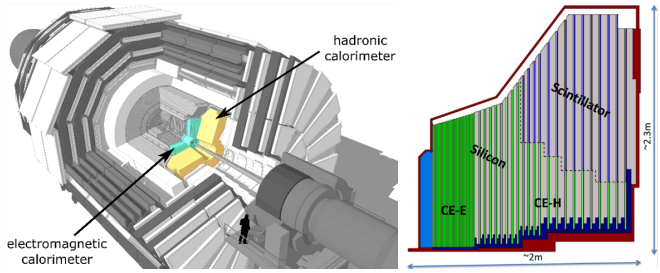
mgr inż. Oliwia Kałuzińska
Elektronika i Telekomunikacja

dr hab. inż. Michał Mazur
Katedra Mikroelektroniki i Nanotechnologii

The European Organization for Nuclear Research (CERN) is in the upgrade phase of the Large Hadron Collider, the largest particle accelerator in the world. One of the upgraded elements is the CMS experiment. It is one of the four detectors located at the Large Hadron Collider. The project of the master's thesis focused on one part of the CMS experiment – the Endcap Calorimeter. It will be replaced with a completely new element, based i.a. on silicon sensors as active material. As part of the diploma thesis, the electrical characterisation and radiation tolerance qualification was performed.

1. Goals and assumptions

The present thesis was pursued at the European Organization for Nuclear Research (CERN). Currently, CERN is in the upgrade phase of the Large Hadron Collider (LHC). The project is referred to as the High-Luminosity LHC (HL-LHC). The aim of the project is to increase the luminosity of the accelerator, which will increase its efficiency and probability for new discoveries This thesis is a part of the project, focusing on upgrading the CMS experiment, especially one part – the Endcap Calorimeter (shown in Figure 5.1a). It is presenting the characterisation of the silicon sensors, which is one of the active materials in the upgraded part as well as qualification of their level of radiation tolerance. The Endcap Calorimeter will be replaced with a completely new element, namely the High Granularity Calorimeter (HGCAL) The cross section of the HGCAL is shown in Figure 5.1b.



(a) the current Endcap Calorimeter – location within CMS detector **(b)** cross section of the new calorimeter design

Figure 5.1. Calorimeter

2. Theoretical introduction

Solid state materials can be divided into three groups: insulators, semiconductors and metals. They are defined by the arrangement of the energy bands – the valence band and the conduction band. In metals, the valence and the conduction band overlap, whereas in semiconductor and insulators, the energy bands are separated by the bandgap. The bandgap energy, the forbidden energy gap, is the energy range between the bottom of the conduction band and the top of the valence band [3].

Position sensitive semiconductor devices were developed for the detection and characterisation of charged particles and are nowadays widely used in experimental particle physics. They exhibit several unique properties that are desirable for high-energy physics, for example: ability of extremely precise position measurement, high readout speed, the simultaneous measurement of energy, position and arrival time and the possibility of integrating detector and readout electronics on a common substrate. The semiconductor material that is most widely used in high-energy physics is silicon [4].

In the absence of impurities, the concentration of electrons n in the conduction band and the holes p in the valence band are equal and the semiconductor is called intrinsic. In order to enhance the conductivity of the intrinsic silicon, additional states in the forbidden gap can be inserted. The process is called doping. Silicon can be doped in two ways: by adding group V element (e.g., arsenic or phosphorus) and thus creating n-type silicon with additional electrons or by adding group III element (e.g., boron) and thus creating p-type silicon with additional holes [4]. The structure that performs a significant function in modern electronics applications is a p-n junction (or diode). It is formed by connecting p-type and n-type materials. When these two regions are brought into contact, electrons will diffuse from the n-side into the p-side and holes from the p-side to the n-side. As a result, a surplus of negative electric charge in the p region and of positive charge in the n region will be created. This generates an electric field that induces a drift counteracting the diffusion. In thermal equilibrium and without external voltage applied, diffusion and drift stabilize and a region depleted from freely moving charge carriers is formed around the boundary of p- and n-sides. The region is called depletion region or space-charge region [3]. The p-n junction can only conduct the current when it is forward biased. However, under reverse bias, some current flows through the depletion region. This current is referred to as

leakage current. It is caused by diffusion of the minority carrier across the depletion region and carrier generation within the depletion region. Irradiation of silicon diodes leads to generation of defects in the silicon bulk. The defects conduce to thermal generation of electron–hole pairs resulting in significant rise in the leakage current, which can affect the sensor performance. It contributes to the noise of the detector which impacts the position resolution [5, 6].

The p-n junction can be transformed into a position sensitive detector by segmenting the front side implant, i.e., creating multiple front side implants instead of one big implant. The readout of the signal is done from all implants separately. A charged particle can traverse the sensor based on p-n junction with segmented implants and generates electron-hole pairs. Charge carriers that are created in the depletion region are collected by the electrodes while the charge produced outside this region is lost due to recombination with free carriers. That is why the silicon sensors are operated under the full depletion, that is, when the reverse bias voltage is high enough that the depletion region extends over the full width of the p-n junction. The minimum value of the reverse voltage at which the junction is fully depleted is called full depletion voltage [5].

The HGCAL silicon sensors will be exposed to high fluences (up to $1.0 \cdot 10^{16} n_{eq}/\mathrm{cm}^2$) resulting in generating bulk damage in the sensor. The radiation damage produced by incident particles can be quantified and scaled by the so-called non-ionising energy loss hypothesis. The NIEL hypothesis assumes that the amount of radiation damage due to lattice displacements produced in the material is linear proportional to the non-ionizing energy loss of the traversing particles. According to the NIEL hypothesis, the change of the leakage current (ΔI) per bulk volume (V) scales linearly with fluence (ϕ) [6]:

$$\Delta I = \alpha \cdot \phi \cdot V \tag{5.1}$$

The proportionality factor α, presented in equation (5.1) is referred to as current related damage rate [4, 7]. The level of the leakage current as well as alpha value are studied in this thesis for the test structures of different thicknesses and exposed to different fluences.

3. The project and realization

The silicon sensors for the HGCAL (also known as full sensors) are produced in hexagonal geometry diced from 8-inch circular wafers by an external company – Hamamatsu Photonics K.K. The structures are mounted on a so-called dicing frame, which has a sticky surface that holds the wafer, surrounded by a plastic frame. There are three active thicknesses of the sensors predicted for the HGCAL: 120 µm, 200 µm and 300 µm. Each sensor is segmented into several hundred diodes also known as cells (or channels, pads) of mainly hexagonal shape. On the remaining pieces of the circular wafer, a so-called test structures were produced, such as single diodes, which are the same structures as a cell on a full sensor. Thanks to their small size, they can be easily measured instead of the full sensors, using infrastructure which is not adapted for full wafers. The full sensor together with the test structures on a dicing frame is shown in Figure 5.2. In this thesis, the electrical characterisation and radiation tolerance qualification of the silicon sensors was presented.

Activities preformed during the thesis project include the electrical characterisation and charge collection study of the neutron irradiated test structures, electrical characterisation of unirradiated full hexagonal sensors and integration of two initially separate setups for characterisation of the test structures.

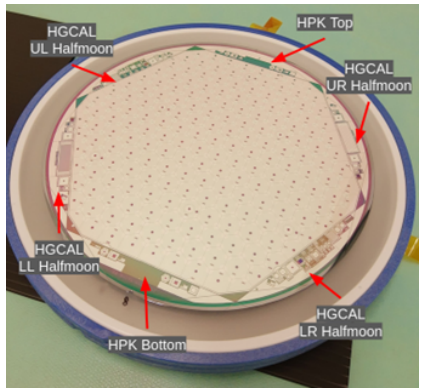
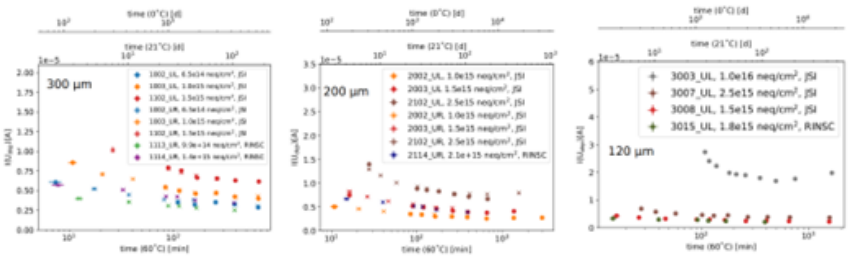


Figure 5.2. HGCAL silicon sensor on a dicing frame

3.1. Electrical characterisation of test structure diodes

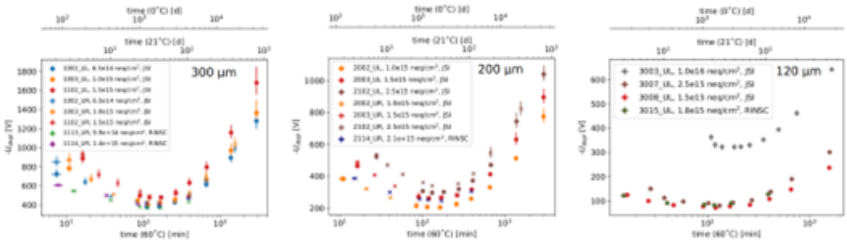
Electrical characterisation of the test structure diodes consist of leakage current (IV) and capacitance (CV) measurements as a function of the bias voltage. The electrical properties of the samples irradiated at two institutes, namely Jožef Stefan Institute (JSI) and Rhode Island Nuclear Science Center (RINSC), are measured and compared, including an isothermal annealing study. The aim of this study is to check whether the fluence estimated for the test structures irradiated at RINSC are accurate. It is done by comparing them to the diodes irradiated at JSI, which reactor has a well-known neutron spectrum and fluence calibration thus can serve as reference. RINSC is a new institute for HGCAL activities and needs to be validated, as it is the only institute where the irradiation of the 8-inch wafer is possible. The leakage current values at the bias voltage equal to depletion voltage as a function of the annealing time at 60 °C are shown in Figure 5.3. Each plot presents sensors of one thickness. Figure 5.3a present results for 300 µm sensors – six irradiated at JSI and two irradiated at RINSC. The samples are irradiated to three fluences at JSI ($6.5 \cdot 10^{14}, 1.0 \cdot 10^{15}$ and $1.5 \cdot 10^{15} n_{eq}/\mathrm{cm}^2$) and determine the leakage current levels at these fluences. Looking at the samples from RINSC, a discrepancy between fluence information can be observed. Figures 5.3b and 5.3c present results for 200 µm and 120 µm sensors, where similar observations can be made.

The Figure 5.4 presents depletion voltage values as a function of the annealing time at 60 °C, each plot for particular sensor's thickness. The first part of the plot shows an effect of beneficial annealing, when the depletion voltage is decreasing. The minimum is reached at around 100 min at 60 °C for all thicknesses. Annealing above this value results in increasing the depletion voltage, which is an effect of reverse annealing. Annealing time at 60 °C can be converted to equivalent annealing time at a different, target temperature, for example room temperature or 0 °C. It can help to predict for how much time the sensor can be annealed during its lifetime at HL-LHC to avoid reverse annealing. At lower



(a) 300 µm **(b)** 200 µm **(c)** 120 µm

Figure 5.3. Leakage current at depletion voltage as a function of annealing time at 60 °C



(a) 300 µm **(b)** 200 µm **(c)** 120 µm

Figure 5.4. Depletion voltage as a function of annealing time at 60 °C

temperatures the annealing effects are slower. By performing the electrical characterisation of the sensors it is possible to calculate the current related radiation damage (α). The alpha value was determined for all annealing steps, separately for samples irradiated at JSI and RINSC. The relation between alpha value and annealing time for all samples is shown in Figure 5.5. It is clearly seen that the alpha values for the samples irradiated at RINSC are lower by a fraction of 2/3 than corresponding values for the diodes irradiated at JSI. The presented comparison of electrical properties between samples irradiated at JSI and RINSC explicitly show a discrepancy of fluence estimation at the two institutes. The real fluence appears to be lower than it is reported.

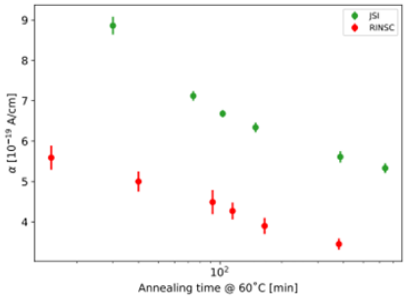
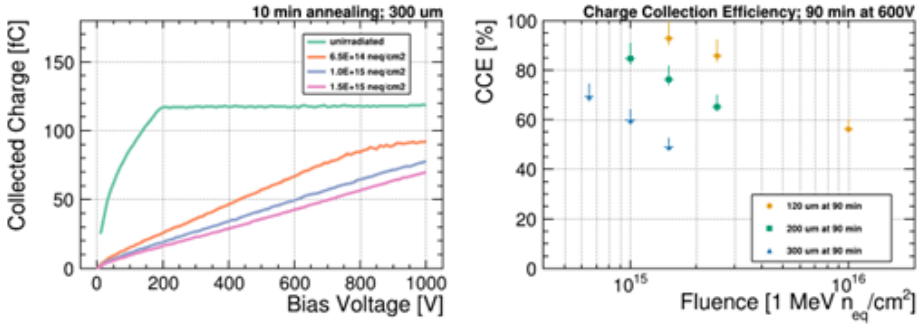


Figure 5.5. Comparison of alpha values for samples irradiated at JSI and RINSC for all annealing steps performed

3.2. Charge collection of test structure diodes

Next to electrical characterisation of the test structures, a charge collection study was performed. In order to obtain the collected charge of the irradiated diodes, laser induced transient currents (TCT) were recorded. All measurements presented in this section were performed at −20 °C, both for irradiated and unirradiated (reference) samples. The collected charge is calculated by integrating the TCT signal over time for each bias voltage. Then, the charge collection as a function of the bias voltage can be plotted as shown in Figure 5.10a. The figure presents the collected charge for 300 µm diodes (one reference and three irradiated samples) at 10 min of the annealing at 60 °C.



(a) Charge collection as a function of the bias voltage for 300 µm samples at 10 min of the annealing at 60 °C

(b) Charge collection efficiency as a function of the fluence at 90 min of the annealing

Figure 5.6. Results of the charge collection study

From the collected charge for all thicknesses and annealing steps, the charge collection efficiency is extracted. It is defined as the ratio of collected charge of an irradiated sample and the collected charge of the reference (unirradiated) sample of the same thickness. The charge collection efficiency at 90 min of annealing as a function of the fluence at the bias voltage of 600 V for all nine measured samples is shown in Figure 6b. The efficiency is decreasing with the fluence. The 300 µm diodes were irradiated to the lowest doses and they have the charge collection efficiency ranging between around 70% and 50%. For the 200 µm diodes, it is between 85% and 65%. The 120 µm samples, which were irradiated up to $1.0 \cdot 10^{16} n_{eq}/\mathrm{cm}^2$, exhibit the charge collection efficiency ranging between around 90% and 60%. One sample for each thickness was irradiated to $1.5 \cdot 10^{15} n_{eq}/\mathrm{cm}^2$. By analyzing this fluence point, it can be observed that the thinner the diode, the higher the charge collection efficiency. Therefore, the 120 µm will be located in the highest radiation regions in the HGCAL.

3.3. Electrical characterisation of full sensors

Electrical characterisation of the full sensor, as for the diodes, includes leakage current and capacitance measurements. After the measurement the results are processed and

analysed. This procedure includes grading of the sensor, which determines whether the sensor fulfills specified requirements.

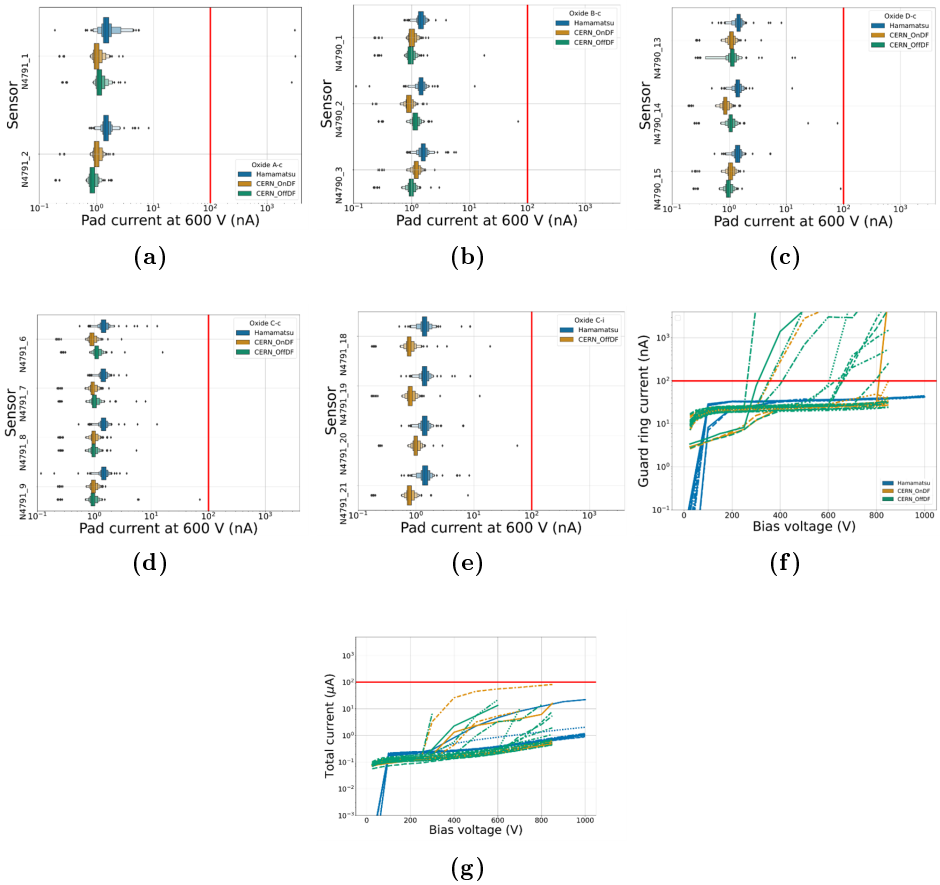


Figure 5.7. Summary of pad current at 600 V, guard ring current and total current for HPK, on DF, off DF

The measurements can be summarized in plots presented in Figure 5.7. They are showing the distribution of the pad current at 600 V in the form of a boxplot, guard ring current and total current as a function of the bias voltage for all tested sensors in three configurations – at external producer, at CERN on dicing frame (on DF) and at CERN removed from dicing frame (off DF). The presented results show that the process of removing the sensors from the dicing frame significantly worsen their properties. The sensors may sustain some mechanical damages during this activity. The effect is especially visible in the guard ring current, which grows notably high. Many of the sensors fail the grading. Therefore, the procedure of removing the sensors from the dicing frame at CERN should be improved or should be performed at Hamamatsu.

The second measurement performed during the electrical characterisation of the full sensor is capacitance (CV) measurement. The sensors are also graded in terms of properties obtained from the CV measurement. In terms of the CV measurement, the results

from CERN before and after dicing frame removal are compared, as Hamamatsu do not measure the capacitance for all channels. To summarize these results, the grading tables are presented – Table 5.1 for CERN on dicing frame and Table 5.2 for CERN off dicing frame. In both cases all the tested sensors passed the CV grading. The results are consistent before and after dicing frame removal.

Table 5.1. CV grading results of the sensors measured at CERN on dicing frame

Sensor ID	V_{dep} corresponding to thickness	Maximum variation of V_{dep} across sensor of ±10%	Thickness variation <10 µm
N4791_1	Passed: 284.9 V < 370 V	Passed: 0.9%	Passed: 2.4 µm
N4791_2	Passed: 290.3 V < 370 V	Passed: 0.7%	Passed: 2.4 µm
N4790_1	Passed: 262.7 V < 370 V	Passed: 1.2%	Passed: 1.2 µm
N4790_2	Passed: 267.7 V < 370 V	Passed: 1.1%	Passed: 1.2 µm
N4790_3	Passed: 251.0 V < 370 V	Passed: 0.9%	Passed: 1.8 µm
N4791_6	Passed: 281.2 V < 370 V	Passed: 0.4%	Passed: 1.2 µm
N4791_7	Passed: 282.3 V < 370 V	Passed: 0.3%	Passed: 0.9 µm
N4791_8	Passed: 283.0 V < 370 V	Passed: 0.4%	Passed: 1.5 µm
N4791_9	Passed: 283.5 V < 370 V	Passed: 0.4%	Passed: 1.2 µm
N4790_13	Passed: 271.3 V < 370 V	Passed: 0.4%	Passed: 1.5 µm
N4790_14	Passed: 269.8 V < 370 V	Passed: 0.5%	Passed: 1.2 µm
N4790_15	Passed: 266.8 V < 370 V	Passed: 0.9%	Passed: 0.9 µm

Table 5.2. CV grading results of the sensors measured at CERN off dicing frame

Sensor ID	V_{dep} corresponding to thickness	Maximum variation of V_{dep} across sensor of ±10%	Thickness variation <10 µm
N4791_1	Passed: 285.8 V < 370 V	Passed: 0.9%	Passed: 2.7 µm
N4791_2	Passed: 293.1 V < 370 V	Passed: 0.7%	Passed: 2.4 µm
N4790_1	Passed: 261.8 V < 370 V	Passed: 1.2%	Passed: 1.2 µm
N4790_2	Passed: 267.6 V < 370 V	Passed: 1.2%	Passed: 1.2 µm
N4790_3	Passed: 251.2 V < 370 V	Passed: 0.9%	Passed: 1.8 µm
N4791_6	Passed: 282.7 V < 370 V	Passed: 0.5%	Passed: 1.2 µm
N4791_7	Passed: 281.6 V < 370 V	Passed: 0.3%	Passed: 1.2 µm
N4791_8	Passed: 283.0 V < 370 V	Passed: 0.4%	Passed: 1.5 µm
N4791_9	Passed: 283.0 V < 370 V	Passed: 0.4%	Passed: 1.5 µm
N4791_18	Passed: 281.1 V < 370 V	Passed: 0.2%	Passed: 1.2 µm
N4791_19	Passed: 281.2 V < 370 V	Passed: 0.3%	Passed: 1.8 µm
N4791_20	Passed: 281.5 V < 370 V	Passed: 0.4%	Passed: 2.4 µm
N4791_21	Passed: 276.2 V < 370 V	Passed: 0.5%	Passed: 1.8 µm
N4790_13	Passed: 271.0 V < 370 V	Passed: 0.4%	Passed: 1.5 µm
N4790_14	Passed: 269.6 V < 370 V	Passed: 0.7%	Passed: 1.2 µm
N4790_15	Passed: 267.4 V < 370 V	Passed: 1.0%	Passed: 0.9 µm

3.4. Integration of IV/CV and TCT+ setups

The last part of the thesis presents an additional project of integrating two separate setups into one common setup. Up to the time of the thesis project, the leakage current and capacitance measurement (IV/CV) and transient current technique (TCT) measurement (where charge collection study was performed) of the test structures were performed with

two different setups. By combining the two setups – creating a new IV/CV - TCT+ setup, faster turnaround can be achieved. Moreover, this approach can lead to better consistency between conducted measurements, as the measurement conditions (temperature, humidity) would be identical between the three measurements. The strategy for the considered setup was to add a new IV/CV part to the existing TCT+ setup. The two setups are connected through a high frequency switch, which is a crucial element in the project. The overview of the development strategy of the setup is shown in Figure 5.8.

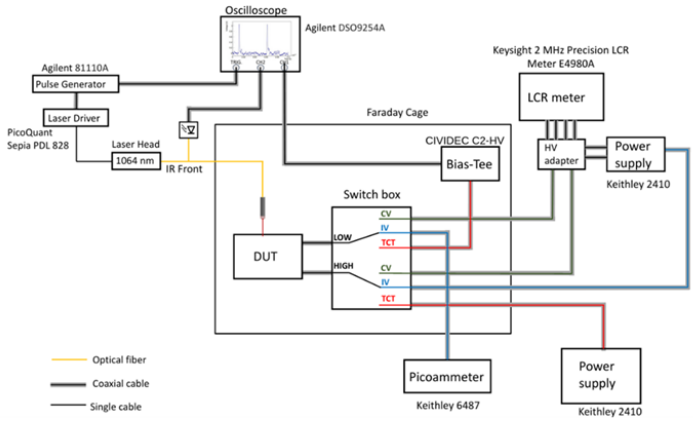
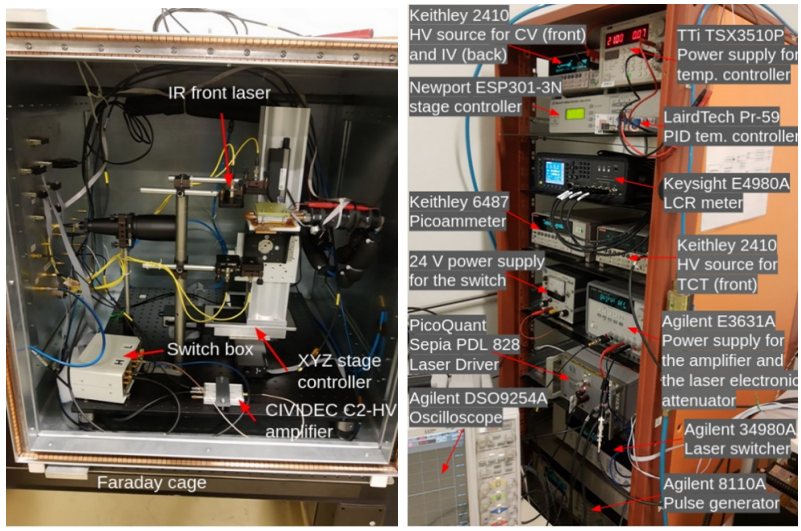


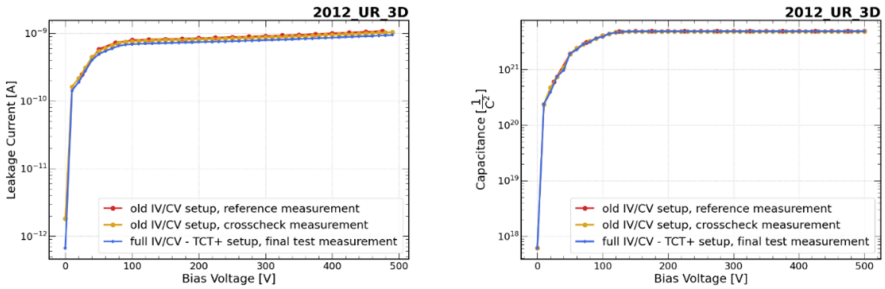
Figure 5.8. New setup diagram developed in thesis, combining the IV/CV and TCT+ setups

The assembling a new IV/CV setup and combining with the existing TCT+ setup was successful. Pictures of the full setup are presented in Figure 5.9. The results obtained with the new combined setup are consistent with reference measurements, as shown in Figures 5.10 and 5.11.



(a) view inside the Faraday cage **(b)** the instruments

Figure 5.9. Pictures of the full IV/CV - TCT+ setup



(a) IV **(b)** CV

Figure 5.10. Final test measurements

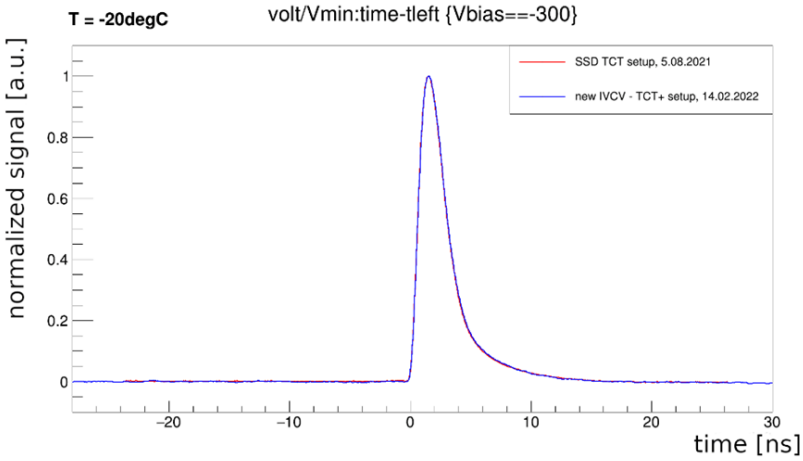


Figure 5.11. TCT final test measurement on the full IV/CV - TCT+ setup compared to the reference measurement performed on the original TCT setup

In summary, a new setup was created by assembling a new IV/CV setup and combining with the existing TCT+ setup. The results obtained with the new combined setup are consistent with reference measurements. The new setup still needs a few upgrades. For example, the compatibility of the IV and CV results at low temperatures has to be validated. Moreover, the software could be combined in one so that it would be easier to use. However, the setup is in a functional state, i.e. it is ready to be used for measurement campaigns of the silicon diodes and other devices of similar size.

4. Summary

The present thesis is devoted to silicon sensor characterisation and radiation hardness qualification for the CMS Endcap Calorimeter Upgrade. The project is part of the High-Luminosity LHC program. The current Endcap Calorimeter will be replaced by the High Granularity Calorimeter, based on i.a. silicon sensors as active material. The silicon sensors are produced in hexagonal geometry diced from 8-inch wafers. On the remaining

pieces of the wafers, test structures are produced. During the time of the thesis work, several activities were performed. They include the electrical characterisation and charge collection study of the test structures, electrical characterisation of full hexagonal sensors and integration of two initially separate setups for characterisation of the test structures.

The electrical characterisation of the test structures were performed on samples that were neutron irradiated at the two institutes: JSI and RINSC and including an isothermal annealing study. The leakage current is decreasing with the annealing time. The beneficial part of the annealing where the depletion voltage was decreasing and the reverse part where the depletion voltage was increasing were clearly seen. The comparison of the electrical properties showed a discrepancy of the fluence estimation between JSI and RINSC institutes. As the JSI is well established irradiation site, it was concluded the fluence estimation at RINSC was not yet satisfactory and could be corrected.

The charge collection study of the test structures was performed on samples of different thicknesses and fluences. The results of the measurements showed that the efficiency of the charge collection is decreasing with the fluence. Moreover, the thinner the diode, the higher the charge collection efficiency. Several steps of the annealing were also performed. It was observed, that charge collection efficiency does not change significantly with annealing time for any thickness.

The electrical characterisation of the full sensors was performed only on the unirradiated samples. They were measured in three configurations: at the producer (HPK), at CERN on dicing frame and at CERN after removing from the dicing frame. After each measurement, the sensor is graded and classified as "passed" or "failed". The grading results for the samples measured at HPK and at CERN on dicing frame were good. However, removing the sensors from the dicing frame significantly degraded their IV properties. The procedure of removing the sensors from the dicing frame at CERN should be improved or should be performed at HPK. All the tested sensors passed the CV grading, both on and off the dicing frame. The further step in the electrical characterisation of the full sensors is to study the properties of the irradiated sensors. Currently, the 8-inch sensors can be irradiated only at RINSC due to their size. That is why the accuracy of the fluence estimation at RINSC should be well established.

The last activity presented in this thesis was the integration of the IVCV and TCT+ setups for the test structures. The motivation of the project was to achieve faster turnaround and better consistency between conducted measurements. The integration was successful, as the results obtained with the new combined setup are consistent with reference measurements, performed on the already existing setups. The setup is in a functional state.

In the next phase of the CMS Endcap Calorimeter Upgrade, the prototypes that show the best performance will be chosen and the producer will start the mass production of the sensors. In total, CMS HGCAL will contain about 3000 8-inch silicon sensors. With this thesis work, the understanding of the silicon sensors prototypes and their radiation tolerance were improved. Lessons learned will be applied in the upcoming production rounds.

References

[1] Paulitsch P., *The Silicon Sensors for the High Granularity Calorimeter of CMS*. Nuclear Inst. and Methods in Physics Research **2020**, A, Vol. 978.

[2] D. Barney, *Overview slide of CE with main parameters*. Available online: `https://cms-docdb.cern.ch/cgi-bin/DocDB/ListBy?topicid=715` (access: 24-11-2021).

[3] Sze S.M., Lee M.K., *Semiconductor Devices: Physics and Technology*, John Wiley & Sons, INC., Singapur, 2012.

[4] Lutz G., *Semiconductor Radiation Detectors*, Springer Science & Business Media, Berlin, Germany, 2007.

[5] Hinger V., *Silicon Sensor Process Quality Control for the CMS Phase-2 Upgrade*. PhD Thesis, TU Vien, Austria, 2021.

[6] Moll M., *Radiation Damage in Silicon Particle Detectors*, PhD Thesis, Universität Hamburg, Germany, 1999.

[7] Quast T., *Qualification, performance validation and fast generative modelling of beam test calorimeter prototypes for the CMS Calorimeter Endcap Upgrade*, PhD Thesis, RWTH Aachen, Germany, 2020.

Ultracienkie membrany dichalkogenków metali przejściowych

mgr inż. Konrad Król
Elektronika

dr hab. inż. Jarosław Serafińczuk
Katedra Nanometrologii

W ramach pracy dyplomowej przeprowadzono badania mechaniczne i elektryczne ultracienkich membran na bazie nowego materiału – MoSSe z grupy Janus-dichalkogenków metali przejściowych (Janus-TMD). Janus-TMD są to materiały opisywane wzorem MXY, gdzie M jest pierwiastkiem metalu przejściowego, natomiast X i Y są chalkogenkami. Według wiedzy autora przeprowadzone w ramach pracy badania są pierwszymi eksperymentalnymi badaniami właściwości mechanicznych i elektrycznych tych materiałów. Uzyskane w ramach pracy wyniki badań, zostały odniesione do dobrze znanych materiałów – MoS_2 oraz $MoSe_2$. Monoatomowe membrany, będące przedmiotem badań wytworzono metodą mechanicznej eksfoliacji oraz umieszczono metodą suchego transferu na mikrownęce wykonanej w podłożu SiO_2/Si. Grubość badanych warstw potwierdzono wykorzystując spektroskopię Ramana. Wykorzystując mikroskopię AFM z sondą Kelvina zmierzono kontaktową różnicę potencjałów oraz obliczono pracę wyjścia. Parametr ten dla MoSSe był największy i wynosił 5,34 eV. Ponadto, badając ugięcie membrany pod wpływem przyłożonej siły oraz wyznaczając moduł Younga, stwierdzono, że parametr ten dla MoSSe ma wartość pomiędzy MoS_2 oraz $MoSe_2$, tzn. wynosi odpowiednio 66, 431,4 i 59,5 GPa.

1. Cel i założenia

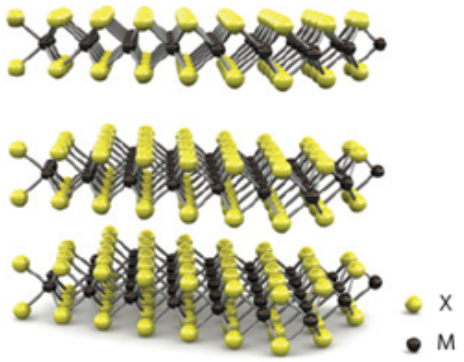
Celem niniejszej pracy jest wytworzenie oraz zbadanie podstawowych właściwości mechanicznych i elektrycznych ultracienkich membran wytworzonych z dichalkogenków metali przejściowych oraz Janus-dichalkogenków metali przejściowych. Główne założenia realizowanej pracy to:

- wytworzenie membran z zastosowaniem metod eksfoliacji mechanicznej oraz umieszczanie struktur nad specjalistycznym podłożem z mikrownękami przy użyciu metody suchego transferu,
- charakteryzacja wytworzonych membran z zastosowaniem metod mikroskopowych,
- porównanie podstawowych właściwości materiału izolowanego od wpływu podłoża.

2. Wstęp teoretyczny

W ostatnich latach następuje dynamiczny rozwój elektroniki i optoelektroniki. Wiąże się on z potrzebą wytwarzania nowych materiałów, które pozwolą rozszerzyć możliwości urządzeń elektronicznych i optoelektronicznych. Jednymi z intensywnie badanymi w ostatnich latach materiałami, są kryształy dwuwymiarowe (2D). Cechują się one wyjątkowymi, niespotykanymi dotąd właściwościami mechanicznymi, optycznymi, termicznymi oraz elektrycznymi. W ramach prowadzonych badań przedstawiono szereg materiałów dwuwymiarowych (2D) z grupy dichalkogenków metali przejściowych (TMD), które wykazują właściwości półprzewodnikowe. Ponadto prowadzone są badania nad nowymi materiałami należącymi do grupy Janus-TMD.

Dichalkogenki metali przejściowych (TMD) są to związki o stechiometrii MX2. TMD zbudowane są z warstwy atomu metalu przejściowego (M) umieszczonej między warstwami atomu chalkogenku (X), co przedstawiono na rysunku 6.1. Monowarstwa składa się z płaszczyzny atomowej X-M-X, w której atomy są połączone wewnętrznie przez silne wiązania kowalencyjne. Objętościowy kryształ TMD składa się z wielu monowarstw, które są połączone ze sobą słabymi wiązaniami van der Waalsa. Do najpopularniejszych TMD zaliczamy MoS_2, WS_2, $MoSe_2$ oraz WSe_2.



Rysunek 6.1. Trójwymiarowy schemat struktury MX2. Atomy metalu przejściowego (M) zaznaczono kolorem czarnym, natomiast atomy chalkogenku (X) kolorem żółtym [1]

Kolejną rodzinę materiałów 2D stanowią Janus-dichalkogenki metali przejściowych (Janus-TMD, ang. Janus Transition Metal Dichalcogenides), która opisywana jest wzorem MXY, gdzie M jest pierwiastkiem metalu przejściowego, natomiast X oraz Y są różnymi chalkogenkami. Do przykładowych materiałów Janus-TMD zaliczamy MoSSe, WSSe, MoSeTe oraz WSTe.

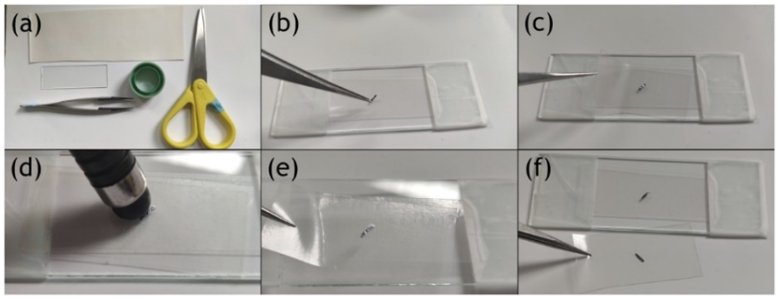
Generalnie półprzewodniki TMD [2] oraz Janus-TMD [3] wykazują zależność charakteru przerwy wzbronionej od grubości struktury. Wraz ze zmniejszaniem się liczby warstw zmienia się przerwa energetyczna ze skośnej w krysztale objętościowym na prostą w pojedynczej warstwie. TMD oraz Janus-TMD pozwalają zejść wymiarowo do pojedynczych warstw atomowych, tj. poniżej 1 nm. Otwiera to wiele możliwości ich zastosowania w urządzeniach elektronicznych i optoelektronicznych, np. w tranzystorach [4], gazo-, chemo- i bioczujnikach [5], fotodetektorach [6] oraz urządzeniach piezoelektrycznych [5].

3. Proces wytwarzania membran oraz metodyka badań

Celem prowadzonych badań było wytworzenie membran rozpiętych nad mikrownęką o średnicy 2,4 µm, wytworzoną w podłożu Si/SiO_2 pokrytym 150 nm warstwą złota oraz zbadanie ich właściwości elektrycznych i mechanicznych. W badaniach realizowanych w ramach niniejszej pracy użyto trzech materiałów: Janus-TMD – MoSSe oraz dwa TMD – MoS_2 i $MoSe_2$. Grubość monowarstwy MoS_2 wynosi 6,2 A [7], $MoSe_2$ – 7,0 A [8], natomiast MoSSe – 3,35 A [9].

3.1. Mechaniczna eksfoliacja

W pierwszym etapie badań należało wytworzyć monowarstwy, a następnie przenieść je na podłoże. Proces wytwarzania monowarstw składa się z etapów przedstawionych na rysunku 6.2. W pierwszym etapie przygotowano dwa prostokątne kawałki taśmy, z których jeden przymocowano do szkiełka mikroskopowego stroną klejącą do góry tak, aby podczas kolejnych etapów nie przemieszczał się. Następnie kawałek wybranego kryształu umieszczono na taśmie przymocowanej do szkiełka (rysunek 6.2(b)) i przykryto drugą taśmą (rysunek 6.2(c)). Ze względu na uwięzione pęcherzyki powietrza pomiędzy taśmami lekko dociskano taśmę do materiału (rysunek 6.2(d)). Następnie górną taśmę odrywano (rysunek 6.2(e)). Proces przyklejania i odrywania taśmy górnej od dolnej powtarzano kilkukrotnie, co powodowało utworzenie wielu płatków o różnej grubości (rysunek 6.2(f)). Po ostatniej eksfoliacji na uzyskane warstwy na taśmie nakładano kawałek polidimetylosiloksanu (PDMS), po czym lekko dociskano i po chwili odklejano. W kolejnym kroku szukano monowarstw, które zostały przeniesione z taśmy samoprzylepnej na PDMS, za pomocą mikroskopu optycznego z użyciem światła przechodzącego.

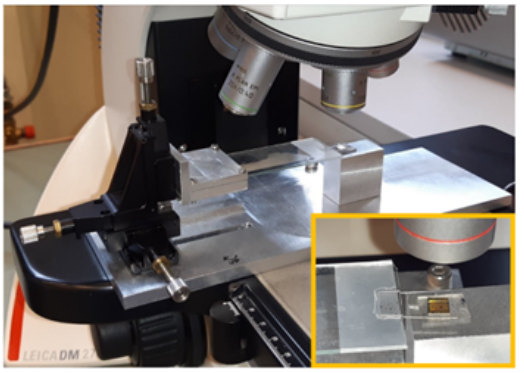


Rysunek 6.2. Etapy eksfoliacji mechanicznej przeprowadzanej w niniejszej pracy: (a) potrzebne materiały, (b) umieszczenie fragmentu kryształu na przymocowanej taśmie samoprzylepnej, (c) umieszczenie drugiego kawałka taśmy na materiale, (d) usuwanie pęcherzyków powietrza, (e) oderwanie górnej taśmy, (f) rozwarstwiony kryształ [10]

3.2. Suchy transfer

Znalezioną pojedynczą warstwę przenoszono metodą suchego transferu [11] na wcześniej oczyszczone plazmowo podłoże. Stanowisko do suchego transferu przedstawione na rysunku 3 składa się z manipulatora trójosiowego, szkiełka mikroskopowego oraz mikroskopu optycznego. PDMS, na którym umieszczona jest monowarstwa, przymocowany jest

częściowo do szkiełka mikroskopowego. Następnie szkło zostało przymocowane do manipulatora trójosiowego płatkiem skierowanym do podłoża w sposób przedstawiony na rysunku 6.3. W celu przeprowadzenia transferu PDMS jest opuszczany tak, aby monowarstwa znajdowała się nad wybraną mikrownęką w podłożu. W kolejnym kroku rozpoczyna się wolne podnoszenie PDMS, ponieważ zbyt szybkie odrywanie struktury spowoduje jej pozostanie na PDMS lub uszkodzenie.



Rysunek 6.3. Stanowisko do suchego transferu z poprawnie przymocowanym PDMS do szkiełka mikroskopowego i umieszczonym nad podłożem

3.3. Spektroskopia Ramana

Spektroskopia Ramana jest skutecznym optycznym narzędziem do określania dokładnej liczby uzyskanych warstw z eksfoliowanego kryształu. Do wzbudzenia użyto lasera półprzewodnikowego o długości fali 532 nm i mocy 300 µW dla TMD oraz 733 µW dla Janus-TMD. Badanie przeprowadzono dla płatków znajdujących się bezpośrednio w kontakcie z podłożem, jak również rozpiętych nad mikrownęką w temperaturze pokojowej. Badanie wykonano w Katedrze Inżynierii Materiałów Półprzewodnikowych.

3.4. Mikroskopia sił z sondą Kelvina

W celu określenia kontaktowej różnicy potencjału (CPD) i wyznaczenia pracy wyjścia (φ) z materiału stosuje się mikroskopię sił atomowych z sondą Kelvina (KPFM). Na podstawie wyników CPD obliczono pracę wyjścia. Aby oszacować φ badanych materiałów należy przeprowadzić dodatkowy pomiar KPFM w celu wyznaczenia pracy wyjścia sondy. W tym celu użyto wysoko zorientowanego pirolitycznie grafitu (HOPG, ang. Highly Oriented Pyrolytic Graphite), którego $\varphi = 4{,}8$ eV. Na podstawie zmierzonego CPD oraz φ sondy oszacowano φ próbki za pomocą wzoru (6.1).

$$\varphi_{probki} = \varphi_{sondy} - e \cdot U_{CPD} \tag{6.1}$$

gdzie:

φ_{probki} oraz φ_{sondy} oznaczają odpowiednio wyjścia próbki i sondy mikroskopu AFM,
e oznacza ładunek elementarny,
U_{CPD} to zmierzona kontaktowa różnica potencjałów.

3.5. Spektroskopia sił – badania mechaniczne

W celu wyznaczenia modułu Younga (E) przeprowadzono pomiary spektroskopii siła–odległość (F–z). Pomiar ten polega na ustawieniu ostrza pomiarowego w punkcie środkowym membrany, a następnie dociskaniu jej. Jednocześnie rejestrowana jest siła docisku ostrza oraz ugięcie materiału. Na podstawie wyników pomiarów oraz wzoru (6.2) wyznaczono moduł Younga.

$$F = \left(\frac{4 \cdot \pi \cdot h^3}{3 \cdot (1 - v^2) \cdot a^2} \cdot E + \pi \cdot S \right) \cdot z \ + \frac{h \cdot E}{q^3 \cdot a^2} \cdot z^3 \tag{6.2}$$

gdzie:
F – przyłożona siła,
z – odkształcenie membrany,
h – grubość membrany,
a – promień membrany,
E – moduł Younga,
v – współczynnik Poissona,
S – naciąg membrany,
q – parametr materiałowy związany ze współczynnikiem Poissona charakterystyczny dla odkształconej membrany wyznaczany ze wzoru (6.3).

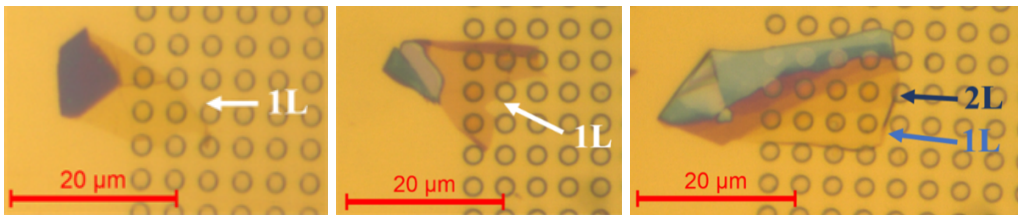
$$q = 1,05 - 0,15 \cdot v - 0,16 \cdot v^2 \tag{6.3}$$

4. Wyniki pomiarów

Wytworzone struktury poddano pomiarom optycznym, elektrycznym oraz mechanicznym.

4.1. Badanie membran mikroskopem optycznym

Wstępną grubość wytworzonych struktur oceniono na postawie barwy – im cieńsza warstwa, tym jest jaśniejsza oraz bardziej przeźroczysta, co pokazano na zdjęciach wykonanych za pomocą mikroskopu optycznego (rysunek 6.4). W tym miejscu należy zwrócić uwagę, że mimo podobnej grubości, każdy materiał ma inną barwę.

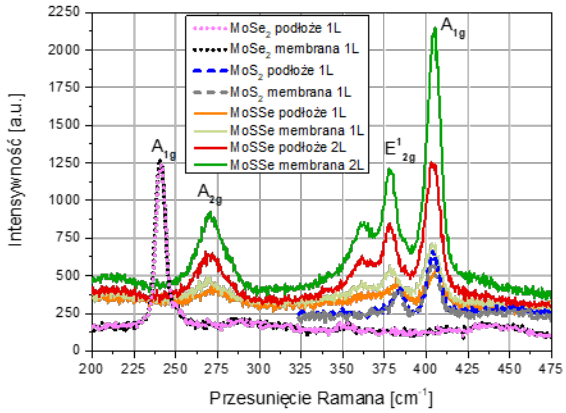


(a) MoS_2 **(b)** $MoSe_2$ **(c)** MoSSe

Rysunek 6.4. Zdjęcia membran wykonanych z różnych materiałów z mikroskopu optycznego

4.2. Badanie warstw za pomocą spektroskopii Ramana

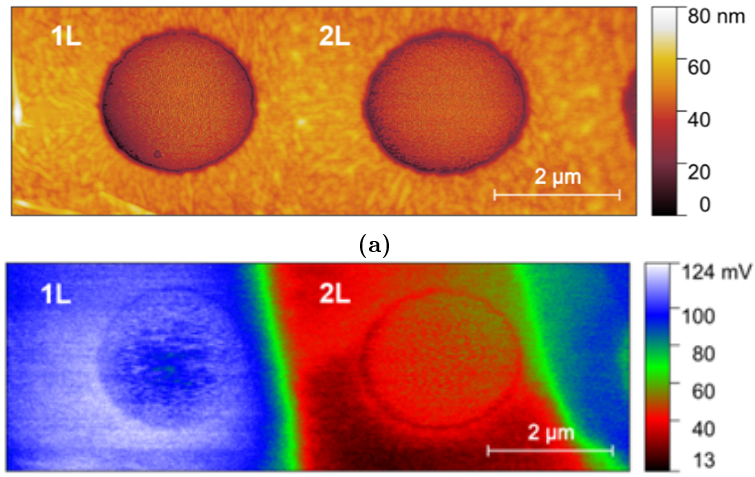
Wyniki pomiarów przeprowadzonych za pomocą spektroskopii Ramana dla Janus-TMD oraz TMD przedstawiono na rysunku 6.5. W widmie MoS_2 występują dwa piki przy 383,3 cm^{-1} oraz 403,8 cm^{-1}, które odpowiadają modowi E12g oraz A1g. Wyniki te są zgodne z wynikami znanymi z literatury [12]. Potwierdza to, że uzyskana membrana jest monowarstwą. W przypadku $MoSe_2$ mod A1g wynosi 240,9 cm^{-1}. Podobnie jak dla MoS_2, wartość tego piku jest zgodna z danymi literaturowymi [13]. Dla MoSSe występują trzy mody – A1g, E12g oraz A2g, które przyjmują dla monowarstwy odpowiednio częstotliwości 404,3 cm^{-1}, 379,9 cm^{-1} oraz 270,5 cm^{-1}, natomiast dla dwuwarstwy – 404,2 cm^{-1}, 377,8 cm^{-1} oraz 270,1 cm^{-1}. Porównując ze sobą wyniki badań użytych materiałów, można zauważyć, że widmo MoSSe jest zbliżone do połączonego widma MoS_2 oraz $MoSe_2$, zwłaszcza w przypadku modu A1g oraz E12g materiału MoS_2 (rysunek 6.5). Na intensywność i częstotliwość modów wpływa kontakt materiału z podłożem oraz liczba warstw, co jest widoczne w przypadku MoSSe (rysunek 6.5). Zwiększa się intensywność, natomiast maleje częstotliwość modów. Jedną z przyczyn powstałych różnic w wynikach mogą być naprężenia struktur powstające nad mikrownęką.



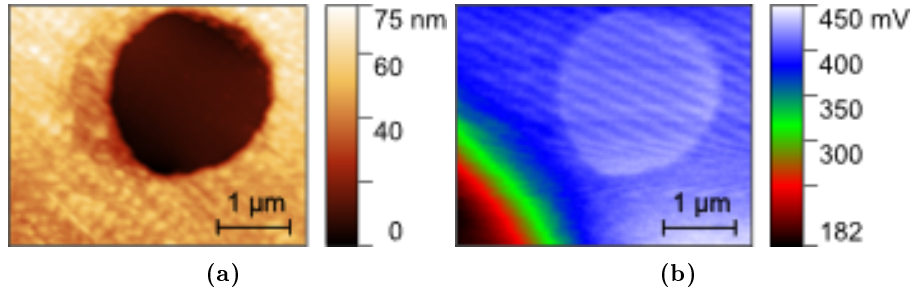
Rysunek 6.5. Spektroskopia Ramana dla monowarstwy (1L) MoS_2, $MoSe_2$ oraz MoSSe i dwuwarstwy (2L) MoSSe

4.3. Badania właściwości elektrycznych

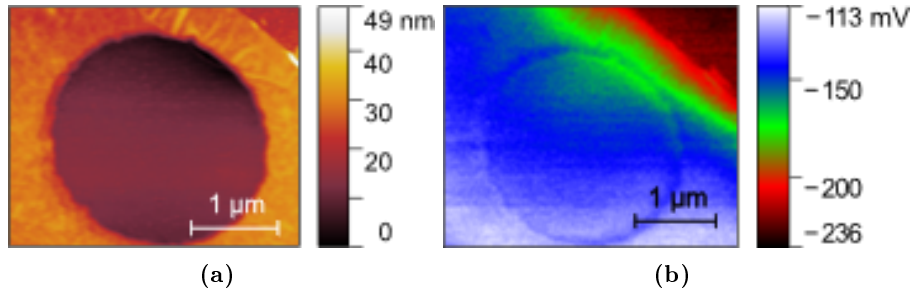
Zmierzoną mapę topografii dla MoSSe, MoS_2 oraz $MoSe_2$ przedstawiono odpowiednio na rysunkach 6.6a, 6.7a oraz 6.8a, natomiast mapę kontaktowej różnicy potencjałów odpowiednio na rysunkach 6.6b, 6.7b oraz 6.8b. Dla każdego materiału uzyskano jednorodne membrany bez większych odkształceń. Ponadto różnica między monowarstwą a dwuwarstwą jest tak mała, że na mapie topografii jest prawie niezauważalna, co można zaobserwować w przypadku MoSSe na rysunku 6.6a. Z tego powodu za pomocą mikroskopii sił atomowych (AFM) nie można jednoznacznie określić grubości uzyskanego płatka. Korzystając z wyników CPD, można zauważyć różną grubość uzyskanej struktury, zwłaszcza między monowarstwą a dwuwarstwą, np. dla MoSSe. CPD dla 1L jest większe od CPD dla 2L. Ponadto podłoże wykazuje niższą CPD od płatków, co można zauważyć na rysunkach 6.7b oraz 6.8b.



Rysunek 6.6. Mapy (a) topografii oraz (b) kontaktowej różnicy potencjałów jedno- (1L) oraz dwuwarstwowej (2L) membrany wykonanej z MoSSe



Rysunek 6.7. Mapy (a) topografii oraz (b) kontaktowej różnicy potencjałów membrany jednowarstwowej wykonanej z MoS_2



Rysunek 6.8. Mapy (a) topografii oraz (b) kontaktowej różnicy potencjałów membrany jednowarstwowej wykonanej z $MoSe_2$

W celu dokładnego określenia wpływu podłoża na właściwości elektryczne, wyznaczono CPD oraz pracę wyjścia dla membrany oraz struktury znajdującej się w bezpośrednim kontakcie z podłożem. Wyniki przedstawiono w tabeli 6.1. Powstające naprężenia na płatek znajdujący się nad mikrownęką wpływają na CPD oraz φ, przy czym różnica ta jest

mniejsza niż 0,02 w obu przypadkach. Nie jest to uzależnione od kontaktu materiału z podłożem, ponieważ φ dla MoSSe 1L znajdującego się nad mikrownęką (5,34 eV) uzyskano większą niż mającego kontakt z podłożem (5,32 eV), natomiast dla MoS_2 – odwrotnie (4,83 eV oraz 4,85 eV). Podobna sytuacja występuje w przypadku CPD.

Tabela 6.1. Kontaktowa różnica potencjałów (CPD) oraz praca wyjścia (φ) dla struktur wykonanych z różnych materiałów oraz różnej grubości (monowarstwa oraz dwuwarstwa)

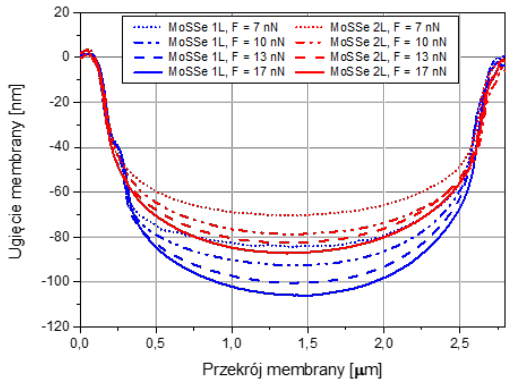
Materiał	Liczba warstw	Miejsce pomiaru	CPD [mV]	φ [eV]
MoS_2	1	membrana	410	4,83
MoS_2	1	podłoże	387	4,85
$MoSe_2$	1	membrana	−144	5,21
$MoSe_2$	1	podłoże	−128	5,20
MoSSe	1	membrana	88	5,34
MoSSe	1	podłoże	106	5,32
MoSSe	1	symulacja	–	5,24 [14]
MoSSe	2	membrana	46	5,02
MoSSe	2	podłoże	35	5,04

Największą CPD uzyskano dla MoS_2 (410 mV), a najmniejszą dla MoSSe 2L (46 mV), przy czym dla $MoSe_2$ jest ujemna (-144 mV). Pracę wyjścia najmniejszą uzyskano dla membrany MoS_2 (4,83 eV), podczas gdy największą dla membrany MoSSe 1L (5,34 eV). Ponadto wyznaczona φ monowarstwy MoSSe jest o 0,1 eV mniejsza w porównaniu z teoretycznymi (5,24 eV) [14]. W przypadku MoSSe CPD oraz φ dla monowarstwy są większe od φ dla dwuwarstwy. Dla membran CPD oraz φ jest większa odpowiednio o 42 mV oraz 0,32 eV, natomiast w przypadku kontaktu struktur z podłożem – 71 mV oraz 0,28 eV. Na tej podstawie można stwierdzić, że wraz ze wzrostem liczby warstw maleje CPD oraz φ.

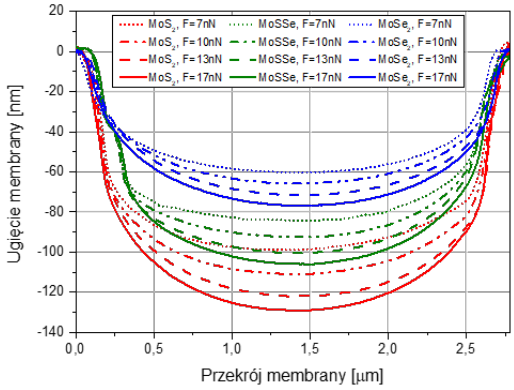
4.4. Badania właściwości mechanicznych

Dla każdego z badanych materiałów zmierzono ugięcie membrany w funkcji przyłożonej siły za pomocą AFM przy przyłożonych siłach równych 7, 10, 13 i 17 nN. Grubość membrany wpływa znacząco na ugięcie membrany. Im cieńsza membrana, tym bardziej ulegają odkształceniom, co przedstawiono dla MoSSe na rysunku 6.9. W przypadku jednowarstwowej struktury najmniejsze odkształcenie wykazuje $MoSe_2$, natomiast największe MoS_2, co przedstawiono na rysunku 6.10. Wielkość ugięcia membran wykonanych z MoSSe znajduje się pomiędzy $MoSe_2$ a MoS_2.

Jedną z właściwości mechanicznych, którą wyznaczono był moduł Younga. Zarówno wynik obliczeń, jak i dodatkowe dane potrzebne do jego wyznaczenia, w tym współczynnik Poissona i grubość membran, umieszczono w tabeli 6.2. Największy moduł Younga wyznaczono dla MoS_2 i wynosi on 431,4 GPa. W porównaniu z grafenem, najbardziej wytrzymałym mechanicznie materiałem, którego E wynosi 501 GPa [15], jest to wartość niższa o 70 GPa. Najmniejszy moduł Younga uzyskano dla $MoSe_2$, przy czym jest on nieznacznie niższy, o 6 GPa, od monowarstwy MoSSe. Ponadto dla dwuwarstwy MoSSe uzyskano prawie trzykrotnie większy E od wyniku uzyskanego dla pojedynczej warstwy. W porównaniu z symulacjami [16] moduł Younga dla monowarstwy MoSSe był o połowę mniejszy.



Rysunek 6.9. Ugięcie membrany jedno- (1L) oraz dwuwarstwowej (2L) wykonanej z MoSSe w zależności od przyłożonej siły



Rysunek 6.10. Ugięcie membrany jednowarstwowej wykonanej z MoSSe, MoS_2 oraz $MoSe_2$ w zależności od przyłożonej siły

Tabela 6.2. Wybrane właściwości wytworzonych membran, w tym wyznaczony moduł Younga

Membrana	Liczba warstw	h [Å]	ν	a [µm]	E [GPa]
MoS_2	1	6,2	0,24	1,2	431,4
$MoSe_2$	1	7,0	0,23	1,2	59,5
MoSSe	1	3,35	0,23	1,2	66,0
MoSSe	1	–	–	–	124,6
MoSSe	2	6,7	0,23	1,2	179,8

5. Podsumowanie

W ramach niniejszej pracy przeprowadzono badania elektryczne i mechaniczne z wykorzystaniem metod spektroskopii AFM monowarstwy i dwuwarstwy z materiałów MoSSe, MoS_2 oraz $MoSe_2$. Grubość wytworzonych struktur potwierdzono za pomocą spektroskopii Ramana. Na podstawie uzyskanych wyników stwierdzono, że widmo MoSSe jest

zbliżone do połączonego widma MoS_2 oraz $MoSe_2$. Stwierdzono, że naprężenia powstające w płatkach umieszczonych nad mikrownęką wpływają zarówno na wyniki optyczne, jak i elektryczne. Zauważono, że na częstotliwość widocznych modów wpływa kontakt materiału z podłożem. Ponadto większą intensywność uzyskuje się dla struktur umieszczonych nad mikrownęką. W porównaniu ze strukturami znajdującymi się w kontakcie z podłożem zauważono, że nie ma jednoznacznej zależności określającej wpływ kontaktowej różnicy potencjałów oraz pracy wyjścia. Zarówno CPD, jak i φ różni się o nie więcej, niż 0,02. Stwierdzono, że MoSSe wykazuje najmniejszą CPD (88 mV) oraz największą pracę wyjścia (5,34 eV). Ponadto φ MoSSe jest mniejsza o 0,1 eV od teoretycznej. Oprócz tego zauważono, że CPD oraz φ maleje wraz ze wzrostem grubości materiału. Po przeprowadzeniu badań mechanicznych stwierdzono, że największe ugięcie membrany uzyskano dla MoS_2, natomiast najmniejsze dla $MoSe_2$. MoSSe przyjmuje wartości pośrednie. Ponadto grubość struktury istotnie wpływa na jej odkształcenie. Monowarstwy ulegają większym ugięciom od dwuwarstw. W przypadku monowarstwy MoSSe moduł Younga jest o połowę mniejszy od teoretycznego i zbliżony do $MoSe_2$. Ponadto w porównaniu z monowarstwą, dla dwuwarstwy zwiększa się on prawie trzykrotnie.

Wyniki przedstawione w niniejszej pracy dyplomowej wskazują, że cel został osiągnięty. Uzyskano dobrej jakości ultracienkie membrany, które posłużyły do dalszych badań właściwości elektrycznych i optycznych, np. piezoelektrycznych. Uzyskane w ramach pracy wyniki są podstawą przygotowywanej publikacji naukowej oraz zostały zaprezentowane w postaci referatu „KPFM and mechanical investigations of ultrathin membranes from MoS_2 and MoSSe" na konferencji 18th International Conference on Diffusion in Solids and Liquids, która odbyła się w dniach 27.06–01.07.2022 r. we Florencji we Włoszech.

Bibliografia

[1] Wang Q.H., Kalantar-Zadeh K., Kis A., Coleman A., Strano M., *Electronics and optoelectronics of two-dimensional transition metal dichalcogenides*, Nature Nanotechnology **2012**, Vol. 7, s. 699–712.

[2] Ravindra N.M., Tang W., Rassay S., *Transition Metal Dichalcogenides Properties and Applications*, Springer Nature Switzerland AG **2019**, s. 333–396.

[3] Wang J., Shu H., Zhao T., Liang P., Wang N., Cao D., Che X., *Intriguing Electronic and Optical Properties of Two-Dimensional Janus Transition Metal Dichalcogenides*, Physical Chemistry Chemical Physics **2018**, Vol. 20, s. 18571–18578.

[4] Liu Y., Zhang S., He J., Wang Z.M., Liu Z., *Recent Progress in the Fabrication, Properties, and Devices of Heterostructures Based on 2D Materials*, Nano-Micro Letters, **2019**, 11(13).

[5] Choi W., Choudhary N., Han G.H., Park J., Akinwande D., Lee Y.H., *Recent development of two-dimensional transition metal dichalcogenides and their applications*, Materials Today **2017**, s. 116–130.

[6] Lopez-Sanchez O., Lembke D., Kayci M., Radenovic A., Kis A., *Ultrasensitive photodetectors based on monolayer MoS_2*, Nature Nanotechnology **2013**, 8, s. 497–501.

[7] Li X., Zhu H., *Two-dimensional MoS_2: Properties, preparation, and applications*, Journal of Materiomics **2015**, 1(1), s. 33–44.

[8] Borodin B.R., Dunaevskiy M.S., Benimetskiy F.A., Lebedev S.P., Lebedev A.A., Alekseev P.A., *Kelvin probe microscopy of $MoSe_2$ monolayers on graphene*, Journal of Physics: Conference Series **2018**, 1124(8).

[9] Jin C., Tamg X., Tan X., Smith S.C., Dai Y., Kou L., *Janus MoSSe monolayer: superior and strain-sensitive gas sensing material*, Journal of Materials Chemistry A **2019**, 7, s. 1099–1106.

[10] Tamulewicz-Szwajkowska A., Zelewski S.J., Serafińczuk J., Kudrawiec R., *Geometric progress in the thickness of exfoliated van der Waals crystals on the example of MoS_2*, AIP Advances **2022**, 025328(12).

[11] Castellanos-Gomez A., Buscema M., Molenaar R., Singh V., Janssen L., Zant H.S.J., Steele G.A., *Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping*, 2D Materials **2014**, 011002(1).

[12] Li X., Shi Y., Li S., Han W., Zhou C., Zhao X., Liang B., *Layer-number dependent reflection spectra of MoS_2 flakes on SiO_2/Si substrate*, Optical Materials Express **2018**, 8(10), s. 3082–3091.

[13] Xia J., Huang X., Liu L.Z., Wang M., Wang L., Huang B, Zhu D.D., Li J.J., Gu C.Z., Meng X.M., *CVD synthesis of large-area, highly crystalline $MoSe_2$ atomic layers on diverse substrates and application to photodetectors*, Nanoscale **2014**, 6, s. 8949–8955.

[14] Zhang X., Song Y., Zhang F., Fan Q., Chen S., Jin Y., Gao S., Xiao Y., Mwankemwa N., *The electronic properties of hydrogenated Janus MoSSe monolayer: a first principles investigation*, Materials Research Expres **2019**, 6(10).

[15] Gajewski K., *Grafenowe przetworniki nanoelektromechaniczne*, Politechnika Wrocławska, Wrocław 2019.

[16] Yang F., Shang J., Kou L., Li C., Deng Z., *Mechanical Behaviors in Janus Transition-Metal Dichalcogenides: A Molecular Dynamics Simulation*, Nanomaterials **2022**, 12(1910).

Wydział Elektryczny

Eksperymentalny robot ramieniowy sterowany za pomocą mikrokontrolera ARM

inż. Szymon Wąchała
Automatyka i Robotyka

dr inż. Krzysztof Dyrcz
Katedra Maszyn, Napędów i Pomiarów Elektrycznych

W niniejszej pracy dyplomowej skupiono się na temacie budowy uproszczonego robota przemysłowego. W poszczególnych rozdziałach została przedstawiona problematyka związana z projektowaniem oraz budową konstrukcji, matematycznym opisie kinematyki robota, doborem elementów układu napędowego oraz stworzeniem środowiska do interakcji z robotem. Ze względu na budżetowe podejście poczynione zostały pewne uproszczenia oraz kompromisy, zapewniając jednak podstawową funkcjonalność w postaci trybu ręcznego – sterowanie za pomocą kontrolera do gier oraz możliwości sterowania robota przez programowanie sekwencji ruchów w dedykowanym do tego języku programowania. Efektem końcowym jest robot o 4 stopniach swobody, dający możliwość liniowego sterowania ruchem chwytaka. Opisane zostały również problemy które pojawiły się podczas próby syntezy regulatorów położenia, czyli problemy w zapewnieniu płynnego ruchu robota.

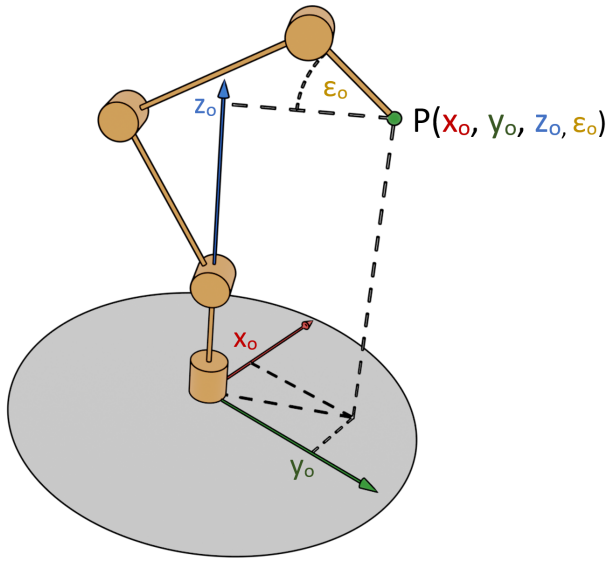
1. Cel i założenia

Celem pracy dyplomowej jest opracowanie i wykonanie eksperymentalnego robota ramieniowego, wykorzystującego w układzie sterowania mikrokontroler z rodziny ARM, oraz realizacja badań na wykonanym modelu. Główne założenia projektowe realizowanej pracy stanowi:

- zapoznanie się z zagadnieniem,
- opracowanie koncepcji robota ramieniowego,
- dobór elementów robota,
- wykonanie prototypu robota i napisanie oprogramowania sterującego,
- wykonanie badań laboratoryjnych,
- redakcja pracy.

2. Wstęp teoretyczny

Istotną informacją dla całej pracy dyplomowej jest definicja robota, a szczególnie robota przemysłowego. Robotem przemysłowym nazywa się wyspecjalizowaną maszynę zdolną do wykonywania konkretnych, często złożonych ruchów dzięki wielu stopniom swobody. Elementem wykonawczym robota jest efektor. Pojęciem tym określa się chwytaki oraz dodatkowe narzędzia, jak np. laser, uchwyt spawalniczy czy pistolet do malowania. Robot wykonuje manewry według odpowiednio zaprogramowanych działań [1]. Podstawowymi elementami robota przemysłowego są: zespół napędowy – odpowiednie układy napędowe wraz z elementami przeniesienia napędu, zespół czujników – układy odpowiedzialne za poprawne poruszanie się oraz reagowanie na inne elementy w przestrzeni otaczającej robota, efektor – układ wykonawczy robota oraz układ sterujący zarządzający całym robotem. Niniejsza praca opisuje robota o określonej strukturze kinematycznej – wykonany robot składa się z 4 przegubów obrotowych.



Rysunek 7.1. Schematyczne przedstawienie struktury kinematycznej robota

W tego typu konstrukcji aby jednoznacznie powiązać punkt stanowiący efektor robota wraz z ułożeniem poszczególnych przegubów, należy określić 4 parametry, a są nimi: współrzędne w układzie kartezjańskim x, y, z punktu wodzącego robota oraz dodatkowo nachylenie ostatniego (względem podłoża) elementu robota względem płaszczyzny podstawy.

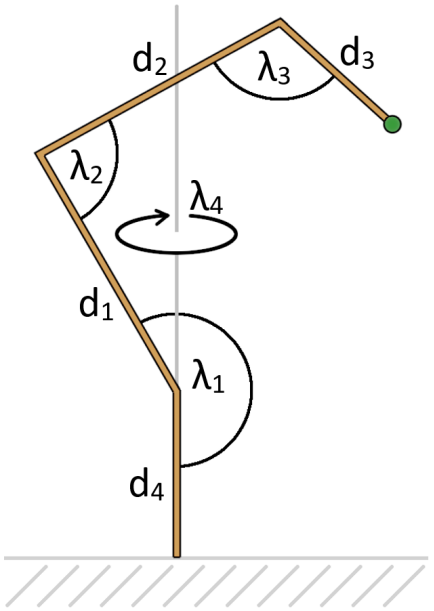
3. Projekt i wykonanie

Złożoność projektu wymagała podzielenia go na pewne etapy, które obejmowały działania powiązane z konkretnymi dziedzinami inżynieryjnymi. Pierwszym etapem była realizacja modelu matematycznego robota w celu określenia możliwości związanych z budżetem

– określenie wymiarów oraz pozostałych parametrów w celu dobrania odpowiednich napędów (które były najdroższym elementem), co stanowiło drugi etap. Kolejną częścią była realizacja mechanicznych aspektów – od projektu konstrukcji przez wykonanie poszczególnych części aż do złożenia wszystkiego w całość. Czwartym etapem była realizacja części elektrycznej, na którą składało się wykonanie szafy sterowniczej wraz z dedykowanym układem mikrokontrolerowym opartym na układzie Teensy 4.1. Ostatnim etapem realizacji budowy było wykonanie środowiska pozwalającego na interakcję z robotem. Po ukończeniu powyższych etapów miały miejsce badania parametrów ruchu, na jakie pozwalał robot.

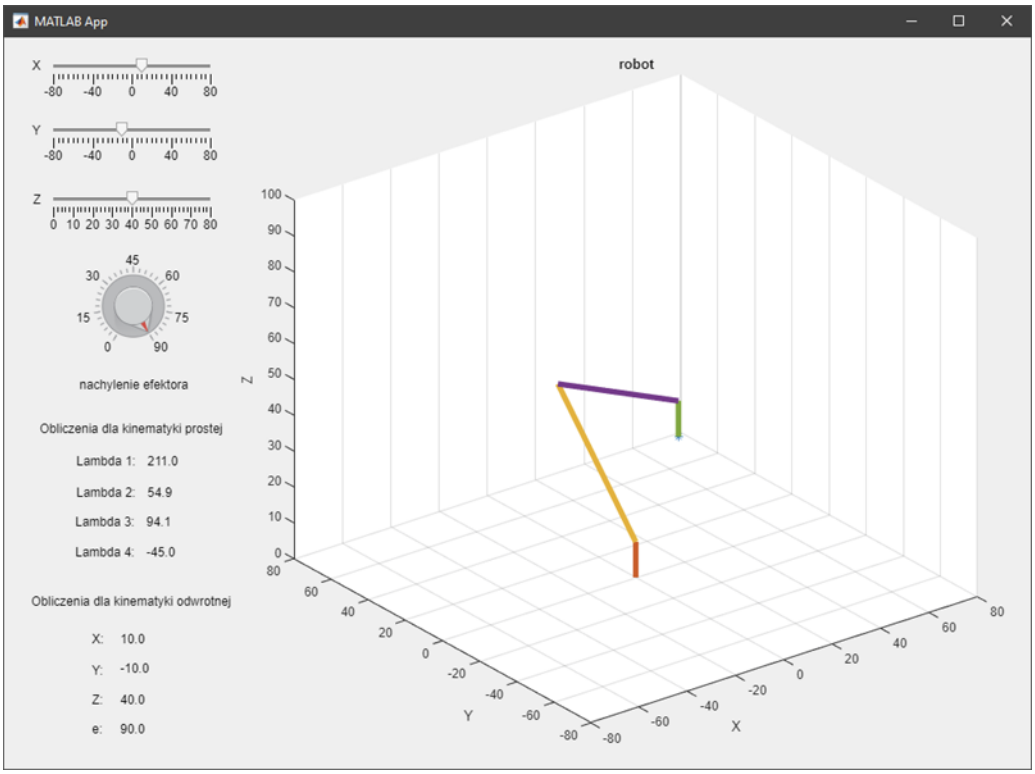
3.1. Opis kinematyki

Wyróżniamy dwa rodzaje kinematyki: prosta – pozwalająca określić położenie efektora na podstawie kątów w poszczególnych przegubach, oraz kinematyka odwrotna, która pozwala określić kąty poszczególnych przegubów dla określonej pozycji efektora.



Rysunek 7.2. Rysunek pomocniczy dla analizy kinematyki robota

Powyższy rysunek ilustruje, które parametry są brane pod uwagę podczas realizacji obliczeń kinematyki. W perspektywie układu sterującego robotem oraz użytkownika, który chce, aby efektor robota znalazł się w punkcie P0, istotna jest kinematyka odwrotna. W celu weryfikacji poprawności wykonanej analizy stworzony został program pomocniczy ilustrujący za pomocą wektorów poszczególne fragmenty ramienia robota. Program ten obliczał oraz przedstawiał pozycję robota na podstawie zadanych parametrów – współrzędnych punktu wodzącego robota oraz na tej podstawie realizował algorytm kinematyki prostej.



Rysunek 7.3. Aplikacja pozwalająca na weryfikację obliczeń kinematyki robota

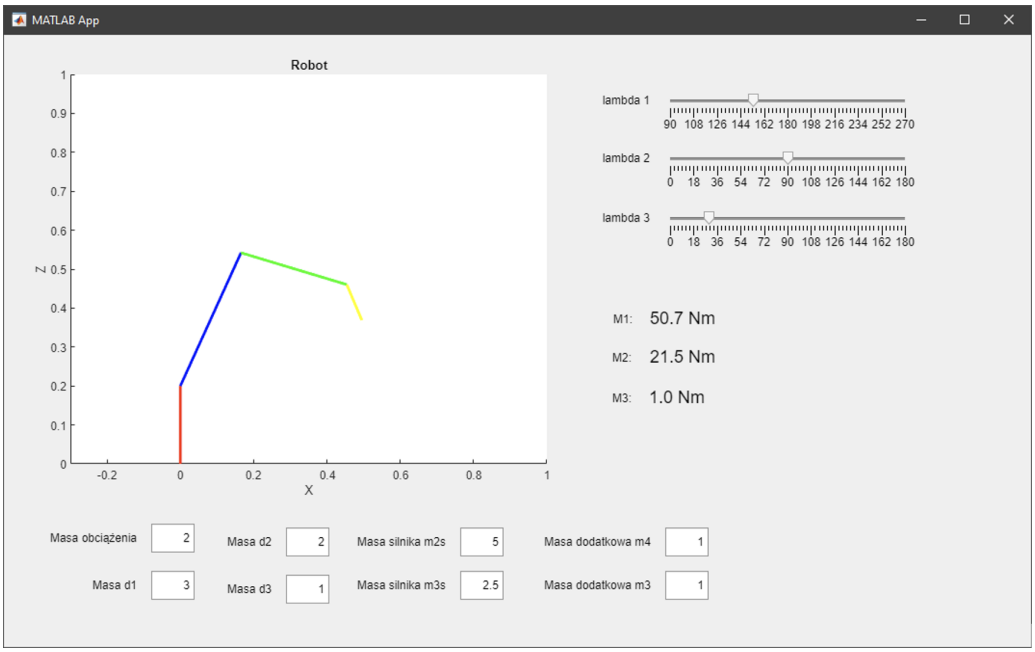
3.2. Dobór układu napędowego

Zagadnienie doboru układu napędowego jest bardzo istotne, ponieważ dobrze dobrany napęd powinien sprostać występujących w danym punkcie obciążeniom, pozwalając na wykonywanie swobodnych ruchów ramienia. Największym ograniczeniem, jakie zostało wzięte pod uwagę był budżet. Jako napęd wykorzystane zostały silniki krokowe.

Aby móc poprawnie dobrać napęd, należy określić, jakie masy będą towarzyszyły poszczególnym segmentom robota. Wybranym materiałem konstrukcyjnym było aluminium, znając wymiary poszczególnych elementów, możliwe było określenie ich masy, a przez to określenie obciążeń. Wstępnie przyjęte zostało podejście określające długości poszczególnych segmentów jako:

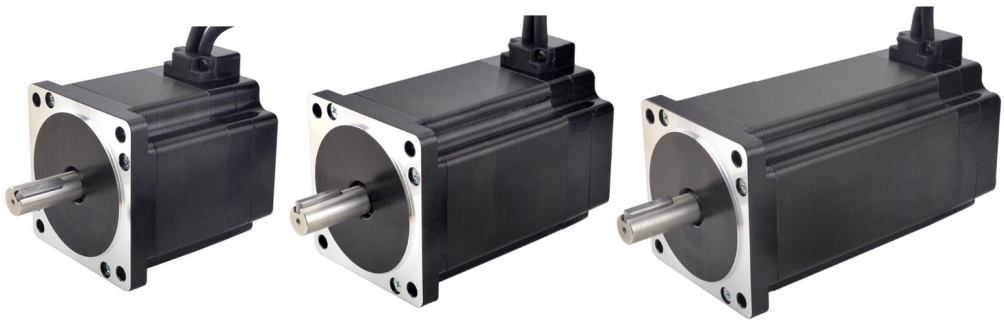
- d1 = 400 mm,
- d2 = 300 mm,
- d3 = 100 mm,
- d4 = 100 mm.

W celu uproszczenia doboru napędu została przygotowana aplikacja umożliwiająca określenie, jakie momenty sił pojawią się poszczególnych przegubach dla dowolnych obciążeń oraz dowolnego ustawienia ramienia robota.



Rysunek 7.4. Aplikacja pozwalająca na obliczenie momentów obciążeń dla każdego z przegubów robota

Pierwotne podejście zakładało, że silnik napędzający daną sekcję robota znajduje się w środku poprzedniej części. Jednak przez powstałe w ten sposób duże momenty obciążeń spowodowały, że wybór padł na przeniesienie napędów ramienia na platformę znajdującą się u podstawy ramienia. Jako jednostki napędowe zostały wybrane trzy rodzaje silników, które przedstawiono na rysunku 7.5.



Rysunek 7.5. Zestawienie silników wykorzystanych w robocie [2]

Są to silniki krokowe opisane standardem NEMA34, umożliwiające pracę w pętli zamkniętej przez wbudowany enkoder. Silnik a) – 4,8 Nm został wykorzystany do sterowania segmentem d3, silniki typu b) – 9 Nm kontrolują segmenty d4 oraz d2, silnik c) – 12 Nm steruje segmentem d1.

3.3. Konstrukcja ramienia

Mając na uwadze dobrane wymiary, silniki oraz koncepcję przeniesienia napędu ku podstawie ramienia, zrealizowany został projekt konstrukcji robota.

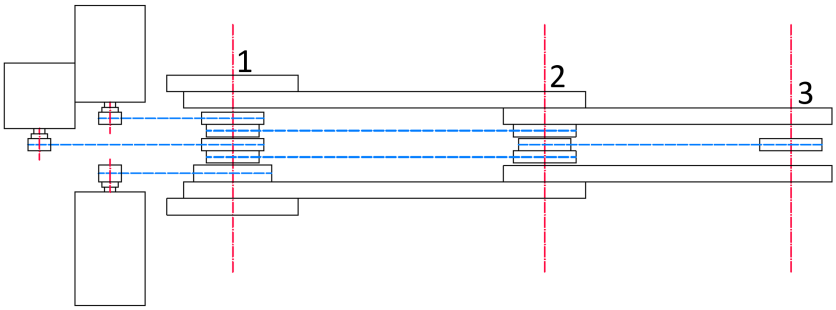


(a) (b)

Rysunek 7.6. Widok opracowanej koncepcji robota

Jak można zauważyć na powyższym rysunku, robot znajduje się na ramie stanowiącej konstrukcję podstawy. Ramię obraca się wraz z platformą, na której znajdują się silniki napędzające poszczególne sekcje robota. Napęd realizujący ten obrót zamontowany jest w podstawie.

Przeniesienie napędu z podstawy do poszczególnych sekcji robota zostało zrealizowane za pomocą pasów zębatych wraz z odpowiednimi kołami zębatymi. Takie podejście pozwoliło na zrealizowanie przekładni zwiększającej moment napędowy.

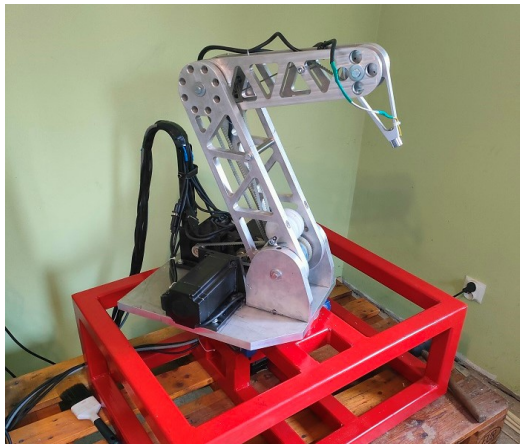


Rysunek 7.7. Schematyczne przedstawienie układu przeniesienia napędu

Powyższy rysunek przedstawia sposób przeniesienia napędu – zostało to zrealizowane za pomocą sprzężonych ze sobą kół zębatych oraz odpowiednich pasów zębatych. Otrzymane w ten sposób przełożenia sprawiły, że uzyskany został moment trzymający w poszczególnych stopniach o wartościach:

- stopień 1: 40 Nm,
- stopień 2: 29 Nm,
- stopień 3: 15 Nm,
- moment trzymający u podstawy: 30 Nm.

Ukończoną konstrukcję przedstawiono na rysunku 7.8.

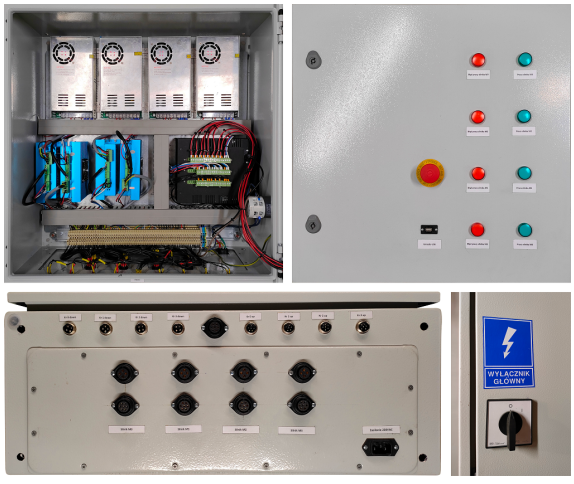


Rysunek 7.8. Ukończona konstrukcja robota

Widoczne na powyższym rysunku kratownice znajdujące się w poszczególnych segmentach służą odciążeniu konstrukcji. Ze względu na konieczność przeprowadzenia dużej ilości kabli do skrzyni elektrycznej na końcu podstawy dla silników znajduje się element prowadzący kable tak, aby nie przeszkadzały w ruchu robota.

Maksymalny udźwig ramienia w obrębie pracy o promieniu do około 50 cm od osi obrotu robota wynosi do 4,5 kg, co pozwala na swobodne manewrowanie przenoszonymi elementami lub narzędziami. Ograniczeniem, które ma istotny wpływ na maksymalny udźwig, jest wykorzystany elektromagnes stanowiący chwytak robota, ograniczający udźwig do 2 kg.

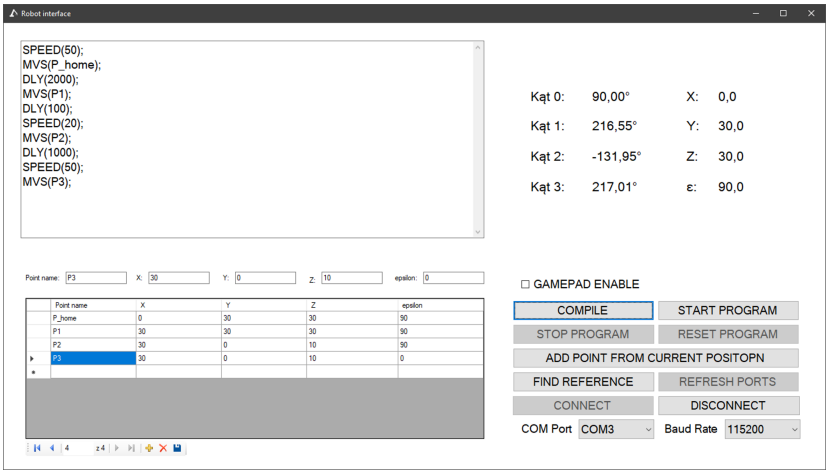
3.4. Układ sterowania



Rysunek 7.9. Główne widoki skrzyni elektrycznej robota

3.5. Środowisko do sterowania robotem

Ważnym założeniem przy realizacji budowy robota było zapewnienie możliwości ręcznego sterowania oraz automatyzacji jego ruchów. Powyższe postulaty zostały zaimplementowane w wykonanej aplikacji komputerowej, pozwalającej na połączenie się z robotem oraz interakcję z nim. Przed przystąpieniem do pracy z robotem należy przeprowadzić procedurę pozycjonowania robota.



Rysunek 7.10. Widok programu do interakcji z robotem

W celu zapewnienia sterowania w trybie ręcznym wykorzystany został kontroler do gier – gamepad, który przez fakt posiadania dwóch drążków analogowych pozwalał na realizację sterowania czterema ze współrzędnych robota. Sterowanie odbywa się z interpolacją liniową. W celu automatycznego sterowania stworzony został dedykowany język programowania ruchów robota inspirowany profesjonalnym językiem MELFA BASIC V.

4. Uzyskane wyniki

Praca polegała na budowie robota o czterech stopniach swobody. Wprowadzony do niej aspekt badawczy polegał na wykonaniu pomiarów dokładności ruchu, w tym celu zostały przeprowadzone 2 rodzaje testów:

- pierwszy z nich badał dokładność, z jaką robot pojawia się w pozycji ustalonej jako pozycja startowa po wykonanym pozycjonowaniu – badanie dokładności pozycjonowania,
- drugi test badał powtarzalność poprzez wielokrotne ruchy z pozycji wyjściowej do pozycji P1 – badanie powtarzalności ruchu.

Badania zostały wykonane precyzyjnym miernikiem zegarowym oraz badana była dokładność ruchu w płaszczyźnie pionowej i poziomej.

Rezultatem badań dokładności pozycjonowania było uzyskiwanie docelowej pozycji ze średnim błędem wynoszącym 0,22 mm w płaszczyźnie poziomej oraz 0,69 w płaszczyźnie pionowej.

W przypadku badania powtarzalności ruchu po dokonanym pozycjonowaniu błąd w płaszczyźnie poziomej wynosił średnio 0,20 mm, natomiast w płaszczyźnie pionowej 0,49 mm.

Ruch w płaszczyźnie poziomej realizowany jest z większą precyzją względem ruchu w płaszczyźnie pionowej, wynika to z faktu, że za ruch w poziomej płaszczyźnie odpowiedzialny był jeden napęd, natomiast w płaszczyźnie pionowej na powstały błąd składały się błędy z pozycjonowania pozostałych trzech napędów.

5. Podsumowanie

Efektem końcowym prac jest robot dający możliwość sterowania ręcznego oraz automatycznego zgodnie z zaprogramowaną sekwencją ruchów. Projekt ten pokazuje, jak w prosty oraz ekonomiczny sposób zbudować robota, który jednak nie jest pozbawiony wad. W trakcie prac nad robotem pojawił się problem związany z regulacją położenia silników pracujących w pętli sprzężenia zwrotnego. Realizowane było to przez dedykowane kontrolery silników, wykorzystujące odczyt z enkoderów zabudowanych w silnikach. Problem dotyczył nastaw regulatora położenia, które fabrycznie nie były dobrane dla układu o tak dużej bezwładności, jakim jest ramię robota. Efektem niepoprawnych nastaw było oscylowanie silników wokół punktów, w których zatrzymywało się ramię. Problem ten dotyczył w największym stopniu napędu segmentu numer d1 (rysunek 7.1) oraz napędu obrotu osi pionowej robota. Ze względu na wykorzystanie sterowników silników, które pochodziły od chińskiego producenta, oraz problemy, a wręcz brak możliwości komunikacji z nimi w celu zmiany nastaw regulatorów, nie udało się uzyskać płynności oraz stabilności w ruchu przy prędkościach powyżej 10% prędkości maksymalnej. Kolejną wadą wynikającą z przyjętych kompromisów jest ograniczone przełożenie, jakie jest efektem wykorzystanego rodzaju przeniesienia napędu, co skutkuje koniecznością pracy silników przy bardzo niskich obrotach. Najlepszym rozwiązaniem byłoby wykorzystanie przekładni o wysokim przełożeniu, jednak idą za tym duże koszty. Wykonane środowisko do pracy z robotem również nie pozwala na zaawansowane sterowanie, a skupia się na umożliwieniu podstawowego sterowania w trybie point to point.

Mimo wielu przyjętych kompromisów, uproszczeń oraz popełnionych błędów projektowych robot spełnia swoje podstawowe zadanie i jest udanym projektem, który pozwolił na rozwinięcie umiejętności inżynierskich w wielu płaszczyznach.

Bibliografia

[1] Kaczmarek W., Panasiuk J., *Robotyzacja procesów produkcyjnych*, PWN, 2017.

[2] *Closed loop stepper motor*. Dostępny online: `https://www.omc-stepperonline.com/closed-loop-stepper-motor/page/1` (dostęp: 20-12-2021).

Programowalny model niewielkiego robota ramieniowego

inż. Michał Zdobych
Automatyka i Robotyka

dr inż. Marcin Pawlak
Katedra Maszyn, Napędów i Pomiarów Elektrycznych

W dzisiejszych czasach zauważa się znaczący wzrost popularności zastosowań robotów w procesach przemysłowych. W związku z tym ważna jest podstawowa wiedza związana z dziedziną robotyki, mikrokontrolerów oraz napędów, aby móc w minimalnym stopniu zrozumieć ich działanie. W tekście zostały poruszone kwestie dotyczące budowy robotów przemysłowych oraz stosowanych w nich układów napędowych. Ponadto opisane zostały podstawy budowy oraz zasady działania mikrokontrolerów na przykładzie platformy programistycznej Arduino. W rozdziale trzecim został opisany proces budowy oraz programowania niewielkiego robota ramieniowego, który był tematem pracy inżynierskiej. Robot został wymodelowany, a następnie wydrukowany za pomocą drukarki 3D. Sterowanie robotem odbywało się za pomocą Arduino Mega oraz aplikacji mobilnej wykorzystującej sygnał Bluetooth.

1. Cel i założenia

Celem poniższej pracy było zaprojektowanie oraz wykonanie modelu niewielkiego robota ramieniowego, który będzie posiadał możliwość zaprogramowania i wykonania dowolnej sekwencji ruchu. Do głównych założeń projektowych należały:

- projekt części mechanicznej robota,
- projekt elektronicznego układu sterującego,
- wykonanie i uruchomienie modelu manipulatora,
- napisanie programu dla sterownika mikroprocesorowego,
- wykonanie testów robota.

2. Wstęp teoretyczny

Aby lepiej zrozumieć omawianą tematykę, należy wyjaśnić podstawowe definicje używanych terminów związanych z robotyką. Robotem nazywa się sterowalną maszynę, która

posiada możliwości manipulacyjno-lokomocyjne oraz wykorzystywana jest do pomocy lub wykonywania pewnych operacji ruchowych, intelektualnych oraz empirycznych człowieka. Kolejnym ważnym terminem jest manipulator, który jest narzędziem przeznaczonym do łapania oraz przenoszenia, przesuwania lub kręcenia przedmiotem [1].

Roboty przemysłowe zbudowane są z 4 podstawowych układów: mechanicznego, napędowego, zasilania i sterowania, które składają się na odrębne zespoły. Układ mechaniczny oraz napędowy wchodzą w skład tak zwanego zespołu ruchu, który także nazywa się manipulatorem. W skład tego zespołu wchodzą: mechanizm (np. przekładnie) oraz układ napędowy (np. silniki) wraz z czujnikami oraz efektorem (końcówką roboczą). Rolę układu sterowania pełni szafa sterownicza, składająca się z pulpitu służącego do uruchamiania robota oraz sterowania ręcznego, a także sterownik, umożliwiający sterowanie oraz programowanie poprzez dodawanie kolejnych punktów pracy. Podstawowymi parametrami technicznymi opisującymi roboty przemysłowe są: udźwig, zasięg, liczba osi z możliwością sterowania, prędkość wykonywanych ruchów, powtarzalność oraz precyzja ruchów. Przedstawiony w niniejszej pracy robot jest robotem przegubowym, ponieważ posiada on pierwsze trzy przeguby obrotowe, licząc od podstawy. Roboty o takiej konstrukcji posiadają wiele zalet, a mianowicie dużą możliwość ruchu po niezbyt dużej przestrzeni roboczej, sztywność konstrukcji, zwinne oraz precyzyjne ruchy [2].

Arduino jest płytką rozwojową z mikrokontrolerem, która posiada wyjście USB wykorzystywane do podłączenia oraz komunikacji z komputerem. Podobnie jak standardowe mikrokontrolery Arduino wyposażone zostało w mikroprocesor, pamięć RAM, pamięć EEPROM lub pamięć FLASH oraz wejścia/wyjścia, które mogą obsługiwać sygnały cyfrowe i analogowe. Wejścia/wyjścia umożliwiają obsługę elektroniki. Przykładowo Arduino można wykorzystać do sterowania silnikami, diodami LED oraz wyświetlaczem LCD. Posiada ono również możliwość odczytywania danych z czujników lub różnych modułów komunikacyjnych, na przykład modułu Bluetooth. Mikrokontroler, który został zamontowany w płytce Arduino, został wykonany przez firmę Atmel, która jest jedną z największych firm odpowiedzialnych za produkcję mikrokontrolerów. Wykorzystane w projekcie Arduino Mega 2560 należy do jednych z najbardziej rozbudowanych modeli z rodziny Arduino [3]. Mikrokontrolery dzięki swoim niewielkim rozmiarom oraz niezbyt dużej mocy obliczeniowej najczęściej są wykorzystywane do sterowania elektroniką. Znajdują one szerokie zastosowanie nie tylko w robotyce, ale również w procesach kontrolno-pomiarowych, automatyce wykorzystywanej w przemyśle oraz w lotnictwie i samochodach [4].

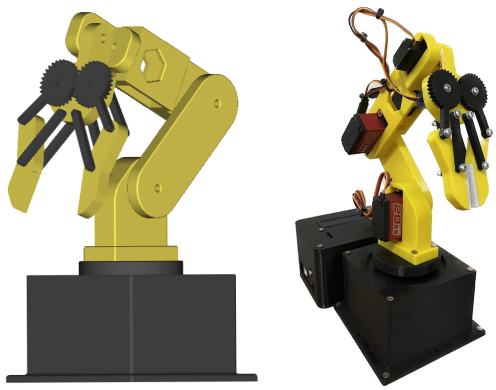
Do komunikacji z urządzeniami współpracującymi mikrokontrolery wykorzystują interfejsy komunikacyjne. Do sterowania robotem zostały wykorzystane interfejsy UART oraz I^2C. Interfejs komunikacyjny UART jest obwodem znajdującym się w mikrokontrolerze, którego celem jest przesyłanie oraz odbieranie informacji. UART tworzą dwa urządzenia współpracujące ze sobą – nadajnik oraz odbiornik. Zadaniem nadajnika jest przetwarzanie informacji z urządzenia odpowiedzialnego za sterowanie (np. procesora), w postaci sekwencyjnej, a następnie wysłanie tej informacji do odbiornika. Do wykonania tej operacji potrzebne są jedynie dwa przewody. Interfejs I^2C jest popularnym standardem wykorzystywanym w urządzeniach elektronicznych używanych na co dzień, elektronice przemysłowej, a także w telekomunikacji. Pozwala łączyć ze sobą dużą liczbę układów, które mogą działać jako podrzędne, czyli slave, oraz nadrzędne, czyli master [4, 5].

3. Projekt i wykonanie

Celem projektu było wykonanie modelu robota ramieniowego. Model ten miał przedstawiać zmniejszoną wersję robota przegubowego. Robot powinien posiadać możliwość wykonania oraz zaprogramowania dowolnej sekwencji ruchu. Przeznaczony jest on do nauki programowania oraz zapoznania z pracą przemysłowego robota ramieniowego. Podobnie jak w większości robotów przemysłowych, programowanie odbywa się poprzez nauczanie. Oznacza to, że robot przemieszcza się do zadanego punktu, współrzędne tego punktu wraz z prędkością są zapisywane, a następnie robot jest zdolny do odtworzenia tych ruchów. Kolejnym podobieństwem jest budowa, ponieważ posiada on również trzy pierwsze przeguby obrotowe, więc można go uznać za robota antropomorficznego. Jego ramię wraz z chwytakiem ma długość 32 cm. Używane silniki w robocie to serwa modelarskie o momencie 20 kg/cm oraz 1,5 kg/cm. Do zasilania robota wykorzystano dwa zasilacze: 5 V DC 4 A oraz 12 V DC 2 A.

3.1. Konstrukcja mechaniczna

Model części mechanicznej robota został wykonany za pomocą programu do modelowania 3D o nazwie FreeCAD 0.18. Po wymodelowaniu elementy te zostały przetworzone do druku za pomocą programu Ultimaker Cura i wydrukowane za pomocą drukarki 3D PRUSA i3 MK3S+. Podczas drukowania zostały użyte filamenty z PLA firmy Prusament w kolorach Pineapple Yellow oraz Jet Black. Rysunek 8.1 przedstawia wykonany model robota ramieniowego wraz z chwytakiem.



(a) (b)

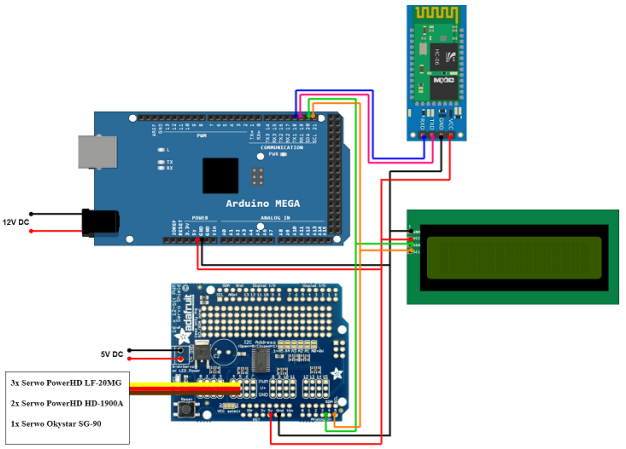
Rysunek 8.1. Wykonany model robota

Podczas modelowania uwzględniono wyważenie ramienia robota poprzez umieszczenie silników w osi obrotu, co znacząco poprawia stabilność całej konstrukcji. Dzięki uwzględnieniu tej cechy robot nie musi być przymocowany do stanowiska, aby móc pracować przy małych obciążeniach. Wynika to również z tego, że podstawa tego modelu jest kwadratowa oraz dosyć ciężka.

3.2. Elektroniczny układ sterujący robota

Rysunek 8.2 przedstawia schemat elektronicznego układu sterującego robotem. Sercem całego układu jest płytka uruchomieniowa mikrokontrolera, czyli Arduino Mega, które zostało zasilone napięciem 12 V DC oraz prądem 2,5 A. Do Arduino został podłączony, za pomocą interfejsu komunikacyjnego I2C, 16-kanałowy sterownik serw Adafruit 1411 wykorzystywany do ustawiania prędkości oraz poruszania serwami. W podobny sposób został włączony do układu wyświetlacz LCD 16×2, na którym będą wyświetlane komunikaty o aktualnym stanie robota. Płytka sterownika jest zasilana napięciem 5 V DC oraz prądem 4 A. Za pomocą portu szeregowego UART podłączono moduł Bluetooth HC-06, który został wykorzystany do połączenia Arduino z telefonem, co umożliwiło sterowanie robotem w czasie rzeczywistym. Moduł ten jest zasilany napięciem 5V DC.

W projekcie robota zostały wykorzystane trzy rodzaje serw o zakresie ruchów od 0° do 180° oraz zostały zasilone napięciem 5 V DC. Trzy serwa PowerHD LF-20MG zostały wykorzystane w pierwszych trzech przegubach. Wynikało to z faktu, że te trzy przeguby będą pracowały pod największym obciążeniem. Serwa te cechuje wysoki moment 20 kg· cm oraz masa 60 g. Kolejne dwa przeguby wykorzystują serwa PowerHD HD-1900A o masie 9 g oraz momencie 1,5 kg· cm, które zostały wybrane głównie ze względu na niewielką masę oraz dosyć duży moment, wystarczający do obrotu chwytaka oraz poruszania chwytakiem. Serwo Okystar Sg-90 zostało użyte do otwierania oraz zamykania chwytaka.



Rysunek 8.2. Elektroniczny układ sterujący [6–8]

3.3. Program sterujący oraz aplikacja mobilna

Napisany program sterujący dla robota umożliwia sterowanie wszystkimi stopniami swobody oraz chwytakiem, prędkością ruchów oraz zapisywaniem ruchów robota wraz z prędkością ich wykonywania. Obsługuje on również przyciski z aplikacji mobilnej odpowiedzialne za: powrót do pozycji domowej, zapis, reset zapisanych pozycji. W trybie odczytu możliwe jest korzystanie z pauzy, przycisku start/stop oraz wyboru opcji prędkości, z której robot może korzystać (z pamięci robota lub ustawionej w aplikacji). Do swojego działania program używa bibliotek związanych ze sterownikiem serw Adafruit 1411, czyli „Adafruit_PWMServoDriver.h” oraz z wyświetlaczem LCD, czyli „LiquidCrystal_I2C.h”.

Dzięki tym bibliotekom możliwe było sterowanie serwami oraz przekazywanie informacji za pomocą wyświetlacza LCD z konwerterem I2C. Wyświetlane są takie informacje, jak:

- „Robot is ready to play!!!" – zaraz po uruchomieniu,
- „Reset saved position" – po usunięciu zapisanej trajektorii ruchu,
- „Return to home position" – powrót do pozycji domowej/startowej,
- „Pause" – pauza,
- „Read mode" – uruchomiony tryb odczytu,
- „Use saved move speed" – gdy robot zostaje ustawiony na korzystanie z zapisanej wartości prędkości,
- „Use speed from controller" – gdy robot zostaje ustawiony na poruszanie się z prędkością ustawianą w aplikacji,
- „Max saved position number" – gdy robot osiągnie maksymalną liczbę zapisanych pozycji,
- „Saved position nr: 1" - wyświetla numer zapisanej pozycji, w tym przypadku 1,
- „Free move mode" – wyłączenie trybu odczytu – swobodne poruszanie robotem.

Aplikacja mobilna sterująca robotem została wykonana za pomocą platformy MIT APP INVENTOR. Jest to strona internetowa, umożliwiająca tworzenie aplikacji mobilnych na urządzenia z systemem Android, używając przy tym języka graficznego typu Scratch. Programowanie w tym języku polega na łączeniu odpowiednich bloczków ze sobą tak, by powstawały warunki, pętle lub zadania do wykonania. Rysunek 8.3 przedstawia okno aplikacji mobilnej odpowiedzialnej za sterowanie robotem. Na rysunku 8.3a przedstawiono wygląd okna zaraz po uruchomieniu aplikacji. Natomiast na rysunku 8.3b przedstawiono okno po połączeniu z modułem Bluetooth.



(a) po uruchomieniu **(b)** po połączeniu z Bluetooth

Rysunek 8.3. Okno aplikacji mobilnej sterującej robotem

Ruch ramieniem robota odbywa się za pomocą suwaków. Wybierając za ich pomocą wartość na osi, określa się pozycję dla każdego z serw w modelu. W podobny sposób

następuje sterowanie prędkością wykonywanych ruchów. Aby zapisać trajektorię ruchów, należy ustawić robot w określonej pozycji i nacisnąć przycisk „Save". Warto zauważyć, że program został tak napisany, że robot zapisuje swoją aktualną pozycję w chwili wciśnięcia przycisku „Save", a nie wartość zadaną. Po wciśnięciu przycisku „Save" zostanie wyświetlony komunikat na wyświetlaczu LCD, informujący o numerze zapisanej pozycji. Aby zapisać kolejną pozycję, należy przestawić ramię robota oraz ponownie użyć przycisku „Save". Aby odtworzyć zapisaną trajektorię ruchów robota, należy wcisnąć przycisk „Start/Stop", który uruchamia odtworzenie jednokrotnie całej sekwencji ruchów. Gdy ten przycisk zostanie użyty w trakcie odczytu pozycji, robot wyjdzie z trybu odczytu i się zatrzyma. Natomiast jeżeli kolejny raz skorzysta się z niego, robot rozpocznie wykonywanie sekwencji ruchów od początku. Wciśnięcie przycisku „Pause" pozwala w trakcie odczytu ruchów zatrzymać robota, a następnie wznowić jego pracę po ponownym jego użyciu. Przycisk „Reset" służy do resetowania zapisanej sekwencji ruchów i umożliwia zapisywanie pozycji od początku. Przełącznik „Save speed?" służy do wyboru prędkości, która będzie obsługiwana podczas odczytywania ruchów. Wybiera się, czy robot ma poruszać się zgodnie z prędkością, z jaką się poruszał podczas zapisu, czy zgodnie z prędkością ustawioną w aplikacji. Przycisk „Home" służy do ustawiania robota w pozycji startowej.

4. Uzyskane wyniki

Wykonany robot został przetestowany w różnych warunkach pracy. Jako pierwsze zostały przeprowadzone testy dotyczące możliwości udźwigu ramienia robota oraz chwytaka. Zanim jednak zostaną one przedstawione warto wspomnieć, że robot waży 622 g, a samo ramię wraz z silnikami i chwytakiem waży 330 g. Długość ramienia robota wraz z chwytakiem wynosi 32 cm i znajduje się na podstawie o wysokości 7,20 cm. Podczas wykonywanych testów został sprawdzony maksymalny udźwig dolnego serwa przy wyprostowanym ramieniu. W trakcie testów na wyprostowanym ramieniu zauważono, że chwytak potrafi pewnie łapać przedmioty o masie do około 60 gramów. Po przekroczeniu tej wartości chwytak nie jest w stanie utrzymać przedmiotu. Jednakże po spięciu szczęki chwytaka opaską zaciskową możliwe było przetestowanie podnoszenia przedmiotów aż do 114 gramów. Po przekroczeniu tej granicy zauważono nienaturalne naprężenie w serwie odpowiedzialnym za obrót chwytaka, zaprzestano więc dalszych testów w obawie przed uszkodzeniem robota. Kolejnym testem było sprawdzenie dokładności odtwarzania ruchów przy różnych prędkościach. Zadanie polegało na zaprogramowaniu sekwencji ruchów polegających na pobieraniu przez robota przedmiotu z jednego miejsca i odkładaniu w miejscu drugim. Po odłożeniu przedmiotu robot wracał do pozycji domowej. Sprawdzono czterokrotnie przy różnych prędkościach poprawność wykonywanych ruchów. Zauważono, że robot niemalże idealnie osiąga zadane pozycje oraz pobiera i odkłada przedmiot w odpowiedni sposób.

5. Podsumowanie

Celem pracy było wykonanie projektu oraz modelu robota ramieniowego o niewielkich rozmiarach. Robot miał cechować się możliwością zapisania oraz wykonania dowolnej sekwencji ruchu. Po zapoznaniu się z literaturą przystąpiono do doboru elektroniki oraz

przygotowania modelu robota i wydrukowania elementów na drukarce 3D. Na tym etapie trzeba było zdecydować o materiale, z jakiego ma być wykonany model. W tym przypadku zdecydowano się na filament PLA, który cechował się dobrą wytrzymałością oraz łatwością drukowania. Po wydrukowaniu modelu można było go złożyć i połączyć z elektronicznym układem sterującym. Następnie przystąpiono do napisania programu sterującego dla Arduino Mega. Po skończeniu pisania programu oraz aplikacji można było uruchomić pierwszy raz model robota. Pojawił się jednak pewien problem, a mianowicie zaraz po podłączeniu układu do zasilania i próbie poruszenia serwem. Problem ten dotyczył serw PowerHD LF-20MG, które pobierały zbyt dużo prądu podczas ruchu. Aby pozbyć się tego problemu, został wlutowany do sterownika serw kondensator o pojemności 560 µF, który został dobrany zgodnie z dokumentacją. Wykonany model robota ramieniowego jest modelem edukacyjnym, posiadającym funkcje zbliżone do funkcji rzeczywistego ramieniowego robota przemysłowego. Do cech wspólnych na pewno można zaliczyć metodę programowania – przez nauczanie, trzy pierwsze przeguby są obrotowe – roboty antropomorficzne oraz oprócz zapisanych ruchów zapisywana jest również prędkość dla danego ruchu. Przygotowany model robota pozwala również wykorzystać zdobytą wiedzę o tematyce mikrokontrolerów oraz robotów i być dobrym materiałem na rozwój w tym kierunku. Na pewno model ten może być rozbudowany o sterowanie oparte na interpolacji liniowej, o dodanie możliwości ruchu poziomego oraz o czujniki, które mogłyby wykrywać obiekty w przestrzeni roboczej robota.

Bibliografia

[1] *Encyklopedia PWN*. Dostępny online: `https://encyklopedia.pwn.pl` (dostęp: 27-09-2022).

[2] Honczarenko J., *Roboty przemysłowe. Budowa i zastosowanie*, Wydawnictwa Naukowo-Techniczne, Warszawa, 2010.

[3] Monk S., *Arduino dla początkujących Podstawy i szkice*, Wydawnictwo Helion, 2014.

[4] Pełka R., *Mikrokontrolery architektura programowanie zastosowania*, Wydawnictwo Komunikacji i Łączności, Warszawa 2000.

[5] Blog Forbot. Dostępny online: `https://forbot.pl/blog/kurs-arduino-uart-komunikacja-z-komputerem-pc-zmienne-id3836` (dostęp: 27-09-2022).

[6] Wokwi.com. Dostępny online: `https://wokwi.com/arduino/new?template=arduino-mega` (dostęp: 27-09-2022).

[7] Sklep Botland. Dostępny online: `https://botland.com.pl/arduino-shield-kontrolery-silnikow-i-serw/2448-sterownik-serw-16-kanalowy-12-bitowy-pwm-i2c-shield-dla-arduino-adafruit-1411-5904422302825.html` (dostęp: 27-09-2022).

[8] Arduino.cc. Dostępny online: `https://create.arduino.cc/projecthub/RucksikaaR/interfacing-the-hc-06-bluetooth-module-with-arduino-f9c315` (dostęp: 27-09-2022).

Zastosowanie platformy Omron Sysmac w układach sterowania wieloosiowych napędów pozycjonujących

mgr inż. Marek Myśluk
Automatyka Przemysłowa

dr inż. Marcin Pawlak
Katedra Maszyn, Napędów i Pomiarów Elektrycznych

W niniejszym artykule przedstawiona została tematyka sterowników programowalnych ze szczególnym uwzględnieniem urządzeń podlegających pod platformę Omron Sysmac. W dalszej części omówiony został proces konstrukcji modelu wieloosiowego napędu pozycjonującego. Następnie na przykładzie działania zbudowanego modelu przedstawiono możliwości zastosowania wybranych funkcjonalności platformy Omron Sysmac. Wygenerowane podczas testów rysunki stanowią odzwierciedlenie działania takich maszyn pozycjonujących, jak frezarki, plotery, maszyny CNC itp. Rozwiązania programowe przedstawione w pracy mogą być zastosowane w znacznie większych oraz bardziej złożonych procesach przemysłowych.

1. Cel i założenia

Celem pracy dyplomowej było przebadanie i analiza możliwości zastosowania platformy Sysmac firmy Omron do realizacji różnych algorytmów sterowania wieloosiowych układów napędowych z silnikami krokowymi. Ponadto w ramach pracy skonstruowany miał być odpowiedni model pozwalający na przeprowadzenie założonych badań. Główne założenia projektowe realizowanej pracy to:

- zaprojektowanie i wykonanie mechanicznego modelu wieloosiowego układu napędu pozycjonującego,
- zaprojektowanie i uruchomienie stanowiska laboratoryjnego ze sterownikiem NJ-301 i modułami impulsowymi serii NX-PG,
- przebadanie wybranych algorytmów sterowania zsynchronizowanych napędów pozycjonujących,
- opracowanie przykładowych programów sterujących na platformie Sysmac.

2. Wstęp teoretyczny

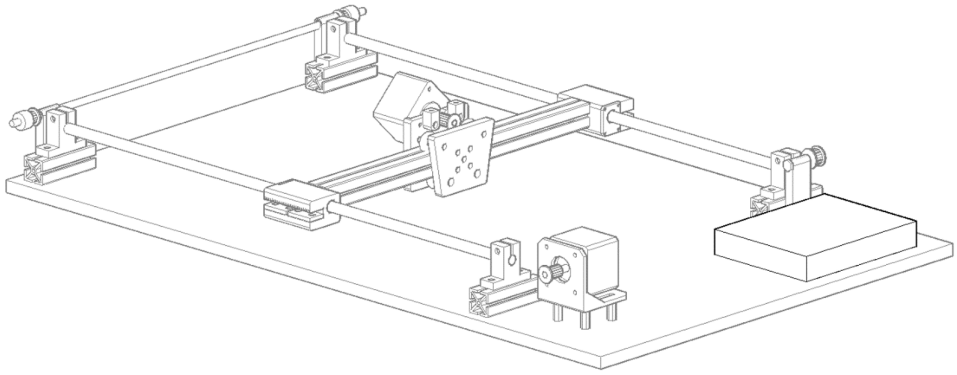
Ciągły rozwój technologiczny oraz postępujący konsumpcjonizm sprawiają, że producenci starają się zoptymalizować swoje procesy produkcyjne pod względem czasu oraz ilości wytwarzanych produktów. Automatyzacji podlegają nie tylko skomplikowane procesy przemysłowe, ale również coraz mniejsze hale czy zakłady, a nawet gospodarstwa domowe. Zjawisko to sprawia, że na przestrzeni lat zaobserwować można znaczący rozwój sterowników programowalnych. Urządzenia te stały się obecnie jednymi z najpopularniejszych aparatów stosowanych do sterowania zautomatyzowanymi procesami technologicznymi, niemal w każdej gałęzi przemysłu. Ich szeroki zakres możliwości pozwala na realizację praktycznie dowolnych rozbudowanych algorytmów sterujących. Duża różnorodność modeli oraz kompaktowa lub modułowa budowa sprawiają, że sterowniki PLC są uniwersalnym narzędziem, którego konfiguracja sprzętowa przystosowana jest do stawianych im zadań czy warunków środowiskowych. Sterowniki PLC, oprócz szerokich możliwości programowalnych, są również kompatybilne z czujnikami, układami pomiarowymi, panelami operatorskimi oraz umożliwiają obsługę układów napędowych, pozwalając nie tylko na sterownie danym procesem, lecz także na zdalną diagnostykę, gromadzenie danych i obsługę systemów bezpieczeństwa. Dzięki temu świetnie sprawdzają się w obsłudze układów z napędami pozycjonującymi.

Sterowniki automatyki maszynowej z serii NJ podlegają pod platformę SYSMAC, która jest najnowszą platformą automatyzacji maszyn firmy Omron. Sterowniki te są wyspecjalizowanymi urządzeniami służącymi do automatyzacji procesów przemysłowych. Charakteryzują się dużą funkcjonalnością, szybkością działania oraz łatwością konserwacji. Sterowniki NJ łączą w sobie możliwości poprzednich wersji sterowników firmy Omron z dodatkowymi funkcjami rozszerzającymi możliwości sterowania maszynami oraz układami napędowymi. W sterownikach platformy SYSMAC w miejsce dedykowanych układów ASIC wykorzystywanych w tradycyjnych sterownikach zastosowano procesory Intel Atom^{TM} lub Core^{TM} oraz systemy RTOS, co znacznie poprawiło wydajność, skróciło fazę projektowania oraz zapewniło szybszy czas reakcji urządzenia. Aby zapewnić najszybszą wymianę danych, platforma wykorzystuje standard EtherCat – nie tylko w układach kontroli ruchu, lecz także w układach wejść/wyjść, bezpieczeństwa oraz wizyjnych.

Omron Sysmac Studio jest zintegrowanym środowiskiem programowania służącym do obsługi sterowników programowalnych podlegających pod platformę SYSMAC oraz innych urządzeń wykonawczych. Łączy ono w sobie obsługę logiki, ruchu, robotyki, paneli HMI, systemów wizyjnych, czujników, zabezpieczeń oraz symulacji 3D. Oprogramowanie jest zgodne z otwartym standardem IEC 61131-3. Pozwala na implementacje programów w języku drabinkowym oraz językach programowania strukturalnego. Poza powszechnie stosowanymi blokami funkcyjnymi posiada bogatą bibliotekę instrukcji do sterowania ruchem – Motion Control. Sysmac Studio pozwala na parametryzację ustawień modułów osi, ustawienia parametrów sieci EtherCat, monitorowanie pracy sytemu sterowania, symulacje oraz edycje parametrów krzywek. W połączeniu z dedykowanymi sterownikami PLC pozwala na sterownie napędami poprzez zsynchronizowanie osi, a więc uzależnienia działania jednej osi od drugiej. Dzięki temu świetnie sprawdza się w sterowaniu rozbudowanymi układami napędów pozycjonujących, znacznie upraszczając proces programowania [1].

3. Projekt i wykonanie

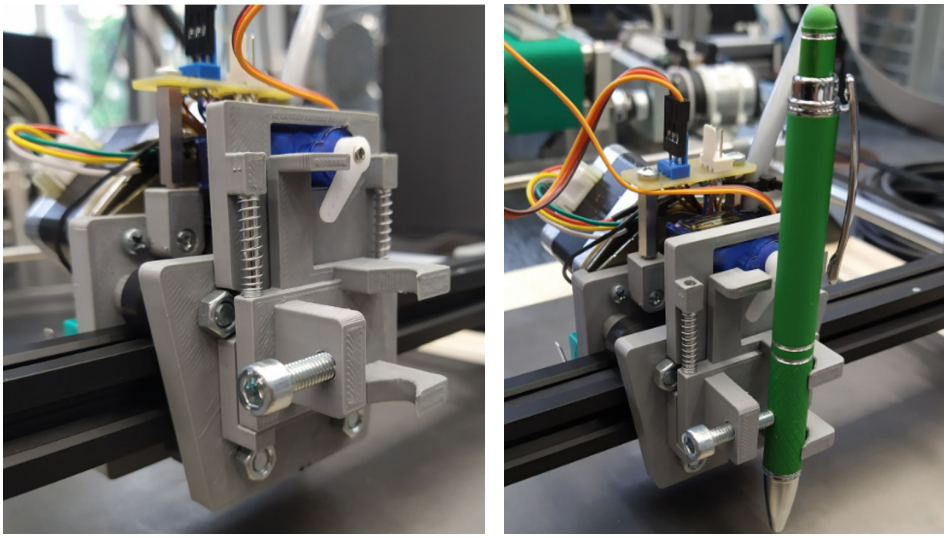
Skonstruowany model miał umożliwić testowanie różnych funkcji platformy Sysmac firmy Omron związanych z układami sterowania wieloosiowych napędów. W związku z tym głównym założeniem było stworzenie makiety posiadającej trzy osie ruchu. Dwie osie napędzane miały być z użyciem silników krokowych, a ich zakres pracy miał wynosić około kilkadziesiąt centymetrów. Zorientowane miały być prostopadle do siebie oraz równolegle do płaszczyzny roboczej. Trzecia oś miała pracować prostopadle do płaszczyzny roboczej, w niewielkim zakresie ruchu. W związku z tym do jej realizacji zastosowany miał być serwomechanizm. Całość napędu miała znajdować się na jednolitej postawie o kształcie prostokąta. Pierwszym krokiem przy konstruowaniu modelu było wykonanie projektu wieloosiowego napędu pozycjonującego, spełniającego założenia projektowe, przedstawiony na rysunku 9.1.



Rysunek 9.1. Projekt wieloosiowego napędu pozycjonującego

3.1. Budowa części mechanicznej

Jako podstawa całej konstrukcji wykorzystana została drewniana sklejka, do której mocowano kolejne elementy konstrukcyjne. Aby zapewnić ruch w pierwszej osi, na modelu zamontowane zostały wałki liniowe osadzone na wspornikach i profilach aluminiowych. Następnie na wałkach tych zamontowano łożyska liniowe umieszczone w specjalnych obudowach, wydrukowanych z użyciem drukarki 3D. Obudowy te zostały zaprojektowane w taki sposób, aby można było zamocować do nich paski klinowe przenoszące napęd z silnika krokowego oraz profil aluminiowy, stanowiący poprzeczkę między wcześniej zamontowanymi wałkami. Następnie wydrukowany został specjalny wózek, posiadający otwory montażowe na silnik krokowy, a także na chwytak. Umieszczenie tego elementu na wcześniej zamontowanym profilu aluminiowym pozwoliło na zrealizowanie napędu w drugiej osi. Następnie wydrukowany został chwytak wykorzystujący serwomechanizm, który następnie przymocowano do ruchomego wózka. Ostatecznie na obszarze roboczym plotera przyklejony został prostokątny płat blachy z czarnej stali. Rysunek 9.2 przedstawia wykonany wózek z zamontowanym chwytakiem.



(a) (b)

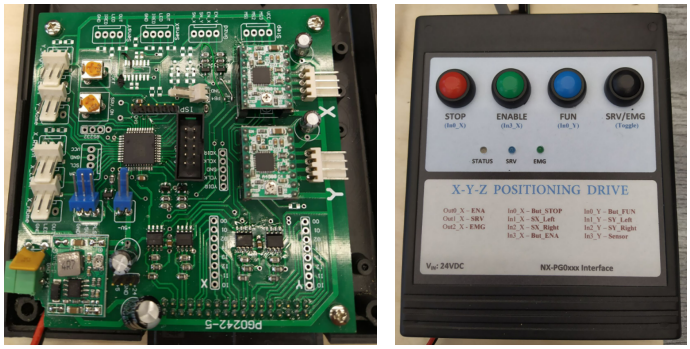
Rysunek 9.2. Gotowy wózek wraz z chwytakiem

3.2. Układ elektroniczny

Podstawą układu elektronicznego modelu jest płytka drukowana (rysunek 9.3). Jest ona uniwersalnym układem pozwalającym na:

- sterowanie silnikami krokowymi,
- sterowanie serwomechanizmem,
- sterowanie elektromagnesem,
- obsługę czujników,
- obsługę przełączników krańcowych,
- obsługę dodatkowych przycisków oraz diod znajdujących się na obudowie układu,
- obniżenie napięcia doprowadzonego ze stanowiska laboratoryjnego do napięcia zasilania poszczególnych elementów roboczych.

Sercem układu jest mikroprocesor, który pozwala na odpowiednie wysterowanie poszczególnych wyjść innych używanych układów scalonych, a także na wygenerowanie sygnału sterującego serwomechanizmem. Ponadto na płytce umieszczone zostały dwa sterowniki Step-stick A4988, które pozwalają na sterowanie silnikami krokowymi. Płytka z układem elektronicznym została umieszczona w obudowie (rysunek 9.3), która została przymocowana do podstawy modelu. Na boku obudowy znajduje się złącze na taśmę przewodów, która służy do połączenia układu ze sterownikiem PLC. Na górze obudowy znajdują się 4 przyciski oraz 3 diody sygnałowe. Funkcje poszczególnych przycisków zostały przedstawione w tabeli 9.1. Ponadto z obudowy wyprowadzone są dwa przewody – czarny oraz czerwony. Są one zakończone męskimi wtykami bananowymi i służą do zasilenia układu stałym napięciem 24 V. Rysunek 9.4 przedstawia gotowy model wieloosiowego napędu pozycjonującego.

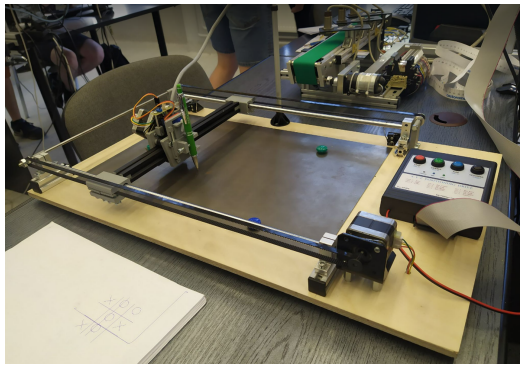


(a) płytka drukowana **(b)** obudowa

Rysunek 9.3. Układ elektroniczny

Tabela 9.1. Funkcje przycisków znajdujących się na obudowie

Przycisk	Funkcja
STOP	Awaryjne zatrzymanie działania modelu. Po jego wciśnięciu wszystkie napędy przestają pracować. Aby ponownie uruchomić układ należy wyłączyć, a następnie ponownie włączyć sygnał „Enable”
ENABLE	Załączenie obwodów mocy układu.
FUN	Pomocniczy przycisk programowalny.
SRV/EMG	Niezależne wymuszenie zadziałania serwomechanizmu lub elektromagnesu



Rysunek 9.4. Gotowy model

4. Uzyskane wyniki

Wszystkie testy przeprowadzone zostały w Laboratorium Sterowania Urządzeniami i Napędami Przemysłowymi Politechniki Wrocławskiej. Celem badań było przetestowanie możliwości zastosowania platformy Omron Sysmac do sterowania wieloosiowym napędem pozycjonującym. Jako element wykonawczy posłużył wcześniej przedstawiony model, wykonany w ramach realizacji pracy. Połączony został on ze sterownikiem NJ301-1100

wyposażonym w moduły szybkich wyjść impulsowych typu NX-PG. Do zaprogramowania sterownika użyte zostało środowisko Omron Sysmac Studio. Przetestowane zostały funkcje platformy z zakresu pozycjonowania do pozycji początkowej, ręcznego sterowania napędami, automatycznego pozycjonowania do wybranego punktu, zastosowania elektronicznych przekładni i krzywek oraz sterowania pogrupowanych osi.

4.1. Sterowanie ręczne

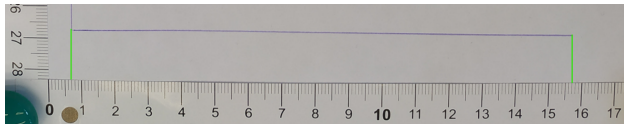
W pierwszej kolejności stworzony został program pozwalający na ręczne sterowanie napędami modelu. Uwzględnia on regulację prędkości przemieszczania się w każdej z osi oraz pozwala użytkownikowi poruszać się w czterech kierunkach: lewo/prawo (w programie negative/positive) w osi X oraz lewo/prawo w osi Y. Do regulacji prędkości oraz ruchu w każdym z kierunków użyto odpowiednio potencjometrów oraz czterech przełączników znajdujących się na stanowisku laboratoryjnym. W programie uwzględniono obowiązek występowania sygnału „enable", aby możliwe było wykonanie ruchu w jakimkolwiek kierunku, oraz zrealizowano instrukcje pozycjonowania każdej z osi do ustawienia początkowego. Z uwagi na to, że w kodzie tym uwzględniono podstawowe funkcje potrzebne do realizacji wszystkich innych testów, posłużył on jako rdzeń, który był rozbudowywany w zależności od zadań, jakie realizować miały kolejne programy.

4.2. Dokładność pozycjonowania

Aby sprawdzić czy wartości wprowadzane w programie przekładają się na działanie napędów, przeprowadzono test polegający na narysowaniu przez model dwóch linii prostopadłych do siebie. Napisany został program, który w pierwszej kolejności ustawia napęd w pozycji 40.40. W tym celu użyto bloków funkcyjnych „MC_MoveAbsolute". Następnie do kartki dociśnięty zostaje długopis pracujący w osi Z. W kolejnym kroku następuje przejazd o 25 cm w prawo wzdłuż osi X. W ten sposób powstaje pierwsza linia. Kolejna linia tworzona jest wzdłuż osi Y poprzez przemieszczenie napędu o 15 cm w prawo. Do zrealizowania tych operacji wykorzystano bloki „MC_MoveRealtive". Ostatecznie serwomechanizm podnosi długopis, a układ wraca do pozycji domowej 0.0. Narysowane linie zostały zmierzone z użyciem miary (rysunek 9.5). Jak widać na rysunku 9.5, ich długość odpowiada tej, którą zaimplementowano w programie, z dokładnością do 1 mm. Można więc uznać, że test przebiegł pomyślnie, a napęd jest pozycjonowany zgodnie z oczekiwaniami.



(a) oś X

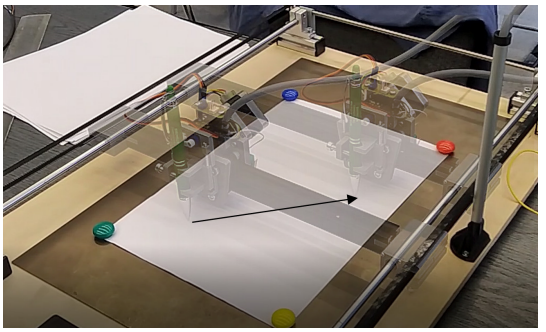


(b) oś Y

Rysunek 9.5. Linie narysowane przez napęd w poszczególnych osiach

4.3. Przekładnie elektroniczne

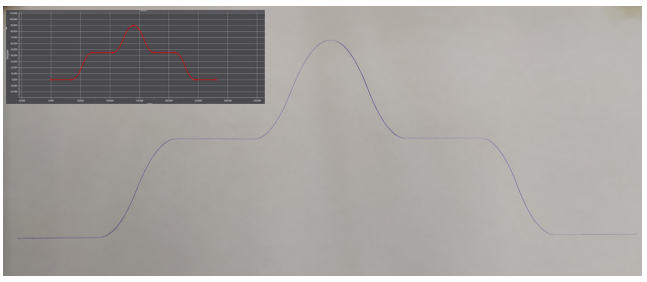
Kolejny test dotyczył wykorzystania funkcji elektronicznej przekładni dostępnej w programie Sysmac Studio. Pozwala ona na stworzenie wirtualnego sprzęgnięcia dwóch napędów. W tym celu należy zastosować bloki funkcyjne „MC_GearIn" oraz „MC_GearOut". W programie ustawić należy, która z osi będzie nadrzędną, a która podrzędną. Następnie podaje się wartość przekładni. Użytkownik będzie miał możliwość operowania jedynie w osi nadrzędnej, natomiast ruch w osi podrzędnej będzie wykonywany automatycznie w taki sposób, jakby między napędami owych osi zamontowana była przekładnia o parametrach odpowiadających tym ustawionym w programie. Rysunek 9.6 przedstawia sposób działania układu przy zastosowaniu elektronicznej przekładni.



Rysunek 9.6. Trasa przebyta przez układ przy zastosowaniu elektronicznej przekładni

4.4. Krzywki

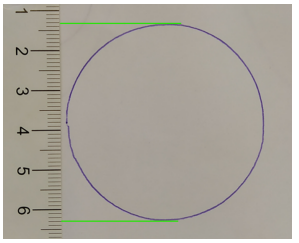
Program Sysmac Studio daje możliwość projektowania krzywek, które następnie można zastosować do sterownia układem napędowym. W tym celu w pierwszej kolejności należy zdefiniować pożądany kształt krzywki. Należy przypisać osie jako nadrzędna oraz podrzędna, a następnie ustawić jednostki, w jakich krzywka będzie interpretowana w danej osi. Przebieg tworzy się poprzez dodawanie kolejnych punktów oraz ustawienie, w jaki sposób przeprowadzona ma zostać interpolacja między nimi (linia prosta, paraboliczna, hiperboliczna itp.). Następnie w sekcji programu należy użyć bloku funkcyjnego „MC_CamIn". Ustawia się w nim, według jakiej krzywki ma odbywać się ruch oraz przypisuje się osie jako „Master" i „Slave". Oznacza to, że operator będzie miał możliwość sterowania jedynie osią nadrzędną, natomiast ruch w osi podrzędnej odbywa się automatycznie, zgodnie z kształtem zadanej krzywki. Efekty przedstawiono na rysunku 9.7.



Rysunek 9.7. Rysunek stworzony przez układ działający w trybie krzywki

4.5. Zsynchronizowanie osi

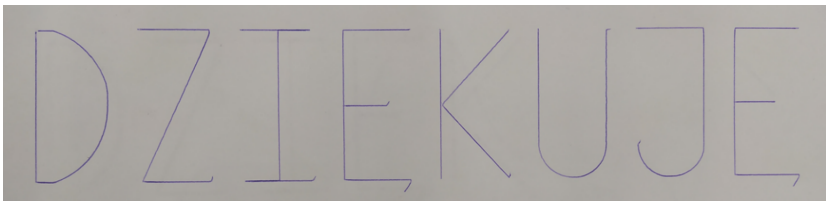
Kolejną funkcjonalnością platformy Sysmac jest możliwość grupowania napędów. Przypisanie kilku osi do jednej grupy pozwala na zsynchronizowanie ich działania, a co za tym idzie – do znacznego uproszczenia algorytmów ich sterowania. Przykładem może być program mający za zadanie zmianę położenia napędu z punktu 0.0 do 35.20. Aby wykonać taki ruch sterując napędami niezsynchronizowanymi, należy użyć dwóch bloków funkcyjnych: jednego odpowiadającego za ruch do położenia 35 w osi X, drugiego odpowiadającego za ruch do położenia 20 w osi Y. Zakładając, że napędy poruszają się z tą samą prędkością i wystartowały one w jednym momencie, szybciej wykona się operacja w osi Y. Dzieje się tak, ponieważ każda z osi działa niezależnie od siebie. Przy przypisaniu osi X oraz Y do jednej grupy, analogiczną operację można uzyskać z użyciem jednego bloku funkcyjnego, w którym wprowadzone zostaną współrzędne 35.20. W tym przypadku jednak osie działają zależnie od siebie, przez co przejazd zostanie wykonany w linii prostej, z punktu początkowego do punktu docelowego. Zarówno napęd w osi X, jak i w osi Y zostanie spozycjonowany w tym samym czasie. Ponadto zastosowanie napędów grupowych pozwala na bardzo proste wykonywanie operacji, które w przypadku sterowania napędami niezależnie od siebie byłyby trudne w realizacji. Przykładem takiej operacji jest interpolacja kołowa. Program Sysmac studio posiada specjalny blok funkcyjny odpowiedzialny za tworzenie okręgów. Przykład jego zastosowania został przedstawiony na rysunku 9.8.



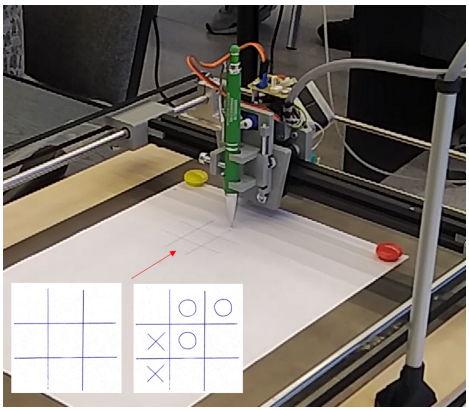
Rysunek 9.8. Okrąg narysowany przez model

4.6. Złożone programy sterujące

Ostatecznie stworzone zostały dwa programy łączące w sobie wszystkie opisane wcześniej funkcjonalności platformy Omron Sysmac. Pierwszy z nich miał za zadanie wysterowanie układu w taki sposób, aby ploter napisał na kartce napis „DZIĘKUJĘ”. Efekty jego działania przedstawiono na rysunku 9.9. Drugi program miał za zadanie narysowanie planszy do gry w kółko i krzyżyk, a następnie umożliwienie użytkownikowi na zdecydowanie o narysowaniu pożądanego znaku na wybranym polu. Przykład działania tego programu przedstawiony został na rysunku 9.10.



Rysunek 9.9. Napis stworzony przez model



Rysunek 9.10. Efekty działania programu do gry w kółko i krzyżyk

5. Podsumowanie

Platforma Omron Sysmac jest przydatnym narzędziem w dziedzinie sterowania napędów pozycjonujących. Bloki funkcyjne oferowane w programie Sysmac Studio pozwalają na realizację prostych operacji, takich jak ręczne sterowanie układem napędowym, czy implikowanie złożonych struktur sterowania automatycznego. Funkcje platformy, takie jak synchronizowanie osi, tworzenie elektronicznych przekładni, projektowanie krzywek, względem których ma odbywać się ruch, czy interpolacje kołowe pozwalają na tworzenie dowolnych programów sterujących, realizujących zarówno skomplikowane, jak i proste procesy technologiczne. Dzięki rozbudowanej bibliotece bloków funkcyjnych „Motion Control" sterowanie ruchem jest znacznie uproszczone. Niejednokrotnie złożone programy sterujące można zastąpić gotowymi blokami funkcyjnymi, których zastosowanie sprowadza się do ustawienia pożądanych parametrów. Dzięki temu korzystanie z platformy jest intuicyjne i wymaga jedynie podstawowej wiedzy z zakresu obsługi sterowników programowalnych.

Skonstruowany model wieloosiowego napędu pozycjonującego stanowi bardzo dobry model dydaktyczny. Jego wielkość jest dobrana w taki sposób, że mieści się on na stanowisku laboratoryjnym, a jednocześnie zapewnia dużą płaszczyznę roboczą. Konstrukcja chwytaka pozawala na zamontowanie na nim zarówno długopisu, jak i elektromagnesu czy sensora. Na podstawie modelu umieszczony został płat blachy z czarnej stali, co pozwala na przytwierdzanie do niej kartek za pomocą magnesów lub zastosowanie czujników indukcyjnych. Dzięki temu model może posłużyć jako pomoc dydaktyczna przy nauce sterowania napędami pozycjonującymi realizującymi różne procesy przemysłowe. Za jego pomocą możliwe jest testowanie oraz nauka zastosowania platformy Omron Sysmac z zakresu sterownia napędami.

Zastosowanie algorytmów sterowania przetestowanych w ramach realizacji pracy przedstawiono na modelu dydaktycznym. Analogiczne rozwiązania mogą zostać wykorzystane przy sterowaniu rzeczywistymi procesami technologicznymi wieloosiowych napędów pozycjonujących, takich jak frezarki, plotery, maszyny CNC itp.

Bibliografia

[1] industrial.omron.pl. Dostępny online: `https://industrial.omron.pl/pl/products/sysmac-studio` (dostęp: 30-09-2022).

Robot z funkcją przenoszenia paczek

inż. Stanisław Piaskowski
Automatyka i Robotyka

dr inż. Paweł Ewert
Katedra Maszyn, Napędów i Pomiarów Elektrycznych

W większości zakładów produkcyjnych ostatnim etapem jest pakowanie, sortowanie i układanie produktów na paletach. Obecnie naturalne jest zastosowanie do tego rodzaju prac robotów. W ramach pracy dyplomowej pod tytułem „Robot z funkcją przenoszenia paczek" skupiono uwagę na sprawdzeniu, z jakimi problemami wiąże się zaprojektowanie i zbudowanie urządzenia umożliwiającego wykonywanie takich czynności. Konstrukcja została zaprojektowana w programie FreeCad. Składa się ona z ramienia zakończonego chwytakiem, umieszczonego na ruchomej podstawie, która umożliwia poruszanie się całości wzdłuż linii prostej. W celu pokazania, jak może wyglądać sortowanie, na przenoszonych paczkach umieszczone zostały etykiety w różnych kolorach, które identyfikowane są przez zainstalowany czujnik. W zależności od barwy paczki są zabierane przez chwytak i umieszczane w odpowiednich miejscach. Model wydrukowano przy pomocy drukarki 3D, a do sterowania zastosowany został mikrokontroler zaprogramowany w środowisku Arduino.

1. Cel i założenia

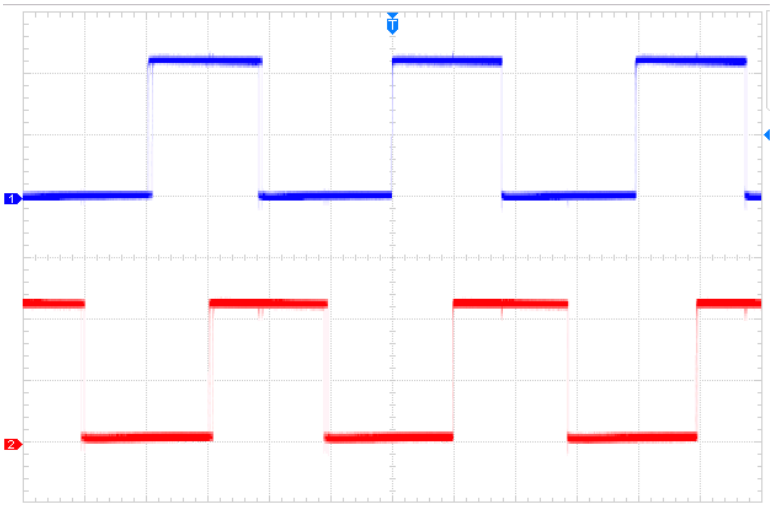
Celem pracy inżynierskiej było zbudowanie robota umożliwiającego bezpieczne chwytanie, przemieszczenie i odkładanie niewielkich paczek. Główne założenia projektowe realizowanej pracy to:

- wybór podzespołów,
- zaprojektowanie i zbudowanie robota,
- opracowanie i wykonanie programu sterującego,
- testowanie opracowanej konstrukcji.

2. Wstęp teoretyczny

Jedną z kluczowych kwestii koniecznych do prawidłowej pracy robota jest dokładność i powtarzalność wykonywanych ruchów. W celu uzyskaniu informacji o wykonywanym przez

wał silnika obrocie stosuje się enkodery. Generują one na wyjściach dwa sygnały prostokątne przesunięte względem siebie o 90 stopni. Dzięki temu przesunięciu możliwe jest określenie również kierunku obrotów. Przykładowy przebieg takich sygnałów przedstawia rysunek 10.1.



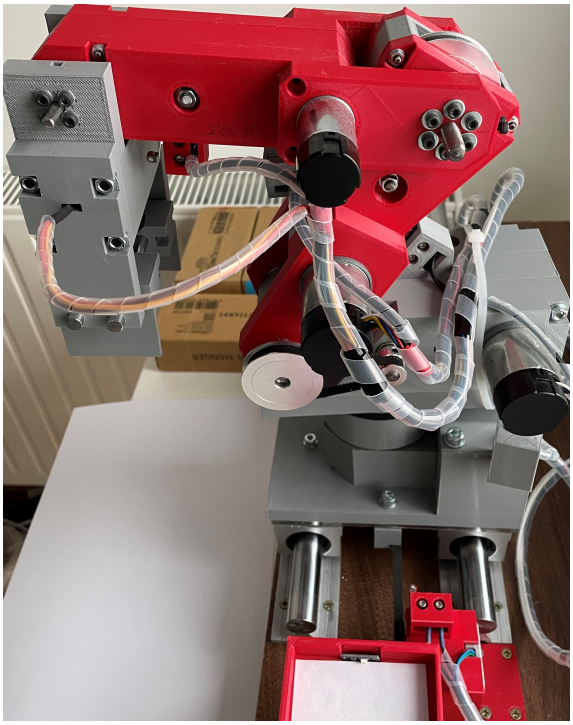
Rysunek 10.1. Przebieg dwóch sygnałów z enkodera zarejestrowany na oscyloskopie [1]

Sygnał z enkodera zmienia się bardzo szybko, a żadna ze zmian nie może być pominięta przez system sterowania silnikiem. W mikrokontrolerach do odczytywania tego typu zmian wykorzystać można przerwania. Wystąpienie przerwania powoduje zatrzymanie wykonywania programu i przejście do procedury jego obsługi. Po wykonaniu przerwania następuje powrót do wykonywania programu, który został przerwany. Wyróżnia się przerwania wewnętrzne, czyli takie, których sygnały pochodzą z wnętrza mikrokontrolera (np. z modułu przetwornika analogowo-cyfrowego), oraz przerwania zewnętrzne, czyli takie, których sygnały pochodzą z zewnątrz mikrokontrolera (np. z enkodera, którego wyjście jest podłączone do odpowiedniego wejścia mikrokontrolera) [1–3].

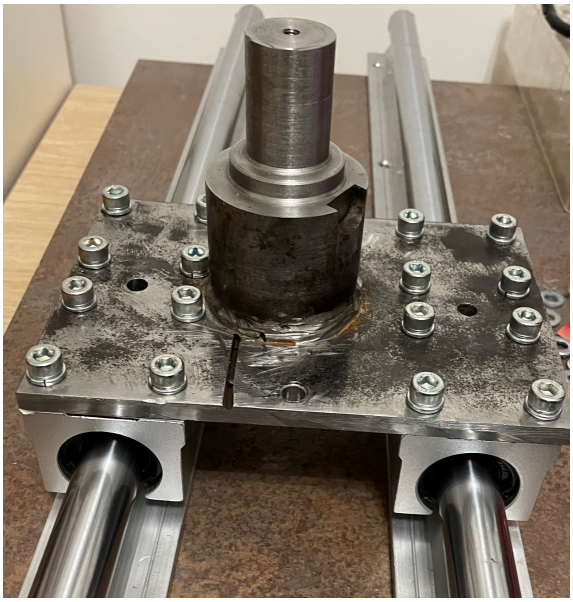
3. Projekt i wykonanie

Tytułowy robot składa się z obrotowej podstawy poruszającej się na prowadnicy liniowej. Na końcu ramienia o długości prawie 50 cm znajduje się chwytak. Przed robotem znajduje się uchwyt na paczki z zainstalowanym czujnikiem koloru. Do napędu wszystkich części ruchomych zastosowano trzy silniki prądu stałego z enkoderami i dwa silniki krokowe. Widoczne na rysunku 10.2 elementy konstrukcji w kolorze szarym i czerwonym zostały zaprojektowane w programie FreeCad, a następnie wydrukowane na drukarce 3d (technologia FDM) [3].

W celu umożliwienia ruchu robota wzdłuż stołu zastosowano dwa wałki, na których znajduje się metalowa podstawa, co przedstawia rysunek 10.3. Obudowy łożysk liniowych są przykręcone do stalowej płytki. Do niej jest również przyspawany wał o trzech różnych średnicach dopasowanych do łożysk [3].

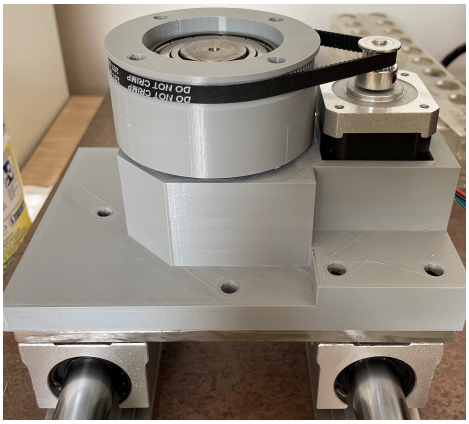


Rysunek 10.2. Widok ogólny robota [3]



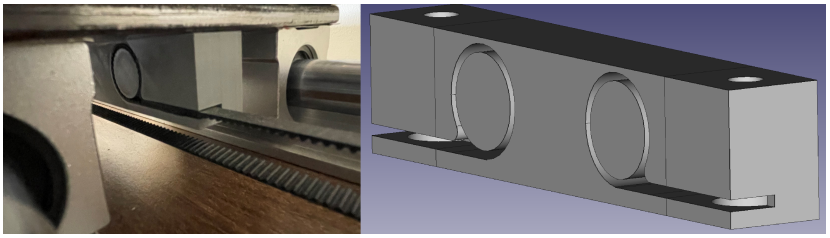
Rysunek 10.3. Prowadnica liniowe [3]

Rysunek 10.4 przedstawia, w jaki sposób silnik krokowy połączony jest z podstawą ramienia, umożliwiając ruch obrotowy [3].



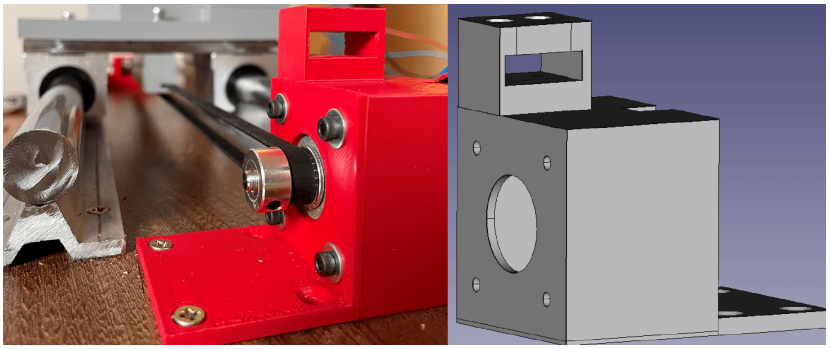
Rysunek 10.4. Przeniesienie napędu z silnika krokowego na koło zębate [3]

Do metalowej części podstawy przykręcony jest element mocujący pasek zębaty, zapewniający możliwość ruchu wzdłuż stołu, co ilustruje rysunek 10.5.



Rysunek 10.5. Mocowanie łączące pasek z ruchomą podstawą [3]

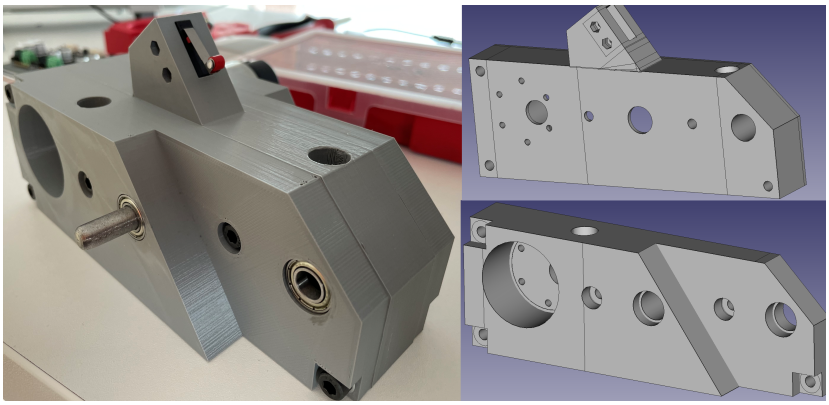
Rysunek 10.6 przedstawia napęd prowadnicy liniowej. Silnik krokowy umieszczony jest na końcu wałków. Na jego wale zamontowana jest zębatka dopasowana do paska zębatego GT2. Na górze obudowy silnika znajduje się miejsce na krańcówkę, dzięki której możliwe jest pozycjonowanie [3].



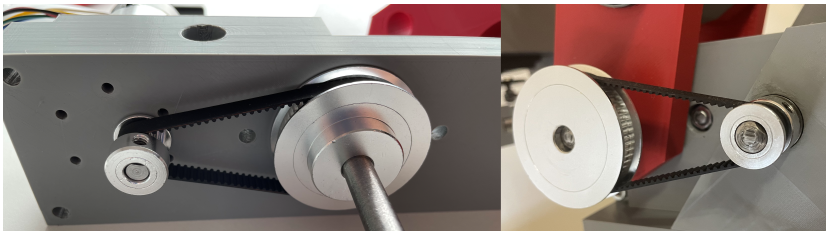
Rysunek 10.6. Obudowa silnika krokowego osi liniowej [3]

Najniżej położoną część ramienia prezentują rysunki 10.7 i 10.8. Silnik prądu stałego z enkoderem powoduje ruch koła zębatego, które poprzez pasek GT2 przenosi siłę na kolejne, większe koło zębate. Następnie poprzez kolejne połączenie paska i zębatki obracane

jest ramię. Wałki osadzone są w łożyskach. Podstawa ramienia składa się z trzech części połączonych za pomocą śrub. Na środkowej części jest miejsce na umieszczenie krańcówki. Kolejne człony ramienia są zbudowane w analogiczny sposób, z tą różnicą, że pasek łączy w nich bezpośrednio silnik z zębatką i członem ramienia [3].

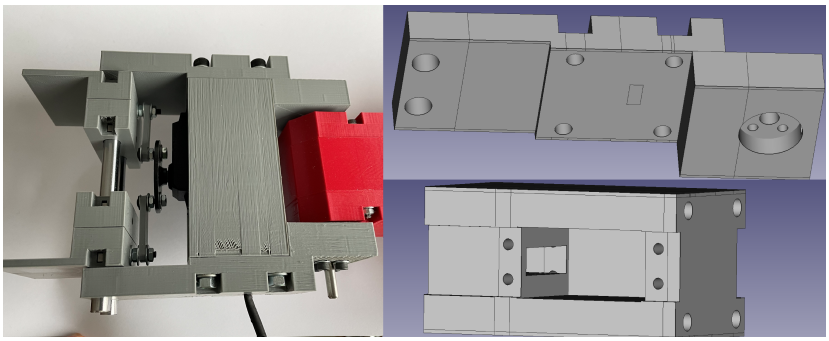


Rysunek 10.7. Podstawa ramienia [3]



Rysunek 10.8. Sposób przeniesienia napędu zlokalizowany w podstawie ramienia [3]

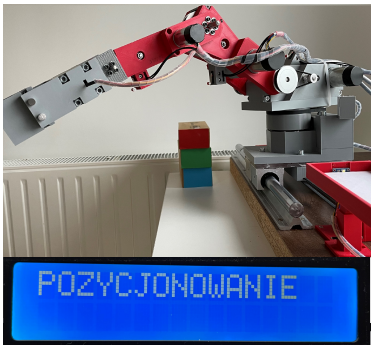
Chwytak znajdujący się na końcu ramienia składa się z pary szczęk, w których znajdują się łożyska liniowe. Dzięki temu możliwy jest ruch szczęk w płynny i równoległy sposób po wałkach. Do napędu zastosowane jest serwo. Rysunek 10.9 obrazuje konstrukcję.



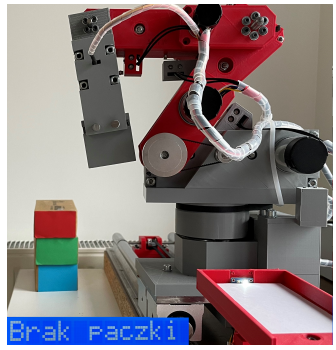
Rysunek 10.9. Chwytak [3]

4. Uzyskane wyniki

Najważniejsze etapy cyklu działania robota przedstawia rysunek 10.10. Po włączeniu zasilania odbywa się pozycjonowanie, czyli ruch wszystkich członów ramienia do pozycji skrajnych (do wyłączników krańcowych), co widać na rysunku 10.10a. Potem ramię obraca się, a podstawa przesuwa się do końca osi liniowej, co kończy proces bazowania. W kolejnym kroku odbywa się oczekiwanie na położenie paczki w uchwycie, sygnalizowane informacją „Brak paczki" na wyświetlaczu, co pokazuje rysunek 10.10b. Kiedy paczka zostanie położona, jej kolor jest identyfikowany na podstawie informacji z czujnika koloru. Następnie ramię wykonuje obrót i zostaje przesunięte na wysokość paczki. Rysunek 10.10c przedstawia chwytak, który obejmuje paczkę. W kolejnym etapie ramię z paczką podnosi się, wykonywany jest obrót i ruch robota w linii prostej na odległość, która zależy od koloru etykiety, a informacja o kolorze jest widoczna na wyświetlaczu (C – czerwony, Z – zielony, N – niebieski), co prezentuje rysunek 10.10d. Po opuszczeniu ramienia (rysunek 10.10e) chwytak zostaje otwarty. Następnie ramię zostaje podniesione (rysunek 10.10f) i robot wraca do pozycji oczekiwania na umieszczenie paczki w uchwycie, co pokazuje rysunek 10.10b. Po położeniu kolejnego kartonika cykl zostaje powtórzony (bez pozycjonowania) [3].



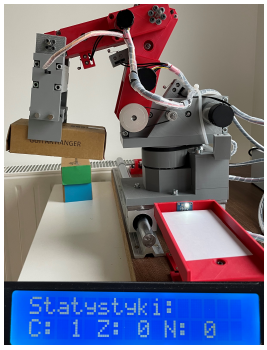
(a) pozycjonowanie



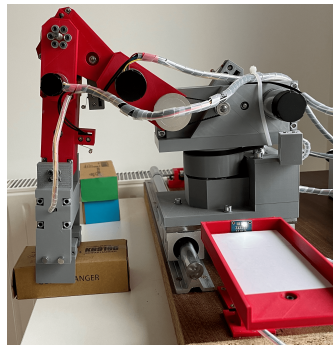
(b) oczekiwanie na paczkę



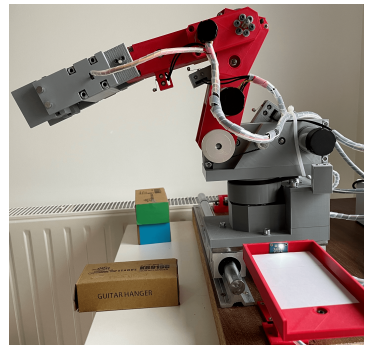
(c) zamknięcie chwytaka



(d) przeniesienie paczki nad miejsce docelowe



(e) położenie paczki



(f) uniesienie chwytaka

Rysunek 10.10. Prezentacja działania

5. Podsumowanie

Zrealizowana praca inżynierska pod tytułem „Robot z funkcja przenoszenia paczek", która wymagała połączenia wiedzy z zakresu modelowania, elektroniki, mechaniki i programowania, unaoczniła interdyscyplinarność robotyki.

Najbardziej czasochłonne okazało się zaprojektowanie i wydrukowanie części, w szczególności tych, w których osadzone zostały łożyska, ze względu na wymagany naciąg pasków przenoszących stosunkowo duże obciążenia. Ze względu na konieczność zapewnienia dużej powtarzalności i dokładności ruchu ramienia nieobojętny był odpowiedni dobór przekładni, która z jednej strony powinna zapewnić duże przełożenie, a z drugiej jak najmniejsze luzy. Dzięki zastosowaniu w silnikach enkoderów uzyskano bardzo dobrą dokładność i powtarzalność ruchów. Nie mniej istotne było również takie ułożenie przewodów, aby nie ograniczały zakresu pracy robota. Warto również zwrócić uwagę na chwytak, który zaprojektowany został w taki sposób, aby trzymane paczki nie przesuwały się i nie wypadały w czasie przenoszenia [3].

Zrealizowana praca spełnia założony cel i zakres, jakim było zaprojektowanie i zbudowanie robota z funkcją przenoszenia paczek. W przyszłośc w celu lepszego dopasowania siły, z jaką zaciskane są szczęki chwytaka, można dodać pomiar prądu pobieranego przez serwo [3].

Bibliografia

[1] Kardaś M., *Mikrokontrolery avr, Język C – podstawy programowania*, wydanie pierwsze, Wydawnictwo Atnel, Szczecin, Polska, 2011.

[2] Sibigtroth James M., *Zrozumieć małe mikrokontrolery*, wydanie pierwsze, Wydawnictwo BTC, Warszawa, Polska, 2003.

[3] Piaskowski S., *Budowa robota z funkcją przenoszenia paczek.* Praca inżynierska, Politechnika Wrocławska, Wrocław, Polska, 2021.

Budowa i sterowanie modelem magnetycznej lewitacji z wykorzystaniem sterownika X20 firmy B&R

inż. Michał Szymecki
Automatyka i Robotyka

dr hab. inż. Piotr Serkies
Katedra Maszyn, Napędów i Pomiarów Elektrycznych

Zabawa magnesami potrafi wciągnąć nawet dorosłego. Założyć można, iż wielu czytelników poniższego artykułu, w dzieciństwie, bawiąc się magnesami, próbowało znaleźć pozycję, w której to magnes mógłby leżeć (bądź też wisieć) nad innym magnesem w sposób stabilny. Nikt się nie przejmował (bo zapewne jeszcze nie słyszał), że już w latach 40. XIX wieku niejaki Samuel Earnshaw powiedział, że tak się nie da. Cel był prosty, magnes ma wisieć nad magnesem stabilnie i koniec. Poniższy artykuł pokaże jak, z wykorzystaniem wielu interesujących narzędzi i rozwiązań, można sprawić, by metalowy obiekt stabilnie zawisł, a właściwie, zaczął lewitować w polu magnetycznym. Omówiony zostanie proces projektowania modelu w formie cyfrowej, bez którego niemożliwe byłoby powstanie części wydrukowanych na drukarce 3D. Przybliżony zostanie czytelnikowi proces tworzenia oprogramowania oraz wytłumaczone, jak z diody i fotorezystora za dosłownie kilka złotych może powstać czujnik potrafiący kontrolować prąd w układzie z precyzją, umożliwiającą utrzymanie ciężkiej, stalowej śruby w powietrzu, w całkowitym bezruchu. Wszystko oparte na sterowniku PLC, bez wykorzystania żadnego z języków programowania sterowników tego typu.

1. Cel i założenia

Celem pracy inżynierskiej było zaprojektowanie i zbudowanie działającego modelu laboratoryjnego, pozwalającego na uzyskanie zjawiska lewitacji magnetycznej. Model oparty miał zostać na sterowniku PLC B&R X20. W zakres prac wchodziło:

- zapoznanie się z zagadnieniem lewitacji magnetycznej,
- zapoznanie się z budową i programowaniem sterowników rodziny X20 firmy B&R,
- opracowanie sposobu wykrywania położenia lewitującego obiektu,
- opracowanie koncepcji układu regulacji i jego implementacji w modelu,
- przeprowadzenie badań eksperymentalnych,
- redakcja pracy.

2. Wstęp teoretyczny

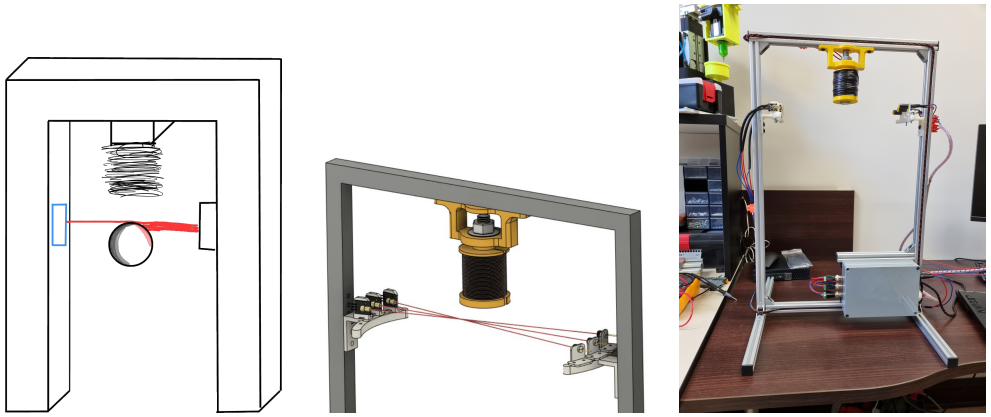
W celu pełnego zrozumienia tekstu przez czytelnika nie zaznajomionego z zagadnieniami z zakresu automatyki czy elektryki poniżej zamieszczone zostały definicje kluczowych pojęć.

1. Sterownik PLC – komputer przemysłowy, wyspecjalizowany do sterowania procesami oraz urządzeniami. Z racji swojej uniwersalności i prostoty w obsłudze znalazły wykorzystywane w bardzo wielu miejscach i dziedzinach życia od sterowania urządzeniami w najprostszych systemach Smart Home po zarządzanie całymi halami produkcyjnymi.
2. Regulator PID – regulator jest urządzeniem lub też elementem programowym, powodującym powrót przebiegu sygnału regulowanego do wartości zadanej po pojawieniu się różnicy między wartością zadaną a wartością sygnału regulowanego [1]. Ze względu na właściwości dynamiczne, regulatory można podzielić na cztery grupy:
   - Regulator proporcjonalny P,
   - Regulator proporcjonalno-całkujący PI,
   - Regulator proporcjonalno-różniczkujący PD,
   - Regulator proporcjonalno-całkująco-różniczkujący PID.
3. Sygnał sterujący jest sumą sygnałów z poszczególnych członów, które po odpowiednim wzmocnieniu lub osłabieniu tworzą regulator dostrojony pod konkretne zastosowanie. To co kluczowe w zastosowaniu regulatorów, to istnienie pętli sprzężenia zwrotnego, w której dochodzi do porównania sygnału zadanego z aktualnym sygnałem na wyjściu układu. Różnica między tymi wartościami to właśnie uchyb.

3. Projekt i wykonanie

Powstawanie modelu można podzielić na kilka etapów: projektowania, konstruowania i programowania. Wcześniejsze stworzenie modelu w wersji cyfrowej pozwoliło uniknąć wielu poprawek i modyfikacji. W projektowaniu modelu niezastąpiony okazał się program Fusion 360, w którym powstał cały cyfrowy odpowiednik. Zbudowanie modelu o takim poziomie dopracowania możliwe było dzięki uczelnianej drukarce 3D. Etapy realizacji projektu przedstawione zostały na rysunku 11.1. W skład części „własnej produkcji" wchodziły:

- karkas,
- mocowanie karkasu do profilu aluminiowego,
- podkładki między karkas a mocowanie,
- uchwyt do śruby z „chwytaczem",
- śruba z „chwytaczem",
- łapy na lasery i fotorezystory,
- uchwyty kątowe do laserów i fotorezystorów,
- tulejki ograniczające światło padające na fotorezystory,
- dysza do chłodzenia magnesu.



(a) wstępna koncepcja **(b)** model 3D **(c)** złożona makieta

Rysunek 11.1. Projekt

Rama skonstruowana została z profili aluminiowych V-Slot o szerokości 20 mm. Dzięki odpowiedniemu zaprojektowaniu wszystkich części model udało się złożyć w kilka godzin, licząc razem z cięciem profili oraz ręcznym nawinięciem 70 m drutu nawojowego o średnicy 1 mm na karkas, co dało ok 338 zwojów. Tak utworzony magnes cechował się wystarczająco dużą siłą, by zasilany prądem ok 5 A, z odległości ok 20 mm przyciągnąć stalową śrubę M12 lub kulkę ze starej myszki komputerowej. Wymieniona wyżej śruba z chwytaczem to specjalnie dorobiona śruba z łbem w kształcie stożka, pokazana na rysunku 11.2. Łapała ona kulkę puszczaną przez magnes.



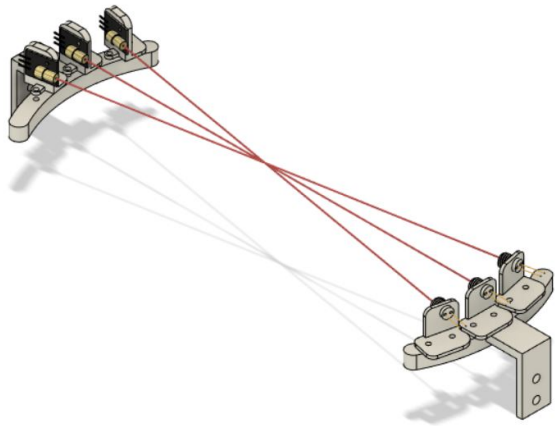
Rysunek 11.2. Chwytacz zastępujący człowieka w łapaniu i podawaniu kulki pod magnes

3.1. Koncepcja czujnika położenia obiektu

Mimo dość wysokich wymagań co do dokładności czujnika mierzącego położenie elementu w polu magnetycznym, jego wykonanie było proste, a części niedrogie. Pomiar położenia kulki opiera się na przesłanianiu przez nią części światła. Światło to pochodzi od diody laserowej, a ilość światła mierzona jest przez leżący naprzeciw diody fotorezystor. Projekt czujnika przedstawiony jest na rysunku 11.3.

W całym czujniku wykorzystane zostały trzy pary dioda–fotorezystor. Każda para położona została na innej wysokości. Taka redundancja pozwoliła na szybkie dodanie do programu części wyłączającej zasilanie magnesu przy braku elementu pod nim. Za to zabezpieczenie odpowiadała dolna para. Górna para wykorzystywana była przy kalibrowaniu regulatora PID – wyłączała zasilanie magnesu, gdy kulka przyciągnięta została całkowicie do magnesu (czyli całkowicie zasłoniła światło diody). Skrajne pary działają w sposób binarny, podczas gdy tylko środkowa para jest „analogowa". Zmiana ilości światła przez kulkę wiązki lasera powodowała zmianę rezystancji na fotorezystorze, a w konsekwencji zmianę napięcia na nim, co też odczytywał sterownik.

Fotorezystory nie są czułe wyłącznie na światło lasera. Duży wpływ na działanie układu miało światło otoczenia – do wytrącenia układy ze stanu stabilności czasem wystarczyło machnięcie ręką w pobliżu czujnika. Rozwiązaniem problemu okazały się krótkie tulejki osłaniające fotorezystory. Ograniczyły one padające na czujniki światło, co w konsekwencji poprawiło stabilność układu i powtarzalność wyników prób.



Rysunek 11.3. Model 3D czujnika położenia obiektu

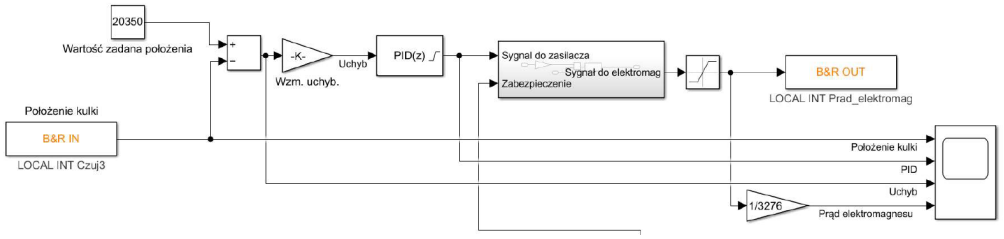
3.2. Połączenia elektryczne

Elektromagnes zasilany jest zasilaczem B&R. Diody oraz fotorezystory wymagały osobnego zasilania – odpowiednio 5 V oraz 10 V. Napięcie 24 V DC pochodzące od zasilacza, dwoma przetwornicami step-down LM2596 zostało dopasowane do wymaganego. Fotorezystory połączone zostały z rezystorami, tworząc w ten sposób dzielniki napięcia. Napięcie na fotorezystorach było sygnałem odczytywanym przez sterownik PLC, a następnie przeliczane na wysokość, na jakiej lewitował przedmiot. Elektromagnes zasilany jest prądem pochodzącym bezpośrednio z modułu MM4456, sterowanym poprzez program.

3.3. Software i sterowanie

Model miał podlegać kontroli ze strony sterownika PLC B&R X20 oraz dodatkowe moduły komunikacyjne, a także moduł zasilania odpowiedzialny za zasilanie elektromagnesu. Dzięki wtyczce Automation Studio Target for Simulink program powstał w całości właśnie

w programie Simulink, skąd został przekonwertowany do środowiska programistycznego Automation Studio, a finalnie na sam sterownik. Dzięki takiemu wykorzystaniu pakietu Simulink, projektowanie układu opartego na regulatorze PID było bardzo łatwe i dawało ogromne możliwości w kwestii szybkiego modyfikowania i konfigurowania programu. Zamiast pisania programu w typowym dla sterowników języku LAD lub ST program został utworzony na podstawie stworzonego w Simulink modelu obliczającego wartości wyjścia na podstawie wartości wejścia. Jako wejście rozumiane jest tu napięcie z fotorezystora w czujniku, natomiast wyjściem jest prąd zasilający elektromagnes. Na rysunku 11.4, poza blokiem regulatora PID zauważyć można także blok odbierający sygnał z zabezpieczeń (o tym pod rysunkiem 11.4), blok ograniczający prąd do ok 5,8 A oraz programowy oscyloskop do podglądania przebiegów.



Rysunek 11.4. Fragment układu sterowania z regulatorem PID

Poza wspomnianym regulatorem PID widocznym na rysunku 11.4 układ został wyposażony w system wyłączania magnesu przy braku elementu pod magnesem oraz część odpowiednio ograniczającą płynący przez elektromagnes prąd (niewidoczne na rysunku). Po zbudowaniu modelu i kilkugodzinnym strojeniu regulatora na jaw wyszły ograniczenia wynikające z materiału użytego do produkcji karkasu. Użyty do druku ABS pod wpływem temperatury wydzielanej przez uzwojenie robił się plastyczny. W obawie o możliwe rozklejanie się warstw w wydrukowanych częściach powstały właśnie ww. systemy wyłączające/ograniczające prąd w magnesie oraz system chłodzenia wentylatorem. Na potwierdzenie słuszności obaw o wysoką temperaturę wykonany został pomiar temperatury wewnątrz karkasu po kilkugodzinnym strojeniu i kalibrowaniu, widoczny na rysunku 11.5.



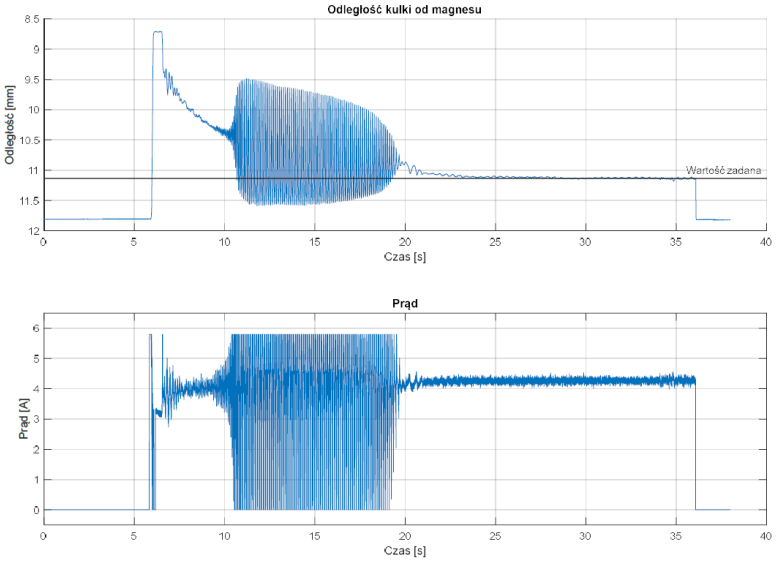
Rysunek 11.5. Zdjęcie multimetru z wynikiem pomiaru temperatury po kilkugodzinnej pracy

4. Uzyskane wyniki

Szukanie nastaw regulatora było w pełni doświadczalnym procesem. Mimo istnienia metod dobierania nastaw w tym przypadku postawiono na szukanie nastaw poprzez obserwację i próby. Był to jeden z bardziej czasochłonnych etapów całego projektu.

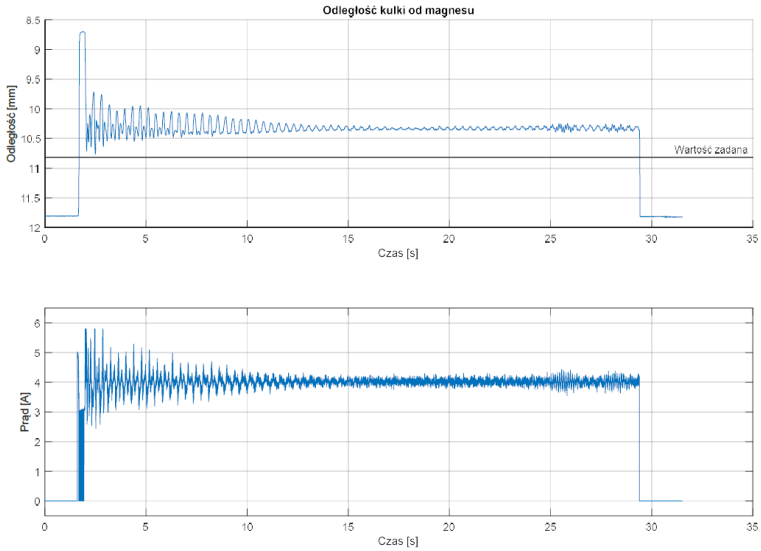
W trakcie badań działania modelu przetestowane zostały dwa rodzaje regulatorów: regulator PID oraz regulator PD. W działaniu tych dwóch modeli można było szybko zauważyć wyraźne różnice, a każdy z nich cechował się wadami oraz zaletami.

Zastosowanie regulatora PID zapewniało lewitację kulki na określonej wcześniej wysokości, drobne drgania podłoża nie wpływały na stabilność układu, nie występował również jakikolwiek limit działania układu – raz podwieszona kulka stabilnie utrzymywała pozycję przez dowolny okres czasu. Jako wyraźną wadę wymienić należy trudność w samym umieszczaniu kulki w polu magnetycznym. Aby magnes „złapał" kulkę, musiała ona zostać ręcznie umiejscowiona w miejscu bliskim ustalonemu i tylko niewielka różnica w położeniu była po czasie niwelowana przez człon całkujący, co można zauważyć na rysunku 11.6.



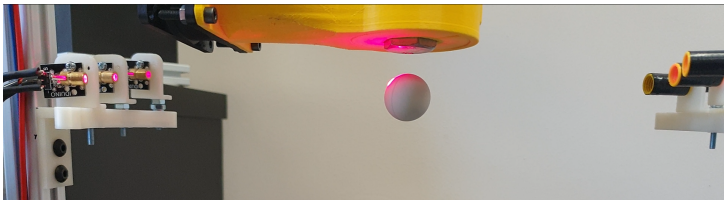
Rysunek 11.6. Wykres położenia kulki i zmian prądu w czasie dla regulatora PID

Działanie układu wyposażonego w regulator PD było znacznie uproszczone. Po podłożeniu kuli pod magnes była ona szybko chwytana i wisiała stabilnie mimo różnicy w wysokości aktualnej i zadanej. Próba pociągnięcia kulki w dół dawała uczucie rozciągania sprężyny – im dalej od magnesu, tym mocniej starał się on przyciągnąć kulkę do siebie, oczywiście w niewielkimi zakresie ruchu. Taki układ sterowania cechował się znacznie wyższą odpornością na wstrząsy oraz łatwością w obsłudze. Różnicę w wysokości osiąganej przez kulkę i wysokości zadanej widać dobrze na rysunku 11.7.



Rysunek 11.7. Wykres położenia kulki i zmian prądu w czasie dla regulatora PD

Widok stalowej kulki nieruchomo zawieszonej w przestrzeni sprawia wrażenie sztuczki magicznej, ponieważ człowiek nie jest przyzwyczajony do widoku przedmiotów unoszących się w powietrzu w kompletnej ciszy i bezruchu. Założenia projektowe należy uznać za spełnione. Niestety, rysunek 11.8 nie oddaje w pełni efektu, jaki sprawia lewitująca kulka widziana na żywo.



Rysunek 11.8. Kulka zawieszona w polu magnetycznym

5. Podsumowanie

Wykorzystanie sterownika PLC jako „mózgu" modelu nie tylko dało ogromne pole do manewru w kontekście budowania i programowania układu, ale też pozwoliło uniknąć nadmiarowej nauki, jak np. poznawanie języka C czy Python, gdyby to zamiast sterownika wykorzystane miało zostać Arduino czy Raspberry Pi. Zastosowany w modelu regulator PID czy PD nie są jedynymi słusznymi sposobami na kontrolowanie modelu lewitacji. W pewien sposób na drodze przypadku zanim regulator PID został odpowiednio wyregulowany, udało się doprowadzić model do stanu, w którym kulka utrzymywana była w powietrzu poprzez „łapanie" i „puszczanie" jej z dużą częstotliwością (prąd o wartości 0 lub ok 5 A – bez stanów pośrednich). Nie był to docelowy, ani zbyt elegancki tryb pracy modelu, drgania kulki były bardzo intensywne. Taki sposób pracy można by w skrócie

opisać jako „włącz magnes, gdy kulka jest za nisko, wyłącz gdy za wysoko". Mimo to, magnes był w stanie utrzymać kulkę w stanie lewitacji dowolnie długi czas.

Ciekawe efekty mogłoby przynieść zastosowanie sieci neuronowej czy też algorytmów genetycznych w celu dobrania optymalnych nastaw regulatora PID, na co model jest gotowy. Przy odpowiedniej modyfikacji samej bramy laserowej zapewne możliwe byłoby poszerzenie zakresu możliwych położeń kulki czy też wprawienie jej w ruch oscylacyjny wokół konkretnego punktu. Zbudowany model może w przyszłości posłużyć do badań nad innymi sposobami sterowania układami wymagającymi szybkich reakcji i precyzji, a możliwość jego modyfikacji jest bardzo duża. Mimo konkretnego zamysłu, według którego projekt był realizowany, często z trudem przychodziło konsekwentne trzymanie się go, by nie porzucić aktualnie testowanej koncepcji na rzecz innej, ciekawszej, jeszcze bardziej zmyślnej, ulepszonej w jeszcze jeden blok całkowania, ograniczania czy mnożenia, z możliwością przełączania trybów pracy. Poza satysfakcją z efektu i przyjemności z samego budowania modelu realizacja projektu dała również inne korzyści, jak jeszcze lepsze zrozumienie sposobu działania regulatora PID oraz opanowanie metod doboru jego nastaw w sposób empiryczny. Wbrew pozorom nie jest to tak trudna sztuka, a tak szybko odpowiadający na modyfikacje model świetnie nadaje się do badań nad regulatorami.

Bibliografia

[1] Rumatowski K., *Podstawy regulacji automatycznej*, Wydawnictwo Politechniki Poznańskiej, Poznań 2008.

Wydział Informatyki
i Telekomunikacji

Design of a TinyML system used to analyze if the face mask is worn properly

inż. Dominik Piątkowski
Informatyka

prof. dr hab. inż. Krzysztof Walkowiak
Katedra Systemów i Sieci Komputerowych

As the COVID-19 pandemic emerged, everyone's attention was brought to the topic of health and safety of the entire human population. It has been proven that wearing a face mask can help limit the spread of the virus. Despite the enormous efforts of people around the world, there still exists a group of people that wear the face mask incorrectly. In order to provide the best level of safety for everyone, face masks must be worn correctly, especially indoors, e.g., in shops, cinemas and theaters. As the security guards can only handle a limited area of the frequently visited objects, intelligent sensors can be used. In order to mount them on the shelves in the shops or near the cinema cash register queues, they need to be capable of battery operation. This restricts the sensor to be as energy-efficient as possible, in order to prolong the battery life of such device. The cost is also a factor, as cheaper device will result in higher accessibility. An interesting and quite novel approach that can answer all these challenges is a TinyML system, that can be defined as a combination of two concepts: Machine Learning (ML) and Internet of Things (IoT). The TinyML approach enables the usage of ML algorithms on boards equipped with low-cost, low-power microcontrollers without sacrificing the classifier quality. The main goal of this thesis is to propose a battery-operated TinyML system that can be used for verification whether the face mask is worn properly. To this end, we carefully analyze several ML approaches to find the best method for the considered task. After detailed analysis of computation and memory complexity as well as after some preliminary experiments, we propose to apply the K-means algorithm with carefully designed filters and sliding window technique, since this method provides high accuracy with the required energy-efficiency for the considered classification problem related to verification of using the face mask. The STM32F411 chip is selected as the best microcontroller for the considered task. Next, we perform wide experiments to verify the proposed ML framework implemented in the selected hardware platform. The obtained results show that the developed ML-system offers satisfactory performance in terms of high accuracy and lower power consumption. It should be underlined that the low-power aspect makes it possible to install the proposed

system in places without the access to power, as well as reducing the carbon footprint of AI-focused industry which is not negligible. Our proposed TinyML system solution is able to deliver very high quality metric values with accuracy, True Positive Ratio (TPR), True Negative Ratio (TNR), precision and recall being over 96% for masked face classification while being able to reach up to 145 days of uptime using a typical 18650 battery with capacity of 2500 mAh and nominal voltage of 3.7 V. The results are obtained using a STM32F411 microcontroller with 100 MHz ARM Cortex M4 which proves that execution of complex computer vision tasks is possible on such low-power devices. It should be noted that the STM32F411 microcontroller draws only 33 mW during the operation.

1. Thesis goal and assumptions

Wearing a face mask is one of the methods of preventing spread of COVID-19 caused by SARS-CoV-2 virus. In order to make the face mask effective, it needs to be worn correctly – it is mandatory to cover both nose and chin. Main goal of this paper is to design a TinyML system based on microcontroller that can be used to analyze if the face mask is worn correctly. Designed system will utilize machine learning algorithms. An important system operation criterion will be minimizing energy consumption. Main design assumptions are as follows:

- collect and process literature and materials regarding the subject of the work,
- determine system requirements,
- create system design,
- conduct computer experiments in order to pick the best algorithm for determining if the face mask is worn correctly,
- prepare system documentation,
- redact bachelor thesis,
- formulate final conclusions.

2. Introduction

Machine Learning (ML) is one of the methods used in the field of Artificial Intelligence (AI). In recent years, the computational complexity of widely used algorithms has been growing exponentially. This is caused by the combination of increasingly faster processor speeds, presence of Big Data and progressively powerful graphics processing units. On one hand, allocating more resources enables training and using state-of-the-art models capable of achieving results similar to humans, or even surpassing them. But on the other hand, training and using such models requires more and more advanced hardware and enormous amounts of electricity. This poses a handful of problems – and one of them is the carbon footprint of AI-focused industry which is not negligible. The popularization of cloud computing is not helping either, as it additionally strains the network infrastructure. One of the solutions to these problems is TinyML – a combination of ML and embedded Internet of Things (IoT) devices.

3. Materials and methods

One of the most important factors is the power consumption that must be kept low in order to create battery-powered solutions. TinyML platforms taken into consideration are shown in Table 12.1.

Table 12.1. Platform comparison

Microcontroler	Price	Speed	Power consumption	Memory size
STM32F103CB	$7.17	72 MHz ARM Cortex M3 without FPU	32 mA @ 72 MHz	128 KB Flash 20 KB RAM
STM32F411CE	$7.83	100 MHz ARM Cortex M4 with FPU	10 mA @ 100 MHz	512 KB Flash 128 KB RAM
STM32H743VI	$15.91	480 MHz ARM Cortex M7 with FPU	132 mA @ 480 MHz	2 MB Flash 1 MB RAM

First two microcontrollers are widely available in the form of small evaluation boards, popular among tinkerers, called Bluepill for STM32F103 and Blackpill for STM32F411. The boards are shown in Figures 12.1 and 12.2. The third microcontroller is not as popular as two previous ones, due to the lack of availability of small form factor board incorporating the STM32H743 chip, although it can be found in expensive, full-sized development boards. Considering all the factors, the STM32F411 chip was picked as the best suitable microcontroller for this task. It is relatively cheap, widely available as a Blackpill evaluation board, fast enough for computational tasks while being the most energy-efficient option out of all candidates. For the task of detecting a face, a sliding window technique combined with K-means classifier was picked, due to the fact that this approach is lightweight enough to fit into the memory of the microcontroller unlike most of the sophisticated models, executes reasonably fast and can be further tuned in order to achieve either better results or higher speed. An example of the sliding window output is shown in figure 12.3.

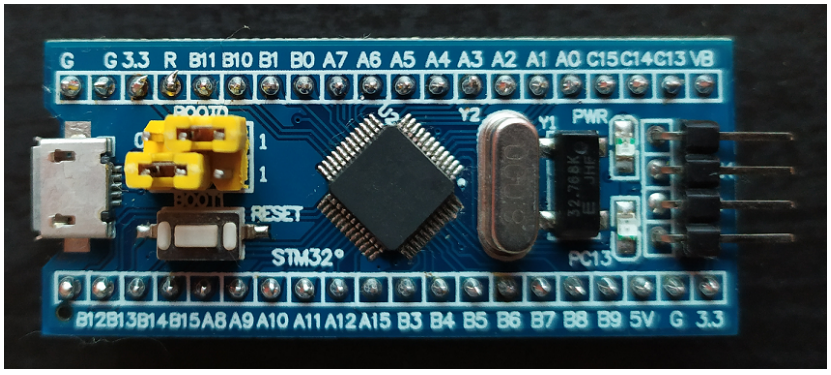


Figure 12.1. Bluepill board with STM32F103CB microcontroller

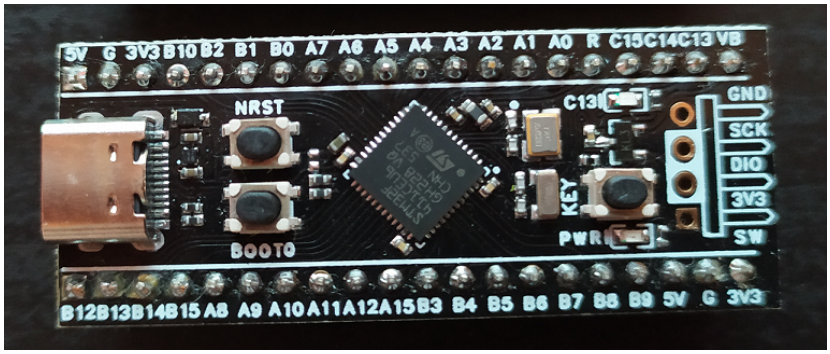


Figure 12.2. Blackpill board with STM32F411CE microcontroller

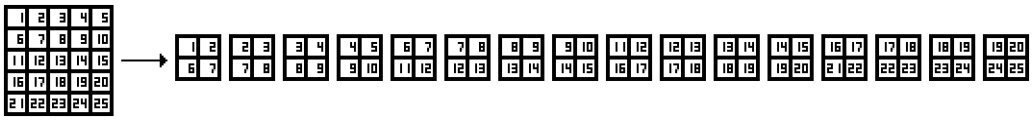


Figure 12.3. Example of the sliding window of size 2×2 with stride value of 1

Data for the classification algorithm consist of preprocessed images from the datasets. The first step is the filtering, which consists of following steps:

- Downscaling the image containing a face to 32×32 px
- Calculating the intermediate color value using the procedure:

```
value = max(R, G, B)
```

- As most of the face masks are cyan, the intermediate color value is changed according to the procedure:

```
if (B > G && G > R && R < 200) value += 50
else value = 0
if (value > 255) value = 255
```

- Applying the threshold filter to the intermediate values with the boundary value of 240 with the procedure:

```
if (value < 240) value = 0
else value = 255
```

The downscaling size of 32×32 px was chosen as it was the smallest size that did not impact accuracy. The example of data preprocessing is shown in Figure 12.4.

The sliding window is used in combination with K-means classifier to detect a face in the image. The image fragment limited by the window is downscaled to 32×32 px in order to pass it to the classifier after applying the grayscale filter. The most computational-heavy task in this method is the downscaling, which must be measured and taken into consideration. Another procedure to measure is the grayscale filter. The sliding window technique measurement results are shown in Table 12.2.

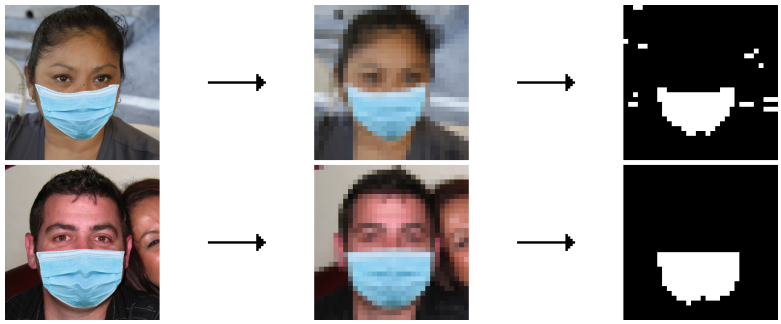


Figure 12.4. Example of the image filtering of the MaskedFace-Net dataset

Table 12.2. Sliding window measurements

Window size	Stride	Window positions	Stride:window	Downscaling time
64×64	16	999	25%	6.920 ms
96×96	24	391	25%	12.95 ms
128×128	32	204	25%	21.30 ms
160×160	40	117	25%	32.00 ms
192×192	48	70	25%	44.96 ms
224×224	48	54	21.43%	60.24 ms
256×256	48	45	18.75%	77.90 ms
288×288	48	40	16.67%	97.80 ms
320×320	48	28	15%	120.0 ms
352×352	48	21	13.64%	144.6 ms
384×384	48	18	12.5%	171.6 ms
416×416	48	10	11.54%	200.8 ms
448×448	48	5	10.71%	232.4 ms
480×480	48	4	10%	266.2 ms

4. Results

The K-means masked face classifier was trained and tested with different parameters and filtering methods in order to create a model with the best results possible. There are 14 tuning approaches with 3 varying parameters: number of centroids, filtering method and if the subset of the masked face set was used. The amount of centroids that was tested is in the range 2–4 due to the existence of 2 major dataset classes (correctly masked face and incorrectly masked face) and 4 total dataset subclasses (correctly masked face, incorrectly masked face – mouth and nose uncovered, incorrectly masked face – nose uncovered, incorrectly masked face - chin uncovered) in order to check which value gives the best results. The 4 types of filtering were tested. The comparison of all proposed tuning approaches is shown in Figure 12.5.

The metrics vary significantly between the tested approaches. The obtained results for the chosen approach (number 4) are of very high quality. The approach number 1, 2, 3, 6 and 7 use a subset consisting only of the images where masks are worn on the chin and a similar amount of randomly picked images where masks are worn correctly. This results in a subset of size 12 226 instead of the complete dataset of 137 013 images.

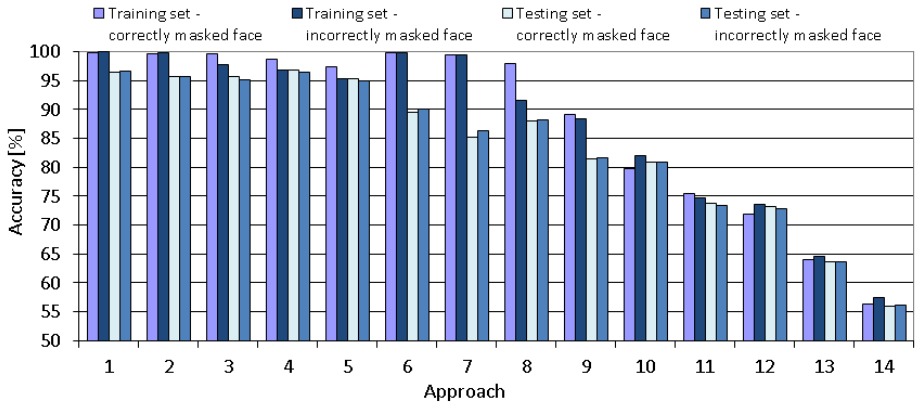


Figure 12.5. Comparison of the accuracy of the K-means classifier vs. parameters and filtering methods

The subset was used to discover the results difference between "easy data" consisting of only two dataset subclasses (correctly masked face and incorrectly masked face – mouth and nose uncovered; the subclasses are balanced in size) and full dataset. The results of all analyzed 14 approaches were validated using the Analysis of Variance (ANOVA) test. For each analyzed approach, the bootstrapping method was applied to obtain 30 results of the accuracy metric. According to the reported values, the null hypothesis is rejected, i.e., there is sufficient evidence to conclude that not all of the means of analyzed 14 approaches are equal and that there are statistical differences between tested approaches. Based on the obtained results, the model from approach 4 was chosen as the best masked face classifier due to the fact that this model performed the best in the test results with accurately classifying the correctly masked face, while being able to achieve almost-best results for accurately classifying the incorrectly masked face, even if only the nose was uncovered. To verify performance of approach 4, we made the t-test comparing approach 4 against other approaches. In all cases, the obtained p-value is <0.00001, what means that the differences between results of approach 4 and results of other approaches are significant. Therefore, in the reminder of the thesis, we present results obtained using approach 4. The masked face classification accuracy results are shown in Table 12.3.

Table 12.3. Masked face classification metrics

Metric	Value
Accuracy of the correct decition of the correctly masked face	96.86%
Accuracy of the correct decition of the incorrectly masked face	96.40%
TPR	96.58%
TNR	96.41%
Precision	96.40%
Recall	96.58%

As the time required to detect a face in image depends on the chosen window size, the results are shown in Table 12.4. The STM32F411 microcontroller is powered by 3.3 VDC

voltage and consumes around 10 mA of current. Time required to classify a masked face is equal to 9.75 ms. Considering the calculated amount of energy required for one face mask operation, the battery life can be estimated. The 18 650 battery with typical capacity of 2 500 mAh and nominal voltage of 3.7 V will be used for calculations. As there are high-efficiency power supply chips available with several µA of quiescent current and over 90% efficiency, the energy loss can be pessimistically modeled as 10%. The singular face mask operation energy consumption results are shown in Table 12.4.

Table 12.4. Face detection time calculation results

Window size	Total downscaling time	Total K-means time	Total detection time
320×320	3360.00 ms	273.0 ms	3633.00 ms
352×352	3036.60 ms	204.8 ms	3241.35 ms
384×384	3088.80 ms	175.5 ms	3264.30 ms
416×416	2008.00 ms	97.5 ms	2105.50 ms
448×448	1162.00 ms	48.8 ms	1210.75 ms
480×480	1064.80 ms	39.0 ms	1103.90 ms

5. Conclusions

Considering the power consumption of the microcontroller, the system could continuously operate for 225 hours. With the example of shop visited by 5000 customers every day, the proposed system uptime without recharging the battery is shown in Table 12.5. This is especially important, as our low-power design makes the mounting of the proposed system wherever it is needed possible, regardless of mains power presence. In the example with a shop, it would be inconvienient to install the mains power cables on market shelves, where the proposed system could be mounted.

Table 12.5. System uptime results

Window size	Stsren uptime
320×320	44.47 days
352×352	49.83 days
384×384	49.48 days
416×416	76.59 days
448×448	132.73 days
480×480	145.48 days

According to the measurements and results shown in this thesis, we have shown that it is possible to create an energy-efficient battery-operated system capable of executing complex computer vision tasks in reasonable time with low amount of available resources by using carefully selected methods and algorithms. This approach can help to reduce the energy usage of the intelligent sensor grids, as well as to decrease the network traffic by replacing the cloud computing where it is worthwhile.

References

[1] Patterson, D., Gonzalez J., Le Q., Liang C., Munguía L.M., Rothchild D., So D., Texier M., Dean J., *Carbon Emissions and Large Neural Network Training*, **2021**, arXiv preprint.

[2] *Machine learning course.* Available online: `https://www.coursera.org/learn/machine-learning` (access: 04-04-2021).

[3] Gowda M., Gowda J., Iyer S., Pawar M., Gaikwad V., *Power Consumption Optimization in IoT based Wireless Sensor Node Using ESP8266*, ITM Web of Conferences **2020**, 32.

[4] Loey M., Manogaran G., Taha M., Khalifa N E., *Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection*, Sustainable Cities and Society **2020**, 65.

[5] Lim H., Ryoo S., Jung H., *Face-Mask Detection with Micro processor*, Journal of the Korea Institute of Information and Communication Engineering **2021**, 25.3, pp. 490–493.

[6] Ng A., *Clustering with the k-means algorithm*, Machine Learning **2012**.

[7] Ng A., *Advice for applying machine learning*, Machine Learning **2011**.

[8] Cabani A., Hammoudi K., Benhabiles H., Melkemi M., *MaskedFace-Net – A dataset of correctly-/incorrectly masked face images in the context of COVID-19*, Smart Health **2020**, 19.

[9] Hammoudi K., Cabani A., Benhabiles H., Melkemi M., *Validating the CorrectWearing of Protection Mask by Taking a Selfie: Design of a Mobile Application "CheckYourMask" to Limit the Spread of COVID-19*, Computer Modeling in Engineering & Sciences **2020**, 124, pp. 1049–1059.

Tożsamość cyfrowa w Polsce: ekosystem, stopień wdrożenia i zastosowanie dowodów osobistych z warstwą elektroniczną

inż. Damian Rybak
Cyberbezpieczeństwo

dr inż. Wojciech Wodo
Katedra Podstaw Informatyki

Powszechne wykorzystanie cyfrowej tożsamości niesie ze sobą liczne korzyści pozwalające na uproszczenie i przyspieszenie wielu procesów przy jednoczesnym zachowaniu wysokiego poziomu bezpieczeństwa. Niestety zagadnienia związane z tą tematyką nie są trywialne, zwykle posiadają złożoną naturę, na którą ma wpływ wiele czynników. Zachodzące zmiany w świecie cyfrowym są dynamiczne, a czasami nawet chaotyczne. Dodatkowo temat tożsamości cyfrowej nie jest tematem popularnym podczas debat publicznych. Z tych względów zagadnienia dotyczące eID dla większości społeczeństwa mogą jawić się jako coś trudnego i niedostępnego. Z tego powodu postanowiono podjąć tę tematykę i dokonać harmonizacji obecnego stanu wiedzy dotyczącego zagadnień związanych z szeroko pojętą tożsamością cyfrową w Polsce. Przedstawienie ekosystemu eID wymaga podejścia interdyscyplinarnego. Analiza rozpoczyna się od zrozumienia fundamentów, jakimi są akty prawne. Później na tej podstawie możliwe jest zaprezentowanie podstawowych aspektów technicznych, które umożliwiają realizację założeń przedstawionych w ustawach. Następnie na bazie zebranej wiedzy możliwe jest przedstawienie i zrozumienie obecnego stanu rozwoju cyfrowej tożsamości. Nie bez znaczenia pozostaje kwesta nowych e-dowodów, które odgrywają bardzo istotną rolę w całym ekosystemie eID.

1. Cel i założenia

Obecnie jesteśmy świadkami ogromnych przemian zachodzących w sferze cyfrowej tożsamości. Niestety mimo faktu, że zagadnienia te dotyczą każdego obywatela, to z rozmaitych powodów temat ten nie jest obecny podczas debat publicznych. Z tego powodu postanowiono podjąć tę tematykę i dokonać harmonizacji obecnego stanu wiedzy dotyczącego zagadnień związanych z szeroko pojętą tożsamością cyfrową w Polsce, ze szczególnych uwzględnieniem dowodów osobistych z warstwą elektroniczną. Do głównych założeń projektowych realizowanej pracy można zaliczyć:

- usystematyzowanie wiedzy dotyczącej zagadnień związanych z tożsamością cyfrową w Polsce,
- przedstawienie obecnego ekosystemu eID ze szczególnym uwzględnieniem dowodów osobistych z warstwą elektroniczną,
- prezentację dostępnych w e-dowodach możliwości i sposobów weryfikacji tożsamości oraz autentyczności dokumentów,
- analizę dostępnych w Polsce rozwiązań wykorzystywanych do zarządzania tożsamością,
- zebranie informacji dotyczących technicznych rozwiązań, mechanizmów i protokołów obecnych w cyfrowej części dowodu.

2. Wstęp teoretyczny

Metody potwierdzania tożsamości znacznie zmieniały się na przestrzeni wieków. Pierwsze dokumenty przypominające swoją formą te wykorzystywane po dzień dzisiejszy zaczęto stosować na początku XX wieku. Obecnie dokumenty tożsamości są wykorzystywane do weryfikowania danych osobowych i potwierdzania tożsamości. Dokumenty tożsamości są powszechnie akceptowalne zarówno w administracji publicznej, jak i w podmiotach komercyjnych.

Wraz z rozwojem usług cyfrowych pojawiła się naturalna potrzeba opracowania rozwiązań umożliwiających wykorzystanie metod identyfikacji elektronicznej. W ten sposób powstały środki identyfikacji elektronicznej, które zawierają dane umożliwiające identyfikację osoby w świecie cyfrowym. Większość obecnie dostępnych form eID (ang. *electronic identification*) podczas jej tworzenia wymaga przeprowadzenia solidnej procedury weryfikacji tożsamości bazującej na tradycyjnych dokumentach tożsamości.

Bardzo ważną rolę w transakcjach zarówno tradycyjnych, jak i tych realizowanych drogą elektroniczną odgrywa podpis. Podpis cyfrowy jest odpowiednikiem podpisu tradycyjnego w świecie cyfrowym. Zgodnie z jednym z ważniejszych aktów prawnych dotyczącym ekosystemu eID w UE – rozporządzeniu eIDAS, (ang. *electronic IDentification, Authentication and trust Services*) podpis cyfrowy to dodatkowe dane w postaci elektronicznej, które są dołączone lub logicznie powiązanie z innymi danymi w postaci elektronicznej.

Do zawierania bezpiecznych transakcji elektronicznych wymagana jest obecność odpowiedniej usługi zaufania. Taka usługa musi być darzona zaufaniem przez oba podmioty biorące udział w transakcji. Usługa zaufania jest odpowiedzialna za tworzenie, weryfikację i walidację m.in. podpisów elektronicznych. Proces walidacji jest niczym innym jak procesem polegającym na weryfikacji i potwierdzaniu ważności podpisu elektronicznego. Bardzo dobrym źródłem wiedzy dotyczącym terminologii wykorzystywanej w zagadnieniach cyfrowej tożsamości jest Przewodnik „Identyfikacja i uwierzytelnienie w usługach elektronicznych" wydany przez Związek Banków Polskich [1].

3. Projekt i wykonanie

Zagadnienia dotyczące cyfrowej tożsamości posiadają złożoną naturę, na którą wpływa wiele aspektów natury prawnej, technologicznej i społecznej. Sytuację dodatkowo kom-

plikuje fakt, że informacje dotyczące tej problematyki są na ogół rozproszone w wielu miejscach, a ich scalenie bardzo często uwidacznia obszary, które były wcześniej pomijane. Z tego względu aby dobrze zrozumieć i przedstawić ekosystem eID, skorzystano z podejścia interdyscyplinarnego. Pracę rozpoczęto od analizy fundamentów, czyli norm prawnych regulujących kwestie e-tożsamości. Następnie skonfrontowano to z obecnym stanem ekosystemu eID w Polsce. Po przedstawieniu roli e-dowodów w ramach systemu identyfikacji elektronicznej skupiono się na zaprezentowaniu wiedzy dotyczącej samych dowodów osobistych. Później dokonano przeglądu narzędzi dostępnych do zarządzania tożsamością ze względu na fakt, że na popularyzację całego ekosystemu mają ogromy wpływ aplikacje wykorzystywane bezpośrednio przez użytkowników końcowych, a w tym przypadku są nimi obywatele.

3.1. Normy prawne regulujące kwestie e-tożsamości

Regulacje prawne są bardzo istotnym elementem mającym ogromny wpływ na rozwój cyfrowej tożsamości. To właśnie normy prawne określają pewne fundamenty przedstawiające ogólny zarys powstającego ekosystemu, wyodrębniają podstawowe podmioty jednocześnie określając ich prawa i obowiązki. Bardzo istotnym faktem wpływającym na analizę rozwoju e-tożsamości w Polsce jest jej członkostwo w Unii Europejskiej. Integracja jest najbardziej widoczna podczas analizy aktów prawnych, które od wielu lat są mocno powiązane z prawodawstwem unijnym. Do najważniejszych aktów prawnych regulacyjnych mających wpływ na kształtowanie systemu e-tożsamości w Polsce można zaliczyć:

- Ustawę z 6 sierpnia 2010 r. o dowodach osobistych [2],
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 910/2014 z 23 lipca 2014 r. w sprawie identyfikacji elektronicznej i usług zaufania w odniesieniu do transakcji elektronicznych na rynku wewnętrznym (eIDAS) [3],
- Ustawę z 5 września 2016 r. o usługach zaufania oraz identyfikacji elektronicznej [4],
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) z 20 czerwca 2019 r. w sprawie poprawy zabezpieczeń dowodów osobistych obywateli Unii i dokumentów pobytowych wydawanych obywatelom Unii i członkom ich rodzin korzystającym z prawa do swobodnego przemieszczania się [5].

Ustawa o dowodach osobistych jest podstawowym aktem prawnym regulującym kwestie związane z dokumentem, który potwierdza tożsamość i obywatelstwo polskie osoby na terytorium RP oraz innych państw członkowskich UE. Dowód osobisty jest obowiązkowy dla każdego pełnoletniego obywatela RP. Z tych względów e-dowód jest bardzo popularnym środkiem identyfikacji elektronicznej wykorzystywanym w Polsce.

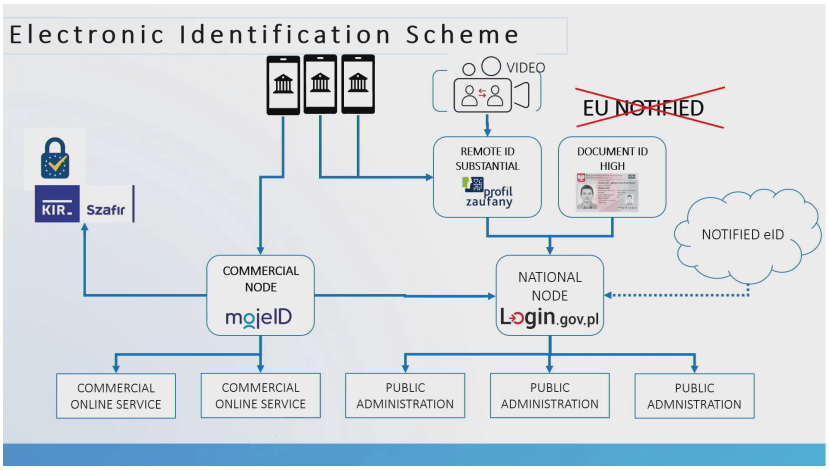
Wprowadzenie Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 910/2014 miało na celu utworzenie jednolitego rynku cyfrowego, jednocześnie minimalizując wiele przeszkód wynikających z rozdrobnienia rynku cyfrowego i braku interoperacyjności między krajami członkowskimi. Akt prawny wprowadza wiele pojęć dotyczących rynku cyfrowego, których przyswojenie nie tylko ułatwia zrozumienie samego rozporządzenia, ale również Ustawy o usługach zaufania oraz identyfikacji elektronicznej, która wprowadza postanowienia rozporządzenia. Bardzo istotnym elementem jest tzw. zasada wzajemności, która umożliwia wykorzystanie notyfikowanych środków identyfikacji elektronicznej we wszystkich krajach wspólnoty.

3.2. Obecny ekosystem związany z tożsamością cyfrową w Polsce

Na mocy ustawy z dnia 5 września 2016 r. o usługach zaufania oraz identyfikacji elektronicznej został powołany krajowy schemat identyfikacji elektronicznej, który znacznie uporządkował powstającą w tym czasie infrastrukturę. Od tego momentu schemat zakładał obecność:

- węzła krajowego – będącego pośrednikiem między dostawcami tożsamości a dostawcami usług. W Polsce taką funkcję pełni platforma login.gov.pl,
- systemów identyfikacji elektronicznej – rozwijanych zarówno przez podmioty publiczne (profil zaufany, profil osobisty) skupione w ramach węzła krajowego, jak i komercyjne, skupione w ramach węzła komercyjnego (system mojeID),
- dostawców usług – systemów umożliwiających realizację usług online (EPUAP, IKP, mObywatel),
- węzła transgranicznego – rozwiązania umożliwiającego integrację z usługami cyfrowymi dostępnymi w innych krajach członkowskich.

Obecny kształt obecnego ekosystemu eID najlepiej przedstawia grafika (rysunek 13.1) zaprezentowana przez Pana Michała Tabora podczas prezentacji pt.: „Trust Services and eID in Poland in relations to the EU", przedstawionej na konferencji EFPE 2020 [6].



Rysunek 13.1. Obecny kształt schematu identyfikacji elektronicznej [6]

3.3. Dowody osobiste

Wprowadzone w 2019 roku dowody osobiste z warstwą elektroniczną rozpoczynają zupełnie nowy rozdział w identyfikacji elektronicznej. Od tego momentu pojawia się cała masa udogodnień, którą warto znać, aby móc w pełni korzystać z dostępnych innowacji. Wraz z nowymi rozwiązaniami pojawiają się również nowe kategorie zagrożeń, które są wykorzystywane przez adwersarzy, dlatego kluczem do zapewnienia bezpieczeństwa jest znajomość rozwiązań i świadomość dotycząca ataków, jakie są z nimi związane.

E-dowód oprócz warstwy wizualnej zawiera warstwę elektroniczną, w której wraz z danymi posiadacza przechowywane są certyfikaty umożliwiające wykorzystanie potencjału oferowanego przez e-dowód w ramach różnych e-usług.

Najnowszy wzór dowodu został wprowadzony 7 listopada 2021. W porównaniu do poprzedniego wzoru (z 2019 roku) nowy model został uzupełniony o odręczny podpis posiadacza dokumentu w warstwie graficznej i odciski palców w warstwie elektronicznej. Warstwa graficzna e-dowodu realizuje takie same możliwości, co jej poprzednie odpowiedniki. Natomiast ogromne korzyści oferuje warstwa elektroniczna, która umożliwia:

- uwierzytelnienie w e-usługach,
- podpisywanie dokumentów elektronicznych,
- odprawę przy automatycznych bramkach granicznych znajdujących się na lotniskach,
- sprawdzenie wiarygodności danych zawartych w warstwie graficznej.

Duże zmiany dotyczą również podejścia do kwestii związanych z odpowiednim zabezpieczeniem dokumentu. Od momentu wprowadzenia e-dowodów oprócz tradycyjnych zabezpieczeń osadzonych w warstwie graficznej pojawia się konieczność zabezpieczenia warstwy cyfrowej. W tym celu wprowadzono szereg nowych zabezpieczeń, do których można zaliczyć:

- uniemożliwienie zdalnego odczytu danych zawartych w warstwie elektronicznej poprzez wymuszenie podania numeru CAN przy każdej próbie nawiązania połączenia z e-dowodem,
- zabezpieczenie kodami PIN certyfikatu podpisu osobistego i certyfikatu identyfikacji i uwierzytelniania,
- blokowanie poszczególnych kodów PIN po trzykrotnym wpisaniu nieprawidłowego kodu,
- możliwość odblokowania zablokowanych kodów PIN po podaniu kodu PUK,
- trwałe zablokowanie odpowiedniego certyfikatu po trzykrotnym błędnym wpisaniu kodu PUK,
- możliwość unieważnienia lub czasowego zawieszenia e-dowodu na stronie gov.pl,
- automatyczną weryfikację autentyczności danych zawartych w warstwie elektronicznej poprzez sprawdzenie integralności umieszczonych w nim danych i proces walidacji certyfikatów.

3.4. Aplikacje przeznaczone do zarządzania cyfrową tożsamością

Za tempo przemian w ekosystemie eID odpowiadają nie tylko przygotowane wcześniej fundamenty prawne i szybkość wprowadzania poszczególnych rozwiązań. Jednak bodźcem mającym niewątpliwie największy wpływ na popularyzację całego ekosystemu miała zmiana przyzwyczajeń społeczeństwa, która była wymuszona niespodziewanym wybuchem pandemii COVID-19.

Ogromny wpływ na popularyzację całego ekosystemu eID wśród społeczeństwa mają realne korzyści wynikające z wykorzystania wprowadzonych rozwiązań. W przypadku e-tożsamości nieodzownym elementem są aplikacje przeznaczone do zarządzania i realnego korzystania z dostępnych możliwości. Do najbardziej popularnej grupy należą aplikacje przeznaczone na smartfony. Istnieją również aplikacje przeznaczone na komputer, jednak nie są tak popularne, ponieważ wymagają zakupienia specjalnego czytnika kart bezstykowych, przez co zapewniają znacznie mniejszą mobilność niż rozwiązania dostępne na smartfony. Do pełnego wykorzystania możliwości oferowanych przez e-dowodu warto sko-

rzystać z aplikacji eDO App. Program umożliwia uzyskanie dostępu do warstwy elektronicznej dowodu osobistego za pośrednictwem modułu NFC dostępnego w większości smartfonów. Aplikacja umożliwia:

- bezpieczne potwierdzenie tożsamości,
- złożenie podpisu osobistego lub kwalifikowanego,
- dostęp do warstwy elektronicznej e-dowodu:
  - wyświetlenie danych z e-dowodu lub paszportu,
  - podpisanie dokumentu,
  - zmianę i odblokowanie kodu PIN.

Bardzo interesującym rozwiązaniem jest aplikacja mObywatel, nazywana cyfrowym portfelem na dokumenty i usługi e-administracji. Umożliwia ona prezentowanie danych zawartych w dokumentach, takich jak:

- dowód osobisty,
- prawo jazdy, dowód rejestracyjny,
- Unijny Certyfikat COVID,
- legitymacja szkolna i studencka.

Za pośrednictwem aplikacji można także uzyskać dostęp do usług umożliwiających:

- sprawdzenie liczby posiadanych punktów karnych,
- realizację eRecepty bez podawania numeru PESEL.

Dane prezentowane w mObywatel mogą być w dowolnym czasie zweryfikowane za pośrednictwem publicznie dostępnej aplikacji mWeryfikator.

3.5. Mechanizmy i protokoły komunikacji z cyfrową częścią dowodu

Z technicznego punktu widzenia e-dowody są niczym innym jak kartami inteligentnymi (ang. *smart cards*). Karty tego typu mimo swoich niewielkich rozmiarów są bardzo skomplikowanymi układami.

W przypadku e-dowodów komunikacja z otoczeniem odbywa się w sposób bezstykowy. Zasilanie i komunikacja z kartą realizowane jest za pomocą fal EM. W takich przypadkach antena ma postać spiralnego lub rozciągniętego po krawędziach karty zatopionego przewodu. Niestety w przypadku Polski nie jest publicznie dostępna żadna dokumentacja opisująca warstwę elektroniczną dowodu osobistego. Jednak na podstawie wiedzy zdobytej podczas studiowania norm prawnych tworzących ekosystem eID wiadomo, że na mocy Rozporządzenia z dnia 20 czerwca 2019 r. w sprawie poprawy zabezpieczeń w dowodach osobistych nowe dokumenty muszą być zgodne z dokumentem ICAO Doc 9303 [7], którego dokumentacja jest publicznie dostępna. Z tego względu znaczna część rozdziału przedstawia szczegóły techniczne zawarte w standardzie opublikowanym przez ICAO:

- struktury danych zawarte w warstwie elektronicznej,
- mechanizmy kontroli dostępu do dokumentu,
- protokoły komunikacji,
- strukturę i przykładowe polecenia APDU wysyłane do karty.

Następnie na podstawie zebranych informacji wykonano próbę komunikacji z warstwą elektroniczną dowodu. Wykonanie krótkiego testu pozwoliło lepiej zrozumieć naturę komunikacji z kartami typu smart card, przeanalizować otrzymane odpowiedzi i potwierdzić,

że dostęp do struktur ICAO nie jest możliwy dla osób postronnych bez wcześniejszego skorzystania z mechanizmów kontroli dostępu.

4. Uzyskane wyniki

Realizacja pracy zakłada harmonizację obecnego stanu wiedzy dotyczącego zagadnień i rozwiązań dostępnych w ramach ekosystemu tożsamości cyfrowej w Polsce. Duża złożoność samego rynku, rozproszenie różnych źródeł informacji i wpływ wielu innych czynników wymusił zastosowanie podejścia interdyscyplinarnego, które łączy zagadnienia natury prawnej, technologicznej i społecznej. Zebranie skrawków informacji i ich logiczne połączenie w całość umożliwiło zaprezentowanie skondensowanej wiedzy dotyczącej ekosystemu ID w bardziej przystępnej formie dla większości społeczeństwa.

5. Podsumowanie

Akty prawne wydane na przestrzeni ostatnich lat mają na celu nie tylko ujednolicenie nazewnictwa w zakresie eID, ale również przedstawienie zarysu architektury, która zapewni odpowiednią wiarygodność i bezpieczeństwo środków umożliwiających identyfikację elektroniczną. Warto podkreślić, że większość z analizowanych aktów prawnych została opracowana na podstawie znanych i popularnych standardów.

Rozwój ekosystemu eID jest mocno uzależniony od danego kraju. Różnice dotyczą wielu czynników, wśród których największą rolę odgrywa przyjęty schemat rynku w zakresie identyfikacji elektronicznej i kwestia podejścia do obowiązkowych dowodów tożsamości w poszczególnych krajach. W przypadku Polski ekosystem eID jest oparty na schemacie federacyjnym, który umożliwia wzajemną współpracę podmiotów publicznych i prywatnych. Schemat ten jest stosowany z powodzeniem w wielu innych krajach należących do UE. Dodatkowo jeden z podstawowych środków identyfikacji elektronicznej jest częścią obowiązkowego w Polsce e-dowodu, co również powinno mieć pozytywny wpływ na popularność e-usług.

Powołany na mocy ustawy krajowy schemat identyfikacji elektronicznej znacznie ułatwia integrację nowych systemów. Niestety system w dalszym ciągu nie został zintegrowany z węzłem transgranicznym. Dodatkowo Polska obecnie nie posiada żadnego notyfikowanego środka identyfikacji elektronicznej. Z obecnie dostępnych rozwiązań tylko umieszczony w e-dowodzie profil osobisty spełnia podstawowe wymagania.

Bardzo ważnym etapem realizacji pracy była prezentacja zabezpieczeń i możliwości oferowanych przez e-dowód. Okazuje się, że metody weryfikacji autentyczności oparte na warstwie elektronicznej dowodu osobistego są znacznie szybsze i zapewniają zdecydowanie wyższe bezpieczeństwo całego procesu.

Każdy obywatel powinien na co dzień w odpowiedni sposób dbać o ochronę swoich danych osobowych, w tym dokumentów tożsamości. Z tego względu w pracy zarówno zamieszczono zbiór dobrych praktyk związanych z codziennym użytkowaniem dokumentów, jak i zaprezentowano schemat reakcji w przypadku sytuacji bardzo stresującej, jaką jest utrata dowodu.

Najbardziej interesujące wyniki dotyczą przeglądu aplikacji wykorzystywanych do zarządzania tożsamością. W przypadku aplikacji mobilnych bardzo ważnym elementem jest odpowiednia weryfikacja danych prezentowanych na ekranie smartfona. Istnieją różne metody weryfikacji, jednak najbardziej skuteczną metodą jest wykorzystanie aplikacji mWeryfikator. Niestety okazuje się, że różnice w implementacji pewnych protokołów uniemożliwiają współpracę systemów iOS i Android. Szczególne trudności mogą wyniknąć podczas próby weryfikowania danych prezentowanych w aplikacji mObywatel za pośrednictwem narzędzia mWeryfikator.

Bibliografia

[1] Kostro M., Tabor M., *Identyfikacja i uwierzytelnienie w usługach elektronicznych*, wydanie pierwsze, Wydawnictwo Związek Banków Polskich, Warszawa, Polska, 2021.

[2] Ustawa z dnia 6 sierpnia 2010 r. o dowodach osobistych; Kancelaria Sejmu, Dz.U. 2010 Nr 167, poz. 1131, 2010.

[3] Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 910/2014 z dnia 23 lipca 2014 r. w sprawie identyfikacji elektronicznej i usług zaufania w odniesieniu do transakcji elektronicznych na rynku wewnętrznym oraz uchylające dyrektywę 199, 2014.

[4] Ustawa z dnia 5 września 2016 r. o usługach zaufania oraz identyfikacji elektronicznej, Kancelaria Sejmu, Dz.U. 2016, poz. 1579, 2016.

[5] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2019/1157 z dnia 20 czerwca 2019 r. w sprawie poprawy zabezpieczeń dowodów osobistych obywateli Unii i dokumentów pobytowych wydawanych obywatelom Unii i członkom ich rodzin korzystającym z prawa do swobodnego przemieszczania się, 2019.

[6] Tabor M., *Trust Services and eID in Poland in relations to the EU*, Europejskie Forum Podpisu Elektronicznego, edycja online, 2020.

[7] ICAO, *Doc 9303 Machine Readable Travel Documents*, wydanie 8, 2021.

Baza danych i warstwa dostępu do danych w aplikacji webowej dedykowanej sprzedaży dzieł sztuki

inż. Jakub Smołka
Informatyka

dr inż. Krzysztof Waśko
Katedra Informatyki i Inżynierii Systemów

Przedmiotem pracy dyplomowej, opisywanej w niniejszym streszczeniu, będącej częścią aplikacji, jest opis procesu projektowania oraz implementacji serwera aplikacji, bazy danych pozwalającej przechowywać dane dotyczące m.in. dzieł sztuki, użytkowników i aukcji, kanału komunikacji między tymi komponentami oraz zabezpieczenie danych, pprzez zastosowanie mechanizmów autoryzacji użytkownika.

W streszczeniu zostały poruszone najważniejsze tematy znajdujące się w pracy dyplomowej. Opisany został cel i założenia aplikacji. We wstępie teoretycznym wytłumaczone zostały najważniejsze pojęcia wymagane do pełnego zrozumienia pracy. Projekt i wykonanie przedstawia analizę rynkową istniejących już rozwiązań i technologii oraz opisuje kolejne kroki wraz z przykładami, prowadzące do osiągnięcia postawionego celu. Na koniec wykonane oprogramowanie zostaje poddane testom, na podstawie których wyciągnięte zostają wnioski.

Cel zostaje osiągnięty, a aplikacja jest zaprojektowana i zaimplementowana w sposób określony w procesach formułowania celu i wymagań projektowych.

1. Cel i założenia

Celem pracy dyplomowej, realizowanej w ramach Zespołowego Przedsięwzięcia Inżynierskiego, jest zapewnienie na rynku miejsca, które będzie zrzeszało młodych, debiutujących artystów, dla których próg wejścia na rynek dzieł sztuki jest ogromny, a także wszelkie instytucje zajmujące się szeroko pojętą sztuką, tj. np. Domy Aukcyjne, prywatne galerie sztuki czy antykwariaty, oraz wszystkich ludzi zainteresowanych nabyciem różnego rodzaju dzieł sztuki. Sposobem osiągnięcia tego celu jest zaprojektowanie oraz implementacja aplikacji webowej dedykowanej sprzedaży dzieł sztuki.

Jednym z najważniejszych elementów takiej aplikacji webowej, nazywanej również portalem, jest odpowiednio zaprojektowany, zaprogramowany oraz zabezpieczony serwer,

składający się między innymi z bazy danych oraz kanału komunikacji między serwerem a interfejsem użytkownika. Tymi elementami szczegółowo zajmuje się autor pracy dyplomowej, a zatem główne założenia projektowe realizowanej pracy to:

- analiza biznesowa projektowanego rozwiązania, a także istniejących już na rynku podobnych rozwiązań,
- analiza technologii serwerowych,
- analiza technologii bazodanowych,
- analiza technologii służących do komunikacji między interfejsem użytkownika i serwerem,
- wybór technologii, które w optymalny sposób pozwolą zrealizować wymagania sformułowane podczas analizy biznesowej,
- projekt oraz implementacja serwera z wykorzystaniem wybranych technologii, pozwalającego na:
  - zarządzanie użytkownikami, dziełami sztuki oraz innymi encjami w bazie danych,
  - płynną i szybką komunikację z interfejsem użytkownika,
  - bezpieczne przechowywanie oraz przesyłanie danych,
- testowanie zaimplementowanego rozwiązania.

2. Wstęp teoretyczny

W celu sprawnego odnajdywania się w tekście zarówno niniejszego streszczenia, jak i samej pracy dyplomowej znacznie pomoże wyjaśnienie stosowanych w tych tekstach następujących pojęć informatycznych:

- Aplikacja webowa (internetowa) – program komputerowy, z którym użytkownik wchodzi w interakcję, głównie z wykorzystaniem przeglądarki internetowej. Znanymi przykładami aplikacji webowych są portale YouTube[1] i Spotify[2].
- Klient – najczęściej przeglądarka internetowa, z której korzysta użytkownik w celu interakcji z aplikacją webową.
- Serwer – oprogramowanie lub urządzenie, zapewniające funkcjonalności i usługi innym programom lub urządzeniom, głównie klientom, poprzez komunikację polegającą na wymianie zasobów. Zasobem mogą być m.in.: pliki, informacje z bazy danych, obrazy.
- Frontend (technologie frontendowe) – technologie związane z interakcją między użytkownikiem a aplikacją webową. Dotyczą głównie tego, co widoczne dla użytkownika, czyli interfejsu graficznego i sposobu interakcji z nim.
- Backend (technologie backendowe) - technologie dotyczące tego, czego użytkownik zwykle nie widzi, czyli związane z serwerem, bazą danych i komunikacją między serwerem a interfejsem użytkownika (frontendem).
- API (ang. *Application Programming Interface*) – połączenie, oprogramowanie lub standard umożliwiający komunikację dwóch urządzeń sieciowych (np. komunikację klienta z serwerem).

Liczba sposobów, na które można skonstruować aplikację webową, rośnie wraz z liczbą nowych technologii webowych na rynku. Logika pierwszych aplikacji internetowych sku-

piona była całkowicie na serwerze. Serwer przesyłał użytkownikom dane i wygląd strony, a także obsługiwał całą interakcję z interfejsem. Wraz z rozwojem języka programowania JavaScript (w skrócie JS) obsługa dużej część interakcji została oddelegowana do klienta (przeglądarki), co pozwoliło znacznie odciążyć serwery i budować dużo bardziej skomplikowane aplikacje. Obecnie popularne technologie frontendowe pozwalają tworzyć aplikacje działające całkowicie po stronie klienta i nie wymagające istnienia konkretnego serwera. Ten ostatni sposób tworzenia aplikacji frontendowych przeznaczony jest dla aplikacji prostych, których głównym celem jest serwowanie użytkownikom statycznych danych i zapewnienie nieskomplikowanej interakcji. Z tego powodu do zaprogramowania aplikacji, o której mowa we wstępie – portalu aukcyjnego dzieł sztuki – wykorzystany został sposób drugi. Aplikacja podzielona jest na interfejs użytkownika (frontend) oraz serwer i bazę danych (backend). Ta druga część, backend, jest przedmiotem pracy dyplomowej, której dotyczy to streszczenie.

3. Projekt i wykonanie

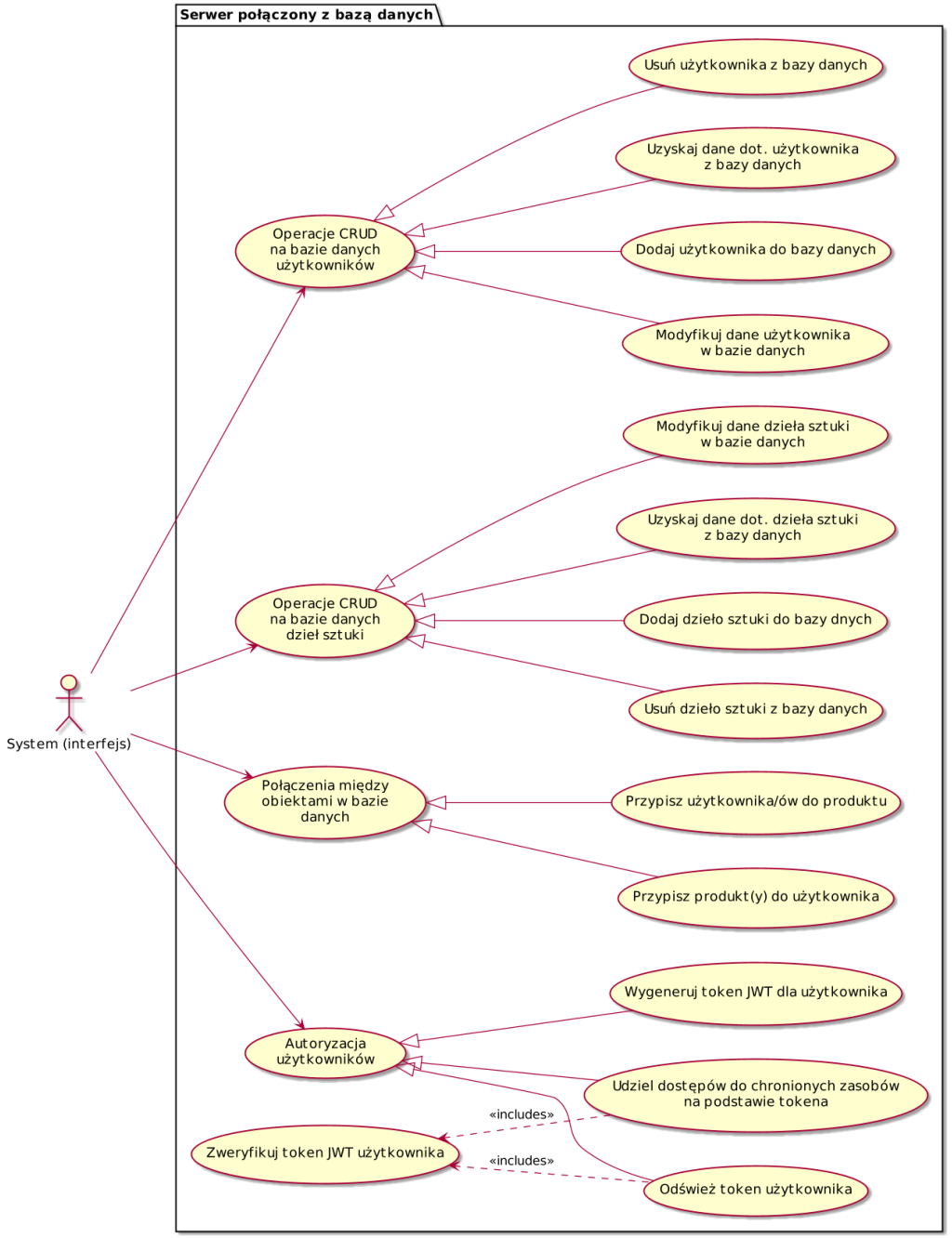
Projekt i wykonanie części serwerowej, a także wszystkich wymaganych komponentów wchodzących w jej skład można podzielić na kilka sekcji, z których każda opisuje ważny etap, kamień milowy, mający duże znaczenie i wpływ na kolejne etapy.

3.1. Wymagania projektowe

Szczegółowo wykonana analiza rynkowa istniejących już portali aukcyjnych dzieł sztuki pozwoliła określić ich wady i zalety. Te z kolei umożliwiły określenie luki, którą tworzona aplikacja ma zapewnić, a także, co najważniejsze, sformułowanie wymagań projektowych dotyczących implementowanego rozwiązania. Najważniejszymi, ogólnie pojętymi funkcjonalnościami portalu jako całości są: zapewnienie możliwości przyjemnego, wygodnego oraz nieskomplikowanego przeglądania, wystawiania, licytowania oraz kupna prac wpisujących się w szeroko pojętą kategorię dzieł sztuki. W części backendowej, która jest kluczowym elementem portalu, wyróżnionych zostało 16 wymagań funkcjonalnych oraz 6 wymagań niefunkcjonalnych. Mówią one między innymi o tym, że:

- Interfejs użytkownika (klient) ma mieć możliwość dostępu do bazy danych poprzez skorzystanie z odpowiedniego kanału komunikacji – API.
- Interfejs użytkownika ma mieć możliwość utworzenia obiektów dotyczących użytkowników, dzieł sztuki, a także innych obiektów powstałych w procesie analizy biznesowej w bazie danych.
- Interfejs użytkownika ma mieć możliwość uzyskania dostępu, modyfikacji i usunięcia utworzonych w bazie danych obiektów.
- Interfejs użytkownika ma mieć możliwość łączenia ze sobą niektórych obiektów w bazie danych.
- Serwer powinien umożliwiać ochronę danych poprzez zastosowanie procesu uwierzytelniania użytkowników.
- Serwer ma zapewniać klientowi możliwość wygenerowania kryptograficznego tokena, którym będzie mógł uwierzytelniać użytkowników.

W procesie analizy rynkowej, oprócz wymagań projektowych, zostały również sformułowane przypadki użycia, które przedstawiono na rysunku 14.1.

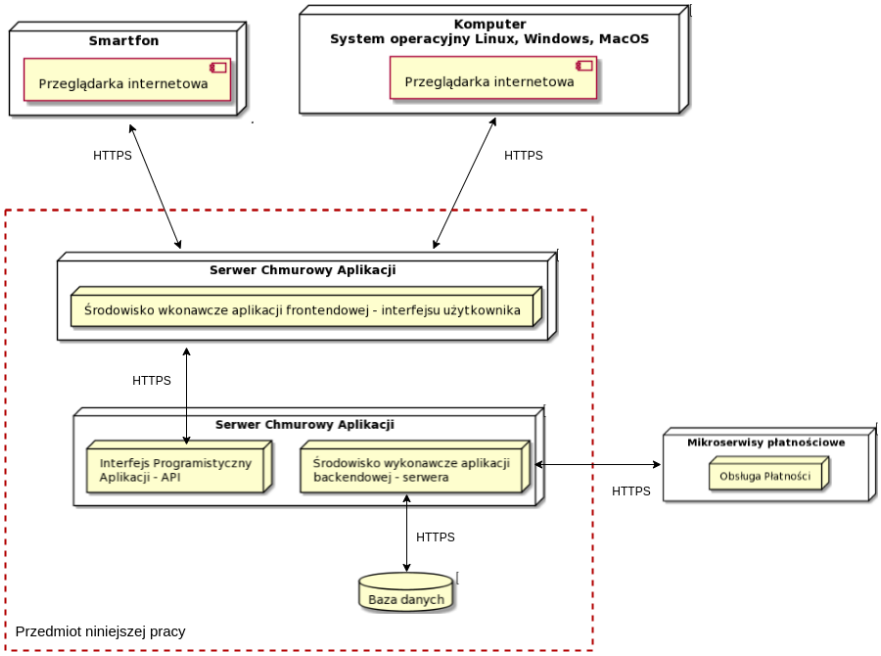


Rysunek 14.1. Diagram przypadków użycia. Wykonanie własne przy pomocy serwera internetowego `www.plantuml.com` [3]

3.2. Architektura systemu

Podczas analizy wymagań projektowych została zaprojektowana architektura aplikacji, która zgodnie z przedstawionymi we wstępnie teoretycznym sposobami konstrukcji aplikacji webowych jest architekturą typu **Klient–Serwer**[4], w której serwer posiadający środowisko wykonawcze aplikacji frontendowej dostarcza użytkownikom smartfonów lub komputerów, z wymienionymi na diagramie systemami, interfejs użytkownika. Serwer ten (nazywany również Systemem) wysyła również różnego rodzaju zapytania związane m.in. z operacjami na bazie danych do serwera posiadającego środowisko wykonawcze aplikacji backendowej poprzez Programistyczny Interfejs Aplikacji (API).

Architektura ta została później, po analizie i wyborze technologii, doprecyzowana i rozszerzona. Podstawową architekturę systemu przedstawioną na diagramie rozmieszczenia prezentuje rysunek 14.2. Rozszerzona architektura wraz z opisem przedstawiona jest w pracy dyplomowej.



Rysunek 14.2. Podstawowa architektura aplikacji, przedstawiona na diagramie rozmieszczenia

3.3. Analiza dostępnych technologii

Choć obecnie istniejące technologie posiadają ogromne możliwości i pozwalają tworzyć szeroką gamę oprogramowania, to odpowiednie ich dobranie do konkretnego projektu jest kluczowe. Odpowiednio dobrane technologie pozwalają utworzyć aplikację w krótkim czasie, zapewnić zaawansowane funkcjonalności, stabilność, a także skalowalność, czyli możliwość swobodnego rozwoju aplikacji w przyszłości.

Najważniejsze technologie, na których bazować będzie cała aplikacja, to technologie serwerowe, bazodanowe oraz te dotyczące API.

Baza danych

Technologie bazodanowe można podzielić ze względu na **typ** oraz **rozmieszczenie**. Podział baz danych ze względu na typ został przedstawiony w tabeli 14.1.

Tabela 14.1. Podział baz danych ze względu na typ

Typ bazy danych	Opis
Relacyjne	Najczęściej używane, posiadają sztywno określone typy i struktury. Dane są uporządkowane w tabelach połączonych ze sobą relacjami.
Obiektowe	Dane przechowywane są w postaci obiektów, zdefiniowanych jako klasy (używane w obiektowych językach programowania).
Relacyjno-obiektowe	Zapewniają obiektowe przedstawienie danych, zachowując jednak bazę relacyjną, jako wewnętrzny mechanizm działania.
Nierelacyjne	Zawierają dane w formie "dokumentów" lub obiektów zawierających pary klucz-wartość[14]. Nie posiada sztywno zdefiniowanych danych oraz relacji, co pozwala na dostosowanie sposobu przechowywania danych w zależności od aktualnych potrzeb.

Podział baz danych ze względu na ich rozmieszczenie przedstawia tabela 14.2.

Tabela 14.2. Podział baz danych ze względu na rozmieszczenie

Rozmieszczenie bazy danych	Opis
Lokalne	Bazy danych, które tworzone są na lokalnym systemie podczas implementacji.
Chmurowe	Bazy danych działające na platformie przetwarzania w chmurze. Nie wymagają posiadania swojego własnego serwera, oferują dużą moc obliczeniową w przetwarzaniu danych oraz stosunkowo duże pojemności, a także relatywnie łatwą integrację z większością najbardziej popularnych technologii na rynku.

Technologie Serwerowe

W pracy dyplomowej zostały porównane najpopularniejsze technologie serwerowe z podziałem na języki programowania służące do ich implementacji. Analizie poddane zostały następujące pozycje:

- **Javascript**
  - Express.js
  - Next.js
  - Nest.js
- **Python**
  - Django
  - Flask
- **Java**
  - Spring Boot
  - Hibernate

Interfejs Programistyczny Aplikacji

W kwestii implementacji komunikacji serwera z interfejsem porównany został standard internetowy REST API z technologią GraphQL.

Tabela 14.3. Porównanie standardu REST API z technologią GraphQL

Standard/Technologia	Opis
REST API [5]	Struktura projektowania API, polegająca na powiązaniu odpowiedniego adresu strony do odpowiednich danych, przyjmująca zestaw kilku prostych reguł o nazwie REST (Representational State Transfer), które opisują sposób komunikacji m.in. pomiędzy klientem i serwerem.
GraphQL API [6]	Serwerowe środowisko uruchomieniowe dla API, posiadające swój własny język zapytań GraphQL. W odróżnieniu od REST API, GraphQL API posiada tylko jeden adres (choć może posiadać więcej), pod który wysyłane zostają wszystkie zapytania. Dane zwracane pod tym adresem zależą od zapytania.

3.4. Wybór odpowiednich technologii

Biorąc pod uwagę założenia projektowe oraz wady i zalety każdej z wymienionych w poprzednim podrozdziale technologii i standardów, wybrane zostały zaprezentowane poniżej narzędzia programistyczne i skrótowo opisane ich najważniejsze zalety.

Baza danych

W odniesieniu do typu bazy danych, wybrana została **nierelacyjna baza danych**, która:

- pozwala na bezproblemową obsługę przyrostowych w szybkim tempie danych, również danych niestrukturalnych oraz o zmieniającej się strukturze,
- zapewnia wysoką elastyczność w kwestii schematu bazy i struktury danych.

W kwestii umieszczenia bazy danych, najlepszą okazała się **chmurowa baza danych**, z następujących względów:

- wykorzystanie tej technologii nie wymaga konfiguracji bazy danych na własnym serwerze,
- pozwala na korzystanie z chmury obliczeniowej, której często nie dorównują małe serwery.

Jeśli chodzi o konkretną technologię, wykorzystana została baza danych **MongoDB** [7], ponieważ:

- jest bazą typu NoSQL oraz oferuje platformę opartą na chmurze,
- pozwala na łatwą integrację z wieloma technologiami języka JavaScript.

Technologia serwerowa

W kwestii technologii serwerowej najlepszy okazał się **Nest.js** [8] ze względu na następujące cechy:

- jest modułowy, co pozwala na integrację ze sobą różnych technologii oraz projektowanie modułów wielokrotnego użytku,
- wykorzystuje popularne, sprawdzone technologie i opakowuje je w warstwę abstrakcji, która pozwala programiście uniknąć wielu błędów podczas projektowania podstawowych funkcjonalności,
- pozwala na dobrą oraz nieskomplikowaną integrację z nierelacyjną, chmurową bazą danych MongoDB oraz wspiera obie, opisane wyżej, architektury API–REST oraz GraphQL.

Interfejs Programistyczny Aplikacji

Biorąc pod uwagę wady i zalety obu opisanych rozwiązań, założenia projektowe oraz wybór nierelacyjnej bazy danych, zastosowana zostanie technologia **GraphQL**, gdyż:

- umożliwia korzystanie z języka zapytań, dzięki któremu interfejs użytkownika może poprosić o konretnie wyspecyfikowane dane,
- oferuje **subskrypcje**, czyli zapytanie specjalnego typu, które tworzy tunel komunikacyjny pomiędzy systemem i serwerem, dzięki któremu serwer może powiadamiać system na bierząco o wszelkich zmianach.

3.5. Implementacja

Do tego momentu zostały przedstawione założenia projektowe, wybrane technologie oraz architektura systemu. W tym rozdziale opisana została konfiguracja zastosowanych technologii, techniczna implementacja postawionych wymagań oraz połączenie serwera z systemem. Zostały również przedstawione fragmenty implementacji wraz z objaśnieniem przedstawianego kodu.

Utworzenie i konfiguracja bazy danych

Pierwszym krokiem w projektowaniu części backendowej było utworzenie klastra danych (miejsca na dane) w chmurowej bazie danych MongoDB oraz skonfigurowanie go w sposób zapewniający maksymalne bezpieczeństwo danych.

W tym celu, zostały zrealizowane następujące kroki:

- zostały utworzone **konta użytkowników** dla członków zespołu, z których każdy otrzymał tylko niezbędne dla niego uprawnienia dostępu i modyfikacji bazy,
- został ograniczony **dostęp sieciowy** do bazy i przydzielony tylko urządzeniom należących do członków zespołu oraz fizycznemu serwerowi, na którym uruchomiona zostanie aplikacja.

Podstawowa konfiguracja serwera i modułów

Kolejnym krokiem implementacyjnym było utworzenie i konfiguracja serwera oraz jego modułów w wybranej technologii **Nest.js**.

Pierwszym modułem, który został utworzony poza głównym modułem serwera, był moduł komunikacji z bazą danych MongoDB. Moduł ten umożliwił przede wszystkim tworzenie i zarządzanie danymi w bazie z poziomu kodu.

Utworzenie modeli w bazie danych

Po podłączeniu bazy do serwera można przejść do tworzenia w niej modeli zaprojektowanych w procesie analizy biznesowej oraz podczas formułowania wymagań projektowych i architektury systemu. Przykładowy model użytkownika, utworzony w kodzie aplikacji, prezentuje rysunek 14.3.

```
@Schema()
export class User {
  _id: MongooseSchema.Types.ObjectId;

  @Prop({ required: true })
  email: string;

  @Prop({ required: true })
  password: string;

  @Prop({ required: true })
  name: string;

  @Prop({ required: true })
  birthDate: string;

  @Prop({ required: true })
  address: string;

  @Prop({ required: true })
  city: string;

  @Prop({ required: true })
  country: string;

  @Prop({ required: true })
  postalCode: string;

  @Prop({ required: false })
  education?: string;

  @Prop({ required: true, default: 0 })
  tokenVersion: number;

  @Prop({ type: [{ type: MongooseSchema.Types.ObjectId, ref: 'ArtProduct' }] })
  artProducts?: ArtProduct[];
}

export const UserSchema = SchemaFactory.createForClass(User);
```

Rysunek 14.3. Model użytkownika w kodzie aplikacji

Rysunek ten przedstawia parametry oraz ich typy, które posiadał obiekt użytkownika w pierwszej fazie tworzenia aplikacji. Adnotacje zaczynające się znakiem **@** to polecenia regulujące sposób, w jaki zostanie utworzone pole w tabeli w bazie. Na przykład ***required: true*** oznacza, że pole jest wymagane i nie będzie możliwe utworzenie użytkownika bez informacji oznaczonej tą adnotacją. W pracy dyplomowej każda przedstawiona instrukcja w kodzie została szczegółowo wytłumaczona.

Implementacja i konfiguracja GraphQL API

Podobnie jak w przypadku połączenia serwera z bazą danych, do skonfigurowania technologii GraphQL potrzebne było napisanie modułu i dołączenie go do modułu głównego serwera. Technologia GraphQL również wymaga zdefiniowania typów danych, co odbywa się w kodzie, w sposób bardzo podobny do definiowania obiektów w bazie danych.

Operacje na bazie danych

Ukończona konfiguracja bazy danych i kanału komunikacji pozwala na przejście do kolejnego, najbardziej obszernego etapu, jakim jest implementacja wszystkich funkcjonalności związanych z operacjami na bazie danych. Do takich funkcjonalności należą na przykład:

- utworzenie użytkownika w bazie danych,
- modyfikacja odpowiednich danych użytkownika,
- usunięcie danych użytkownika z bazy,
- przypisanie użytkownika jako autora danego dzieła sztuki poprzez utworzenie odpowiedniego połączenia między obiektem użytkownika i dzieła w bazie.

Funkcjonalności te zostały zdefiniowane w przypadkach użycia i wymaganiach funkcjonalnych. Ich dokładny opis, a także opis procesu ich implementacj, z przykładami kodu źródłowego znajduje się w pracy dyplomowej.

Bezpieczeństwo danych i autoryzacja

Ostatnim i bardzo ważnym krokiem jest zabezpieczenie danych w bazie i zastosowanie procesu autoryzacji użytkownika. Ten proces jest niezwykle istotny w przypadku aplikacji komercyjnych, ponieważ jakikolwiek wyciek danych często prowadzi do ogromnych strat finansowych i wizerunkowych. Pierwszym krokiem w kierunku bezpieczeństwa informacji jest **szyfrowanie** wszystkich wrażliwych danych. Dlatego też m.in. hasła, które użytkownik podaje w procesie rejestracji, a następnie używa do zalogowania, są szyfrowane z wykorzystaniem **kryptograficznej funkcji skrótu**. Proces autoryzacji użytkowników odbywa się za pomocą **tokena JWT** (ang. Json Web Token) [9]. Jest to pewien standard, bardzo często używany w sieci, którego działanie polega na emisji zaszyfrowanych tokenów podpisanych kluczem kryptograficznym i zawierających informacje na temat ich posiadaczy. Przykładowe zastosowanie tokena JWT w komunikacji sieciowej można przedstawić w kilku krokach.

1. Użytkownik loguje się do portalu.
2. Serwer, po zaakceptowaniu danych użytkownika, generuje dla niego zaszyfrowany token.
3. Użytkownik (często bez wiedzy o tym) przy każdej interakcji z serwerem przesyła swój token.
4. Serwer rozszyfrowuje token użytkownika i weryfikuje na podstawie danych zawartych w tokenie tożsamość użytkownika.
5. Po poprawnej weryfikacji serwer udostępnia swoje zasoby (np. listę dzieł sztuki z bazy danych) użytkownikowi.

Token może posiadać różne dane, najczęściej są to: dane identyfikujące użytkownika w bazie danych, termin wygaśnięcia tokena, czyli czas, po którym przestanie być ważny, oraz kryptograficzny podpis uniemożliwiający jego sfałszowanie.

4. Testy i uzyskane wyniki

Aplikacja serwerowa została poddana szeregowi różnych testów. Większość z nich to testy manualne, przeprowadzone z wykorzystaniem środowiska testowego, udostępnianego przez technologię GraphQL. Środowisko to pozwala na wykonywanie różnego rodzaju zapytań, co pozwoliło sprawdzić nie tylko poprawność działania wszystkich funkcjonalności, ale też bezpieczeństwo danych. Dodatkowo sama baza danych była regularnie sprawdzana pod kątem różnego rodzaju wycieków w celu osiągnięcia wysokiej pewności co do bezpie-

czeństwa danych. Wyniki testów pozwoliły uznać aplikację za gotową do uruchomienia na serwerze i wypuszczenia w świat.

5. Podsumowanie

Celem niniejszej pracy było zaprojektowanie oraz implementacja funkcjonalności zapewniających możliwość przechowywania oraz wykonywania operacji na danych w bazie danych, jak również umożliwienie komunikacji systemu z serwerem oraz zabezpieczenie wrażliwych danych.

Cel niniejszej pracy został osiągnięty, co z kolei w znacznej mierze umożliwiło osiągnięcie głównego celu projektu jako całości. Dzięki temu, że każdy członek zespołu osiągnął wyznaczony cel, aplikacja jest gotowa do użytku i zapewnia wiele wymaganych funkcjonalności.

Dzięki wykorzystaniu technologii bazodanowych i serwerowych opisanych w niniejszej pracy aplikacja umożliwia przechowywanie w bezpieczny sposób dużych ilości danych dotyczących m.in. użytkowników i dzieł sztuki. Algorytmy licytacji umożliwiają przeprowadzanie na portalu aukcji, a mikroserwisy płatnościowe pozwalają na dokonanie sprzedaży. Odpowiednia konfiguracja chmurowego serwera pozwala na dostarczenie aplikacji każdemu użytkownikowi, który posiada dostęp do Internetu, a interfejs graficzny zapewnia użytkownikowi platformę, poprzez którą, w sposób prosty i wygodny, osiągnie swój cel dotyczący prezentacji lub zakupu sztuki.

Implementacja wszelkiego rodzaju oprogramowania często weryfikuje podjęte przez programistę, lub zespół programistów, decyzje projektowe. W przypadku projektu, którego częścią jest niniejsza praca, zdecydowana większość decyzji okazała się celna.

- Ujednolicenie projektu pod względem języka programowania ułatwiło współpracę w zespole, ponieważ pomimo używania różnych technologii, łatwiej było wspólnie podjąć pewne decyzje projektowe.
- Wybór nierelacyjnej bazy danych pozwala na łatwą skalowalność aplikacji w przyszłości.
- Wybór technologii chmurowych pozwoli w łatwy sposób zmigrować system oraz serwer, jeśli taka konieczność nastąpi.
- Wybór technologii chmurowych w kwestii bazy danych w pewien sposób zapobiega utracie danych, ponieważ technologie chmurowe często używają systemów rozproszonych.
- Wykorzystanie modularnej technologii serwerowej Nest.js oraz technologii GraphQL do komunikacji również zapewnia wysoką skalowalność.

Bibliografia

[1] Apliacja Webowa Youtube. Dostępna online: `https://www.youtube.com` (dostęp: 15-04-2022).

[2] Apliacja Webowa Spotify. Dostępna online: `https://www.spotify.com/pl/` (dostęp: 15-04-2022).

[3] Serwis internetowy pozwalający tworzyć diagramy. Dostępny online: `https://plantuml.com` (dostęp: 25-02-2022).

[4] Opis architektury Klient–Serwer. Dostępny online: `https://pl.wikipedia.org/wiki/Klient-serwer` (dostęp: 15-04-2022).

[5] Opis standardu REST, dotyczącego API, w języku angielskim. Dostępny online: `https://en.wikipedia.org/wiki/Representational_state_transfer` (dostęp: 15-04-2022).

[6] Serwis i internetowa dokumentacja technologii GraphQL. Dostępna online: `https://graphql.org` (dostęp: 15-04-2022).

[7] Platforma dostępowa i dokumentacja internetowa technologii bazodanowej MongoDB. Dostępna online: `https://www.mongodb.com` (dostęp: 15-04-2022).

[8] Dokumentacja internetowa technologii serwerowej Nest.js. Dostępna online: `https://nestjs.com` (dostęp: 15-04-2022).

[9] Dokumentacja oraz dekoder tokenów JWT. Dostępny online: `https://jwt.io` (dostęp: 15-04-2022).

Zastosowanie filtrów morfologicznych w redukcji zakłóceń w obrazach cyfrowych

mgr inż. Dawid Siuda
Informatyka Techniczna

prof. dr hab. inż. Olgierd Unold
Katedra Informatyki Technicznej

Praca opisuje algorytmy korzystające z mechanizmów morfologii matematycznej, wykorzystywane w filtrach odszumiania obrazów. Szczegółowo opisano w niej działanie filtrów morfologicznych i miękkich filtrów morfologicznych wraz z ich operatorami, tj. erozja, dylatacja, otwarcie i domknięcie. Pokrótce opisano również działanie algorytmów genetycznych. Po wstępie merytorycznym zaproponowano wykorzystanie algorytmu genetycznego w celu znalezienia jak najlepszego zbioru parametrów autorskiego filtru morfologicznego, powstałego na bazie podobnych filtrów znanych literaturze. Po przedstawieniu implementacji i uruchomieniu algorytmu genetycznego otrzymano zbiór rozwiązań, który szczegółowo przeanalizowano. Po ich analizie wyselekcjonowano jeden najlepszy zbiór parametrów i przy ich zastosowaniu przebadano powstały filtr pod kontem wydajności (rozumianej jako najefektywniejsze pozbywanie się szumu z obrazów). Uzyskane wyniki zestawiono z wynikami innych filtrów znanych w literaturze. Otrzymany filtr okazał się najlepszy w odsiewaniu szumu na poziomie 40–60% i uzyskał także zadowalające wyniki dla obrazów zaszumionych w 10%, 20% i 30%. Dla 70% i 80% wyniki jego działania są porównywalne ze zwykłym filtrem medianowym.

1. Cel i założenia

Celem pracy jest zaimplementowanie i wykorzystanie algorytmu genetycznego do znalezienia zbioru parametrów autorskiej funkcji filtrującej. Dana funkcja wykorzystuje do filtracji obrazu odpowiednio współpracujące ze sobą nawet cztery miękkie filtry morfologiczne. Algorytm genetyczny po znalezieniu zbioru parametrów wykorzystuje go do uruchomienia funkcji filtrującej. Wyniki działania danej funkcji oceniane są przy wykorzystaniu czterech różnych miar. Cały proces przeszukiwania przestrzeni rozwiązań, czasami nazywany również uczeniem, wykorzystuje zbiór małych obrazów o rozmiarze nieprzekraczającym 20×20 pikseli.

2. Filtry morfologiczne

Miękką morfologię matematyczną wprowadzili Koskinen, Astola i Neuvo [15] w 1991 roku. W tym podejściu stosuje się ważoną statystykę porządkową zamiast operacji minimum lub maksimum. W przeciwieństwie do zwykłych filtrów morfologicznych nie wybieramy najmniejszej lub największej wartości z otoczenia danego piksela, lecz i-tą – najmniejszą lub największą wartość, gdzie i jest parametrem wejściowym algorytmu filtracji. Główną różnicą w stosunku do standardowej morfologii jest podział elementu strukturyzującego na dwie części; twardy środek i miękką granicę. Liczby w twardej środkowej części mają wagę większą niż jeden, która jest ustalana przez tzw. parametr rangi, nazywany też często parametrem powtórzenia. Liczby w części miękkiej granicy elementu strukturalnego mają wagi równe jeden. Podsumowując, niektóre elementy z oryginalnej nieposortowanej listy uczestniczą z wagą większą niż jeden, tj. są dodawane do zbioru, z którego wybierany jest i-ty element więcej niż jeden raz. Miękka morfologia posiada takie same operatory jak zwykła morfologia matematyczna, czyli:

- miękka erozja,
- miękka dylatacja,
- miękkie otwarcie,
- miękkie zamknięcie.

3. Projekt i wykonanie

Główną częścią aplikacji badawczej była implementacja algorytmu genetycznego. Zasada działania tego algorytmu przedstawiona jest w książce [2]. Implementacja algorytmu wykorzystana w przeprowadzonych badaniach została skonfigurowana w następujący sposób:

- początkowa populacja składa się z 50 losowo generowanych osobników;
- wykorzystano krzyżowanie jednolite („równomierne"), gdzie każdy gen (liczba) jest wybierany losowo z jednego z odpowiednich genów chromosomów rodzicielskich;
- osobniki do krzyżowania wybieramy poprzez selekcję turniejową, dzielimy całą populację na dwuelementowe podzbiory, następnie z każdego zbioru wybieramy osobnika o lepszym wskaźniku przystosowania;
- po każdym pokoleniu najlepszy osobnik jest zachowywany, gdy reszta zostaje usunięta;
- prawdopodobieństwo mutacji w chromosomie ustawione jest na 10%;
- algorytm kończy się po 1000 iteracji;
- do oceny przystosowania danego osobnika użyto średniej z 4 miar:
  - średni błąd bezwzględny (wzór (15.1))

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{(i)} - \hat{y}_{(i)}\right| \tag{15.1}$$

gdzie:
MAE – średni błąd bezwzględny,
n – liczba obserwacji,

i – kolejne obserwacje,
$y_{(i)}$ – obserwacje empiryczne (uzyskane w badaniu),
$\hat{y}_{(i)}$ – obserwacje prognozowane na podstawie modelu;

– średni błąd kwadratowy (wzór (15.2))

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - \hat{y}_{(i)}\right)^2 \tag{15.2}$$

gdzie:
MSE – średni błąd kwadratowy,
n – liczba obserwacji,
i – kolejne obserwacje,
$y_{(i)}$ – obserwacje empiryczne (uzyskane w badaniu),
$\hat{y}_{(i)}$ – obserwacje prognozowane na podstawie modelu;

– błąd kształtu (wzór (15.3))

$$f_3 = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{(i',j')\in w} |(X(i,j) - X(i^{'},j^{'}) - (Y(i,j) - Y(i^{'},j^{'})))^2|^y \tag{15.3}$$

gdzie:
f_3 – średni błąd kształtu,
N – szerokość obrazu,
M – wysokość obrazu,
i – kolejna linia obrazu,
j – kolejny piksel z linii obrazu,
X – obraz wzorcowy,
Y – obraz porównywany,
y – stała porządku warzonego błędu kształtu (zazwyczaj równa 2);

– błąd jasności (wzór (15.4))

$$f_4 = \left(\frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} X(i,j)\right) - \left(\frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} Y(i,j)\right) \tag{15.4}$$

gdzie:
f_4 – błąd jasności,
N – szerokość obrazu,
M – wysokość obrazu,
i – kolejna linia obrazu,
j – kolejny piksel z linii obrazu,
X – obraz wzorcowy,
Y – obraz porównywany.

Opis chromosomu

Aby algorytm był w stanie rozwiązać problem, który został przed nim postawiony, wymagane jest stworzenie odpowiedniej struktury danych, która będzie mogła zostać wykorzystana jako reprezentacja chromosomu. Struktura stworzona na potrzeby niniejszej pracy badawczej składa się z pięciu opisanych poniżej elementów:

- **Typ filtru** zawiera informacje o tym, jaki typ filtru jest wykonywany w każdym z 4 etapów filtrowania. Możliwe wartości to: dylatacja, erozja, otwarcie i domknięcie.
- **Typ sąsiedztwa** pozwala określić, ile punktów z sąsiedztwa piksela będzie używane do wyznaczenia jego nowej wartości. Możliwe opcje to sąsiedztwo 4-punktowe (czterospójne) oraz sąsiedztwo 8-punktowe (ośmiospójne).
- **Element strukturalny** zawiera korekty wartości punktów otoczenia przetwarzanego piksela. Każdy z dziewięcioelementowego wektora przypisany jest jednemu elementowi strukturalnemu dla każdego z czterech etapów.
- **Wektor przynależności elementów do twardego środka otoczenia piksela** przechowuje wartości prawda/fałsz, określające, czy punkt z otoczenia przetwarzanego piksela należy do jego tzw. twardego centrum (wartość „True") czy do miękkiej granicy (wartość „False").
- **Wektor indeksów porządku** przechowuje informacje określające, ile razy elementy z twardego środka będą umieszczane w zbiorze, na podstawie którego wyznaczona zostanie nowa wartość przetwarzanego piksela.

Opis funkcji celu

Funkcja celu działa na podstawie sumy czterech wartości funkcji błędów: średni błąd bezwzględny, średni błąd kwadratowy, błąd kształtu i błąd jasności opisanymi wzorami (15.1)–(15.4). Błąd powinien być maksymalnie niski, więc – z założenia – poszukujemy minimum globalnego, co oznacza, że im dana suma błędów jest niższa, tym osobnik jest lepiej przystosowany, czyli filtr zbudowany na jego podstawie daje lepsze rezultaty.

Opis przeprowadzonych badań

Podczas implementacji programu zwrócono uwagę na kilka miejsc w algorytmie, których zmiana niosła nadzieję na polepszenie wyników jego działania. Dlatego też podczas pisania aplikacji zaimplementowano możliwość zmiany sposobu działania algorytmu poprzez odpowiednie sparametryzowanie wywoływanych funkcji, a część badań oparto na manipulowaniu właśnie tymi parametrami. Badanie przeprowadzono więc poprzez modyfikację 5 parametrów modyfikujących sposób działania filtrów, potoku filtrów lub sposób generowania nowych osobników w trakcie działania algorytmu genetycznego. Konkludując, przeprowadzono badania przez zbadanie wszystkich permutacji ustawień parametrów, takich jak:

- **Współczynnik wymiany genów** (Przebadane wartości: 0,2; 0,5.). Określa, jaka część chromosomu jest modyfikowana podczas operacji krzyżowania. Dla parametru 0,5 podczas krzyżowania nowy osobnik jest tworzony na podstawie 50% genów każdego z rodziców (następuje wymiana połowy genów pomiędzy kopiami chromosomów biorących udział w krzyżowaniu). W przypadku ustawienia tego parametru

na wartość 20% podczas krzyżowania nowy osobnik tworzony jest na podstawione 80% genów jednego rodzica i 20% genów drugiego z rodziców (następuje wymiana 20 procent informacji między kopiami chromosomów biorących udział w krzyżowaniu). Podczas wstępnych testów zauważono, że duże modyfikacje w chromosomie na zaawansowanym etapie działania algorytmu genetycznego nie przyczyniają się do znajdowania nowych minimów lokalnych w przeszukiwanej przestrzeni rozwiązań.

- **Liczba filtrów w potoku** (Przebadane wartości: 2, 4). Parametr ten odpowiada za maksymalną liczbę możliwych filtrów w potoku filtrów. Dla liczby filtrów równej 2 do potoku możliwe jest dodanie i wykonanie maksymalnie dwóch operatorów morfologicznych, natomiast w przypadku parametru równego 4 liczba operatorów morfologicznych możliwych do użycia wzrasta do 4 i pokrywa się z liczbą filtrów wykorzystanych w pracy [4]. Uwzględnienie zmiany liczby operacji w filtrze miało na celu przebadanie, czy za duża liczba etapów filtrowania obrazu nie ma na niego negatywnego wpływu. Dodatkową opcją tego parametru jest możliwość powiększenia zbioru elementów, z którego wyliczana jest mediana, jeśli flaga odpowiedzialna za zastąpienie potoku filtrów wyliczaniem mediany jest włączona.
- **Zastąpienie potoku filtrów poprzez wyznaczenie mediany** (Możliwe wartości: true, false). Parametr ten pozwala na sterowanie sposobem działania zbioru filtrów. Gdy parametr jest ustawiony na „false", zaimplementowany algorytm wykonuje zbiór filtrów przekazanych za pośrednictwem chromosomu w potoku, jeden po drugim. Pozwala to na stopniową filtrację obrazu poprzez kompensatę strat spowodowanych jednym filtrem poprzez wykonanie kolejnego filtru w potoku. Wykonywanie filtrów w potoku stymuluje dane filtry do współpracy i może mieć znaczny wpływ na podział „obowiązków" każdego z filtrów w potoku. Natomiast gdy flaga odpowiedzialna za używanie mediany jest włączona, to filtry nie wykonują się w potoku, lecz w pełni niezależnie. Tutaj również możliwy jest podział obowiązków przez filtry, każdy z nich może się wyspecjalizować w filtrowaniu konkretnego rodzaju szumu. Finalnie z wyników przetworzenia konkretnego piksela przez każdy z filtrów wybierana jest mediana. W założeniu ma to na celu odrzucenie skrajnych wartości ze zbioru wyników działania filtrów, które pojawiły się tam za sprawą działania operatora, który nie poradził sobie z szumem w danym obszarze.
- **Użycie zmiennych korygujących** (Możliwe wartości: true, false). Parametr ten umożliwia włączenie do obliczeń filtrów morfologicznych, zmiennych korygujących. Zadaniem wspomnianych zmiennych jest wprowadzenie wstępnej korekty wartości pikseli biorących udział w działaniu filtru morfologicznego. W przypadku operacji dylatacji wartość korygująca, np. równa -9 jest dodawana do wskazanego piksela. W przypadku erozji wartość np. -9 jest odejmowana od wskazanego piksela. Istnieją badania [5–7], które wskazują, że taki sposób działania ma potencjał na polepszenie wyników działania filtrów.
- **Użycie rozkładu normalnego do generowania wartości korygujących** (Możliwe wartości: true, false). Podczas wstępnych testów programu zastanawiano się nad wpływem występowania skrajnych wartości zmiennych korygujących na działanie filtrów. Aby to sprawdzić, postanowiono zaimplementować możliwość znacznego zmniejszenia prawdopodobieństwa ich wystąpienia i wykorzystano w tym celu roz-

kład normalny. Gdy opisany tutaj parametr jest włączony, to wartości korygujące przypisane do wszystkich nowych osobników są generowane z prawdopodobieństwem odpowiadającym rozkładowi normalnemu z środkiem w zerze i wartościami skrajnymi -255 i 255. W założeniu powinno to stymulować powstawanie lepiej przystosowanych osobników już podczas tworzenia populacji początkowej, ale także podczas tworzenia nowych osobników na początku każdego nowego pokolenia.

4. Poszukiwanie najlepszego zbioru parametrów stworzonego filtra

Wykonując badania dla wszystkich permutacji parametrów opisanych w sekcji „Opis przeprowadzonych badań", otrzymano 320 wyników. Z tego zbioru wybrano 50 najlepszych filtrów i porównano ich ustawienia. Zestawienie liczby użytych poszczególnych parametrów dla wyselekcjonowanego zbioru przedstawiono w tabeli 15.1. Widać tutaj wyraźnie, że najlepsze rozwiązania znajdowano, gdy stosunek wymiany genów podczas krzyżowania wynosił 20%, używano czterech filtrów przetwarzających obraz w potoku i nie używano wartości korygujących. To, czy wartości korygujące były wyznaczane przy użyciu rozkładu normalnego, nie miało większego znaczenia. Spowodowane jest to faktem, że w większości przypadków zmienne korygujące są nieużywane podczas działania filtru.

Tabela 15.1. Tabela zawierająca statystyki użytych parametrów w 50 najlepszych uruchomieniach algorytmu genetycznego

Prawdopodobieństwo krzyżowania genu	Ile razy użyto
0.2	29
0.5	21
Liczba filtrów w potoku	Ile razy użyto
2	9
4	41
Czy użyto mediany zamiast potoku	Ile razy użyto
Tak	6
Nie	44
Czy użyto wartości korygujących	Ile razy użyto
Tak	17
Nie	33
Czy użyto rozkładu Gaussa	Ile razy użyto
Tak	26
Nie	24

5. Analiza najlepszego znalezionego filtru

Po kilkugodzinnym działaniu programu badawczego, najlepszym znalezionym filtrem okazał się osobnik o parametrach:

- wektor filtrów = {„Dylatacja", „Otwarcie", „Erozja", „Otwarcie"},
- rodzaj współpracy filtrów: połączenie szeregowe (potok),
- typy sąsiedztwa = {4, 8, 8, 8},

- użycie wartości korygujących = false,
- wartości korygujące = {},
- tablica wektorów przynależności punktów do twardego środka:

$$\left\{ \begin{bmatrix} \text{Nie} & \text{Nie} & \text{Nie} \\ \text{Tak} & \text{Nie} & \text{Tak} \\ \text{Nie} & \text{Tak} & \text{Nie} \end{bmatrix}, \begin{bmatrix} \text{Tak} & \text{Nie} & \text{Tak} \\ \text{Tak} & \text{Nie} & \text{Tak} \\ \text{Nie} & \text{Tak} & \text{Nie} \end{bmatrix}, \begin{bmatrix} \text{Nie} & \text{Tak} & \text{Nie} \\ \text{Tak} & \text{Tak} & \text{Tak} \\ \text{Tak} & \text{Nie} & \text{Nie} \end{bmatrix}, \begin{bmatrix} \text{Nie} & \text{Tak} & \text{Nie} \\ \text{Tak} & \text{Tak} & \text{Tak} \\ \text{Nie} & \text{Nie} & \text{Tak} \end{bmatrix} \right\}$$

- wektor parametrów rang: {8, 7, 2, 3}.

Przedstawione parametry użyto do stworzenia funkcji filtrującej. Wyniki działania tej funkcji przedstawiono w dalszych etapach pracy.

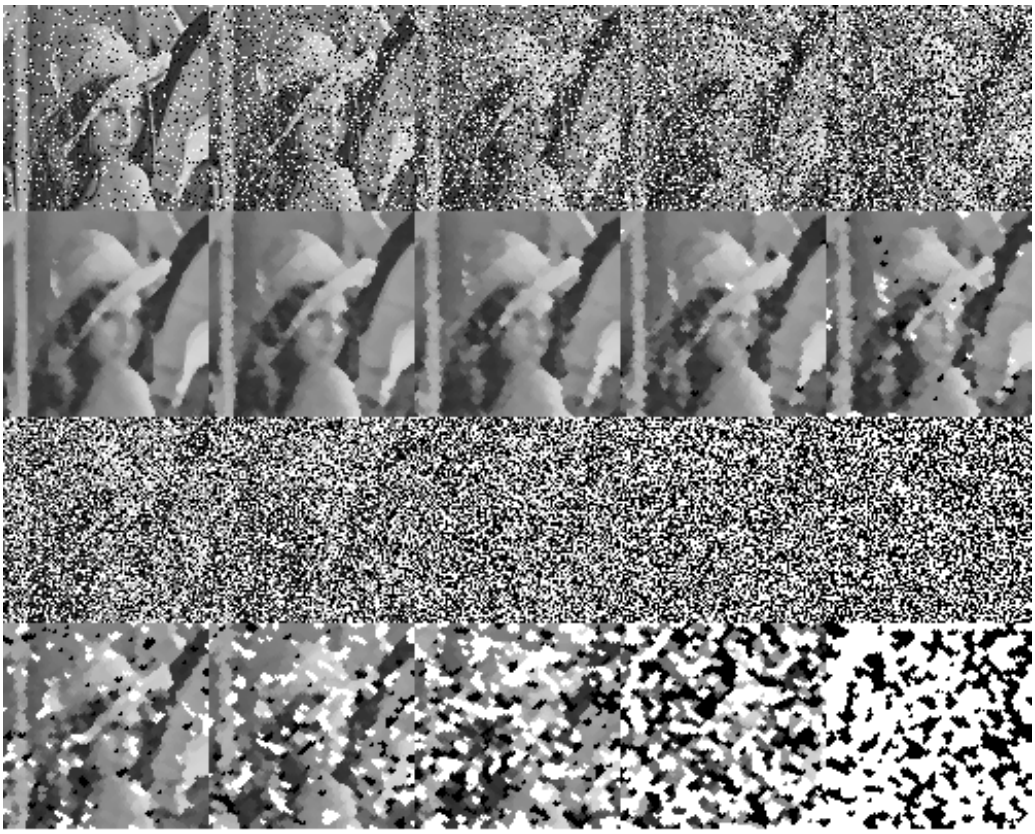
5.1. Wyniki działania filtru przedstawione na obrazie lena

Podstawowe obrazy używane do oceny jakości działania filtrów miały rozdzielczość 100×100 pikseli. Przykładową taką grafiką była fotografia „lena" przedstawiona na rysunku 15.1.



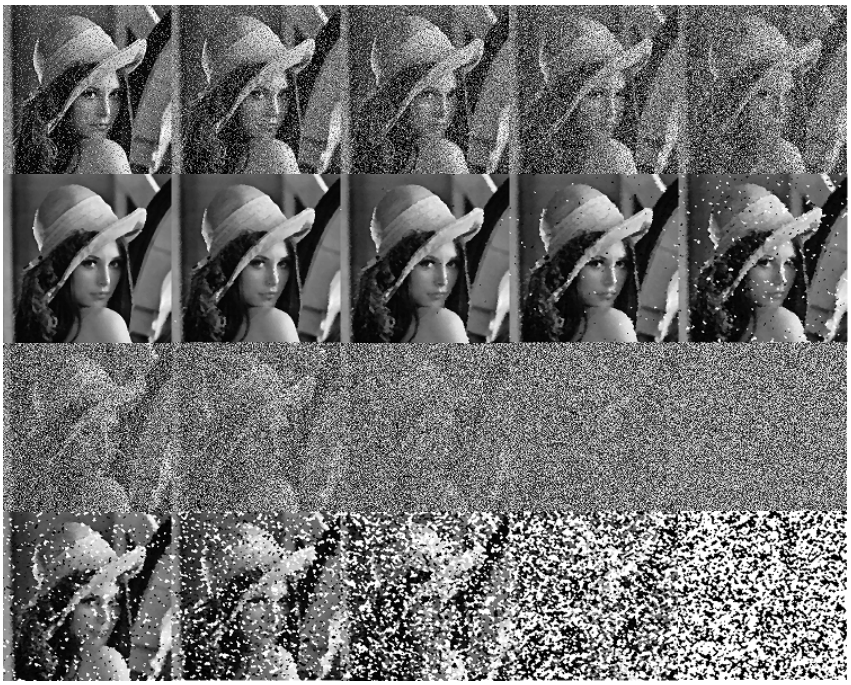
Rysunek 15.1. Nieprzetworzony obraz lena (rozmiar 100×100 px)

W celach badawczych znaleziony filtr wykorzystano do odszumowania grafiki „lena". Wyniki tego odszumiania przedstawiono na rysunku 15.2. Grafika przedstawia zaszumione obrazy szumem typu sól–pieprz. Szum ten, znany również jako szum impulsowy, objawia się modyfikacją kolorów pikseli obrazu cyfrowego, zmieniając intensywność losowych pikseli na wartość minimalną (kolor czarny) lub maksymalną (kolor biały) [1]. Pierwsze zdjęcie, umieszczone w lewym górnym rogu, przedstawia obraz zaszumiony w 10%, a w każdym kolejnym obrazie poziom szumu powiększa się o 10% aż do poziomu 100% dla ostatniego zdjęcia w trzecim wierszu. Zdjęcia znajdujące się pod fotografiami z wprowadzonymi zakłóceniami przedstawiają obraz po odfiltrowaniu. Przedstawiona grafika wyraźnie pokazuje, że operacja filtrowania znacznie poprawia czytelność obrazów, jednakże gdy poziom szumów przekroczy 60%, to operacja filtracji nie poprawia w wyraźny sposób czytelności obrazu. Obraz z szumem na poziomie 100% przedstawiono w celu pokazania zachowania algorytmu dla obrazu składającego się wyłącznie z czystego szumu. Widać tam wyraźnie działanie filtrów morfologicznych oraz to, jak próbują one doszukiwać się kształtów w przetwarzanych obrazach i wygładzać ich krawędzie bądź też uzupełniać brakujące piksele.

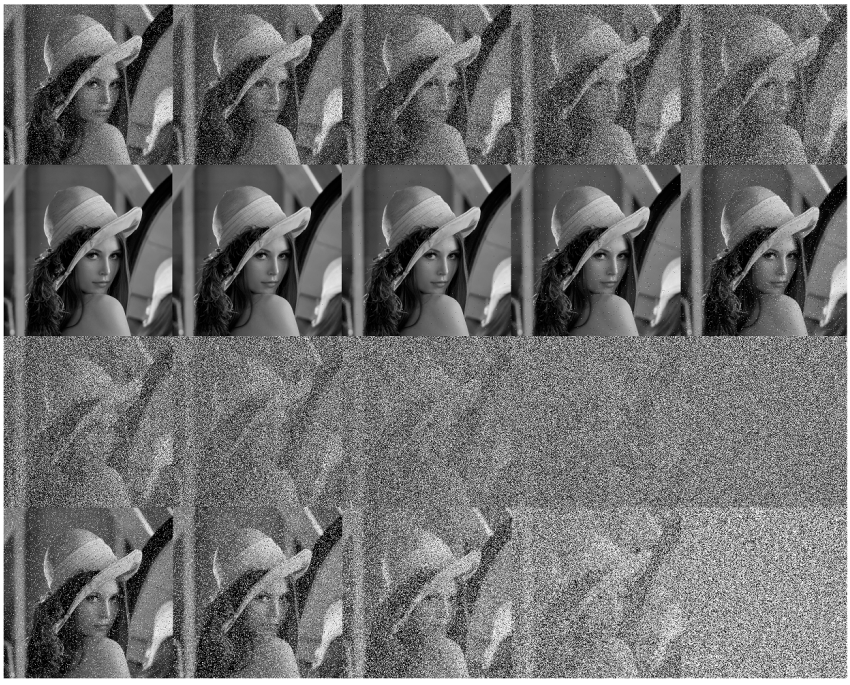


Rysunek 15.2. Wyniki działania powstałego filtru morfologicznego na obrazie lena (szum 10%–100%, rozmiar 100×100 px)

Obrazy o małych rozmiarach, takie jak przedstawione powyżej, mają kilka znaczących wad, mianowicie – ich czytelność potrafi być mocno utrudniona z powodu małej liczby pikseli opisujących ich zawartość. Ponadto, takie zdjęcia często powstają w wyniku skalowania zdjęć o większej rozdzielczości, co polega na uśrednianiu wielu pikseli do jednej wartości. Taki zabieg sprawia często to, że po przeskalowaniu obrazu jeden piksel opisuje wiele elementów przedstawionych na oryginalnym zdjęciu. Fakt ten sprawia, że funkcje filtrujące mają problem z wyznaczaniem nowych wartości pikseli utraconych w wyniku nałożenia szumu na obraz. Dlatego też filtry graficzne, a co za tym także idzie – miękkie filtry morfologiczne, dużo lepiej pokazują swoje działanie na obrazach z większą liczbą pikseli. Takim obrazem jest np. wspomniana już wcześniej fotografia „lena", ale o rozdzielczości 300×300 pikseli. Grafika 15.3 przedstawia, wyniki filtracji obrazu „lena" o rozmiarze 300×300 pikseli. Można tutaj zauważyć, że obraz o większej rozdzielczości dużo łatwiej poddaje się regeneracji w procesie filtracji. Na przedstawionej grafice widać, że znaleziony filtr w dużym stopniu poprawia jakość obrazów z szumem na poziomie 10–40%. Dla obrazów z szumem na poziomie 50–60% filtr także w znacznym stopniu poprawia czytelność obrazu, lecz pojawiają się tam już pewne artefakty. Dla wyższych poziomów zaszumienia widać znaczną poprawę jakości spowodowaną działaniem filtrów, lecz obraz nawet po operacji filtrowania jest trudny do odczytania.



Rysunek 15.3. Wyniki działania powstałego filtru morfologicznego na obrazie lena (szum 10%–100%, rozmiar 300×300 px)



Rysunek 15.4. Wyniki działania powstałego filtru morfologicznego na obrazie lena (szum 10%–100%, rozmiar 1960×1960 px)

Przeprowadzono również badania na zdecydowanie większych obrazach. Grafika 15.4 przedstawia wyniki działania powstałego filtru morfologicznego na obrazie „lena" o rozdzielczości 1980×1980 pikseli. Dla tak dużej rozdzielczości obraz potrafi być czytelny nawet po odfiltrowaniu szumu na poziomie 70%. Warto tutaj podkreślić, co zostało zauważone również w poprzednim zestawieniu, że obrazy z szumem na poziomie 10–40% mogą być rekonstruowane do tego stopnia, że wstępnie porównując obraz oryginalny i odzyskany, można nie zauważyć różnicy. Dla obrazów o większym stopniu zaszumienia po filtracji widać już zauważalne artefakty na odfiltrowanym obrazie.

6. Porównanie z innymi filtrami

Na dalszym etapie prac postanowiono porównać wydajność stworzonego filtru z wynikami z publikacji naukowych opisanych w rozdziale drugim mojej pracy magisterskiej oraz z wydajnością innych filtrów znanych z literatury. Do porównania wybrano filtr medianowy, który jest jednym z popularniejszych filtrów używanych w grafice. Jego działanie opiera się na zastępowaniu utraconych wartości medianą wyznaczoną z sąsiedztwa przetwarzanego punktu. Dokładny opis jego działania można znaleźć w książce [8]. Kolejnym rozpatrzonym filtrem był adaptacyjny ważony filtr medianowy [9]. W celach badawczych użyto implementacji danych filtrów dostępnych w pakiecie Image processing Toolbox dla programu Matlab [11].

Niektóre wyniki filtrów użytych do porównania były kopiowane bezpośrednio z publikacji. Owe publikacje nie przedstawiały wyników MSE i MAE dla szumu powyżej 20%, dlatego większość wyników jest nieznana. Reszta wyżej opisanych filtrów była testowana na obrazie „lena" o rozmiarach 300×300 pikseli. Wszystkie te filtry były oceniane tymi samymi implementacjami funkcji oceny, co filtr morfologiczny opracowany w tej pracy dyplomowej. Na obrazie 15.2 przedstawiono wysokość średniego błędu bezwzględnego dla każdego z porównywanych filtrów. Opis kolumn wygląda następująco:

- kolumna „Median" odpowiada wartości błędu dla filtru medianowego,
- kolumna „MWF" przechowuje wartości błędów dla adaptacyjnego filtru medianowego,
- kolumna „GF" przechowuje wyniki dla filtru gaussowskiego,
- kolumna „GA2" zawiera wyniki miękkiego filtru morfologicznego opisanego w publikacji „Soft morphological filter optimization using a genetic algorithm for noise elimination" [5],
- kolumna „TS" zawiera wyniki dla miękkiego filtru morfologicznego opisanego w publikacji „Optimization of Soft Morphological Filter based on Tabu Search" [12],
- kolumna „GW" zawiera wyniki dla miękkiego filtru morfologicznego opisanego w publikacji „Optimization of Soft Morphological Filters with Parallel Annealing-Genetic Strategy" [13],
- kolumna „Soft morphology pipeline"' zawiera wyniki potoku miękkich filtrów morfologicznych, który jest efektem badań przeprowadzonych w tej pracy dyplomowej.

Porównując zaprezentowane wyniki można zauważyć, że omawiany potok filtrów wykazuje najlepsze właściwości ze wszystkich zaprezentowanych filtrów dla szumu na poziomie 50–60 %, a w pozostałych zakresach niewiele odbiega od reszty.

Tabela 15.2. Tabela wartości średniego błędu bezwzględnego dla różnych filtrów

Poziom zaszumienia	Median	AVM	GF	Soft morphology pipeline	GA2	TS	GW
0.1	1.89	0.27	5.29	2.46		2.06	4.01
0.2	2.23	0.66	7.22	2.65	2.52	3.31	6.77
0.3	3.06	1.32	9.20	2.93			
0.4	4.59	2.65	11.38	3.46			
0.5	7.55	5.13	13.35	4.88			
0.6	12.30	9.52	15.39	8.30			
0.7	19.41	15.67	17.94	14.70			
0.8	28.30	24.26	19.77	25.74			
0.9	39.07	36.02	22.36	38.54			

Tabela 15.3 prezentuje średnie wartości błędu kwadratowego dla poszczególnych filtrów. Również tutaj zaprezentowany filtr okazał się najlepszy do odszumiania obrazu z szumem na poziomie 40–50%. Filtr też pokazuje porównywalne wyniki z innymi filtrami dla szumów na poziomie 30% i 60%. Kolumna „UGA" zawiera wyniki dla miękkiego filtru morfologicznego opisanego w publikacji "Optimization design of soft morphological filters based on improving genetic algorithm" [14].

Tabela 15.3. Tabela wartości średniego błędu kwadratowego dla różnych filtrów

Poziom zaszumienia	Median	AVM	GF	Soft morphology pipeline	GA2	UGA	TS	GW
0.1	49.17	7.69	148.38	75.77		156.90	49.90	153.65
0.2	77.06	30.96	230.60	86.11	31.91	255.70	91.20	261.64
0.3	187.19	98.41	349.22	105.59				
0.4	425.36	272.82	501.62	164.08				
0.5	902.93	654.47	676.07	378.33				
0.6	1716.54	1380.65	890.19	982.45				
0.7	2978.20	2413.59	1181.60	2143.92				
0.8	4530.26	3886.23	1448.73	4214.93				
0.9	6466.42	5956.17	1826.26	6576.30				

W tabeli 15.4 przedstawiono wielkości błędu jasności. Widać tutaj, że filtr ma porównywalne wartości błędów z adaptacyjnym ważonym filtrem medianowym, który okazał się być najlepszy w tej kategorii.

Tabela 15.4. Tabela wartości błędu jasności dla różnych filtrów

Poziom zaszumienia	Median	AWF	GF	Soft morphology pipeline
0.1	26.25	3.00	143.70	37.87
0.2	26.29	5.37	291.17	23.36
0.3	11.10	1.46	456.05	5.48
0.4	37.12	23.95	611.22	32.15
0.5	103.96	81.19	736.73	127.61
0.6	230.78	197.67	882.79	370.94
0.7	459.17	379.18	1052.18	821.04
0.8	711.63	672.30	1180.11	1529.60
0.9	1165.19	981.79	1387.62	2223.93

W tabeli 15.5 przedstawiono ostatnią rozpatrywaną miarę oceny jakości filtru, czyli błąd kształtu. Najlepszy okazał się filtr gaussowski. Potok miękkich filtrów morfologicznych okazał się być najlepszy dla szumu na poziomie 30%.

Tabela 15.5. Tabela wartości błędu kształtu dla różnych filtrów

Poziom zaszumienia	Median	AWF	GF	Soft morphology pipeline
0.1	0.06	0.01	0.09	0.07
0.2	0.08	0.03	0.09	0.08
0.3	0.19	0.10	0.10	0.09
0.4	0.40	0.27	0.10	0.13
0.5	0.79	0.62	0.10	0.24
0.6	1.38	1.20	0.11	0.51
0.7	2.18	1.94	0.11	0.94
0.8	3.06	2.88	0.12	1.55
0.9	4.07	4.02	0.12	2.15

W tabeli 15.6 przedstawiono sumy wszystkich błędów. Autorski miękki filtr morfologiczny okazał się najlepszy podczas filtrowania obrazów z szumem na poziomie 40–60%. Uzyskał także zadowalające wyniki dla obrazów zaszumionych w 10%, 20% i 30%. Dla 70% i 80% wyniki jego działania są porównywalne ze zwykłym filtrem medianowym.

Tabela 15.6. Tabela reprezentująca wartości sumy błędu dla różnych filtrów

Poziom zaszumienia	Median	AWF	GF	Soft morphology pipeline
0.1	77.36	10.96	297.45	116.16
0.2	105.66	37.03	529.09	112.20
0.3	201.54	101.30	814.56	114.08
0.4	467.48	299.70	1124.32	199.81
0.5	1015.22	741.41	1426.25	511.06
0.6	1960.99	1589.04	1788.47	1362.21
0.7	3458.97	2810.38	2251.83	2980.61
0.8	5273.24	4585.66	2648.72	5771.82
0.9	7674.75	6978.00	3236.36	8840.92

7. Podsumowanie

W ramach pracy dyplomowej przeanalizowano literaturę naukową traktującą o filtrach graficznych, a w szczególności o filtrach morfologicznych. Po głębszej analizie prac postanowiono skupić się na miękkiej odmianie filtrów morfologicznych i – bazując na publikacji „Soft morphological filter optimization using a genetic algorithm for noise elimination" [5] – zaimplementować algorytm pozwalający na maksymalne wykorzystanie ich możliwości. Dodatkowo zaproponowano kilka autorskich rozwiązań mających na celu wspieranie tworzenia coraz lepszych filtrów na dalszych etapach działania aplikacji badawczej. Aby znaleźć zbiór parametrów wejściowych miękkich filtrów morfologicznych, wykorzystano algorytm genetyczny. W tym celu stworzono aplikację badawczą, która implementowała cztery parametryzowane operatory miękkiej filtracji morfologicznej, tj. erozję, dylatację, otwarcie i domknięcie. Zaimplementowano także cztery funkcje oceny zgodności oryginalnego obrazu z obrazem po filtracji. W dalszej części pracy zajęto się implementacją algorytmu genetycznego służącego do poszukiwań najlepszego, 88-genowego chromosomu, reprezentującego zbiór parametrów stworzonej funkcji filtrującej. W kolejnym kroku wykorzystano zaimplementowany algorytm do przeszukiwań przestrzeni rozwiązań dla 32 różnych kombinacji parametrów początkowych i pięciu różnych obrazów. Podczas badań otrzymano 320 wyniki i poddano je szczegółowej analizie. W trakcie analizy znaleziono

i opisano trendy panujące w zbiorze najlepszych wygenerowanych filtrów, przedstawiono ich cechy wspólne i wyselekcjonowano najlepszy zbiór parametrów. Wyselekcjonowany zbiór wykorzystano do stworzenia funkcji filtrującej i empirycznie przedstawiono wyniki jej działania. Na ostatnim etapie pracy porównano otrzymane wyniki z wynikami innych znanych z literatury filtrów graficznych.

Filtr opracowany podczas badań opisanych w tej pracy osiągał bardzo dobre rezultaty, a z szumem na poziomie 40–60% radził sobie nadzwyczaj dobrze w porównaniu do innych filtrów. Podczas badań zauważono także, że filtry graficzne radzą sobie zdecydowanie lepiej z obrazami o rozdzielczości pozwalającej na zapisywanie obiektów znajdujących się na nich w postaci więcej niż jednego piksela, a najlepiej w postaci kilkunastu lub kilkudziesięciu.

Zaprojektowana i stworzona aplikacja badawcza ma duży potencjał rozwojowy. Jednym z czynników, który mógłby zostać przebadany, jest wpływ dynamicznie zmieniającego się współczynnika wymiany genów na powstające osobniki lub wpływ wielkości i kształtu elementu strukturalnego na jakość filtracji.

Bibliografia

[1] Tadeusiewicz R., Korohoda P., *Komputerowa analiza i przetwarzanie obrazów*; Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.

[2] Michalewicz Z., *Algorytmy genetyczne + struktury danych = program ewolucyjne*, Wydawnictwo Naukowo-Techniczne, Warszawa 1999.

[3] Tumański S., *Technika pomiarowa*, Wydawnictwo Naukowe PWN, Warszawa 2016.

[4] Abreu E., Lightstone M., Mitra S.K., Arakawa K., *A new efficient approach for the removal of impulse noise from highly corrupted images, Image Processinga*, IEEE, **1996**.

[5] Erçal T., Özcan E., Asta S., *Soft morphological filter optimization using a genetic algorithm for noise elimination*, IEEE, **2014**.

[6] Li D., Chongzhao H., *Research on Edge Detection of Gray-Scale Image Based on Mathematical Morphology Algorithm and Rough Sets*, IEEE, **2009**.

[7] Urbach E., Wilkinson M., *Efficient 2-D Gray-Scale Dilations and Erosions with Arbitrary Flat Structuring Elements*, IEEE, **2006**.

[8] Zieliński T., *Cyfrowe przetwarzanie sygnałów od teorii do zastosowań*, Wydawnictwo Komunikacji i Łączności, Warszawa 2005.

[9] Chen T., Wu H.R., *Adaptive impulse detection using center-weighted median filters*, IEEE, **2001**.

[10] Akansu A., Haddad R.A., *A class of fast Gaussian binomial filters for speech and image processing*, IEEE, **1991**.

[11] Dokumenctacja implementacji filtru medianowego dla aplikacji „Matlab". `https://www.mathworks.com/help/images/ref/medfilt2.html` (dostęp: 29-05-2022).

[12] Zhao C., Liang G., Ning H., *Optimization of Soft Morphological Filter based on Tabu Search*, IEEE, **2006**.

[13] Tian Y., Zhao C., *Optimization of Soft Morphological Filters with Parallel Annealing-Genetic Strategy*, IEEE, **2010**.

[14] Chunhui Z., *Optimization design of soft morphological filters based on improving genetic algorithm*, IEEE, **2003**.

[15] KoskinenL., Astola J., *Soft morphological filters*, San Diego 1991.

Modelowanie interakcji społecznych na podstawie strumienia danych

mgr inż. Michał Wiczk
Informatyka Stosowana

dr hab. inż. Radosław Michalski
Katedra Sztucznej Inteligencji

Praca przedstawia różne sposoby rozwiązania problemu reprezentacji danych dotyczących interakcji społecznych. Jest to istotne ze względu na fakt, że tego typu interakcji dokonujemy coraz więcej. Z tego powodu pojawia się problem, w jaki sposób powinniśmy takie dane reprezentować, aby ułatwić ich analizę i przetwarzanie. Interakcje, które otrzymujemy w formie na przykład strumienia zdarzeń, można nanieść na strukturę matematyczną, jaką jest graf, reprezentujący sieć społeczną. Dzięki swojej strukturze sieciowej zasadne jest zastosowanie popularnych ostatnio technik osadzania w grafach. Pozyskane tak informacje w postaci wektorowej stanowią dane wejściowe dla modeli uczenia głębokiego. W pracy porównano różne metody modelowania interakcji społecznych opartych na sieciach neuronowych i rozszerzonych o różne techniki osadzania w grafach statycznych. Zastosowano także autorskie podejście przez wprowadzenie grafów dynamicznych. Wyniki z przeprowadzonych eksperymentów ukazują minimalną przewagę modeli wzbogaconych o etap osadzania i grafy dynamiczne nad pierwotną metodą modelowania.

1. Cel i założenia

Celem przeprowadzanego i opisanego w niniejszym streszczeniu projektu badawczo-rozwojowego jest dokonanie analizy już istniejących i opracowanie własnych metod modelowania interakcji społecznych na podstawie strumieni danych. Jego przeznaczeniem jest chęć poszerzenia znanej wiedzy dotyczącej modelowania relacji społecznych. Przed przystąpieniem do badań dokonano opracowania teoretycznego. W wyniku analiz otrzymano pięć modeli, które stanowiły podstawę do badań porównawczych. Następnie zaplanowano metodykę dla eksperymentów. Ich szczegółowe założenia przedstawiono poniżej:

- sprawdzenie wpływu zastosowania różnych technik osadzania zdarzeń na jakość modelu interakcji społecznych,
- porównanie skuteczności metod osadzania w grafach do modelowania interakcji społecznych,

- porównanie skuteczności zastosowania grafów statycznych i dynamicznych w technikach osadzania do modelowania interakcji społecznych.

Wyniki badań szczegółowo omówiono oraz zobrazowano w formie wykresów i tabel. Dokonano także krytycznej oceny eksperymentów. Wskazano możliwości dalszego rozwoju badań w tej tematyce.

2. Wstęp teoretyczny

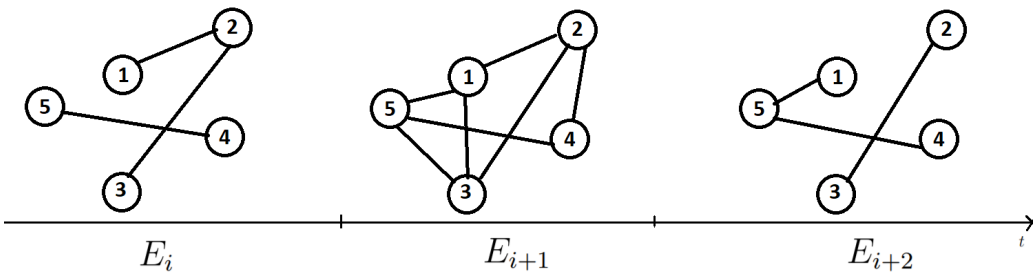
Interakcje międzyludzkie można definiować jako dowolny proces, na który składają się wzajemne reakcje między co najmniej dwiema osobami. Ich analiza pozwala na wnioskowanie dotyczące organizacji społeczeństwa i jego struktury [1].

2.1. Strumienie danych, grafy i procesy punktowe

Strumienie danych stanowiące źródło informacji dla analizowanych modeli określane są jako statystyczny opis, który reprezentuje przybywanie zgłoszeń (zdarzeń) do systemu [2]. Otrzymywane dane z łatwością można nanieść na strukturę grafu. Oprócz tradycyjnej definicji grafu G jako zbioru wierzchołków V oraz krawędzi E istotnym pojęciem są grafy temporalne (czasowe). Stanowią one specjalny rodzaj grafów, w których zbiór krawędzi zmienia się w kolejnych jednostkach czasu. Formalnie graf czasowy opisuje równanie (16.1). Istotę tego typu struktury ilustruje rysunek 16.1.

$$G_T = (G, f_T) \tag{16.1}$$

gdzie:
G_T – graf temporalny,
G – graf (bazowy); $G = (V, E)$, gdzie $E = \sum_{t=0}^{T} E_t$,
$f_T : T \to E_t$ – funkcja oznaczania czasu,
T – czas (okres) definiowany przez graf temporalny (może być nieskończony),
t – krok czasowy $t \in [0; T]$.



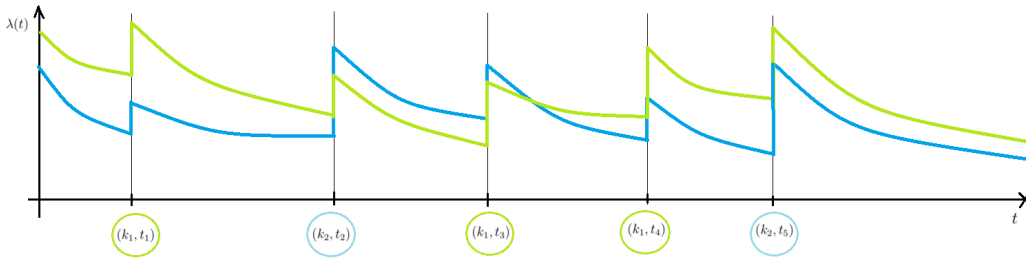
Rysunek 16.1. Przykład grafu czasowego

Interakcje społeczne w ogólnej postaci są zdarzeniami. W celu modelowania tego typu zjawisk często wykorzystuje się procesy punktowe (ang. *point process*). Procesy punktowe

to zbiory punktów losowo umieszczonych w dowolnej n-wymiarowej przestrzeni matematycznej. Istnieje kilka rodzajów procesów punktowych.

2.2. Neuronowe procesy Hawkesa

Dla rozważanej dziedziny najlepszym modelem są procesy punktowe Hawkesa nazywane również procesami „samoistniejącymi" (ang. *self-existing*). Oznacza to, że przeszłe zdarzenia zwiększają prawdopodobieństwo (intensywność) wystąpienia zdarzenia w przyszłości. Taki wpływ charakteryzuje się następującymi cechami: dodatniość wartości wpływu, addytywność względem przeszłych zdarzeń oraz wykładniczy zanik wartości wpływu wraz z upływem czasu. Co więcej, wydarzenia danego typu nie muszą powodować wzrostu wartości funkcji intensywności (wzbudzać) tylko tego samego rodzaju zdarzenia. Charakterystykę tych właściwości przedstawia wykres na rysunku 16.2. Na wykresie widoczne są wszystkie trzy cechy procesów punktowych Hawkesa.



Rysunek 16.2. Przykład wykresu funkcji intensywności dla procesów punktowych Hawkesa z naniesionymi pięcioma zdarzeniami różnego typu k ($k \in k_1, k_2$) (na podstawie [3])

Jednak w przypadku niektórych zjawisk cechy charakterystyczne dla procesów Hawkesa mogą stanowić jego ograniczenie. W przytoczonej pracy [3] zaproponowano modyfikację ograniczeń dotyczących dodatniości wartości wpływu i addytywności względem przeszłych zdarzeń. Modyfikacje zostały przeprowadzone w taki sposób, aby przy niezmienionej strukturze modelu zwiększyć jego wyrazistość i obszar zastosowania. Implementację opisanych modyfikacji umożliwia wykorzystanie sztucznej, rekurencyjnej sieci neuronowej LSTM (Long Short Term Memory). Architektura LSTM przetwarza sekwencje danych z dodatkowo wprowadzonym sprzężeniem zwrotnym. Rozwiązuje ona podstawowe problemu zwykłych sieci neuronowych, tj. wstecznie propagowane gradienty mogą zanikać (ang. *vanish*), a więc dążyć do zera lub eksplodować (ang. *explode*) i dążyć do nieskończoności. Sieci LSTM posiadają strukturę podobną do łańcucha kopii tradycyjnej sieci neuronowej. Z tą różnicą, że w komórce pamięci (wewnętrzny moduł sieci) zapisuje się dany stan w pewnym kroku czasowym. Dzięki temu uzyskano możliwość nauki złożonych zależności między czasem, kolejnością i liczbą zdarzeń [4]. Ostatecznie otrzymany model nazwano procesami neuronowymi Hawkesa (ang. The Neural Hawkes Process) [3].

2.3. Grafowe, neuronowe procesy Hawkesa

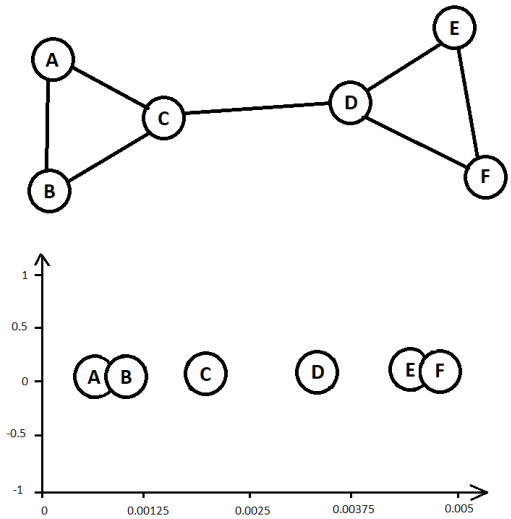
W artykule [5] również rozważa się problem modelowania zdarzeń z wykorzystaniem procesów punktów Hawkesa opartego na sieciach neuronowych. Rozwiązanie będące obiektem badań przedstawionych w cytowanej pracy rozszerza jednak model „Neural Hawkes Process" o informacje przestrzenne i wzajemne wpływy między różnymi wydarzeniami. W tym celu wykorzystuje on metodę osadzania danych o strukturze grafu. Techniki osadzania sieci (ang. *network embedding*) są stosowane w celu przekształcenia wielowymiarowej sieci w nisko wymiarową przestrzeń wektorową, którą jesteśmy w stanie wykorzystać w większości zadań analizy. Całość metody składa się z etapów osadzania grafu i uczenia modelu opartego na sieci neuronowej.

Pierwsza technika osadzania grafu polega na wykorzystaniu splotowych sieci neuronowych do otrzymania cech przestrzennych grafów zbudowanych w oparciu o pozyskane dane zdarzeń. Struktura sieci neuronowej została określona w oryginalnej pracy jako grafowa sieć splotowa (ang. *Graph Convolutional Network*). Ogólny model został nazwany jako procesy neuronowe Hawkesa z metodą osadzania zdarzeń opartą na sieci splotowej. Głównym celem związanym z siecią GCN jest wygenerowanie reprezentacji dla każdego z wierzchołków grafu, korzystając zarówno z ich cech, jak i cech wierzchołków sąsiadujących [5]. Jako parametry wejściowe sieć otrzymuje macierz sąsiedztwa dla wierzchołków grafu zdarzeń. Konieczne jest także wyznaczenie macierzy cech wierzchołków i macierzy stopni grafu. Na wyjściu sieci otrzymujemy wektor cech przestrzennych dla wierzchołków rozpatrywanego grafu zdarzeń. Dzięki tej technice informacje o wierzchołkach grafu i ich cechach ukazane są na przestrzeni euklidesowej. Zobrazowanie działania techniki osadzania grafów z wykorzystaniem GCN pokazano na rysunku 16.3.

Inną techniką osadzania grafu jest metoda weg2vec (ang. *weighted event graph to vector*) opisana w pracy [6]. Algorytm oparty jest na technikach uczenia nienadzorowanego. Rozważany problem jest częściowo tożsamy z metodą analizy skupień. Opiera się ona na strukturalnych i czasowych podobieństwach zdarzeń. Przedstawiona metoda odwzorowania wychwytuje ukryte cechy struktury i podobieństwa między zdarzeniami obejmującymi różne węzły w różnym zakresie czasu. Bierze pod uwagę zarówno przeszłe, jak i przyszłe zdarzenia. Całość metody weg2vec opiera się na zbudowaniu ważonego grafu zdarzeń z wejściowego grafu temporalnego, próbkowaniu sąsiedztw w ważonym grafie zdarzeń oraz przekazaniu sąsiedztw do modelu Skip-Gram. Ostateczny model określono jako procesy neuronowe Hawkesa z metodą osadzania zdarzeń opartą na technice weg2vec.

Drugi etap modelu grafowych i neuronowych procesów Hawkesa jest podobny do modelu przedstawionego w pracy [3] i podrozdziale 2.2. z modyfikacją, polegającą na tym, że oprócz sekwencji zdarzeń reprezentowanych przez zbiór par (znacznik zdarzenia k, czas zdarzenia t) do modelu przekazuje także informacje o cechach przestrzennych zdarzeń g otrzymaną z procesu osadzania grafu, co pokazano we wzorze (16.2).

$$S = (k_l, t_l, g_l)_{l=1}^{L} \tag{16.2}$$



Rysunek 16.3. Uproszczony przykład działania metody osadzania grafów z wykorzystaniem GCN

2.4. Wykorzystanie grafów czasowych

Wykorzystanie grafów czasowych stanowi autorską modyfikację dla metod ukazanych w powyższych podrozdziałach i wprowadza nowe podejście do rozważanego problemu osadzania interakcji. W przytoczonych modelach cechy przestrzenne g były obliczane dla statycznego grafu zdarzeń, zawierającego zarówno wszystkie wierzchołki, jak i krawędzie stanowiące informacje o zastałych interakcjach. Wprowadzona modyfikacja polega na zastąpieniu grafu statycznego grafem temporalnym, w którym cechy przestrzenne g^p są wyliczane dla kolejnych okien czasowych p, na które zostaje podzielony ogół zdarzeń. Liczba okien czasowych jest stała dla wszystkich sekwencji przekazywanych do modelu. Ich długość może być różna, ponieważ jest zależna od wartości znacznika czasowego pierwszego i ostatniego zdarzenia w sekwencji Sp. Liczba interakcji w danych oknach czasowych (nawet w obrębie tej samej sekwencji) może być różna.

$$S = (k_l, t_l, g_l^p)_{l=1}^{L} \tag{16.3}$$

3. Projekt i wykonanie

Rozdział opisuje przyjętą metodykę przeprowadzonych badań i stanowi opis użytych zbiorów danych oraz wykorzystanych narzędzi.

3.1. Zbiory danych

W celu wykonania niezbędnych badań przygotowano dwa zbiory danych. Zbiory różnią się zarówno liczbą interakcji, liczbą ich rodzajów, jak i liczbą osób zaangażowanych w analizowane relacje społeczne.

Zbiór danych NetHealth jest jednym z rezultatów wieloletniego projektu badawczego dotyczącego relacji społecznych i zdrowia realizowanego przez Centrum Nauki o Sieci i Danych na Uniwersytecie Notre Dame w stanie Indiana w Stanach Zjednoczonych [7]. Jako drugi zbiór danych zdecydowano się wybrać sztuczny zbiór wygenerowanych interakcji, gdyż jest to popularny zabieg w podobnych pracach [3, 5]. Umożliwia on kontrolowanie cech charakterystycznych zbioru przez zmianę wartości parametrów generatora. Pozwala na przetestowanie rozmaitych sytuacji rozkładu danych, co nie zawsze jest możliwe w przypadku zbiorów rzeczywistych.

3.2. Narzędzia i środowisko badawcze

Implementację oraz badania przeprowadzono z wykorzystaniem języka programowania Python w środowisku Jupyter. Do implementacji sztucznej sieci neuronowej użyto biblioteki PyTorch. Kod źródłowy w formie notatników w formacie .ipynb stworzony w ramach tej pracy został umieszczony na platformie github.com i jest dostępny pod zamieszczonym w przypisie adresem[1].

3.3. Metodyka badań

W rezultacie analizy teoretycznej do przeprowadzonych badań wyszczególniono następujące modele:

- Neuronowy Hawkes (Neural Hawkes),
- Neuronowy Hawkes z metodą osadzania interakcji opartą na sieci splotowej (GCN Neural Hawkes),
- Czasowy, neuronowy Hawkes z metodą osadzania interakcji opartą na sieci splotowej Temporal GCN Neural Hawkes),
- Neuronowy Hawkes z metodą osadzania interakcji opartą na technice weg2vec (Weg2vec Neural Hawkes),
- Czasowy, neuronowy Hawkes z metodą osadzania interakcji opartą na technice weg2vec (Temporal Weg2vec Neural Hawkes).

Dla każdej z instancji parametrów badania funkcja dokonująca ewaluacji modelu została uruchomiona dwa razy, a otrzymane wyniki stanowią średnią arytmetyczną z ich wartości.

W badaniu można wyróżnić następujące czynniki mające wpływ na obiekt:

1. parametry poddane badaniom
   (a) użyty zbiór danych (zbiór NetHealth, zbiór sztuczny), więc pośrednio także liczba rodzajów interakcji etypes,
   (b) użyty model (nn, gcn, tgcn, w2v, tw2v),
   (c) $\mathrm{card}(S_i)$ – liczba interakcji w sekwencji,
   (d) p – liczba okien czasowych
2. parametry których wartości przyjęto jako stałe
   (a) δt – parament, którego wartość warunkuje fakt δt-przyległości interakcji,
   (b) nb – liczba środowisk interakcji,
   (c) s – liczność każdego ze środowisk interakcji,
   (d) α – współczynnik udziału wag ścieżki oraz δt-przyległości,

[1] `https://github.com/micwiccode/social-interactions`

(e) n_{epoch} – liczba epok ucznia sztucznej sieci neuronowej,
(f) n_{batch} – liczba próbek podawanych uczeniu jednocześnie,
(g) n_{hidden} – liczba cech w stanie ukrytym sieci neuronowej,
(h) $f(x)$ – użyta funkcja transformująca w sieci Neural Hawkes,
(i) σ – funkcja aktywacji dla sieci LSTM (przyjęto jako ten parametr funkcję sigmoidalną),
(j) $n_{gcn_{hidden}}$ – liczbę neuronów w stanie ukrytym sieci splotowej GCN.

Główny eksperyment badawczy stanowi badanie skuteczności modeli wyrażonej jako wskaźniki jakości – średniej poprawności przewidywanych rodzajów przyszłych interakcji (wzór (16.4)). Istotnym wynikiem badań jest także wartość uśrednionego błędu średniokwadratowego przewidywanego czasu interakcji (wzór (16.5)) oraz całkowity czas budowania modelu (wzór (16.6)).

$$Q = \frac{\sum\limits_{S_i \in C} \frac{card(S_i^p)}{card(S_i)}}{card(C)}, Q \in [0,1] \tag{16.4}$$

gdzie:

C – zbiór sekwencji przyszłych interakcji,
S_i^p – zbiór interakcji, których rodzaj zdarzenia został poprawnie przewidziany przez model,
S_i – zbiór wszystkich interakcji w sekwencji i.

$$MSE(t_{pred}) = \frac{\sum\limits_{S_i \in C} \frac{\sum\limits_{j \in S_i} (\hat{t}_j - t_j)^2}{card(S_i)}}{card(C)} \tag{16.5}$$

gdzie:

C – zbiór sekwencji przyszłych interakcji,
$\hat{t}_j$ – rzeczywisty czas wystąpienia interakcji j,
t_j – przewidywany czas wystąpienia interakcji j,
S_i – zbiór wszystkich interakcji w sekwencji i.

$$t_{sum} = t_{prep} + t_{train} \tag{16.6}$$

gdzie:

t_{sum} – całkowity czas budowania modelu z uwzględnieniem czasu uczenia sieci neuronowej oraz czasu potrzebnego na zdefiniowanie sekwencji interakcji (w tym osadzania zdarzeń),
t_{prep} – czas potrzebny na zdefiniowanie sekwencji interakcji (w tym osadzanie zdarzeń),
t_{train} – czas uczenia sieci neuronowej.

4. Uzyskane wyniki

W ramach badań udało się przeprowadzić wszystkie zaplanowane eksperymenty. Na początku przeanalizowano skuteczność modeli w zadaniu predykcji rodzaju przyszłych interakcji. Następnie dokonano badania faktu wykorzystania różnych technik osadzania grafów oraz zastosowania grafów czasowych. W ostatnim etapie porównano procesy uczenia się modeli. Dla każdego z tych badań opisano szczegółowo rezultaty opatrzone wykresami i tabelami oraz wyciągnięto zasadne wnioski.

Dla każdej z par (model, zbiór danych) zdefiniowano kombinacje wartości badanych parametrów długości sekwencji oraz liczby okien czasowych. Wyniki pokazano w tabelach 16.1 oraz 16.2.

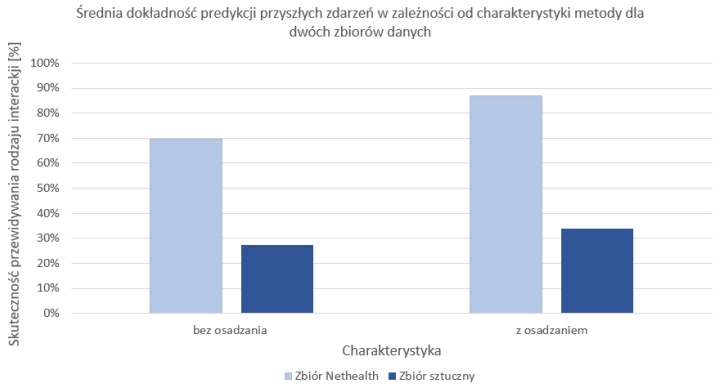
Tabela 16.1. Zbiorcze wyniki badań wraz z wszystkimi wartościami badanych parametrów dla zbioru danych NetHealth

Lp.	Model	card(S_i)	p	Q	$MSE(t_{pred})$	t_{prep} [s]	t_{train} [s]	$t_{sum}[s]$
1	nn	750	–	0.8394	3210,68	48,36	15676,22	15724,58
2	nn	3000	–	0.5526	8278,94	53,73	10887,79	10941,52
3	gcn	750	–	0.8700	3210,65	4963,87	8955,70	13919,57
4	gcn	3000	–	0.8820	8278,23	12033,42	9189,15	21222,57
5	tgcn	750	3	0.8759	3210,65	2022,19	8686,65	10708,84
6	tgcn	3000	6	0.9120	8278,21	4790,94	10008,24	14799,18
7	tgcn	750	6	0.8297	3210,64	1921,44	9773,05	11694,49
8	tgcn	3000	3	0.9121	8278,85	4920,89	10260,96	15181,85
9	w2v	750	–	0.8768	3210,77	3984,54	10022,81	14007,35
10	w2v	3000	–	0.7583	8278,17	5889,70	12350,95	18240,65
11	tw2v	750	3	0.8793	3210,70	3084,33	9221,12	12305,45
12	tw2v	3000	6	0.8618	8278,86	4542,54	10658,57	15201,11
13	tw2v	750	6	0.8694	3210,70	2927,78	8738,68	11666,46
14	tw2v	3000	3	0.9120	8278,32	4666,23	10595,69	15261,92

Tabela 16.2. Zbiorcze wyniki badań wraz z wszystkimi wartościami badanych parametrów dla sztucznego zbioru danych

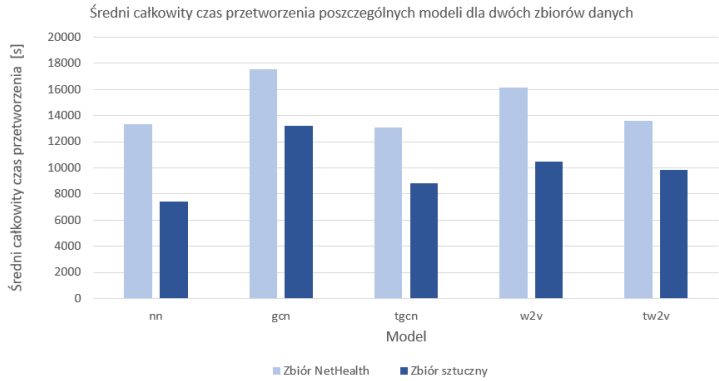
Lp.	Model	card(S_i)	p	Q	$MSE(t_{pred})$	t_{prep} [s]	t_{train} [s]	$t_{sum}[s]$
15	nn	750	–	0.2722	3,83	58,43	6827,34	6885,77
16	nn	3000	–	0.2763	8,41	55,20	7882,20	7937,40
17	gcn	750	–	0.4537	3,83	1847,23	7006,43	8853,65
18	gcn	3000	–	0.2653	8,40	8499,61	9033,81	17533,41
19	tgcn	750	3	0.4749	3,83	1203,95	6434,34	7638,29
20	tgcn	3000	6	0.2738	8,41	2651,59	7339,49	9991,08
21	tgcn	750	6	0.4195	3,82	1223,81	7282,84	8506,65
22	tgcn	3000	3	0.3335	8,40	1883,42	7245,05	9128,47
23	w2v	750	–	0.3115	3,82	1897,50	7873,73	9771,22
24	w2v	3000	–	0.2777	8,40	2267,70	8957,98	11225,68
25	tw2v	750	3	0.3763	3,83	1342,25	6458,05	7800,30
26	tw2v	3000	6	0.2732	8,47	2122,70	8697,49	10820,19
27	tw2v	750	6	0.3001	3,84	1546,78	7946,78	9493,55
28	tw2v	3000	3	0.2884	8,40	2375,31	8863,13	11238,43

Wyniki zobrazowane na rysunku 16.4 ukazują, że rozszerzenie metod osadzania o autorski pomysł wykorzystania grafów dynamicznych zamiast ich statycznych odpowiedników przyniósł minimalną poprawę rezultatów w zakresie predykcji rodzaju przyszłych interakcji. Średnia skuteczność przewidywania rodzaju zdarzenia dla modelu z techniką osadzania wykorzystującą grafy dynamiczne okazała się o około 4 pp. lepsza od technik na bazie grafów statycznych dla zbioru NetHealth. W przypadku zbioru sztucznego przewaga ta jest nieco mniejsza (dokładnie 2 pp. na korzyść grafów temporalnych).



Rysunek 16.4. Średnia dokładność predykcji przyszłych zdarzeń w zależności od zastosowanego rodzaju grafu w dwóch zbiorach danych

Na rysunku 16.5 widać, że wprowadzenie dodatkowego etapu osadzania grafu do budowanego modelu powoduje znaczne wydłużenie czasu budowania modeli. W przypadku podstawowego modelu nn ten uśredniony całkowity czas wyniósł ok. 13 333 sekund dla zbioru NetHealth oraz ponad 7411 sekund dla zbioru sztucznego. Zaś w przypadku modeli gcn i w2v ten parametr osiągał wartości odpowiednio ponad 17 000 i 16 000 sekund dla zbioru NetHealth oraz ok. 13 300 i 10 500 sekund dla zbioru sztucznego. Jednak wprowadzanie grafów temporalnych i podziału sekwencji na okna czasowe zmniejszyło czas potrzebny na przeprowadzenie procesu osadzania grafu. Różnice te wynoszą od 600 sekund do nawet 4500 sekund w zależności od zbioru danych na korzyść modeli wykorzystujących grafy dynamiczne (tgcn, tw2v).



Rysunek 16.5. Średni całkowity czas przetworzeniai dla dwóch zbiorów danych

Wnioski z przeprowadzonych eksperymentów można streścić do następującej postaci:

- zastosowanie technik osadzania grafów zarórno temporalnych, jak i statycznych pozwala na uzyskanie lepszych rezultatów w zadaniu przewidywania przyszłych rodzajów interakcji,
- modele rozszerzone o technikę osadzania grafów mogą pozwolić na zmniejszenie zjawiska przetrenowania bazowego modelu sieci,
- modele cechujące się krótszą sekwencją interakcji wykazały lepsze wyniki w postawionym zadaniu predykcji rodzaju zdarzenia,
- w zadaniu przewidywania czasu interakcji modele wykorzystujące techniki osadzania okazały się lepsze tylko w przypadku mniej zbalansowanego zbioru danych NetHealth,
- zastosowanie technik osadzania grafów wiąże się ze znacznym zwiększeniem całkowitego czasu budowania modelu,
- autorskie podejście do zastosowania grafów temporalnych poprzez podział sekwencji na okna czasowe pozwala zmniejszyć całkowity czas budowy modelu oraz zwiększyć minimalnie jakość modelu w zadaniu przewidywania rodzaju zdarzenia,
- w zdaniu przewidywania rodzaju zdarzenia lepsze okazały się modele cechujące się mniejszą liczbą okien czasowych $p = 3$. Jednocześnie brak zastosowania takiego podziału, ($p = 1$ dla grafów statycznych) przynosił gorsze rezultaty niż modele o dużej liczbie podziałów $p = 6$,
- modele oparte na technikach osadzania w grafach (w szczególności modele o krótszej sekwencji interakcji) szybciej się uczą i już we wcześniejszej fazie uczenia wykazują wysoki poziom skuteczności predykcji.

5. Podsumowanie

W ramach realizowanej pracy przeprowadzono szereg eksperymentów badawczych. Ich wykonanie zostało poprzedzone analizą obecnego stanu sztuki w zakresie modelowania zdarzeń. Dokonano krytycznej oceny znanych modeli w celu znalezienia możliwych opcji udoskonalenia i rozwoju. Na tej podstawie zaproponowano autorską modyfikację dla znanych metod modelowania interakcji. Dzięki temu pracę można określić jako innowacyjną. Charakteryzuje się ona także aspektem badawczym dzięki zaplanowaniu i przeprowadzeniu wnikliwych eksperymentów dla tych modeli.

W ramach badań udało się przeprowadzić wszystkie zaplanowane eksperymenty. Na początku przeanalizowano skuteczność modeli w zadaniu predykcji rodzaju przyszłych interakcji. W kolejnym etapie dokonano badania faktu wykorzystania różnych technik osadzania grafów. Następnie dokonano obserwacji związanych z autorskim podejściem zastosowania grafów czasowych. W ostatnim etapie porównano procesy uczenia się modeli. Dla każdego z tych badań opisano szczegółowo rezultaty opatrzone wykresami i tabelami oraz wyciągnięto stosowne wnioski.

W wyniku przeprowadzonych eksperymentów okazało się, że zastosowanie różnych metod osadzania grafów, jak i wykorzystanie okien czasowych i grafów temporalnych może poprawić działanie bazowego modelu Neural Hawkes opartego na sieci neuronowej LSTM w zadaniu przewidywania rodzaju i czasu przyszłych zdarzeń. Jednak modyfikacja

wprowadzająca etap osadzania grafu wiąże się z kosztami, tj. znacznym zwiększeniem czasu budowania całości modelu. Autorskie podejście polegające na zastosowaniu grafów temporalnych (okien czasowych dla sekwencji) pozwoliło na zmniejszenie czasu budowania modelu. Wprowadzenie tych elementów wpłynęło także na zachowanie modeli w procesie uczenia samej sieci. Pozwoliło to w pewnych przypadkach na uniknięcie sytuacji zbytniego przeuczenia modelu.

Prowadzone badania można kontynuować w szerokim zakresie. Możliwie jest wykorzystanie różnych kombinacji innych metod modelowania opartych niekoniecznie na sieci LSTM. Niewykluczone jest na przykład wykorzystanie sieci GRU. Przytoczony model sieci neuronowej można by połączyć z innymi technikami osadzania struktur sieciowych. Temat „graph embeddingu" w ostatnich latach zyskuje bardzo na popularności i pojawia się wiele nowych metod w tej tematyce.

W ramach kontynuacji badań możliwe jest także przeprowadzenie bardziej obszernych eksperymentów w zakresie dostrajania hiperparametrów sieci neuronowej. Wartościowe byłoby przetestowanie modelu na innych zbiorach danych. W przypadku dysponowania większymi zasobami obliczeniowymi można by zastosować opisany model na większej liczbie interakcji, stosując inny podział na sekwencje czy okna czasowe. Analiza interakcji społecznych i tym samym zdarzeń wydaje się stawać coraz bardziej popularnym tematem badawczym. Jest to odpowiedź na rosnącą liczbę urządzeń elektronicznych w naszym życiu, przez co większość naszego czasu i interakcji przenosimy do rzeczywistości wirtualnej.

Bibliografia

[1] American Psychological Association, Dostępny online: `https://dictionary.apa.org/social-interactions` (dostęp: 10-01-2022).

[2] Margara A., Rabl T., *Definition of Data Streams Encyclopedia of Big Data Technologies*, **2018**, s. 1–4.

[3] Mei H., Eisner J., *The neural hawkes process: A neurally selfmodulating multivariate point process*, Advances in neural information processing systems **2017**.

[4] *Understanding LSTM Networks*, Dostępny online: `https://colah.github.io/posts/2015-08-Understanding-LSTMs/` (dostęp: 10-01-2022).

[5] Du H., Zhou Y., Ma Y., Wang S., *Astrologer: Exploiting graph neural Hawkes process for event propagation prediction with spatio-temporal characteristics*, Knowledge-Based Systems **2021**.

[6] Torricelli M., Karsai M., Gauvin L., *Event embedding for temporal networks*, Scientific Reports **2020**.

[7] Sepulvado B., Wood M., Fridmanski E., Wang C., Chandler M., Lizardo O., Hachen D., *Predicting homophily and social network connectivity from dyadic behavioral similarity trajectory clusters*, Social Science Computer Review **2022**, s. 195–211.

Analiza aktywności podmiotów stanowiących cyberzagrożenie klasy State-Sponsored

mgr inż. Mateusz Stafiniak
Cyberbezpieczeństwo

dr inż. Wojciech Wodo
Katedra Podstaw Informatyki

Przyglądając się rosnącej liczbie ataków w cyberprzestrzeni, warto pochylić się nad zagadnieniem motywacji adwersarzy. Z uwagi na największy procentowy udział ataków o charakterze zarobkowym we wszystkich atakach możemy odnieść wrażenie, że w cyberprzestrzeni dochodzi jedynie do wyłudzeń, oszustw czy kradzieży danych. Nie jest to jednak cały obraz współczesnych cyberzagrożeń, a w okresie nazywanym przez geopolityków pauzą geostrategiczną (rozluźnienie napięć pomiędzy mocarstwami w efekcie zakończenia Zimnej Wojny) cyberprzestrzeń może dostarczyć nowoczesne metody wpływu na stosunki międzynarodowe. Praca dotyczy zgłębienia tematu cyberataktów w kontekście geostrategicznym oraz wypracowanie metody analizy źródeł celem poszerzania wiedzy o charakterze techniczno-strategicznym. Na podstawie zebranych danych z wielu różnych źródeł wnioskowano o zależnościach między obserwowanymi działaniami. Do tego celu wykorzystano osiągnięcia dziedziny cyberbezpieczeństwa, a następnie nałożono na zdobyte informacje konteksty geopolityczne zdarzeń celem uzyskania pełnego obrazu cyberzagrożeń. Wysnuto również szereg wniosków dotyczących specyfiki działań grup klasy State-Sponsored oraz postawiono predykcje przyszłych zdarzeń.

1. Cel i założenia

Celem tej pracy jest przedstawienie obrazu współczesnego świata wirtualnego z perspektywy realizacji interesów geostrategicznych państw i pokazanie, że zachodzi relacja między działaniami podejmowanymi w cyberprzestrzeni przez podmioty zagrożenia a wydarzeniami budującymi geopolityczny kontekst państw, których wspomniane działania dotyczą. Na potrzeby rozważań zawartych w tej pracy powstała hipoteza: *Cyberprzestrzeń jest miejscem realizacji ofensywnych działań umotywowanych geopolitycznie.* Przyglądając się działaniom i operacjom, jakie wykonuje złośliwe oprogramowanie, jesteśmy w stanie wykluczyć lub potwierdzić jedną z dwóch motywacji atakujących: destrukcja lub zarobek.

W przypadku technicznego wykazania destrukcyjności złośliwego oprogramowania, możliwe jest przypisanie go do pewnego podmiotu zagrożenia, z którym skojarzone są już inne działania. Dokładając do tego sytuację geopolityczną państw, uzyskujemy szerszy kontekst działań pozornie ściśle technicznych.

2. Wstęp teoretyczny

Współczesny świat obserwuje dynamiczny wzrost znaczenia cyberbezpieczeństwa. Rozwój metod i praktyk stosowanych przez atakujących stał się na tyle istotnym zagadnieniem, że na rynku pojawiły się zupełnie nowe klasy narzędzi zwiększających stopień przygotowania organizacji do podjęcia walki ze złośliwym podmiotem cyberzagrożenia. Do nich należy zaliczyć przede wszystkim następców popularnych kiedyś systemów antywirusowych – EDR i XDR [1], systemy agregacji informacji klasy SIEM [2] czy narzędzia automatyzacji i orkiestracji procesów bezpieczeństwa [3]. Wspomniane klasy narzędzi muszą być w czasie rzeczywistym zasilane informacjami na temat konkretnych zagrożeń. Gałęzią cyberbezpieczeństwa odpowiedzialną za śledzenie aktywności podmiotów zagrożenia, zbieranie wskaźników naruszenia bezpieczeństwa, przypisywanie konkretnych wydarzeń określonym grupom zajmuje się tzw. wywiad przeciw cyberzagrożeniom.

3. Projekt i wykonanie

W pracy zebrano wiedzę na temat aktywności grup cyberprzestępców w kontekście realizacji przez nie interesów geopolitycznych. Przedstawiony został proces zbierania i analizy informacji dotyczących badanego zagadnienia oraz wnioskowania na ich podstawie.

3.1. Metodyka

Wykorzystując wiele różnych źródeł wykonano przekrojową analizę na rzecz odtworzenia wydarzeń z lat 2008-2022 oraz wskazano metody pozyskiwania informacji istotnych dla badanego zagadnienia. Szczególnie dzisiaj, w czasie trwającej inwazji Federacji Rosyjskiej na Ukrainę, istotnym jest pochylenie się nad kwestią realizacji interesów geopolitycznych przy wykorzystaniu cyberprzestrzeni. Podstawowym działaniem w ramach niniejszej pracy jest analiza informacji pochodzących z czterech klas źródeł: artykułów naukowych, artykułów pozanaukowych, biuletynów badaczy oraz platformy Cyber-Threat Intelligence. Kolejnym etapem jest selekcja informacji dotyczących aktywności badanych grup zarówno pod kątem technicznym (taktyki, metody i techniki), jak i pod kątem operacyjnym, a także przeprowadzenie analizy źródeł opisujących techniczne aspekty ataków i geopolityczny kontekst ich wystąpienia.

3.2. Wykonanie

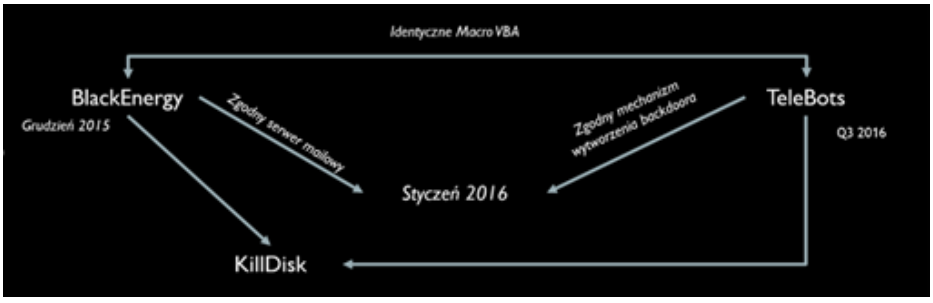
W połowie 2008 roku dochodzi do starcia między Federacją Rosyjską a Republiką Gruzji, które przejdzie do historii pod nazwą II Wojna w Osetii Południowej. Na podstawie analiz dostarczonych przez amerykański think-tank United States Cyber Consequences Unit [4]

ustalono pewne cechy charakterystyczne rywalizacji, która równolegle miała miejsce w cyberprzestrzeni. W innych źródłach [5, 6] potwierdzono, że złośliwe oprogramowanie użyte do budowania sieci botnet, która posłużyła do wyprowadzenia ataku klasy DDoS, należy do rodziny BlackEnergy – względnie prostego toolkitu sprzedawanego w rosyjskim cyberpodziemiu w tamtych czasach. Z czasem BlackEnergy ewoluowało i w wersji drugiej postawiło na funkcjonalności, które sprawiają, że można o BEv2 powiedzieć, iż jest to narzędzie APT [7]. Owe funkcjonalności to:

- Przeprojektowanie architektury oprogramowania na modułową,
- Wyrafinowany rootkit,
- Silne szyfrowanie komunikacji zwrotnej z C&C.

Następnym kontekstem geopolitycznym, w ramach którego prowadzono dalszą analizę jest Wojna Hybrydowa na Ukrainie (2014–2022). Rok rozpoczęcia tego konfliktu jest również tym samym rokiem, w którym trzecia odsłona BlackEnergy została zauważona i opisana m.in. przez Roberta Lipovskiego – badacza z firmy ESET. Niedługo potem, w grudniu 2015 r. przeprowadzono ataki przy użyciu trzeciej odsłony malware'u BlackEnergy. Według źródeł ataki te celowały w ukraiński sektor medialny oraz energetyczny (dostawcę energii elektrycznej). Analiza źródeł skłania ku wnioskom, że działanie z użyciem BEv3 umotywowane jest nie zarobkiem, a chęcią wyrządzenia możliwie jak największych szkód w sektorach krytycznych (z punktu widzenia funkcjonowania nowoczesnego państwa) – informacyjnym i energetycznym. Ostatecznie w wyniku tej kampanii połowa mieszkańców miasta Iwano-Frankiwsk (czyli ok. 700 tys. ludzi) doświadczyła przerwy w dostawie prądu. To miasto nie jest pojedynczym przypadkiem, bowiem ESET, powołując się na swoją telemetrię, twierdzi, że inne firmy energetyczne w kraju również zostały zaatakowane [8].

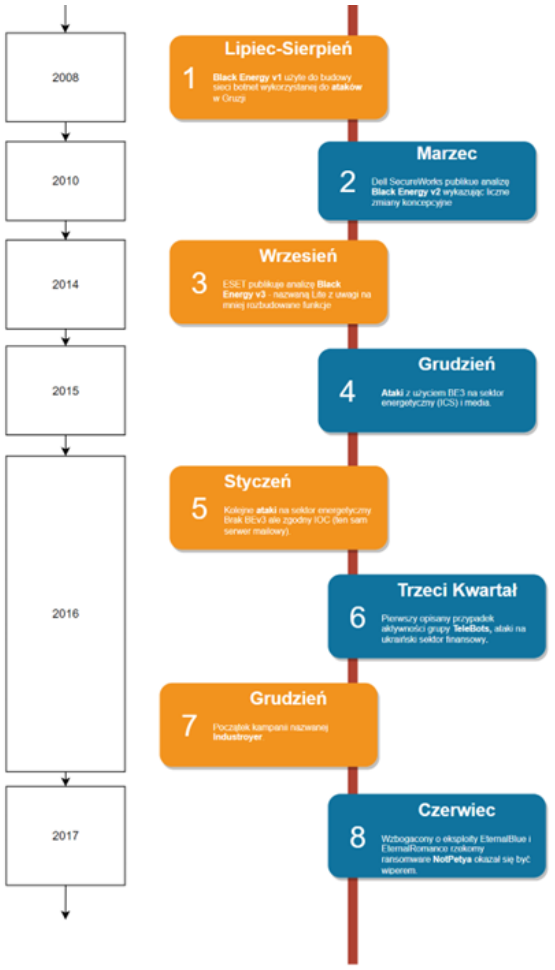
W styczniu 2016 roku miał miejsce zupełnie inny atak, a część wspólna sprowadza się do użycia tego samego serwera mailowego w ataku phishingowym [9]. Szerszy obraz natomiast budują informacje zebrane i przeanalizowane w kontekście aktywności grupy TeleBots w trzecim kwartale 2016 roku, bowiem metoda budowy złośliwego oprogramowania zdaje się być bliźniaczo podobna do tej zastosowanej w styczniu tego samego roku. W obu przypadkach użyto PyInstallera do budowy pliku wykonywalnego. Pierwszą fazą ataku było nakłonienie użytkowników do uruchomienia makra programu Microsoft Excel (język VBA) – zadziwiająco podobnego makra w obu przypadkach. Graficzną reprezentację ustalonych powiązań prezentuje rysunek 17.1.



Rysunek 17.1. Ustalone powiązania

W dalszej części dokumentu przyjrzano się zachowaniu malware Industroyer. W grudniu 2016 roku doszło do kolejnego ataku w cyberprzestrzeni, której celem stały się podmioty ukraińskiego sektora energetycznego. Autorzy (ESET [10], Dragos [11]) raportów dotyczących tych wydarzeń zauważają rozległą wiedzę twórców złośliwego oprogramowania w zakresie specyfiki Przemysłowych Systemów Kontroli (ang. Industrial Control Systems), szczególnie w przypadku wykorzystania ich w procesach sterowania systemami wytwarzania prądu elektrycznego. ESET stawia tezę, że wydaje się być mało prawdopodobne, aby ktoś opracował takie złośliwe oprogramowanie bez dostępu do specyficznego i dziedzinowego sprzętu używanego w tego rodzaju środowiskach.

W dalszej kolejności przyjrzano się kampanii NotPetya – jednego z największych ataków w historii, groźnego za sprawą wykorzystania exploita EternalBlue, a istotnego również z innego powodu: był to Wiper udający Ransomware. Omawiane w tekście wydarzenia zostały naniesione na oś czasu, którą przedstawia rysunek 17.2.



Rysunek 17.2. Wydarzenia opisywane w tekście nałożone na oś czasu

W końcowej części dokumentu zestawiono wnioski, obserwacje i ustalenia z wiedzą na temat najnowszych ataków, jakie próbowano przeprowadzić (lub przeprowadzono) na Ukrainie po 24 lutego 2022 r.

W działaniach, które miały miejsce w ostatnich miesiącach, można dostrzec pewne podobieństwa z tymi, które przedstawione zostały w tym dokumencie. Nie powinny one być rozpatrywane w kategorii dowodu na udział dokładnie tych samych osób lub grup we wszystkich operacjach w ostatnich latach, natomiast sprawia, że obraz rosyjskich cyberzagrożeń zdaje się być pełniejsze.

15 stycznia 2022 r. [12] Microsoft publikuje notatkę dotyczącą nowego podmiotu zagrożenia, który nazywa DEV-0586. Grupa odpowiedzialna jest za atak (z użyciem autorskiego malware WhisperGate), który jest ciekawy z perspektywy prowadzonych tutaj rozważań z dwóch powodów:

- Jest to wiper, który udaje ransomware. Podobny jest więc miejscami do NotPetyi,
- Proces niszczenia jest selektywny i specyficzny – to buduje powiązanie z Industroyerem

Według raportu Microsoftu z końca kwietnia 2022 r. [13] na 8 kwietnia 2022 r. zaplanowane zostało uruchomienia nowego malware'u o nazwie Industroyer2, który został nazwany w ten sposób, ponieważ zarówno przez ESET, jak i Microsoft jest rozpoznany jako następca opisywanego tutaj Industroyera z 2016 r. Jest to wersja z jednej strony okrojona (wykorzystuje komunikacje protokołami jedynie z jednego standardu – IEC 60870-5-104), a z drugiej strony rozwinięta, gdyż wykorzystywany payload jest rozbudowaną wersją tego, który odkryto w 2016 r. Microsoft w swoim raporcie podaje również obszerne informację na temat koincydencji zdarzeń zarówno w cyberprzestrzeni, jak i w ramach konwencjonalnych działań zbrojnych. Z raportu wynika, że działania dotyczące tej samej klasy celów były prowadzone przez podmioty cyberzagrożeń z tygodniowym wyprzedzeniem względem działań militarnych. Jest to sytuacja bardzo podobna, z jaką do czynienia miała Gruzja w 2008 r., o jakiej informował wspomniany wcześniej US Cyber Consequences Unit. Twórcy raportu oceniają, że już w marcu 2021 r. rozpoczął się proces przygotowań do wojny w cyberprzestrzeni po stronie podmiotów zagrożeń „sprzymierzonych z Rosją", do zapewnienia sobie bezpiecznego i łatwego dostępu ukraińskiej infrastruktury, celem zdobywania informacji wywiadowczych lub prowadzenia destrukcyjnych ataków. Microsoft podaje, że w tym czasie doszło m.in do:

- W marcu 2021 r. grupa Nobelium rozpoczęła kampanię phishingową na masową skalę skierowaną przeciwko podmiotom starającym się zyskać międzynarodowe poparcie dla działań Ukrainy przeciw Rosji. Dodatkowo grupa Nobelium wielokrotnie próbowała uzyskać dostęp do firm obsługujących klientów rządowych w państwach członkowskich Paktu Północnoatlantyckiego;
- W połowie roku 2021 grupa DEV-0586 dostała się do infrastruktury przedsiębiorstwa informatycznego budującego system zarządzania zasobami dla ukraińskiego Ministerstwa Obrony. Naruszyła również systemy bezpieczeństwa organizacji działających w sektorze komunikacji i transportu;
- Około 93% wszystkich wspieranych przez Rosję ataków zaobserwowanych w usługach online Microsoftu w 2021 roku było wymierzonych w państwa członkowskie NATO, w szczególności w Stany Zjednoczone, Wielką Brytanię, Norwegię, Niemcy

i Turcję;
- W sierpniu 2021 r. grupa Actinium przerowadziła kampanię spear-phishingową. Tym razem celowano bezpośrednio w wybranych pracowników humanitarnych i doradców wojskowych przebywających na Ukrainie;
- W sierpniu 2021 r. grupa Strontium próbowała rozpracować zabezpieczenia organizacji z sektora obronności;
- Pod koniec 2021 r. DEV-0586 uzyskało dostęp do infrastruktury dostawcy usług IT na Ukrainie – firmy Kitsoft – aby w styczniu 2022 r. przeprowadzić atak na kilku klientów tego przedsiębiorstwa.

W tym miejscu warto zwrócić uwagę na fakt, że US-CCU w swoim raporcie dotyczącym kampanii związanej z wojną w Gruzji informowało o wysokim poziomie przygotowania adwersarzy do ataków. W tym przypadku mamy podobny scenariusz, aczkolwiek prawdopodobnie zrealizowany na większą skalę. Wiedzę o planowanej inwazji w marcu 2021 r. raczej na pewno posiadała jedynie nieliczna grupa ludzi związanych z Siłami Zbrojnymi Federacji Rosyjskiej, co pozwala z prawdopodobieństwem graniczącym z pewnością stwierdzić, że podmioty zagrożenia, które wzięły udział w cyberwojnie przeciw Ukrainie, są klasy State-Sponsored.

4. Podsumowanie

Nie ulega wątpliwości, że w cyberprzestrzeni mamy do czynienia z działaniami o charakterze realizacji interesów geopolitycznych. Jedną z istotnych obserwacji jest zmiana częstości ich występowania. Bez wątpienia w ciągu ostatnich czternastu lat liczba ataków o szczególnych (geopolitycznych) motywacjach adwersarzy wzrosła. Społeczność Cyber-Threat Intelligence z roku na rok rozpoznaje kolejne przykłady takich działań, efektywnie pogłębiając wiedzę na temat tej klasy zagrożeń oraz rozwijając metody ich analizy. W tym miejscu należy również zwrócić uwagę na kolejny istotny wniosek: dynamicznie rosnący poziom wyrafinowania ataków. Już przywoływane w tym dokumencie wydarzenia zdają się być koronnym dowodem na potwierdzenie tej tezy (a w całym spektrum ataków tej klasy incydentów jest znacznie więcej). W 2008 roku mieliśmy do czynienia z budowaniem sieci botnet do prowadzenia ataków typu DDoS. Osiem lat później doczekaliśmy się rozbudowanego i dziedzinowego złośliwego oprogramowania zaprojektowanego z myślą o atakach na przemysłowe systemy sterowania. Atakujących charakteryzuje już nie tylko determinacja, ale również bardzo wysokie kompetencje w bardzo wąskich dziedzinach (jak np. protokoły ICS). Wniosek: będzie coraz trudniej i coraz intensywniej.

Stosunki międzynarodowe i relacje dyplomatyczne są niezwykle napięte w dzisiejszych czasach, a zarówno polscy decydenci polityczni, jak i społeczeństwo od samego początku wojny konwencjonalnej jednoznacznie wspiera jedną ze stron. Należy się więc spodziewać kolejnych ataków, dużo bardziej śmiałych (przez co również prostszych do rozpoznania), także na infrastrukturę znajdującą się na terenie Rzeczpospolitej Polskiej. Wiedza dotycząca historii grup State-Sponsored, a także charakterystycznych metod i praktyk przez nie przedsiębranych, powinna towarzyszyć zespołom bezpieczeństwa operatorów infrastruktury krytycznej oraz usług kluczowych, w zasadzie każdemu zespołowi bezpieczeństwa, gdyż być może czeka nas cyberwojna totalna.

Bibliografia

[1] Gartner, *Endpoint Detection and Response (EDR) Solutions Reviews and Ratings*, Dostępny online: https://www.gartner.com/reviews/market/endpoint-detection-and-response-solutions (dostęp: 07-08-2022).

[2] Gartner, *Security Information and Event Management (SIEM) Reviews and Ratings*, Dostępny online: https://www.gartner.com/reviews/market/security-information-event-management (dostęp: 07-08-2022).

[3] Gartner, *Security Orchestration, Automation and Response Solutions Reviews Reviews and Ratings*, Dostępny online: https://www.gartner.com/reviews/market/security-orchestration-automation-and-response-solutions (dostęp: 08-08-2022).

[4] United States Cyber Consequences Unit, *Overview of the Cyber Campaign Against Georgia in August 2008*, US-CCU, 2009.

[5] ThreatSTOP, *Security Report Black Energy*, Dostępny online: https://f.hubspotusercontent10.net/hubfs/2548414/Case%20Studies/ThreatSTOP_BlackEnergy.pdf (dostęp: 03-08-2022).

[6] PlohmannD., Enders S., *Malpedia*, Dostępny online: https://malpedia.caad.fkie.fraunhofer.de/details/win.blackenergy (dostęp: 03-08-2022).

[7] Stewart J., *BlackEnergy Version 2 Threat Analysis*, Dostępny online: https://www.secureworks.com/research/blackenergy2 (dostęp: 03-08-2022).

[8] Robert L., Cherepanov A., *BlackEnergy trojan strikes again: Attacks Ukrainian electric power industry*, Dostępny online: https://www.welivesecurity.com/2016/01/04/blackenergy-trojan-strikes-again-attacks-ukrainian-electric-power-industry/ (dostęp: 07-08-2022).

[9] Cherepanov A., *The rise of TeleBots: Analyzing disruptive KillDisk attacks*, Dostępny online: https://www.welivesecurity.com/2016/12/13/rise-telebots-analyzing-disruptive-killdisk-attacks/ (dostęp: 08-08-2022).

[10] Cherepanov A., *WIN32/INDUSTROYER A new threat for industrial control systems*, Dostępny online: https://www.welivesecurity.com/wp-content/uploads/2017/06/Win32_Industroyer.pdf (dostęp: 07-08-2022).

[11] Dragos, *CRASHOVERRIDE Analysis of the Threat to Electric Grid Operations*, Dostępny online: https://www.dragos.com/wp-content/uploads/CrashOverride-01.pdf (dostęp: 07-08-2022).

[12] Microsoft, *Destructive malware targeting Ukrainian organizations* Dostępny online: https://www.microsoft.com/security/blog/2022/01/15/destructive-malware-targeting-ukrainian-organizations/ (dostęp: 03-08-2022).

[13] Microsoft, *Special Report: Ukraine. An overview of Russia's Cyberattack activity in Ukraine* Microsoft, Redmond 2022.

Wydział Podstawowych
Problemów Techniki

Konstrukcja urządzenia do monitorowania przebiegu treningu siłowego z funkcją pomiaru EMG

inż. Joanna Zapała
Inżynieria Biomedyczna

prof. dr hab. Krystian Kubica
Katedra Inżynierii Biomedycznej

Przedmiotem pracy jest konstrukcja urządzenia do monitorowania przebiegu treningu siłowego z funkcją umożliwiającą powierzchniowe badanie elektromiograficzne (sEMG). Przed przystąpieniem do projektowania układu pomiarowego dokonano przeglądu urządzeń znajdujących się na rynku. Następnie skonstruowano urządzenie i przetestowano je podczas kilkutygodniowego treningu siłowego mięśnia ramienno–promieniowego: za pomocą funkcji umożliwiającej wykonywanie elektromiografii powierzchniowej badano sygnał elektromiograficzny pochodzący od kurczącego się w czasie ćwiczeń mięśnia. Wyniki pomiarów wraz z analizą przedstawiono w ostatnim rozdziale pracy. Skonstruowane urządzenie zostało wykorzystane do pilotażowych badań sygnału EMG pochodzącego z mięśnia ramienno-promieniowego, ponieważ mięsień ten jest jednym z powierzchownych, silnych mięśni przedramienia. Położony jest tuż pod skórą, a do jego badania można zastosować elektromiografię powierzchniową wykonaną za pomocą skonstruowanego urządzenia. Po przeprowadzeniu badań zauważono, że mięsień ramienno-promieniowy wykazywał największą aktywność podczas uginania przedramienia w chwycie młotkowym oraz nachwycie, dlatego w trakcie efektywnego treningu oporowego wzmacniającego siłę mięśnia mogą być wykonywane te ćwiczenia.

1. Cel i założenia

Celem pracy jest konstrukcja urządzenia z funkcją pomiaru sygnałów mioelektrycznych. Urządzenie to będzie wykorzystane w pilotażowych badaniach oceny wpływu treningu mięśnia na sygnał EMG. Zadania do wykonania to:

- projekt układu pomiarowego bazującego na analogowym czujniku EMG, współpracującym z platformą Arduino,
- konstrukcja układu pomiarowego,
- badanie zarejestrowanych w trakcie treningu mięśnia sygnałów,
- analiza uzyskanych wyników.

2. Wstęp teoretyczny

W tym rozdziale zostaną przedstawione najważniejsze informacje teoretyczne dotyczące elektromiografii, zakłóceń występujących podczas przeprowadzania badań EMG oraz przegląd dostępnych na rynku elektromiografów.

2.1. Elektromiografia i elektrody stosowane w sEMG

Elektromiografia polega na rejestracji sygnału elektrycznego płynącego z mięśni. W trakcie badania elektromiograficznego rejestruje się potencjały czynnościowe płynące z powierzchni całych mięśni lub wybranych jednostek motorycznych [1]. Istnieją dwa rodzaje elektromiografii różniące się sposobem pomiaru: elektromiografia powierzchniowa (sEMG) oraz elektromiografia cienkoigłowa (eEMG). Płaskie elektrody używane w elektromiografii powierzchniowej można podzielić na żelowe i suche. W diagnostyce najczęściej stosuje się jednorazowe elektrody żelowe, jednak w przypadku badań zajmujących długi czas wygodne jest stosowanie suchych elektrod wielorazowych. Obniża to także znacząco koszty eksperymentu. Elektrody suche mogą mieć postać sztywnych płaskich płytek albo być umieszczone na elastycznym materiale, który ściśle przylega do skóry [2]. W przypadku powszechnie stosowanych elektrod żelowych Ag/AgCl konieczne jest zastosowanie elektrolitu, dzięki czemu impedancja takich elektrod jest mniejsza niż impedancja elektrod suchych [3].

2.2. Artefakty występujące w elektromiografii

W przebiegu sygnału EMG mogą pojawić się różnego rodzaju zakłócenia (artefakty). Przykładowe artefakty i sposoby ich eliminacji:

- nakładanie się z szumem zasilania – szum uziemienia z sieci oraz słabo uziemione urządzenia mogą wpływać na rejestrowany sygnał (dlatego trzeba uziemić urządzenia znajdujące się w pobliżu miejsca, w którym wykonuje się pomiar i zastosować filtr wycinający częstotliwość sieciową, która w Polsce wynosi 50 Hz),
- offset sygnału – jest widoczny na przykład wtedy, gdy jedna z elektrod przemieści się w czasie trwania badania lub w trakcie kalibracji badany spiął mięśnie zamiast je rozluźnić; do usunięcia offsetu można zastosować filtr górnoprzepustowy,
- przesunięcie linii podstawowej – pojawia się, kiedy przewody są źle zamocowane lub na mięsień wywierany jest ucisk; zwykle problem rozwiązuje się przez ponowne dokładne oczyszczenie skóry oraz przymocowanie elektrod i przewodów [4].

2.3. Przegląd dostępnych na rynku elektromiografów

Na rynku dostępnych jest wiele urządzeń do pomiarów elektromiograficznych. Mogą one służyć zarówno do monitorowania pacjenta hospitalizowanego, jak i być przez niego używane samodzielnie w domu. Elektromiografy są popularne nie tylko w zastosowaniach medycznych, ale również w badaniach sportowców do oceny postępów oraz jakości treningów. Porównano ze sobą funkcje trzech urządzeń:

- Nu-Tek Maxi Plus 1 – posiada funkcję biofeedback EMG, która pozwala na monitorowanie aktywności mięśni na przykład w trakcie gry sterowanej za pomocą impulsów nerwowych, może także wspomagać rehabilitację [5],
- MyoPlus 2 – wspomaga rehabilitację układu nerwowego i mięśniowego dzięki połączeniu metody biofeedback z elektrostymulacją funkcjonalną, która przyczynia się do polepszenia przewodnictwa nerwowego [6],
- BIO-EMG – bazuje na metodzie biofeedback, podczas pomiaru można stymulować mięśnie, a następnie badać ich aktywność; urządzenie może pracować w trybie poniżej progu oraz utrzymując planowany poziom, co odróżnia je od dwóch poprzednich; jako jedyne ze wspomnianych urządzeń może służyć pacjentom cierpiącym na nadmierną aktywność mięśniową: łagodzi ból związany z chorobą wynikający z nadmiernego skurczu mięśni [7].

3. Projekt i konstrukcja urządzenia do monitorowania przebiegu treningu siłowego

Zaprojektowane i skonstruowane urządzenie do monitorowania przebiegu treningu siłowego ma za zadanie mierzyć tętno ludzkiego serca oraz rejestrować sygnał elektromiograficzny w trakcie trwania treningu oporowego. Pomiary tętna oraz sygnału elektromiograficznego mogą być stosowane do określania jakości treningu, stopnia wytrenowania organizmu, a także szybkości jego regeneracji po wykonanej sesji treningowej [8]. Projekt urządzenia do monitorowania przebiegu treningu siłowego przeniesiono najpierw na płytkę stykową, aby sprawdzić poprawność połączeń oraz działanie poszczególnych elementów elektronicznych, a następnie wykonano projekt płytki PCB w programie Altium Designer w postaci nakładki na moduł kompatybilny z Arduino Uno Rev3 i skonstruowano urządzenie.

3.1. Funkcjonalności urządzenia

Urządzenie posiada trzy funkcje włączane przyciskami bistabilnymi. Umożliwiają one:

- pomiar sygnału EMG i przesyłanie danych z czujnika do urządzenia z technologią Bluetooth, nie starszą niż 2.0,
- pomiar ilości skurczów mięśnia w trakcie trwania treningu – liczenie ilości powtórzeń danego ćwiczenia i wyświetlanie wyniku na wyświetlaczu LCD; funkcja ta pozwala uniknąć pomyłki w ilości wykonanych powtórzeń i w pełni koncentrować się na technicznie poprawnym wykonywaniu ćwiczenia. Ponieważ każde ćwiczenie wywołuje w mięśniu inną odpowiedź i mierzony sygnał elektromiograficzny ma inną amplitudę, konieczne jest skalibrowanie urządzenia pod konkretną osobę, mięsień i wykonywane ćwiczenie,
- pomiar tętna ludzkiego serca (rysunek 18.1) i wyświetlanie wyniku na LCD wraz z informacją dotyczącą intensywności treningu (aby skorzystać z tej funkcji, należy wprowadzić swój wiek do programu dedykowanego skonstruowanemu urządzeniu i napisanemu w środowisku Arduino IDE). Intensywność treningu jest określana przy użyciu wzoru Karvonena [9] i jej podziale na strefy [10].

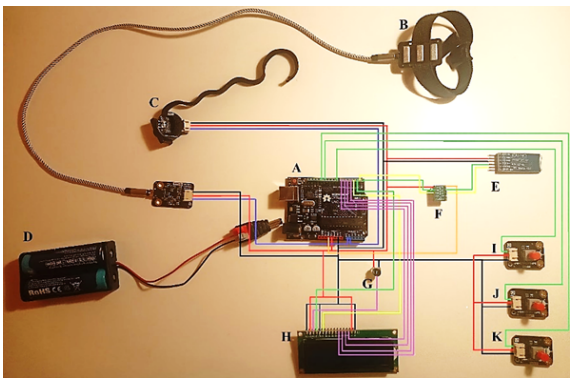


Rysunek 18.1. Pomiar wartości tętna ludzkiego serca wykonany za pomocą skonstruowanego urządzenia [źródło własne]

W skład urządzenia wchodzą: czujnik do pomiaru tętna ludzkiego serca SEN0203 DFRobot, czujnik do pomiaru sygnału EMG SEN0240 DFRobot & OYMotion, płytka bazowa z mikrokontrolerem ATmega 328P DFRduino UNO R3, moduł Bluetooth Classic HC-05, dwa przyciski bistabilne, wyświetlacz LCD, akumulatory Li-Ion oraz elementy elektroniczne umożliwiające prawidłowe zintegrowanie układu. Zastosowanie modułu Bluetooth oraz zasilania w postaci akumulatorów zapewniają mobilność urządzenia.

3.2. Projekt płytki PCB w postaci nakładki na DFRduino

Połączenia między elementami układu pomiarowego zostały zrealizowane w postaci nakładki PCB na płytkę bazową. Nakładkę zaprojektowano w programie Altium Designer i wykonano metodą fotochemiczną na dwustronnym laminacie światłoczułym. Schemat połączeń na nakładce przedstawiono na ilustracji poniżej (rysunek 18.2). Grubość ścieżek na płytce dwustronnej o wymiarach 74 × 58 mm wynosi 1,3 mm, a średnica padów 1,8 mm.



Rysunek 18.2. Schematyczne połączenia elementów skonstruowanego układu. A – DFRduino; B – czujnik EMG; C – czujnik tętna; D – akumulatory; E – moduł Bluetooth; F – konwerter napięć; G – potencjometr; H – wyświetlacz LCD; I, J, K – przyciski bistabilne [źródło własne]

3.3. Oprogramowanie

DFRduino UNO R3 zaprogramowano w Arduino IDE jako Arduino Uno Rev3. Arduino IDE to środowisko umożliwiające między innymi pisanie programów, wgrywanie ich na płytki bazowe oraz przedstawianie wyników na monitorze portu szeregowego lub kreślarce. Oprócz Arduino IDE do części pomiarowej niniejszej pracy związanej z badaniem sygnału EMG zarejestrowanego podczas treningu siłowego zostały wykorzystane następujące programy: monitor portu szeregowego PuTTY oraz AcqKnowledge 4.2 ułatwiający analizę uzyskanych danych. Dane zarejestrowane przez czujnik EMG zostały zapisane do pliku o rozszerzeniu .csv za pomocą PuTTY, a następnie otworzone w AcqKnowledge.

3.4. Porównanie sygnału uzyskanego za pomocą SEN0240 z sygnałem uzyskanym za pomocą BIOPAC

Urządzenie MP36 BIOPAC wykorzystywane jest do przetwarzania danych w trakcie trwania badań związanych na przykład z fizjologią człowieka. Można je zastosować do wykonania: elektromiografii, elektrokardiografii, elektroencefalografii, elektrookulografii i innych. MP36 współpracuje z dedykowanym oprogramowaniem AcqKnowledge i umożliwia analizę danych w czasie rzeczywistym. Charakteryzuje się II klasą ochronności przeciwporażeniowej oraz klasyfikowane jest jako urządzenie medyczne, a bezpieczeństwo jego stosowania jest zgodne z normą IEC 60601-1. W celu porównania urządzeń za pomocą każdego z nich wykonano trzy serie po trzy próby statyczne z obciążeniem polegające na oparciu łokcia lewej ręki na stole i uniesieniu dłoni. Następnie na dłoni układano sztangielki o różnych masach, w kolejności od najmniejszej do największej i rejestrowano sygnał pochodzący z mięśnia ramiennopromieniowego. Wyniki pomiarów przedstawiono w tabeli 18.1.

Tabela 18.1. Porównanie wyników pomiarów uzyskanych przez MP36 oraz SEN0240 podczas wykonywania próby statycznej dla różnych obciążeń mięśnia ramienno– promieniowego

Urządzenie	Obciążenie	Średnia RMS [mV]	Przyrost średniej RMS [mV]	[%]
MP36	2 kg	0,0260	–	–
	3 kg	0,0363	0,0104	39,9
	5 kg	0,0658	0,0295	81,1
SEN0240	2 kg	0,0039	–	–
	3 kg	0,0093	0,0055	141,9
	5 kg	0,0255	0,0161	173,1

Jak widać w tabeli 18.1, przyrost wartości RMS względem poprzedniego pomiaru jest większy w przypadku SEN0240, ale mimo to amplitudy obu sygnałów wzrastają wraz ze wzrostem obciążenia. Wartości RMS są większe dla MP36 w stosunku do SEN0240. Obliczone dla SEN0240 wartości RMS dla poszczególnych obciążeń są niższe od wartości RMS uzyskanych za pomocą MP36 o następujące wartości procentowe:

- dla obciążenia 2 kg: 85,2%,
- dla obciążenia 3 kg: 74,3%,
- dla obciążenia 5 kg: 61,3%.

Zgodnie z przypuszczeniami (związanymi z większą impedancją elektrod suchych wykorzystanych w przypadku pomiarów wykonywanych czujnikiem SEN0240 w porównaniu do impedancji elektrod żelowych zastosowanych do pomiaru za pomocą MP36) pierwiastek średniokwadratowy z amplitudy uzyskanego sygnału jest mniejszy w przypadku użycia elektrod suchych i czujnika SEN0240. Sygnał uzyskany za pomocą SEN0240 zachowuje się w podobny sposób jak sygnał zarejestrowany przez MP36, dlatego uznano, że za pomocą SEN0240 można porównywać między sobą sygnały EMG uzyskane podczas trwania treningu siłowego i na tej podstawie oceniać jego postęp.

3.5. Ocena postępu treningu siłowego

Rozdział jest poświęcony pomiarom wykonanym w trakcie trwania kilkutygodniowego planu treningowego – wybranym ćwiczeniom, sposobie pomiaru oraz analizie uzyskanych wyników. Pomiary wykonano na mięśniu ramienno-promieniowym lewej ręki osoby zdrowej, bez dysfunkcji narządów ruchu. Na skórze zaznaczono miejsce, na którym każdorazowo mocowano elektrody pomiarowe (rysunek 18.3). Warto zaznaczyć, że w projekcie SENIAM zawierającym informacje o złotych standardach stosowanych w elektromiografii nie uwzględniono miejsca mocowania elektrod do pomiaru sygnału EMG płynącego z mięśnia ramienno-promieniowego. Skorzystano jednak z ogólnej zasady, która mówi o tym, że elektrody powinny znajdować się nad najszerszą częścią brzuśca mięśnia [11] i sugerowano się publikacjami naukowymi [12].



Rysunek 18.3. Próba detekcji badanego mięśnia wraz z zaznaczonym obszarem, nad którym umieszczano elektrody [źródło własne]

W planie treningowym znalazły się 3 ćwiczenia dynamiczne wykorzystujące następujące akcesoria oporowe: hantlę o masie 3 kg (do uginania przedramion ze sztangielkami w nachwycie oraz chwycie młotkowym), i narzędzie wzmacniające chwyt o oporze 15 kg oraz jedno ćwiczenie statyczne polegające na trzymaniu sztangielki dłonią opartą łokciem o płaską powierzchnię. Każde z ćwiczeń w różnym stopniu angażuje do pracy mięsień ramienno-promieniowy. Podczas sesji treningowych wykonywano 4 serie danego ćwiczenia, każda po 12 powtórzeń. Odpoczynek pomiędzy seriami i poszczególnymi ćwiczeniami

trwał 1 minutę. Przed każdą sesją wykonywano także próbę maksymalnego dobrowolnego skurczu (MVC), która według zaleceń z literatury [4] polegała na oparciu przedramienia na stole i przytrzymaniu uniesioną na wysokość 20 cm dłonią (bez ściskania) sztangielki o masie 5 kg oraz możliwie maksymalnym spięciu badanego mięśnia w czasie 15 sekund. Następnie sztangielkę z pomocą innej osoby usuwano z dłoni, a całe przedramię opierano na blacie w celu odpoczynku. Wyniki pomiarów powtarzano trzykrotnie w odstępie minuty w celu uniknięcia zmęczenia mięśnia przed każdym treningiem [13]. Treningi wykonywano co drugi dzień.

3.6. Procedura pomiaru

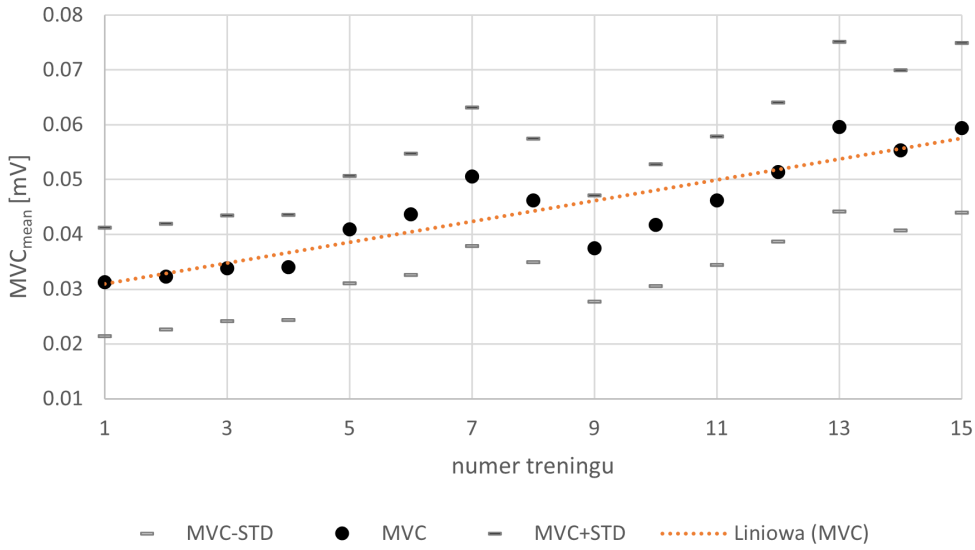
Stosowana procedura pomiaru obejmuje: krótką rozgrzewkę, odłączenie od sieci urządzeń elektrycznych wokół miejsca pomiaru, oczyszczenie skóry w miejscu styku z elektrodami: usunięcie włosów i odtłuszczenie w celu zmniejszenia impedancji, mocowanie elektrod w najszerszym miejscu brzuśca mięśnia ramienno-promieniowego (oznaczono na skórze miejsce przyczepu elektrod), mocowanie do skóry przewodu za pomocą plastra w celu zminimalizowania artefaktów wynikających z poruszania nim podczas pomiarów, przygotowanie akcesoriów oporowych, metronomu oraz stopera, nawiązanie połączenia Bluetooth między układem pomiarowym a docelowym urządzeniem (komputerem), otworzenie emulatora terminala PuTTY, zaznaczenie odpowiednich ustawień oraz określenie docelowego miejsca na zapisany plik z pomiarami, wciśnięcie przycisku rozpoczynającego pomiar i przesyłanie danych, wykonanie próby MVC, wykonanie serii ćwiczeń dynamicznych, wykonanie ćwiczenia statycznego, zakończenie rejestracji sygnału przez ponowne wciśnięcie przycisku, zapisanie danych do pliku o rozszerzeniu .csv.

Po zapisaniu wyników pomiaru do pliku o rozszerzeniu .csv sygnał jest już przefiltrowany (użyto zaprogramowanych w Arduino IDE filtrów z biblioteki producenta czujnika: dolnoprzepustowego o częstotliwości granicznej 150 Hz, górnoprzepustowego o częstotliwości granicznej 20 Hz oraz wycinającego częstotliwość sieciową 50 Hz [**?**]) i wzmocniony na drodze sprzętowej. Utworzony plik następnie otworzono w programie AcqKnowledge. Dzięki oprogramowaniu wyznaczono RMS (pierwiastek średniokwadratowy) sygnału, a następnie znormalizowano go do % MVC dzięki wykorzystaniu możliwości wyliczania średniej z danej części sygnału.

4. Uzyskane wyniki

Badania pokazały, że wraz z postępem treningu rośnie wartość MVCmean (średnia wartość RMS dla odcinka sygnału, w którym zaobserwowano najwyższą wartość tego parametru podczas wykonywania próby MVC), która pod koniec trwania eksperymentu jest większa aż o 89,7% od pomiaru uzyskanego przed rozpoczęciem treningu (rysunek 18.4).

Próba MVC jest wykonywana poprawnie przez osoby aktywne fizycznie, na przykład przez sportowców, natomiast prawidłowe wykonanie tej próby przez osoby nietrenujące jest znacznie trudniejsze [4], co potwierdzają uzyskane wyniki pomiarów. Wraz z czasem trwania treningu siłowego zwiększa się siła mięśni i przyzwyczajają się one do dużego wysiłku wymaganego w trakcie wykonywania próby MVC.



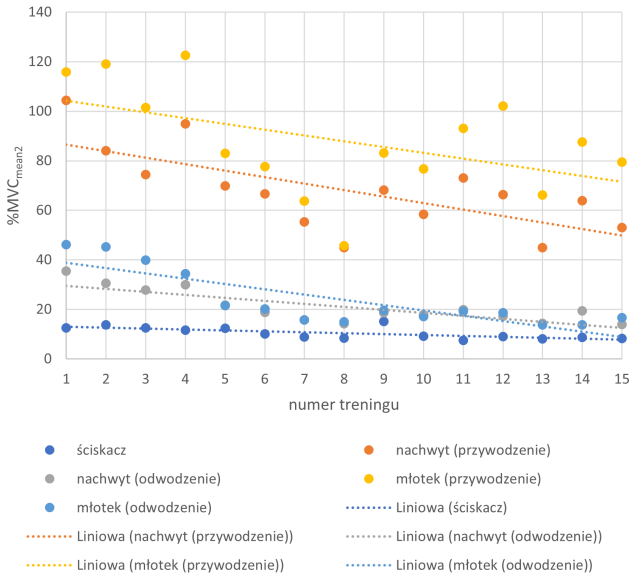
Rysunek 18.4. Zmiana MVCmean podczas trwania treningu oporowego [źródło własne]

4.1. Próba statyczna

W trakcie trwania treningu nie zmieniano wartości obciążenia, liczby powtórzeń oraz długości odstępu czasowego między poszczególnymi seriami i ćwiczeniami. Mięśnie wraz z postępem treningu szybko przyzwyczajają się do obciążenia oraz charakteru ćwiczeń, dlatego ważne jest, aby z czasem zmieniać obciążenie na większe, zwiększać liczbę powtórzeń lub skracać czas odpoczynku między seriami. Wyniki pomiarów uzyskane podczas próby statycznej i znormalizowane do % MVC sięgają wartości powyżej 100%. Ponieważ próba statyczna nie jest tak męczącym ćwiczeniem jak ćwiczenia dynamiczne, można przypuszczać, że próba MVC była wykonywana niepoprawnie ze względu na brak przyzwyczajenia mięśni do wykonywania dużego wysiłku. Wątpliwa pozostaje również sama forma wykonywania ćwiczenia zaproponowana w dostępnej literaturze, dlatego niezwykle istotne jest przeprowadzanie eksperymentów dotyczących dobrania prawidłowego ćwiczenia umożliwiającego uzyskanie maksymalnej wartości amplitudy sygnału elektromiograficznego pochodzącego z mięśnia ramienno-promieniowego.

4.2. Ćwiczenia dynamiczne

Wyniki analizowano pod kątem postępu w trakcie wykonywania danego ćwiczenia w czasie trwania eksperymentu, a także porównano ćwiczenia między sobą. Wpływ każdego z ćwiczeń na uzyskany sygnał EMG jest inny (rysunek 18.5). Niektóre z ćwiczeń bardziej angażują mięsień ramienno-promieniowy (co rozumiane jest jako zwiększona aktywność mięśnia), a kilka z nich prawdopodobnie jedynie w niewielkim stopniu wpływa na badany mięsień albo zarejestrowany sygnał pochodzi z mięśni położonych głębiej lub znajdujących się w jego bliskim sąsiedztwie.



Rysunek 18.5. Znormalizowane do % MVC wyniki pomiarów (% MVCmean2) uśrednione po wszystkich seriach danego ćwiczenia wykonywanego w danej sesji treningowej [źródło własne]

W przypadku ćwiczeń słabo angażujących mięsień ramienno-promieniowy (ściskacz i odwodzenie) wyniki znormalizowane do % MVC pozostają prawie na stałym poziomie, natomiast ten sam parametr dla faz ruchu związanych z odwodzeniem przedramienia nieznacznie wzrasta. Z przeprowadzonych badań wynika, że fazy ruchu związane z przywodzeniem przedramienia do klatki piersiowej najmocniej angażują mięsień ramienno-promieniowy. Najlepszym ćwiczeniem spośród wykonanych jest uginanie przedramion z ciężarem w chwycie młotkowym. Konkretnie to ćwiczenie powoduje, że z mięśnia rejestrowane są sygnały o najwyższej amplitudzie, to znaczy mięsień w trakcie jego wykonywania musi użyć najwięcej siły, aby przyciągnąć dłoń z ciężarkiem do piersi, zatem do wzmocnienia badanego mięśnia najlepiej wykonywać to ćwiczenie. Również uginanie przedramion z ciężarem w nachwycie angażuje mięsień ramienno-promieniowy bardziej niż ćwiczenie ze ściskaniem dłonią akcesorium oporowego oraz fazy ruchu związane z odwodzeniem przedramienia. Po znormalizowaniu wyników do % MVC okazuje się, że wraz z upływem czasu ćwiczenia coraz słabiej angażują mięsień ramienno-promieniowy. Jest to spowodowane brakiem zmiany obciążenia na większe. Aby móc dalej zwiększać siłę mięśnia, należy zwiększyć obciążenie, liczbę powtórzeń, serii lub zmniejszyć odstępy czasowe pomiędzy nimi (to znaczy zastosować progresywne przeciążenie), ponieważ widoczny spadek % MVCmean2 spowodowany jest przyzwyczajeniem się mięśnia do monotonnego charakteru ćwiczeń.

5. Podsumowanie

W niniejszej pracy na podstawie studiów literaturowych dotyczących skurczu mięśni i elektromiografii powierzchniowej przedstawiono przykłady dostępnych na rynku elektromio-

grafów oraz zaprojektowano i skonstruowano urządzenie ułatwiające trening siłowy. Wykonano także pomiary sEMG podczas trwania kilkutygodniowego eksperymentu i opracowano je pod kątem doboru najlepszego ćwiczenia do wzmocnienia mięśnia ramienno-promieniowego. Ogólnodostępna literatura zawiera niewiele informacji na temat tego mięśnia. Brak rzetelnych informacji na temat umieszczenia elektrod pomiarowych (również w projekcie SENIAM) dodatkowo utrudnia przeprowadzanie eksperymentów związanych z mięśniem ramienno-promieniowym. Zauważono także, że ćwiczenie zaproponowane do wykonania próby MVC może nie być prawidłowo dobrane. Pokazano również, że mięsień ramienno-promieniowy wykazywał największą aktywność podczas uginania przedramienia w chwycie młotkowym oraz nachwycie, dlatego w trakcie efektywnego treningu oporowego wzmacniającego siłę mięśnia mogą być wykonywane te ćwiczenia. Mięsień ramienno-promieniowy wykazywał niewielką aktywność podczas ściskania dłonią akcesorium oporowego, co może oznaczać, że nie jest on zaangażowany podczas tego typu ruchu. Stanowi jednak ważną część mechanizmu odpowiadającego za przywodzenie przedramienia do klatki piersiowej, szczególnie z przedramieniem ułożonym w pozycji pośredniej między supinacją i pronacją.

Bibliografia

[1] Zahak J.M., *Signal acquisition using surface EMG and circuit design considerations for robotic prosthesis*, Computational Intelligence in Electromyography Analysis-A Perspective on Current Applications and Future Challenges **2012**, 18, s. 427–448.

[2] Fayyaz P., Pourkheyrollah H., Jahanshahi A., Ghafoorifard H., *Highly conformable stretchable dry electrodes based on inexpensive flex substrate for long-term biopotential (EMG/ECG) monitoring*, Sensors and Actuators A: Physical **2019**, 295, s. 678–686.

[3] Yulin F., Zhao J., Dong Y., Wang X., *Dry electrodes for human bioelectrical signal monitoring*, Sensors **2020**, 20.13.

[4] Peter K., *ABC EMG. Praktyczne wprowadzenie do elektromiografii kinezjologicznej*, TECHNOMEX Spółka z o.o., Gliwice, Polska, 2007.

[5] *Urządzenia Nu-Tek do treningu i diagnostyki ETS/EMG Biofeedback*, Dostępny online: `diamedica.pl/product-pol-794-Program-Nu-Tek-System-do-diagnostyki-i-treningu-ETS-EMG.html` (dostęp: 22-05-2021).

[6] *Myo Plus broszura*, Dostępny online: `innomed.pl/img/Grafiki-OpisyEMG/~%5BPDF~%5DBroszuraSEMG_2017.pdf` (dostęp: 22-05-2021).

[7] *BIO-EMG broszura*, Dostępny online: `www.bardomed.pl/aparaty-przenosne-do-fizykoterapii/709-aparat-do-emg-biofeedback-bio-emg.html?gclid=CjwKCAjwqcKFBhAhEiwAfEr7zZanLqD71SSy8LTVmpwOunwwq37u_Em76Kw7ESn8TBgooz35OnO4OBoCDJsQAvD_BwE` (dostęp: 22-05-2021).

[8] *Wpływ treningu sportowego na wytrzymałość mięśni obszernych przyśrodkowych i szybkość powysiłkowej regeneracji u biegaczy*, Dostępny online: `https://ppm.umlub.pl/info/article/UML00f07d0408314646bb776497356329b2/` (dostęp: 06-05-2021).

[9] She J., Nakamura H., Makino, K., Ohyama Y., Hashimoto H., *Selection of suitable maximum-heart-rate formulas for use with Karvonen formula to calculate exercise intensity*, International journal of automation and computing **2015**, 12.1, s. 62–69.

[10] *Heart Rate Training Zones*, Dostępny online: `www.brianmac.co.uk/hrm1.htm` (dostęp: 05-12-2021).

[11] Merlo A., Bò M., Campanin I., *Electrode Size and Placement for Surface EMG Bipolar Detection from the Brachioradialis Muscle: A Scoping Review*, Sensors **2021**, 21.21.

[12] Buzzi J., Ferrigno G., Jansma J., De Momi E., *On the value of estimating human arm stiffness during virtual teleoperation with robotic manipulators*, Frontiers in neuroscience **2017**, 11.

[13] Zebis M., Sørensen M., Lauridsen H., Bencke J., Andersen C., Carlsbæk J., Jespersen P., Kallehauge A., Andersen L., *Electromyography Evaluation of Bodyweight Exercise Progression in a Validated Anterior Cruciate Ligament Injury Rehabilitation Program: A Cross-Sectional Study*, American journal of physical medicine & rehabilitation **2019**, 98.11.

Wpływ wybranych czynników zaburzających na dokładność pomiaru czujnika oksymetrycznego

inż. Dominik Żyłka
Inżynieria Biomedyczna

dr hab. inż. Agnieszka Ulatowska-Jarża
Katedra Inżynierii Biomedycznej

Wykorzystanie czujników pulsoksymetrycznych jako nieinwazyjnej metody pomiaru saturacji krwi staje się coraz bardziej popularne. Wynika to z łatwości ich użycia oraz możliwości przeprowadzenia pomiarów w warunkach domowych. W pracy zbadano wpływ czynników zaburzających na dokładność pomiarów czujnika oksymetrycznego. Wykonano model takiego czujnika, oraz zaproponowano fantom, na którym przeprowadzono część pomiarów. Dokładność pomiaru modelu sprawdzono, testując go wraz z dostępnym urządzeniem medycznym. Nie wykazano statystycznie znaczących różnic w pomiarach między testowanymi urządzeniami ($p > 0{,}05$). Do zbadanych czynników zaburzających należały: obecność dodatkowego źródła światła, pomalowanie paznokci lakierem (bezbarwny, czerwony, czarny) oraz przesunięcie palca w obrębie czujnika. Pomiary przeprowadzono na palcu wskazującym autora pracy oraz na wykonanym fantomie palca ze słoniny świńskiej. Wykazano brak istotnego wpływu ($p > 0{,}05$) obecności lakieru do paznokci na wartości saturacji krwi. W przypadku pomiarów w dodatkowym świetle wykazano znaczący wpływ ($p < 0{,}05$), ale zmiany wartości saturacji spowodowane tym czynnikiem, okazały się mniejsze niż przyjęty błąd pomiarowy (2%). Największe różnice zauważono przy zmianie położenia palca w obrębie czujnika ($p < 0{,}001$). Wartości, o które różniły się pomiary, znacznie przekraczały przyjęty błąd pomiarowy. Badania na fantomie palca nie przyniosły zamierzonych efektów. Nie udało się uzyskać statystycznie istotnych danych. Na tej podstawie wyciągnięto wnioski, ze wśród zbadanych czynników największy wpływ na dokładność pomiaru saturacji krwi ma położenie palca w obrębie czujnika.

1. Cel i założenia

Celem pracy było opracowanie i konstrukcja modelu czujnika oksymetrycznego wraz z konstrukcją fantomu palca, na którym przeprowadzane będą pomiary. Postarano się także wykazać różnice w pomiarach w zależności od czynników zaburzających. Głównymi założeniami projektowymi realizowanej pracy były:

- przygotowanie modelu urządzenia oraz fantomu palca,
- przeprowadzenie pomiarów z użyciem czynników zaburzających,
- analiza wyników i opisanie wpływu wybranych czynników zewnętrznych na dokładność wyników.

2. Wstęp teoretyczny

Nośnikiem tlenu w organizmie człowieka jest hemoglobina. Dzięki możliwości łączenia się z tlenem umożliwia ona transport tlenu z płuc do tkanek. Hemoglobinę związaną z tlenem określa się mianem oksyhemoglobiny (HbO_2), natomiast odtlenowana hemoglobina nazywana jest deoksyhemoglobiną bądź po prostu hemoglobiną odtlenowaną lub zredukowaną (Hb). Saturacja w ogólnym pojęciu oznacza nasycenie cieczy gazem. W medycynie określa ona stopień wysycenia krwi tętniczej tlenem, czyli stosunek udziału hemoglobiny utlenowanej do hemoglobiny całkowitej. Stosunek ten oznaczany jest symbolem SpO_2 i wyrażany procentowo wzorem:

$$SpO_2 = \frac{HbO_2}{HbO_2 + Hb} \cdot 100\% \tag{19.1}$$

Zasada działania pulsoksymetru oparta jest na pomiarze absorpcji promieniowania przez hemoglobinę utlenowaną oraz deoksyhemoglobinę. Deoksyhemoglobina silniej pochłania światło czerwone, podczas gdy oksyhemoglobina lepiej absorbuje promieniowanie z zakresu podczerwieni. Dla długości fali równej około 660 nm obserwuje się minimum absorbancji dla oksyhemoglobiny, natomiast maksimum przypada dla długości fali z zakresu podczerwieni. Własność tę wykorzystano w czujnikach oksymetrycznych, używając diody emitującej światło czerwone o długości fali 660 nm oraz diody emitującej promieniowanie podczerwone o długości fali 940 nm [1].

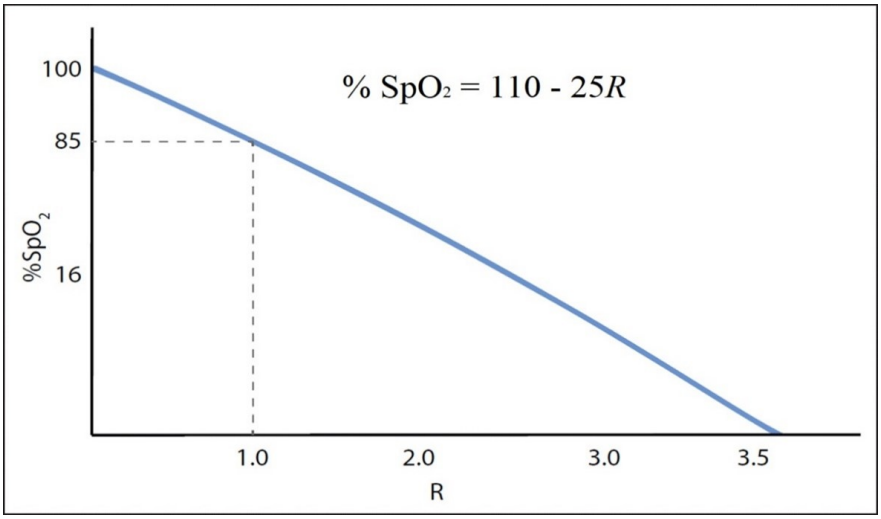
Pochłanianie światła zarówno czerwonego, jak i podczerwonego nie odbywa się tylko przez krew tętniczą, ale zachodzi również w obrębie mięśni, kości czy tkanki tłuszczowej. Aby obliczyć absorbancję wywoływaną jedynie przez krew tętniczą, zakłada się, że pozostałe tkanki nie powodują zmian absorpcji w czasie i traktuje się je jako składowe stałe (ang. DC) Pochłanianie promieniowania przez krew tętniczą traktuje się jako składową zmienną (ang. AC), o częstości zmian równej częstości skurczów serca. Ta część promieniowania, która nie uległa pochłonięciu przez tkanki, zamieniana jest następnie przez fotodetektor na sygnał elektryczny. Znając wartości składowej zmiennej oraz składowej stałej dla dwóch długości fal, możliwe jest obliczenie współczynnika R wyrażanego wzorem:

$$R = \frac{AC_{660}}{DC_{660}} / \frac{AC_{940}}{DC_{940}} \tag{19.2}$$

We wzorze tym skrót AC_{660} oznacza absorbancję składowej zmiennej sygnału dla długości fali równej 660 nm, natomiast DC_{660} to wartość absorbancji składowej stałej sygnału dla tej samej długości fali. Analogiczne oznaczenia zastosowano w przypadku oznaczeń składowych stałych przy użyciu promieniowania o długości fali 940 nm [2].

Współczynnik R wiąże się z saturacją krzywa kalibracyjna. Wyznaczana jest ona em-

pirycznie, na podstawie pomiarów wykonywanych na grupie ochotników. Poprzez porównanie wyników otrzymanych za pomocą gazometrii oraz obliczonego współczynnika R wykreśla się krzywą kalibracyjną oraz wyznacza się odpowiadające jej równanie. Na jego podstawie wyznacza się później procentową wartość saturacji krwi (rysunek 19.1) [3].



Rysunek 19.1. Przykładowa krzywa kalibracyjna

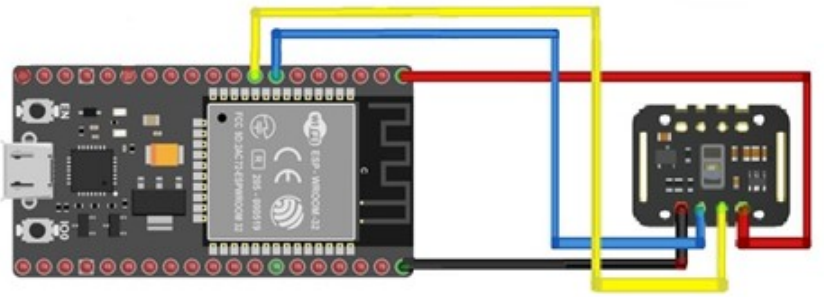
3. Projekt i wykonanie

Ze względu na ograniczone możliwości komercyjnie dostępnych pulsoksymetrów zdecydowano się na wykorzystanie własnoręcznie skonstruowanego w tym celu modelu czujnika oksymetrycznego. Wykonanie własnego urządzenia ułatwiło zbadanie czynników zaburzających oraz późniejszą analizę danych.

Przed przystąpieniem do pomiarów czujnik został przetestowany w kierunku prawidłowości wskazań. W tym celu wykonano kontrolne pomiary saturacji krwi, za pomocą komercyjnie dostępnego urządzenia medycznego, a następnie porównano je z tymi uzyskanymi przy użyciu skonstruowanego urządzenia. Wyniki nie odbiegały znacząco od siebie.

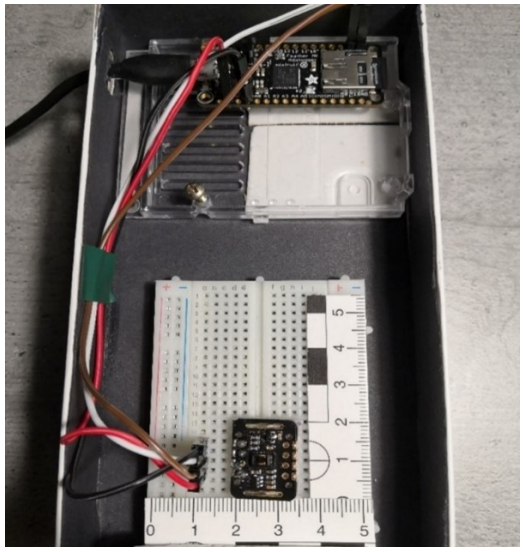
3.1. Schemat i budowa własnego urządzenia

Do budowy oksymetru wykorzystano dwa moduły: moduł do pomiaru tętna i zawartości tlenu w krwi z układem MAX30102 oraz moduł Adafruit Feather M0 z mikrokontrolerem Cortex M0+. Całość umieszczono na płytce stykowej oraz połączono przewodami, a następnie podłączono do komputera. Uproszczony schemat połączenia przedstawiono na rysunku 19.2.



Rysunek 19.2. Schemat czujnika oksymetrycznego

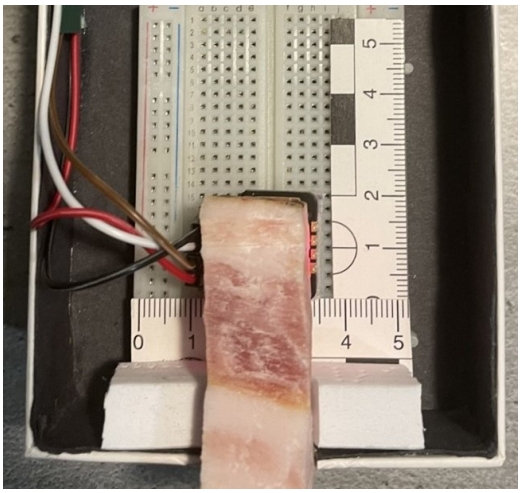
Czujnik umieszczono w specjalnie przygotowanej obudowie (rysunek 19.3). Jej wnętrze wyłożono ciemną, matową tkaniną w celu minimalizacji niepożądanego odbicia światła. Na brzegu obudowy umieszczono uchwyt, ułatwiający przyłożenie palca oraz zapewniający stabilizację podczas pomiarów. Do pomiarów przesunięć palca oraz fantomu w obrębie czujnika służyła naniesiona na płytkę stykową taśma miernicza ze skalą.



Rysunek 19.3. Zdjęcie wykonanego modelu czujnika oksymetrycznego

3.2. Fantom palca ludzkiego

Do przeprowadzenia części badań zdecydowano się na wykonanie fantomu ludzkiego palca. Wykorzystano w tym celu świńską słoninę, którą uformowano na kształt prostopadłościanu, rozmiarami odpowiadającego wymiarom ludzkiego palca. Wybór takiego materiału pokierowany był zbliżonymi właściwościami fizycznymi do właściwości fizycznych ciała ludzkiego. Wykonany fantom miał służyć do zasymulowania idealnego pomiaru. Eliminował tym samym problem polegający na występowaniu nieuniknionych ruchów dłonią podczas pomiarów. Co najważniejsze zaś, umożliwiał precyzyjne ułożenie go na czas pomiarów. Zdjęcie przygotowanego modelu przedstawiono na rysunku 19.4.



Rysunek 19.4. Słonina zwierzęca wykorzystana jako fantom palca

3.3. Metodyka pomiarów

Właściwą część pomiarów przeprowadzono w trzech etapach. W pierwszym z nich zbadano wpływ obecności światła otoczenia w trakcie wykonywanych pomiarów. Drugą część stanowiły pomiary wykonane na palcu, którego paznokieć pomalowany był dodatkowo lakierem (bezbarwnym, czerwonym oraz czarnym). W trzeciej części sprawdzono wpływ przesunięcia palca w obrębie czujnika na dokładność wskazań saturacji krwi. Przesunięcia dokonywano w dwóch płaszczyznach, za każdym razem z przesunięciem o 5 milimetrów względem położenia prawidłowego. Pomiary przeprowadzono również na przygotowanym w tym celu fantomie ludzkiego palca i wykonano w identyczny sposób. Badanie saturacji krwi, wykonano na osobie zdrowej, niemającej problemów z oddychaniem. Pomiary wykonywane były w pozycji siedzącej, na palcu wskazującym lewej dłoni.

4. Uzyskane wyniki

W celu wykazania ewentualnych różnic między pomiarem kontrolnym a pomiarami badawczymi wykorzystano jednoczynnikową analizę wariancji (ANOVA). Do wykazania różnic między wpływem czynników zaburzających na wyniki pomiarów wykorzystano test post-hoc (test Dunna). Czynniki uznawano za różniące się między sobą, gdy wartość parametru $p < 0{,}05$.

4.1. Pomiary na palcu dłoni

W celu dokładnej analizy wpływu czynników posłużono się tabelą 19.1, gdzie zestawiono wyniki uzyskane przy pomocy analizy wariancji, współczynnikiem p, świadczącym o mocy statystycznej danego testu oraz z procentową zmianą wartości średnich saturacji krwi. Największą zależność pomiarów, wykazano dla pionowych zmian położenia palca podczas pomiarów. Współczynnik p był znacząco mniejszy od założonego minimum ($p < 0{,}001$), poniżej którego uznawano wpływ czynnika za znaczący. Również procentowa

zmiana w wartości średniej saturacji, wyniosła najwięcej ze wszystkich i ukształtowała się na poziomie 7,5%. Mniejsze, choć również znaczące zmiany ($p < 0{,}05$), zaobserwowano w przypadku przesunięć w osi poziomej. Zauważyć można, że dla zmiany oświetlenia, różnica względna między pomiarami, pomimo bycia znaczącym statystycznie, jest mniejsza niż przyjęty błąd pomiarowy (2%). W przypadku testowania obecności lakierów do paznokci a dokładności wskazań saturacji nie zaobserwowano znaczącego wpływu ($p > 0{,}005$) tych czynników na ostateczny wynik.

Tabela 19.1. Porównanie czynników zaburzających wraz z ich wpływem na wynik saturacji

	Wpływ na pomiar	p-value	Zmiana wartości względem pomiaru odniesienia [%]
Brak oświetlenia	TAK (zmiana mniejsza niż błąd pomiarowy)	0,0037	1,1
Lakier bezbarwny	NIE	0,21	−0,4
Lakier czerwony	NIE	0,21	−0,6
Lakier czarny	NIE	0,21	−0,2
Przesunięcie pionowe w górę	NIE	1,0	0,3
Przesunięcie pionowe w dół	TAK	<0,001	−7,5
Przesunięcie poziome w prawo	TAK	0,048	−2,6
Przesunięcie poziome w lewo	TAK	0,014	−3,1

4.2. Pomiary na fantomie palca

Znalezienie odpowiedniego materiału i wykonanie modelu palca okazało się dużym wyzwaniem. Materiał ten musiał wykazywać się podobnymi do ludzkiej tkanki właściwościami, przede wszystkim wykazywać podobne maksima absorbcji dla danych długości promieniowania emitowanego przed diody LED czujnika.

Przeprowadzone pomiary nie pozwalały jednak uznać tego materiału za właściwy do modelowania zachowania ludzkiej tkanki. Zaobserwowano niską wartość saturacji oraz brak korelacji między poszczególnymi warunkami zaburzającymi. Na tej podstawie zdecydowano się nie brać ich pod uwagę do dalszej analizy. Wykonany ze słoniny fantom palca odrzucono zatem jako nienadający się do pomiarów. Dalsza analiza i wnioski oparta została tylko i wyłącznie na pomiarach wykonanych na palcu dłoni.

5. Podsumowanie

Najważniejsze i kluczowe podczas pomiarów saturacji krwi okazuje się być położenie palca na czujniku. Skrajne przyłożenie powoduje, że tylko część wychodzącego promieniowania transmituje bezpośrednio przez tkankę. Pozostała jego część może ulegać niekontrolowanym załamaniom na krańcu palca, zaniżając tym samym ostateczny wynik. Pozostałe czynniki, takie jak obecność dodatkowego źródła światła czy pomalowane lakierem paznokcie, okazują się nie wprowadzać statycznie istotnego błędu do pomiarów. Zbadanie wpływu oświetlenia jako czynnika zaburzającego znajduje swoje uzasadnienie w kilku publikacjach. W pierwszej z nich wykazano, że w zależności od wykorzystanego czujnika błąd pomiarowy, przekraczający przyjęty przez autorów próg 4%, występuje przez 1 do 11% czasu trwania pomiaru. Brak informacji o dokładnych wskazaniach i odchyleniach od normy (Trivedi, 1997) [4]. W badaniu przeprowadzonym 6 lat później na grupie 45 ochotników nie wykazano wpływu większego niż 0,5% dodatkowego źródła światła na wynik pomiarów ($p < 0{,}05$) (Fluck, 2003) [5]. Wynikać może to z dwóch powodów. Pierwszym z nich jest udoskonalenie układu pomiarowego. W produkowanych dzisiaj czujnikach obecny jest specjalny system kompensujący ewentualny nadmiar promieniowania z innego źródła niż diody LED. Drugim powodem może być zmiana konstrukcji samego urządzenia. Współczesne pulsoksymetry posiadają specjalną obudowę, która nie tylko unieruchamia palec, ale także ogranicza dostęp światła w głąb czujnika. Nowszych badań nie znaleziono.

Do przyjętych zaleceń dotyczących przeprowadzania właściwego pomiaru saturacji, należy również informacja o posiadaniu niepomalowanych paznokci podczas pomiaru na palcu dłoni. W badaniu przeprowadzonym w 2003 r. (Mallory, 2003) wykazano, że niewielki (około 2%), ale statystycznie znaczący ($p < 0{,}05$) wpływ na pomiary wykazały lakier w kolorze czarnym oraz brązowym [6]. Późniejsze badanie (Rodden, 2007) opisuje wpływ aż 10 kolorów do lakierów [7]. Wynikiem badania jest obserwacja, że lakiery powodują zmianę saturacji mniejszą niż 1%, mieszczącą się tym samym w dopuszczalnym błędzie pomiarowym urządzenia. Podobne wnioski przeczytać można w pracy z roku 2013 (Yönt, 2013), gdzie uzyskane pomiary wskazują na brak klinicznie istotnego wpływu [8]. Wszystkie powyższe wnioski potwierdza również praca przeglądowa (Adedimeji, 2017), w której autorzy przeanalizowali osiem różnych badań na przełomie 5 lat i stwierdzili, że lakier do paznokci nie ma istotnego wpływu na dokładność odczytów pulsoksymetru [9]. Identyczne do powyżej opisywanych wyników uzyskano w pracy. Stwierdzono brak istotnego ($p < 0{,}05$) wpływu testowanych lakierów do paznokci na dokładność wskazań czujnika oksymetrycznego.

Prawidłowe położenie palca na powierzchni czujnika bywa często pomijane w zaleceniach przez fakt, że duża część czujników posiada zabezpieczenie polegające na niewykonywaniu pomiarów w przypadku niewykrycia przyłożenia. Możliwe jednak jest, uzyskanie błędnego odczytu w przypadku nieprawidłowo przyłożonego czujnika do ciała. Błędy takie pojawiać się mogą w przypadku urządzeń zakładanych samodzielnie, bez pomocy wykwalifikowanego personelu medycznego.

W celu zasymulowania podobnej sytuacji sprawdzono, jak skrajne położenia palca wpływają na odczyt saturacji przez urządzenie. Okazało się, że jest to czynnik mający statystycznie oraz klinicznie istotne znaczenie. Zmiana saturacji w przypadku nieprawidło-

wego przyłożenia palca do czujnika przekraczała dopuszczalny błąd pomiarowy i największa była w przypadku przesunięcia palca poniżej dolnej krawędzi czujnika. Przesunięcia palca w osi poziomej również niosły za sobą istotny błąd pomiarowy, jednak mniejszy niż w poprzednim przypadku. Podobnych badań testujących wpływ tego czynnika zaburzającego nie znaleziono.

Podsumowując, samodzielny pomiar saturacji krwi może dostarczyć wielu informacji na temat staniu zdrowia pod warunkiem, że jest wykonany prawidłowo. Stosując się do zaleceń producentów urządzeń, możliwe jest uniknięcie błędów pomiarowych.

Bibliografia

[1] Metcalfe B., Iravani P., Graham-Harper-Cater J., Bowman R., Stirling J., Wilson P., *A Cost-Effective Pulse Oximeter Designed in Response to the COVID-19 Pandemic*, J. Open Hardw. **2021**, 5(1).

[2] Chan E. D., Chan M. M., Chan M. M., *Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations*, Respiratory Medicine **2013**, 107(6), s. 789–799.

[3] Milner Q.J.W., Mathews G.R., *An assessment of the accuracy of pulse oximeters*, Anaesthesia **2012**, 67(4), s. 396–401.

[4] Trivedi N.S., Ghouri A.F., Shah N.K., Lai E., Barker S.J., *Effects of motion, ambient light, and hypoperfusion on pulse oximeter function*, J. Clin. Anesth. **1997**, 9(3), s. 179–183.

[5] Fluck R.R., Rrt M., Rrt C.S., Crt G.F., Crt B.K., Engbretson B., *Does Ambient Light Affect the Accuracy of Pulse Oximetry?* **2003**.

[6] Chan M.M., Chan M.M., Chan E.D., *What Is the Effect of Fingernail Polish on Pulse Oximetry?*, Chest **2003**, 123(6), s. 2163–2164.

[7] Rodden A.M., Spicer L., Diaz V.A., Steyer T.E., *Does fingernail polish affect pulse oximeter readings?*, Intensive Crit. care Nurs. **2007**, 23(1), s. 51–55.

[8] Hakverdioğlu Yönt G., Akin Korhan E., Dizer B., *The effect of nail polish on pulse oximetry readings*, Intensive Crit. care Nurs. **2014**, 30(2), s. 111–115.

[9] Titus J., *Does Nail Polish Have Significance on Pulse Oximetry Readings?: A Literature Review*, STTI **2017**.

Zastosowanie uczenia maszynowego do klasyfikacji stadiów snu

inż. Krzysztof Adamczyk
Inżynieria Biomedyczna

dr hab. inż. Mirosław Łątka
Katedra Inżynierii Biomedycznej

Celem pracy było opracowanie algorytmu klasyfikacji stadiów snu na podstawie ilościowej analizy EEG z użyciem uczenia maszynowego. Skuteczne rozpoznawanie stadiów snu pełni istotną rolę zarówno w badaniach klinicznych, jak i w diagnostyce niektórych schorzeń. W pracy porównano trzy modele uczenia maszynowego: sieć neuronową MLPClassifier, drzewo klasyfikacyjne DecisionTreeClassifier oraz algorytm KNeighborsClassifier. Jako kryterium klasyfikacji użyte zostały średnie moce falkowe sygnałów EEG dla wybranych częstotliwości. Dodatkowo, w celu wyrównania treningowych zbiorów danych, wykorzystano technikę SMOTE, która umożliwia tworzenie syntetycznych próbek pozwalających na zwiększenie skuteczności działania modeli uczenia maszynowego. Najbardziej skuteczne okazały się modele KNeighborsClassifier oraz DecisionTreeClassifier, które w większości rozróżnień stadiów snu wykazały skuteczność powyżej 80%.

1. Cel i założenia

Celem pracy było opracowanie algorytmu klasyfikacji stadiów snu na podstawie ilościowej analizy EEG z użyciem uczenia maszynowego. Badania snu pełnią istotną rolę zarówno w diagnostyce, jak i w celach badawczych. Skuteczna klasyfikacja snu umożliwia znacznie dokładniejsze rozpoznawanie wielu zaburzeń snu oraz poszerza wachlarz potencjalnych badań, które pozwolą na dokładniejsze zbadanie wielu mechanizmów związanych ze snem [1, 2]. Proces klasyfikacji stadiów snu związany jest z wieloma utrudnieniami. Jako że odbywa się ona zazwyczaj na podstawie analizy elektroencefalogramów, jeden z problemów stanowi szum charakterystyczny dla sygnałów bioelektrycznych [3]. Istotną przeszkodą są również często szybkie oraz niejednoznaczne przejścia między stadiami [4].

2. Wstęp teoretyczny

W poniższych podrozdziałach przedstwione zostały najważniejsze opisy ułatwiające zrozumienie podjętego w pracy tematu.

2.1. Sen

Sen jest niezwykle ważny pod względem fizjologicznym, ponieważ umożliwia on wzmożoną regenerację tkanek [5], odpowiednią pracę układu hormonalnego oraz usuwanie neurotoksycznych substancji z tkanki mózgu. Z psychologicznego punktu widzenia sen jest niezwykle ważny dla utrzymywania pozytywnego nastroju, koncentracji, sprawności motorycznej, chęci do działania, rozwagi w podejmowaniu decyzji oraz kreatywności [6]. Wskutek niedoboru snu można doświadczyć wielu objawów związanych nie tylko z aspektem psychologicznym. W skrajnych przypadkach występować mogą znaczące zaburzenia funkcjonowania organizmu o daleko idących skutkach.

Problemy ze zdrowym snem dotyczą bardzo dużej części populacji. Są one również skorelowane z wieloma chorobami, które określane są jako cywilizacyjne. Wśród najczęściej występujących patologii snu znajdują się:

- bezsenność,
- zespół bezdechu sennego,
- zespół niespokojnych nóg,
- narkolepsja,
- parasomnie.

Każde z wymienionych wyżej schorzeń może doprowadzić do niedoboru snu i znacznie obniżyć jakość życia.

Systematyka stadiów snu przedstawia się następująco:

- NREM I – pierwsze, najkrótsze oraz najpłytsze stadium snu wolnofalowego, w którym świadomość nie zostaje zniesiona całkowicie i występuje przewaga fal β.
- NREM II – drugie stadium snu wolnofalowego, charakteryzujące się całkowitym zniesieniem świadomości oraz przewagą fal θ.
- Sen głęboki – zaliczają się do niego stadia NREM III oraz NREM IV, z których to drugie rozpoznawane jest na podstawie przewagi fal δ.
- REM – stadium powiązane z szybkim ruchem gałek ocznych, w którym organizm wypoczywa mniej oraz występują najintensywniejsze marzenia senne.

Proces snu zaczyna się w momencie wejścia w stadium NREM I, następnie następuje przejście w stadium NREM II, po którym następuje całkowite pogłębienie snu. W trakcie dalszego trwania procesu następują cykliczne przejścia między snem głębokim oraz stadium REM.

2.2. Elektroencefalografia

Elektroencefalografia (EEG) to badanie służące poznaniu czynności bioelektrycznej mózgu. Sygnał generowany jest przez aktywność neuronów, a odbierany przez elektrody znajdujące się na powierzchni skóry głowy pacjenta. Standardowo używa się dziewiętnastu elektrod (dwie z nich mają charakter referencyjny), ale niektóre badania przewidują

ich znacznie więcej. Sygnał EEG może zostać rozłożony na składowe o różnych częstotliwościach. Określane są one jako fale lub rytmy. Najważniejszymi składowymi snu są:

- Fale β – mają częstotliwość 14–30 Hz i najbardziej odpowiadającym im stanem jest skupienie, choć są też obserwowane podczas wpływu niektórych leków.
- Fale α – mają częstotliwość 8–12 Hz i najbardziej odpowiadający im stan to wypoczynek w trakcie czuwania.
- Fale θ – mają częstotliwość 4–7 Hz i pojawiają się podczas snu oraz w stanie medytacji.
- Fale δ – mają częstotliwość poniżej 4 Hz i występują w trakcie głębokiego snu.

W pracy wykorzystane zostały badania polisomnograficzne, czyli takie w których oprócz badania EEG dokonuje się rejestracji wielu innych sygnałów biologicznych. Badanie takie trwa całą noc.

2.3. Technologie oraz narzędzia zastosowane w pracy

Najważniejszym narzędziem matematycznym zastosowanym w pracy była transformacja falkowa, która umożliwiła wyznaczenie średnich mocy falkowej sygnału dla różnych częstotliwości. Wielkości te stanowiły parametry klasyfikacji stadiów snu. Przebieg tej operacji przedstawia równanie:

$$X(\tau, s) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-\tau}{s} \right) dt \tag{20.1}$$

gdzie:

s – parametr skali,
τ – rozmiar okna czasowego,
ψ – odpowiednio dobrana funkcja macierzysta.

W pracy porównane zostały trzy algorytmy uczenia maszynowego pod kątem ich skuteczności w omawianym zagadnieniu.

Pierwszym z nich była sztuczna sieć neuronowa, której sposób działania odzwierciedla pracę mózgu. Podstawową jednostką jej budowy jest sztuczny neuron, czyli funkcja programistyczna oparta na matematycznym modelu działania biologicznego neuronu. Posiada on swoją funkcję aktywacji będącą odpowiednikiem potencjału czynnościowego w komórkach nerwowych. Pierwsza warstwa sztucznych neuronów odpowiada za pobór danych, ostatnia za zwrot danych, a pośrednie (ukryte) za ich przetwarzanie. Aby klasyfikator oparty na sieci neuronowej działał poprawnie, należy odpowiednio dobrać parametry klasyfikacji, funkcję aktywacji oraz rozmiar i liczbę warstw neuronów w sieci. Słuszność doboru wszystkich tych aspektów musi zostać zweryfikowana empirycznie. W pracy wykorzystano obiekt klasy MLPClassifier [7].

Drugim algorytmem klasyfikacji było drzewo klasyfikacyjne klasy DecisionTreeClassifier [8]. Graficzna reprezentacja strukturą przypomina drzewo, którego węzłom odpowiadają warunki, a gałęzie określają, jak klasyfikowana próbka odnosi się do tych warunków. Z każdym węzłem próbka przypisywana jest do podzbioru o ściślej określonych cechach. Ostatni podzbiór determinuje klasę próbki. Warunki klasyfikacji określane są na podstawie zbioru treningowego.

Jako ostatni algorytm wykorzystano obiekt KNeighborsClassifier [9]. Klasyfikuje on próbkę poprzez porównanie jej ze zbiorem treningowym. Próbki traktowane są jako punkty w przestrzeni n-wymiarowej, gdzie n oznacza liczbę parametrów próbki. Miarą podobieństwa próbek jest ich odległość od siebie w tej przestrzeni. Algorytm znajduje zbiór najbliższych „sąsiadów" próbki klasyfikowanej i przypisuje ją klasie, która ma w tym zbiorze najliczniejszą reprezentację.

W trakcie trwania snu niektóre stadia trwają znacznie dłużej niż inne. W konsekwencji wykorzystany w pracy zbiór danych był bardzo nierówny (próbki powiązane z niektórymi stadiami występowały znacznie liczniej). Takie zjawisko może stanowić czynnik obniżający skuteczność działania algorytmów uczenia maszynowego. W celu wyrównania zbioru danych wykorzystano obiekt SMOTE. Technika ta pozwala na tworzenie syntetycznych próbek. Podobnie jak wyżej wymieniony algorytm klasyfikacji, SMOTE traktuje próbki jako punkty w przestrzeni, a nowe tworzy na odcinkach łączących punkty tej samej klasy. Dzięki temu syntetyczne próbki posiadają cechy obu punktów na końcach odcinka.

Cała część wykonawcza klasy wykonana została w języku Python, który pozwolił na odczytanie poszczególnych sygnałów EEG z plików o formacie EDF oraz na wykorzystanie wszystkich omówionych w tym rozdziale obiektów.

3. Projekt i wykonanie

W pracy użyte zostały badania jedenastu pacjentów. Jedno badanie trwało około 9 godzin. Sygnał odbierany był z 19 elektrod, a częstotliwość próbkowania wynosiła 200 Hz. Do każdego z badań załączony był hypnogram, który umożliwił podział sygnału na dwudziestosekundowe fragmenty z przypisanymi im stadiami snu. Na podstawie tych materiałów powstały dokumenty w formacie CSV, w którym każdy wiersz odpowiadał fragmentowi badania i zawierał:

- numer będący oznaczeniem uwzględnianych stadiów snu,
- średnie moce falkowe (z użyciem funkcji macierzystej Complex Morlet) dla częstotliwości: 3 Hz, 6 Hz, 10 Hz, 14 Hz, 20 Hz, 25 Hz, 30 Hz, dla kanałów C3 i C4,
- średnią moc falkową (z użyciem funkcji macierzystej Mexican Hat) dla częstotliwości z przedziału 2,5 Hz–3,5 Hz z krokiem 0,1 Hz, dla kanałów C3 i C4.

Powstałe dokumenty zawierały trzy następujące podziały:

- czuwanie – sen
- NREM – REM
- sen płytki – głęboki.

W celu weryfikacji poprawności klasyfikacji zastosowana została walidacja krzyżowa, w której każdy z zestawów danych został podzielony na pięć podzbiorów. Jako miarę skuteczności wykorzystano średnią skuteczność dla każdego podziału na zbiór treningowy i testowy. Ponieważ w uczeniu maszynowym przeszkodę w osiągnięciu optymalnych wyników stanowić może nadmiar parametrów klasyfikacji, podjęto również próbę redukcji ich liczby o połowę w celu weryfikacji słuszności ich doboru.

4. Uzyskane wyniki

W tabeli 20.1 zebrane zostały dane dotyczące skuteczności działania testowanych algorytmów klasyfikacji.

Tabela 20.1. Skuteczność algorytmów klasyfikacji

	Czuwanie – sen	NREM – REM	Sen płytki – sen głęboki
Sieć neuronowa – wszystkie parametry	82,69%	75,92%	78,94%
Drzewo klasyfikacyjne – wszystkie parametry	87,65%	82,11%	84,74%
KNeighborsClassifier – wszystkie parametry	89,25%	80,58%	84,18%
Sieć neuronowa – wybrane parametry	81,81%	74,79%	76,92%
Drzewo klasyfikacyjne – wybrane parametry	83,36%	79,70%	81,76%
KNeighborsClassifier – wybrane parametry	82,26%	78,74%	82,49%

Każde rozróżnienie przebiegło z większą skutecznością, gdy wykorzystane zostały wszystkie zaproponowane parametry. Korzyść z użycia pełnego zestawu parametrów była często nieznaczna, lecz w niektórych przypadkach wynosiła ponad 5 punktów procentowych. Algorytm KNeighborsClassifier wykazał największą skuteczność w rozróżnieniu snu oraz czuwania, gdzie jego skuteczność wyniosła ponad 89%. Drzewo klasyfikacyjne było najbardziej pasującym narzędziem do rozróżnienia stadiów NREM oraz REM. Oba te modele osiągnęły podobne wyniki w rozróżnieniu snu płytkiego oraz głębokiego. Sieć neuronowa okazała się najmniej skutecznym modelem. W większości rozróżnień jej skuteczność nie przekroczyła 80% i w żadnym z nich nie była ona najlepszym z zaproponowanych rozwiązań.

5. Podsumowanie

Otrzymane w pracy wyniki uzasadniają wykorzystanie zaproponowanych algorytmów, ale nie pozwalają wyłonić, który można uznać za najbardziej skuteczny w obrębie całego omawianego zagadnienia, ponieważ żaden z nich nie okazał się najbardziej pasujący do wszystkich uwzględnionych celów jednocześnie. W żadnym przypadku nie osiągnięto poprawy wyniku związanej z redukcją liczby parametrów klasyfikacji. Świadczy to o słuszności doboru parametrów. Skuteczność przedstawionych rozwiązań potwierdza ich potencjalne zastosowanie w szybkiej i obiektywnej klasyfikacji snu, zwłaszcza z uwzględnieniem możliwości ich udoskonalenia.

Bibliografia

[1] Malhi G., Man J., *Depression*, The Lancet **2018**, 392(10161), s. 2299–2312.

[2] Marin-Alejandre B., Abete I., Cantero I., Riezu-Boj J., Milagro F., Monreal J., Elorz M., Herrero J., Benito-Boillos A., Quiroga J., Martinez-Echeverria A., Uriz-Otano J., Huarte-Muniesa M., Tur J., Martínez J., Zulet M., *Association between Sleep Disturbances and Liver Status in Obese Subjects with Nonalcoholic Fatty Liver Disease: A Comparison with Healthy Controls*, Nutrients **2019**, 11(2), s. 322.

[3] Singh Y., Singh S., Ray A., *Bioelectrical Signals as Emerging Biometrics: Issues and Challenges*, ISRN Signal Processing **2012**, s. 1–13.

[4] Cui Z., Zheng X., Shao X., Cui L., *Automatic Sleep Stage Classification Based on Convolutional Neural Network and Fine-Grained Segments*, Complexity **2018**, s. 1–13.

[5] Elkhenany H., AlOkda A., El-Badawy A., El-Badri N., *Tissue regeneration: Impact of sleep on stem cell regenerative capacity*, Life Sciences **2018**, 214, s. 51–61.

[6] Short M., Louca M., *Sleep deprivation leads to mood deficits in healthy adolescents*, Sleep Medicine **2015**, 16(8), s. 987–993.

[7] *Prezentacja graficzna sieci neuronowej*. Dostępny online: `https://moonbooks.org/media/images/src/neural-network-01.PNG` (dostęp: 24-12-2021).

[8] *Dokumentacja klasy DecisionTreeClassifier*. Dostępny online: `https://www.kite.com/python/docs/sklearn.tree.DecisionTreeClassifier` (dostęp: 01-01-2022).

[9] *Dokumentacja klasy KNeighborsClassifier*. Dostępny online: `https://www.kite.com/python/docs/sklearn.neighbors.KNeighborsClassifier` (dostęp: 01-01-2022).

Ocena wykorzystania sygnału EEG jako narzędzia do testowania funkcjonowania poznawczego

inż. Kinga Kaczmarek
Inżynieria Biomedyczna

dr inż. Wioletta Nowak
Katedra Inżynierii Biomedycznej

W pracy podjęto próbę zbadania, w jaki sposób zadania poznawcze wpływają na rytmy EEG rejestrowane podczas jego wykonywania oraz jak można wykorzystać tę wiedzę do badań nad otępieniami czołowo-skroniowymi. Podczas badania wykorzystano system neurofizjologiczny Neuron-Spectrum-5 umożliwiający pomiar aktywności elektrycznej mózgu oraz specjalistyczne, dedykowane oprogramowanie, które pozwoliło na obserwację rejestrowanego sygnału w czasie rzeczywistym, zapis oraz późniejszą analizę uzyskanych wyników. Dodatkowo wykorzystano cztery rodzaje testów neuropsychologicznych: test MoCA oraz MoCA-B jako badanie przesiewowe, a także test kolorowych wyrazów Stroppa i przygotowane na podstawie testu SLUMS zadanie oceniające pamięć logiczną. W pierwszej wykonano test MoCA w dwóch wersjach – badający wiele funkcji poznawczych jednocześnie. Na podstawie analizy uzyskanych wyników dla poszczególnych elektrod dla pojedynczych uczestników oraz dla całej grupy wyciągnięto wnioski o obszarach mózgu odpowiedzialnych za dane zadanie kognitywne. Na tej podstawie wybrano kolejne dwa testy, które umożliwią badanie aktywności mózgu w części czołowej i skroniowej.

1. Cel i założenia

Celem pracy dyplomowej była ocena wykorzystania sygnału EEG jako narzędzia do testowania funkcjonowania poznawczego z wykorzystaniem aparatu EEG Neuron-Spectrum-5 w badaniu otępień czołowo-skroniowych. W pracy podjęto próbę zbadania, w jaki sposób zadania poznawcze wpływają na rytmy EEG rejestrowane podczas jego wykonywania. Dodatkowo postawiono hipotezę mówiącą o tym, że płat czołowy odpowiada za funkcje językowe oraz uwagę, a płat skroniowy za pamięć. Na tej podstawie przyjęto założenie, że do wykrycia otępienia czołowo-skroniowego mogą służyć takie testy neuropsychologiczne, jak test kolorowych wyrazów Stroopa oraz ocena pamięci logicznej na podstawie kilku krótkich historii.

2. Wstęp teoretyczny

2.1. Elektroencefalografia

Elektroencefalografia (EEG) jest badaniem polegającym na pomiarze aktywności elektrycznej mózgu za pomocą specjalistycznego urządzenia zwanego elektroencefalografem. Do rejestracji potencjałów elektrycznych wykorzystywane są także elektrody najczęściej umieszczane na powierzchni skóry głowy pacjenta. Szczególnym rodzajem badania elektroencefalograficznego jest elektrokortykografia (ECoG). W przypadku tego badania elektrody umieszcza się na powierzchni kory mózgowej. Potencjały czynnościowe, potencjały postsynaptyczne oraz depolaryzacja neuronów stanowią źródło sygnałów, które generuje neuron. W elektroencefalografii raczej nie rejestruje się potencjałów czynnościowych. Wynika to z faktu, że potencjały te wytwarzają zbyt krótkie lokalne impulsy prądowe o małym polu. W zapisie sygnału EEG można zauważyć głównie potencjały postsynaptyczne. Związane jest to z faktem, że są znaczenie dłuższe od potencjałów czynnościowych oraz mają większe pole elektryczne [1].

2.2. Rytmy EEG

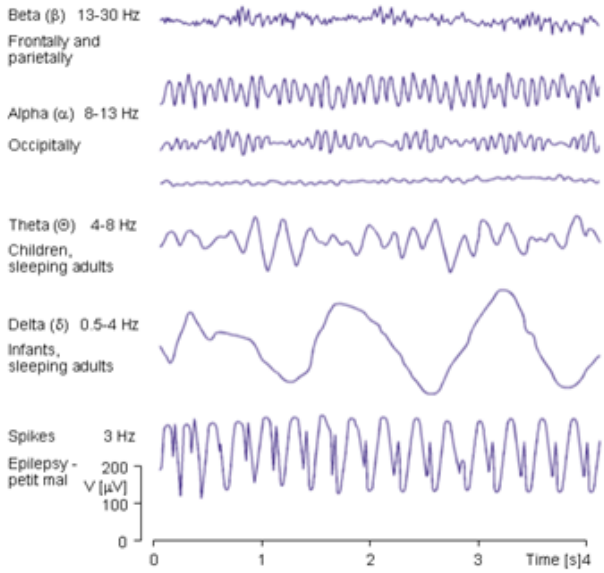
Na wartości norm oraz zakresów parametrów związanych z rejestrowanym sygnałem EEG wpływ mają m.in. wiek oraz stan zdrowia pacjenta. W przypadku osoby dorosłej znajdującej się w spoczynku, mającej zamknięte oczy podstawowym rytmem jest rytm alfa. Charakteryzuje się on amplitudą sygnału o wartości 20–100 µV oraz częstotliwością zmian od 8 Hz do 13 Hz [1]. Posiada sinusoidalny przebieg, a najwyższe amplitudy występują w części potylicznej. Zanika w momencie skupienia uwagi lub w czasie wzmożonego wysiłku poznawczego. Ma bardzo duże znaczenie w analizie snu, ponieważ wskazuje na czuwanie poprzedzające sen. Występujące w czołowo-centralnych częściach mózgu fale beta o częstotliwości zmian od 13 Hz do 30 Hz pojawiają się w momencie skupienia uwagi, pobudzenia lub pracy umysłowej [2]. Pasmo fal beta można dodatkowo podzielić na: wolne fale beta o częstotliwości od 12 do 15 Hz, średnie pasmo beta o częstotliwości od 15 do 18 Hz oraz powyżej 19 Hz szybkie fale beta [6]. Rytm theta występujący w linii środowej. Dominacja tego rytmu w stanie czuwania może świadczyć o senności lub nieprawidłowościach. Charakteryzuje się częstotliwością zmian w zakresie od 4 Hz do 7 Hz, a także amplitudą sygnału o wartości 30 µV [1]. Rytm ten dominuje w czasie snu bądź głębokiej medytacji [2]. Zakres częstotliwości fali delta mieści się w przedziale od 0,5 Hz do 4 Hz. Rytm ten związany jest z głębokim snem oraz głęboką medytacją. Rytmem związanym z aktywnością ruchową jest rytm gamma. Zakres związanych z nim fal jest dość szeroki, ponieważ obejmuje przedział od 31 Hz, aż do 100 Hz [1].

Zwyczajowo do rejestracji zapisu EEG wykorzystuje się od 19 do 21 elektrod rozmieszczonych na powierzchni głowy w odpowiedni sposób. Najpopularniejszy schemat rozlokowania elektrod na skórze głowy nosi nazwę systemu 10-20. Elektrody umieszcza się pod gumowym czepkiem w celu ich utrzymania na odpowiednim miejscu. Litery w skrócie danej elektrody pochodzą od łacińskich nazw obszarów anatomicznych czaszki:

- FP to elektrody przedczołowe,
- F oznacza elektrody czołowe,

- C to elektrody centralne,
- T natomiast to elektrody umieszczane w części skroniowej,
- P odpowiadają płatowi ciemieniowemu,
- O znajdują się w okolicy potylicznej,
- A natomiast to elektrody uszne.

Cyfra odpowiada lokalizacji na danej półkuli mózgu. Numery parzyste odpowiadają elektrodom umieszczonym na prawej półkuli, natomiast numery nieparzyste na lewej [5].



Rysunek 21.1. Przykładowe fale występujące w sygnale EEG [4]

2.3. Funkcje poznawcze

Wśród procesów poznawczych wyróżniamy dwa typy: elementarne oraz złożone [6]. Elementarne procesy poznawcze to przede wszystkim:

- Uwaga (wpływa na selekcję docierających informacji),
- Percepcja (odpowiada za informacje odbierane z otoczenia, możemy wyróżnić percepcję smakową, wzrokową, węchowa, zmysł równowagi),
- Pamięć (umożliwia przechowywania informacji słuchowej, wzrokowej, dotykowej),
- Kontrola poznawcza (pojęcie związanym z kontrolowaniem, jak przebiegają procesy poznawcze).

Natomiast wśród złożonych procesów poznawczych wyróżnia się:

- Myślenie (łączenie wiadomości poznanych z bodźcami oraz informacjami z otaczającej rzeczywistości),
- Język (system znaków i sposób komunikowania się za ich pomocą),
- Funkcje wykonawcze (procesy, które mają za zadanie odpowiadać za przemyślane i sensowne reakcje jednostki).

Płat czołowy odpowiada między innymi za mowę oraz odpowiednie wypowiadanie słów, zatem związany jest z funkcjami językowymi, a przede wszystkim z fluencją. Płat

przedczołowy jest odpowiedzialny za planowanie i tworzenie długoterminowych celów oraz za dopasowywanie pewnych zachowań do związanych z nimi emocji, łączy więc w sobie kilka funkcji poznawczych. Płat skroniowy odpowiada za zapamiętywanie informacji oraz pamięć leksykalną, jego działanie skupia się więc głównie na funkcji poznawczej, jaką jest pamięć. Płat ciemieniowy odpowiada za percepcję przestrzenną oraz motorykę, natomiast płat potyliczny za percepcję wzrokową [7].

2.4. Otępienie skroniowo-czołowe

Otępienie skroniowo-czołowe (FTD) jest postępującym z biegiem lat uszkodzeniem ośrodkowego układu nerwowego. Głównie objawia się w postaci zmian zwyrodnieniowych w obrębie płata czołowego i przednio-skroniowego [8]. Wyróżnić można trzy podstawowe odmiany otępień czołowo-skroniowych. Należą do nich:

- postać czołowa otępienia (wariant czołowy),
- postać skroniowa otępienia (wariant skroniowy),
- postępująca afazja w przebiegu której dominują zanurzenia funkcji językowych [8].

Otępienie skroniowo-czołowe przyjmuje także dwa warianty fenotypowe, takie jak:

- wariant behawioralny, w których dominują zaburzenia zachowania spowodowane powstawaniem patologicznych zmian w części przedczołowej,
- wariant językowy charakteryzuje się problemami z mową [8].

Wśród najbardziej popularnych objawów schorzenia można wyróżnić [8]:

- zaburzenia zachowania,
- zaburzenia mocy,
- zmiany osobowości,
- zmienne nastroje,
- zachowania nieadekwatne do sytuacji,
- pobudliwość,
- brak krytycznej oceny własnego zachowania.

3. Projekt i wykonanie

Grupa biorąca udział w badaniu składała się z 10 osób, wśród których było sześć kobiet oraz czterech mężczyzn w wieku od 19 do 28 lat. Były to osoby zdrowe, które nie miały zdiagnozowanych chorób na tle neurologicznym.

3.1. Narzędzia pomiarowe oraz testowe

Urządzeniem pomiarowym wykorzystywanym podczas badania był system neurofizjologiczny Neuron-Spectrum-5 umożliwiający pomiar aktywności elektrycznej mózgu. Głównym elementem aparatu pomiarowego jest 41-kanałowa głowica. Ważną częścią całego systemu pomiarowego jest także specjalistyczne, dedykowane oprogramowanie Neuron-Spectrum instalowane na PC, które umożliwia obserwację rejestrowanego sygnału EEG w czasie rzeczywistym, zapis przebiegu badania, a także późniejszą analizę uzyskanych wyników. Elektrody EKG połączone z aparatem Neuron-Spectrum-5 zostały wykorzystane podczas pomiaru dodatkowo w celu zbadania rytmu serca w trakcie wykonywania zadań poznawczych. Zarejestrowany sygnał EKG pozwolił na usunięcie artefaktów w sygnale EEG pochodzących od aktywności elektrycznej serca.



Rysunek 21.2. Aparat neurofizjologiczny Neuron-Spectrum-5

Podczas przeprowadzanych badań wykorzystywane są cztery rodzaje testów neuropsychologicznych. Pierwszy etap badania składa się z połączonego testu MoCA w wersji 8.1 oraz testu MoCA-B. Połączenie tych testów zapewnia kompleksowe zbadanie wszystkich funkcji poznawczych. Na podstawie pierwszej części eksperymentu, pełniącej rolę badania przesiewowego, wyróżnia się rodzaje zadań najsilniej oddziałujących na sygnał elektryczny mózgu w płacie czołowym oraz skroniowym. Test bada różne funkcje poznawcze człowieka, takie jak: pamięć, koncentracja, uwaga, orientacja, zdolności kalkulacyjne, myślenie koncepcyjne oraz funkcje językowe, wykonawcze, a także wzrokowo-przestrzenne. Zmodyfikowany test składa się z 18 zadań. Maksymalna liczba punktów do uzyskania wynosi 42. Wynik 37 lub więcej jest uważany za prawidłowy. Na tej podstawie wybiera się funkcje poznawcze, za które odpowiada część mózgu będąca głównym przedmiotem analizy. W drugiej części badania wykorzystano dwa testy neuropsychologiczne związane ze zwiększoną czynnością mózgu w płacie czołowym i skroniowym. Test kolorowych wyrazów Stroopa składa się z 3 części. Polega na odczytaniu/podaniu nazw kolorów zgodnie z poleceniem osoby przeprowadzającej badanie. Składa się ze 100 wyrazów podanych w pięciu kolumnach po 20 wyrazów w każdej. Słowa zapisane w kolumnach są nazwami trzech kolorów: zielony, czerwony i niebieski. Każda z części testu trwa po 45 sekund. Czas został dostosowany do wieku grupy badawczej; w przypadku osób z otępieniami czołowo-skroniowymi należy czas badania wydłużyć. Wyrazy pojawiają się na ekranie komputera i samoczynnie znikają po 45 sekundach. Wtedy pojawia się kolejna lista słów. Część pierwsza polega na odczytaniu nazw trzech kolorów zapisanych czarnym atramentem. W drugiej części od badanego oczekuje się identyfikacji koloru na podstawie użytego atramentu. Trzecia część wymaga podania nazwy koloru na podstawie atramentu, niezgodnie z zapisem nazwy. Wyniki testu zależą od sprawnej uwagi, test bada także funkcje językowe. Pomysł na test trzeci został zaczerpnięty na podstawie przeanalizowania zadania z testu SLUMS. Jako przydatne do oceny otępień skroniowo-czołowych wybrano ostatnie zadanie, dotyczące oceny pamięci logicznej. Zadanie stworzone na potrzeby eksperymentu składa się z trzech

krótkich historyjek. Następnie na podstawie historyjek sforułowano sześć pytań, po dwa pytania do każdej. Test polegał na skupieniu uwagi przez uczestnika badania na czerwonej kropce wyświetlanej na ekranie, stanowiącej punkt fiksacji oraz uważnym słuchaniu trzech tekstów. Po zakończeniu czytania przeprowadzający badanie robi krótką przerwę, a następnie czyta wolnym tempem kolejne pytania. Pomiędzy pytaniami występują przerwy na odpowiedź badanego.

3.2. Faza przygotowawcza

Eksperyment rozpoczęto od doboru narzędzi pomiarowych i testowych. Wyboru dokonano na podstawie przeglądu dostępnych źródeł literatury. Specjalistyczne oprogramowanie Neuron-Spectrum zainstalowane na PC pozwoliło także na optymalny dobór montażu elektrod oraz parametrów próbkowania i filtracji sygnału. Podczas badania wykorzystywano 17 elektrod, w tym 14 elektrod pomiarowych, 2 elektrody referencyjne oraz jedną elektrodę jako uziemienie. Wśród typów wykorzystywanych elektrod można wyróżnić: elektrody grzybkowe służące jako elektrody pomiarowe, zamocowane na uszach elektrody klipsowe pełniące funkcję elektrod referencyjnych oraz elektrodę miseczkową służącą za masę. Częstotliwość próbkowania sygnału w trakcie badania wynosiła 500 Hz, dodatkowo zastosowano również filtry cyfrowe: filtr sieciowy, filtr górnoprzepustowy o częstotliwości granicznej 0,1 Hz, a także filtr dolnoprzepustowy o częstotliwości granicznej 35 Hz. Podczas analizy sygnału zauważono, że wybrane filtry są niewystarczające. Dodatkowo zastosowano jeszcze filtr pasmowo-przepustowy o zakresie częstotliwości 8 Hz–35 Hz. Przed przystąpieniem do pomiaru właściwego uczestnikowi badania na powierzchnię skóry głowy w kilku miejscach zostaje nałożony żel, a następnie specjalny czepek do zamocowania elektrod zbierających sygnał EEG. Oprócz elektrod EEG zamocowanych zgodnie ze stworzonym montażem dodatkowo przypinane są również do rąk i nóg elektrody EKG. Czas potrzebny na przypięcie wszystkich elektrod wynosił około 30 minut. Ważnym elementem jest także sprawdzenie impedancji elektrod. Pomiar taki umożliwia oprogramowanie Neuron-Spectrum. Dążymy do uzyskania podobnych wartości impedancji dla wszystkich elektrod, jednocześnie nie większych niż 11–12 kΩ.

3.3. Pomiar właściwy

Przed przystąpieniem do badania uczestnik usadawia się wygodnie na krześle. Następnie jest proszony o uważne zapoznanie się z instrukcją i wypełnienie formularza świadomej zgody na udział w badaniu. Instrukcja zawiera informacje o badaniu oraz o sposobie przetwarzania danych zebranych podczas badania. Badany zostaje także poproszony, aby podczas badania nie wykonywał gwałtownych ruchów mogących zaburzyć pomiar lub odczepienie elektrod, a także o wykonywanie poleceń osoby przeprowadzającej badanie. Kolejno osoba badana proszona jest o wypełnienie kwestionariusza osobowego. Zawiera on pytania mające na celu zebranie najistotniejszych informacji o osobie badanej, które mogą mieć wpływ na pomiar EEG. Między innymi osoba badana musiała odpowiedzieć ile godzin spała poprzedniej nocy, czy odczuwa zmęczenie, czy cierpi na bezsenność lub miewa inne problemy ze snem, a także czy choruje na migrenę lub przewlekłe bóle głowy. Następnie uruchamiany jest zapis EEG. Sygnał rejestrowany jest przez specjalistyczne

oprogramowanie w sposób ciągły, z uwzględnieniem przerw między kolejnymi zadaniami, ułatwiających interpretację uzyskanych wyników. Początkowo osoba badana jest proszona o zamknięcie oczu i zrelaksowanie się przez około 1–2 minuty. Pomiar ten będzie pomiarem odniesienia dla innych sygnałów uzyskiwanych przy wykonywaniu określonego zadania poznawczego. Kolejno następuje właściwa część pomiarowa. Rozpoczyna się przeprowadzeniem pierwszej części testu, czyli połączonego testu MoCA wersja 8.1 oraz MoCA-B. W programie zaznaczane są za pomocą specjalnych znaczników momenty zamknięcia, otwarcia oczu oraz czas uwzględniony na przeczytanie polecenia. Po zakończeniu pierwszej części testu osoba prowadząca badanie prosi o zamknięcie oczu i zrelaksowanie się. Pierwsza część testu trwa około 15–20 minut w zależności od badanego. W drugiej części pomiaru osoba badana wykonuje test kolorowych wyrazów Stroopa. Ta część testu trwa około 3 minut. Ostatni test trwa około 2–3 minut. Po zakończeniu pomiaru przeprowadzany jest wywiad z osobą badaną dotyczący jej odczuć, a także poziomu zmęczenia psychicznego.

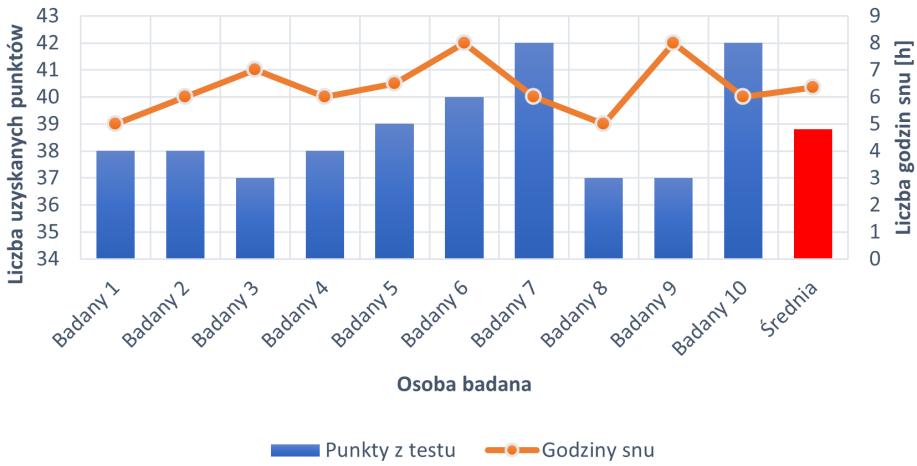
4. Uzyskane wyniki

Analizę uzyskanych podczas przeprowadzanego badania wyników rozpoczęto od analizy testu MoCA w wersji 8.1 połączonego z testem MoCA-B w wersji papierowej. Natomiast wyniki uzyskane w postaci zapisu EEG analizowane były przez ich podział na dwie części. W pierwszej części analizy pod uwagę bierze się wyłącznie test MoCA. Na jej podstawie wybrane zostały zadania, które miały największy wpływ na częstotliwość sygnału w części czołowo-skroniowej mózgu. Taka analiza pozwoliła ustalić, jakie funkcje poznawcze człowieka zlokalizowane są w płacie czołowym i skroniowym. Analizę pierwszej części badania można przeprowadzić na dwa sposoby. Pierwszy sposób polegał na analizie zmian sygnału EEG na poszczególnych elektrodach dla pojedynczych osób. Wariant ten uwzględnia cechy osobnicze, które różnią się między jednostkami. Pozwala ustalić, czy dla danego zadania można zaobserwować powtarzalny schemat między jednostkami. Drugi wariant zakładał natomiast próbę analizy całościowej, bez uwzględniania różnic osobniczych pomiędzy poszczególnymi badanymi. W drugiej części analizy pod uwagę brany jest test kolorowych wyrazów Stroopa oraz ocena pamięci logicznej na podstawie zadania z trzema krótkimi tekstami. Tutaj ważne są jedynie elektrody zbierające sygnał z części czołowej i skroniowej kory mózgowej, a więc trzy elektrody przedczołowe oraz trzy elektrody zlokalizowane na wysokości płata czołowego i dwie zbierające sygnał z części skroniowej. Pierwszy sposób analizy polega na badaniu zmian sygnału EEG na poszczególnych elektrodach dla każdej z osób, żeby uwzględnić różnice międzyosobnicze. Drugi to próba analizy sygnału na poszczególnych elektrodach dla całej grupy osób badanych jednocześnie. W trzecim wariancie sprawdzono czy występują różnice między dwiema częściami grupy badanej. Podzielono ją więc na kobiety oraz mężczyzn i dokonano analizy.

4.1. Porównanie wyników testu MOCA wersja 8.1 połączonego z MOCA-B

Wyniki testu neuropsychologicznego w wersji papierowej służą do potwierdzenia, że grupę badawczą stanowiły osoby zdrowe, bez wyraźnych oznak otępienia skroniowo-czołowego.

Zadania można podzielić na grupy. Poszczególne grupy umożliwiają zbadanie innych funkcji poznawczych człowieka. Maksymalna liczba punktów z testu wynosi 42, natomiast minimalna 37. Są jednak pewne odstępstwa od tej skali w zależności od poziomu edukacji osób badanych. Nie uwzględniono jednak ich w trakcie wykonywania analizy z uwagi na to, że wszystkie osoby biorące udział w badaniu posiadają wykształcenie wyższe lub średnie. Wśród wyników najczęściej pojawia się 38 punktów. Liczba godzin snu nie miała mimo wszystko istotnego wpływu na uzyskane punkty. Osoby, które spały krócej, uzyskiwały lepsze rezultaty od osób śpiących po 7–8 godzin. Średnia liczba godzin snu w grupie osób badanych wynosiła 6,4 h.

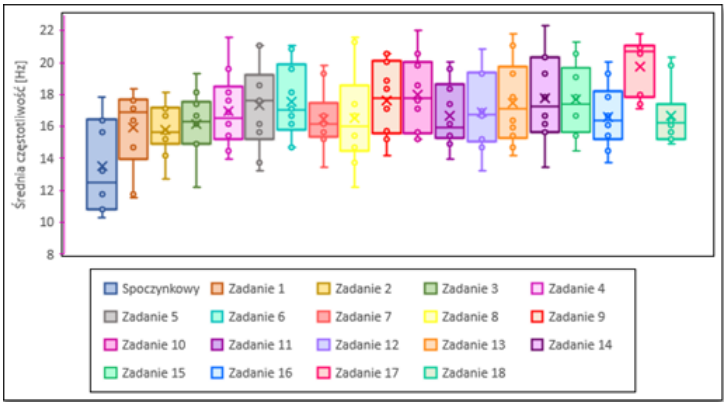


Rysunek 21.3. Liczba uzyskanych punktów w teście neuropsychologicznym MoCA oraz liczba godzin snu dla każdej z osób badanych, a także średnia tych dwóch składowych

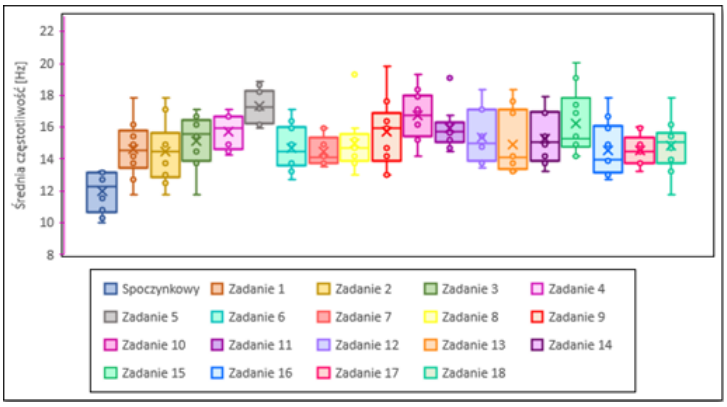
4.2. Analiza zmian sygnału EEG na poszczególnych elektrodach dla testu MOCA oraz MOCA-B

Najpierw dokonano analizy wyników uzyskanych dla badania przesiewowego, którym był test MoCA w wersji 8.1 połączony z wersją MoCA-B. Analizując uzyskany sygnał EEG, skupiono się przede wszystkim na analizie częstotliwościowej. Mimo iż starano się podczas pomiarów uzyskać zbliżone wartości impedancji elektrod, to jednak nie udało się całkowicie wyeliminować różnic między nimi. Wybierając analizę amplitudową, należałoby się liczyć z błędami w rejestracji napięć sygnałów mimo takiej samej aktywności mózgu dla danego zadania. Biorąc pod uwagę wszystkie elektrody wykorzystane w eksperymencie, największe wartości częstotliwości średniej w przypadku elektrod przedczołowych zarejestrowano w przypadku zadań badających uwagę oraz funkcje językowe. Patrząc łącznie na elektrody zbierające sygnał elektryczny mózgu z przedniej części płata skroniowego, największą aktywność mózgu dla każdej z tych elektrod zaobserwowano w przypadku zadań badających pamięć. Dla elektrod badających aktywność w płacie czołowym mózgu zaobserwować można przewagę zadań dotyczących uwagi oraz funkcji językowych, zarówno dotyczących fluencji, jak i nazywania. Kolejno analizując sygnał EEG rejestrowany z części

znajdującej się na granicy płata czołowego oraz ciemieniowego, można dostrzec, że najwyższe wartości częstotliwości zarejestrowano dla zadań badających percepcję wzrokowo-przestrzenną oraz uwagę. W przypadku elektrod ciemieniowych największy wpływ na częstotliwość miały percepcja wzrokowo-przestrzenna oraz uwaga.



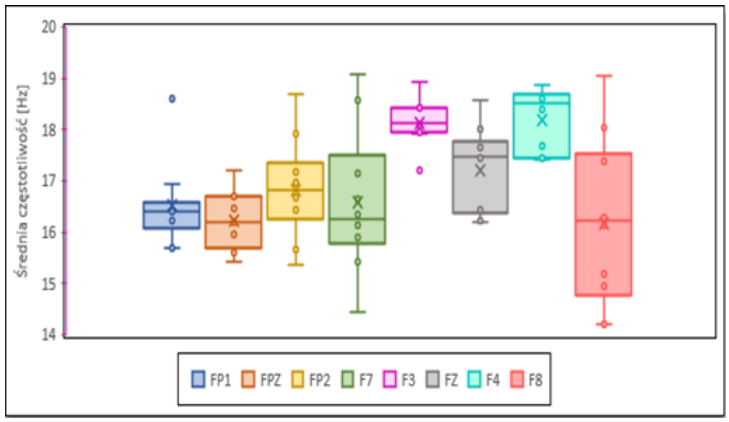
Rysunek 21.4. Zakres średnich częstotliwości dla całej grupy badanej dla elektrody F7



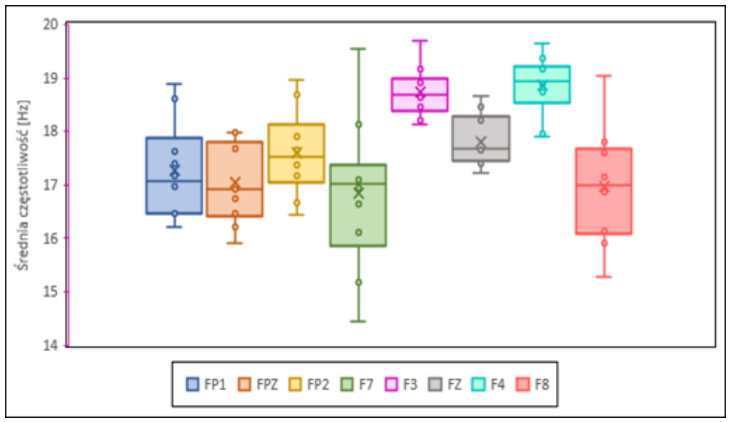
Rysunek 21.5. Zakres średnich częstotliwości dla całej grupy badanej dla elektrody F4

4.3. Analiza zapisu EEG dla testu kolorowych wyrazów Stroopa oraz oceny pamięci logicznej

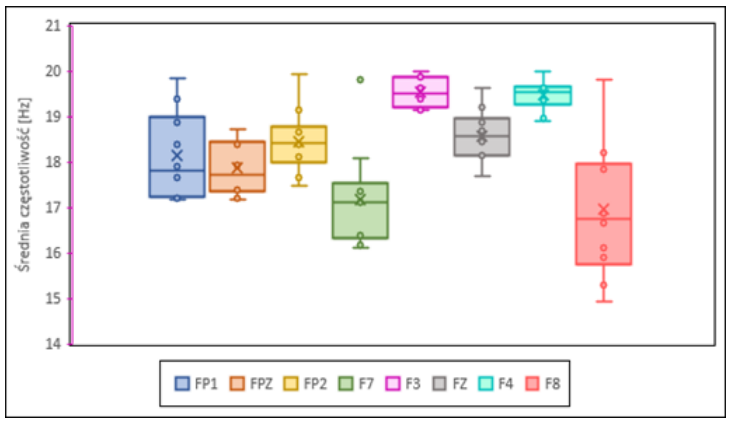
W drugiej części doświadczenia wykonano test kolorowych wyrazów Stroopa oraz zadanie oceniające pamięć logiczną. Do analizy wybrano w tym przypadku tylko elektrody przedczołowe, czołowe oraz przednio-skroniowe. Na zaprezentowanych poniżej wykresach przedstawiono, jak zmienia się częstotliwość średnia w przypadku elektrod czołowych i przednio-skroniowych dla całej grupy badanej w przypadku trzeciej części testu kolorowych wyrazów Stroopa oraz zadania oceniającego pamięć logiczną.



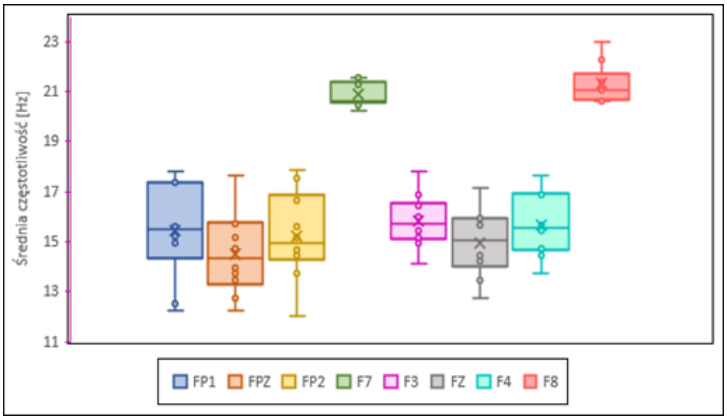
Rysunek 21.6. Zakres średnich częstotliwości dla całej grupy badanej dla każdej z elektrod w przypadku pierwszej części testu kolorowych wyrazów Stroopa



Rysunek 21.7. Zakres średnich częstotliwości dla całej grupy badanej dla każdej z elektrod w przypadku drugiej części testu kolorowych wyrazów Stroopa



Rysunek 21.8. Zakres średnich częstotliwości dla całej grupy badanej dla każdej z elektrod w przypadku trzeciej części testu kolorowych wyrazów Stroopa



Rysunek 21.9. Zakres średnich częstotliwości dla całej grupy badanej dla każdej z elektrod w przypadku zadania oceniającego pamięć logiczną osób badanych

Analizując uzyskane wykresy dla wszystkich trzech części testu kolorowych wyrazów Stroopa, można zauważyć, że w każdej części największe wartości częstotliwości średniej przeważają dla elektrod czołowych F3 oraz F4. Można również zaobserwować, że wraz ze wzrostem poziomu trudności testu Stroopa zwiększa się różnica między wartościami średnich częstotliwości dla elektrod czołowych i przedczołowych. W przypadku zadania oceniającego pamięć logiczną badanego można wyraźnie zauważyć przewagę dużo wyższych częstotliwości średnich rejestrowanych przez elektrody przednio-skroniowe w porównaniu do tych wartości, które rejestrują elektrody czołowe.

4.4. Analiza zmian sygnału EEG na elektrodach czołowych i przednio-skroniowych dla dwóch grup badanych

Na podstawie wyników badań dla obu płci trudno dostrzec tendencję zmian wartości średniej częstotliwości między badanymi grupami. Można jedna zauważyć, ze w przypadku elektrod F7 oraz F8 to mężczyźni podczas wykonywania zadań zarówno z testu kolorowych wyrazów Stroopa, jak i zadania oceniającego pamięć logiczną uzyskują wyższe częstotliwości. Za to w przypadku kobiet na pozostałych elektrodach, to znaczy na elektrodach czołowych oraz przedczołowych, uzyskano wyższe częstotliwości.

4.5. Analiza spektrum mocy

Na przedstawionych rysunkach obrazujących mapy spektrum mocy przedstawiono, jak zmienia się moc podczas wykonywania testu kolorowych wyrazów Stroopa oraz zadania oceniającego pamięć logiczną dla jednej z osób badanych. Kolorem czerwonym oznaczone są strefy mózgu o największej mocy sygnału, natomiast im kolor zimniejszy, tym mniejsza moc. Obszarowi o najmniejszej mocy odpowiada kolor niebieski. W przypadku pozostałych badanych tendencja zmian i lokalizacji maksymalnych częstotliwości była bardzo zbliżona do wybranego przykładu.



Rysunek 21.10. Mapa spektrum mocy w przypadku testu kolorowych wyrazów Stroopa



Rysunek 21.11. Mapa spektrum mocy w przypadku zadania oceniającego pamięć logiczną

Na podstawie ukazanych map spektrum mocy można zauważyć, że w przypadku testu kolorowych wyrazów Stroopa obszarem mózgu o największej mocy sygnału jest płat czołowy, natomiast w przypadku zadania oceniającego pamięć logiczną zdecydowanie najwyższe wartości mocy rejestrowane są w płacie skroniowym.

5. Podsumowanie

Badania elektroencefalograficzne z zastosowaniem testów neuropsychologicznych stanowią bardzo duże wyzwanie logistyczne ze względu na bardzo dużą różnorodność zadań kognitywnych. W przeprowadzanym badaniu starano się wybrać zadania, które nie wymagają od uczestnika ruchu, aby nie zakłócić przebiegu sygnału w znaczący sposób. Jednak ze względu na wybór testu MoCA, który został zastosowany do badań przesiewowych, nie udało się całkowicie uniknąć części związanej z percepcją przestrzenną.

Poza powyższym czynnikiem na sygnał elektroencefalograficzny wpływ mógł mieć stan psychiczny osób badanych. Część badanej grupy odczuwała zmęczenie przed wykonywanym testem, co mogło mieć wpływ na ich uwagę oraz koncentrację. Kolejnym zakłóceniem były warunki panujące w laboratorium. Część z osób badanych w rozmowie przeprowadzonej po zakończeniu pomiaru mówiło, że słyszało dźwięki z zewnątrz, które rozpraszały ich uwagę i utrudniały skupienie się na zadaniu.

Ważnym elementem mającym także wpływ na analizowany sygnał EEG jest aparatura pomiarowa. Wszystkie odczyty, jakie uzyskano podczas badania charakteryzują się bardzo dużą mocą w zakresie fal delta. Fale te występują w zapisie EEG tylko w czasie głębokiego snu oraz medytacji. Widoczne są także w sygnale u osób z różnymi chorobami neurodegeneracyjnymi. Zatem obecność fal delta nie jest pożądana w sygnale oraz nie była w nim spodziewana. Jej pojawienie się zostało uznane za artefakt. Jest ono prawdopodobnie związane z aktywnością elektryczną serca.

Pierwsza część pomiaru będąca badaniem przesiewowym pozwoliła zaobserwować, jak oraz jakie funkcje poznawcze wpływają na częstotliwość sygnału w płacie czołowym oraz skroniowym. Uzyskane rezultaty sugerują, że płat skroniowy odpowiada za funkcje pa-

mięciowe. Znaczny wzrost częstotliwości w tym obszarze dla zadań 6 oraz 17 świadczy o zaangażowaniu tego obszaru mózgu w wykonywane przez uczestnika zadanie. Wartości uzyskanych częstotliwości podczas tych zadań są w zakresie 19–21 Hz i mają związek z dużym wysiłkiem umysłowym. Zgodnie z wcześniejszą wiedzą dotyczącą płata skroniowego oraz jego zadań potwierdzono, że jest on obszarem odpowiadającym za pamięć werbalną. W przypadku płata czołowego oraz obszaru przedczołowego można wysnuć wniosek, że ten rejon mózgu odpowiada przede wszystkim za funkcje językowe oraz uwagę. Zadania poznawcze dotyczące funkcji językowych, do których zaliczamy zarówno te badające fluencję, jak i dotyczące nazewnictwa, powodują wzrost częstotliwości do wartości około 17–19 Hz w zależności od osoby badanej. Tak samo jest w przypadku uwagi. Zgodnie z wcześniejszą wiedzą obszar czołowy odpowiada za mowę oraz za logiczne rozumowanie i koncentrację. Zatem uzyskane wyniki potwierdzają wcześniejszą wiedzę co do spodziewanych wyników eksperymentu. U kilku osób występują też fale SMR, szczególnie w części czołowej i przedczołowej. Zgodnie ze stanem obecnej wiedzy jest to normalne i wśród części populacji w stanie relaksu dominują niskie fale beta w przedniej części mózgu.

Druga część eksperymentu dotyczyła potwierdzenia postawionej na wstępie hipotezy, że test kolorowych wyrazów Stroopa oraz zadanie oceniające pamięć logiczną są odpowiednie do zbadania części czołowej oraz skroniowej mózgu w celu wykorzystania ich w przyszłości do badania otępień w tym obszarze. Na podstawie uzyskanych rezultatów wysunięto następujące wnioski:

- Zadanie mające na celu ocenę pamięci logicznej wpływa na znaczny wzrost częstotliwości w obszarze przednio-skroniowym mózgu.
- W przypadku testu kolorowych wyrazów Stroopa wzrost częstotliwości następuje głównie w przypadku sygnału rejestrowanego przez elektrody czołowe oraz przedczołowe. Wraz z każdą częścią testu następuje większy wzrost częstotliwości. Test niezgodny stanowi większy wysiłek poznawczy i sprawia więcej trudności, stąd takie rezultaty badania. W płacie czołowym zarejestrowano częstotliwości powyżej 20 Hz odpowiadające zakresowi dla fal beta2. Świadczy to o dużym zaangażowaniu tej części mózgu w wykonywane zadanie.

Uzyskane rezultaty podczas badania mogą stanowić podstawę do dalszych badań. Ze względu na zbyt małą grupę osób badanych nie mogą stanowić one jednoznacznego potwierdzenia hipotezy. Ze względu na trudności napotkane podczas przygotowań do badania i sytuację epidemiologiczną nie udało się przeprowadzić badań dla szerszej grupy. Aby potwierdzić wnioski wyciągnięte na podstawie eksperymentu, należałoby rozszerzyć grupę uczestników o osoby z otępieniami czołowo-skroniowymi w celu porównania wyników między grupą osób chorych oraz zdrowych. Dodatkowo można poprawić także warunki przeprowadzanego badania. Ze względu na bardzo dużo bodźców rozpraszających wyniki mogły być częściowo zaburzone, stąd warto byłoby przeprowadzić eksperyment w izolowanym od bodźców zewnętrznych pomieszczeniu. Należałoby także zwiększyć grupę badawczą oraz kontrolną do liczby, która pozwoli stworzyć wiarygodne dane statystyczne. Dobrze byłoby także uprościć test kolorowych wyrazów Stroopa do mniejszej liczby słów lub zwiększyć ilość czasu ze względu na to, że jeśli tak duże trudności napotkały osoby zdrowe, mógłby on być niemożliwy do wykonania przez osobę chorą.

Bibliografia

[1] Hulewicz A., Jukiewicz M., *Analiza sygnałów EEG na potrzeby interfejsu mózgkomputer*, Acta Bio-Optica et Informatica Medica. Inżynieria Biomedyczna **2014**, 20(3), s. 137–143.

[2] Broniec A., Chodak J., *Sterowanie prostym urządzeniem elektrycznym za pomocą sygnału EEG*, Automatyka/Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie **2009**, 13, s. 1059–1067.

[3] Durka P., Duszyk A., Kamiński M., Kuś R., Malinowska U., Mikuła I., Żygierewicz, J., *Elektryczny ślad myśli*, Uniwersytet Warszawski, Warszawa, Polska, 2012.

[4] *Rytmy EEG*. Dostępny online: `http://wikidyd.iem.pw.edu.pl/attachments/PodstBioinz/EEG` (dostęp: 11-12-2021).

[5] Zyss T., *Technika digitalizacji położenia elektrod w badaniu elektroencefalograficznym EEG*, Acta Bio-Optica et Informatica Medica. Inżynieria Biomedyczna **2011**, 17(1), s. 72–78.

[6] Lezak M.D., Howieson D.B., Loring D.W., Fischer J.S., *Neuropsychological assessment*, Oxford University Press, USA 2004.

[7] Modrak M., *Pamięć sensoryczna czyli myśleć ciałem. Doskonalenie zasobów pamięci zmysłowej*, Difin SA, Warszawa, Polska, 2016.

[8] Moryś J.M., Moryś J., Bidzan L., Herman-Sucharska I., Urbanik A., Lass P., Guzińska K., *Otępienie czołowo-skroniowe: ujęcie interdyscyplinarne*, Oficyna Wydawnicza AFM, Kraków, Polska, 2012.

Oficyna Wydawnicza
Politechniki Wrocławskiej
Wybrzeże Wyspiańskiego 27
50-370 Wrocław
ISBN: 978-83-7493-226-4

