

Roman REGULSKI  
Piotr OWCZAREK  
Dominik RYBARCZYK  
Paweł BACHMAN  
Jarosław GOŚLIŃSKI

## STEROWANIE GŁOSOWE MANIPULATOREM ELEKTROHYDRAULICZNYM

**STRESZCZENIE** *W artykule przedstawiono nowatorskie rozwiązanie sterowania manipulatorem elektrohydraulicznym za pomocą komend głosowych. W ramach pracy opracowano interfejs na komputer klasy PC oraz na platformę mobilną Android. Obydwa rozwiązania pozwalają w prosty sposób sterować pozycją manipulatora z wykorzystaniem technologii rozpoznania mowy. Głównym celem tej pracy jest przedstawienie możliwości i użyteczności systemów rozpoznawania mowy pod kątem sterowania urządzeniami.*

**Słowa kluczowe:** *sterowanie głosowe, rozpoznawanie mowy, manipulator elektrohydrauliczny*

### 1. WSTĘP

---

Systemy automatycznego rozpoznawania mowy znajdują coraz więcej zastosowań w dziedzinie mechatroniki i robotyki. Stają się nieodzownym elementem inteligentnych interfejsów typu człowiek-maszyna. Pozwalają one użytkownikom urządzeń komunikować się z nimi w najbardziej naturalny sposób – za pomocą mowy. W ostatnich latach można zauważyć próby opracowania takich systemów w rozwiązaniach przeznaczonych dla przemysłu [6, 7, 8, 9].

---

**mgr inż. Roman REGULSKI, mgr inż. Piotr OWCZAREK,  
mgr inż. Dominik RYBARCZYK, mgr inż. Jarosław GOŚLIŃSKI**  
e-mail: roman.regulski@doctorate.put.poznan.pl; piotr.owczarek@put.poznan.pl;  
dominik.rybarczyk@put.poznan.pl; jaroslaw.a.goslinski@doctorate.put.poznan.pl

Politechnika Poznańska, ul. Piotrowo 3, 60-965 Poznań

**dr inż. Paweł BACHMAN**  
e-mail: p.bachman@eti.uz.zgora.pl

Uniwersytet Zielonogórski, ul. Licealna 9, 65-417 Zielona Góra

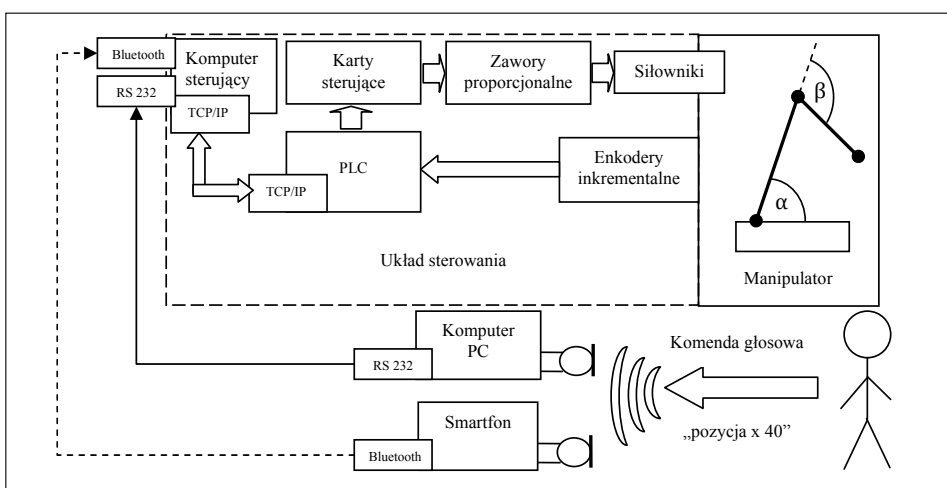
PRACE INSTYTUTU ELEKTROTECHNIKI, zeszyt 263, 2013

Rozpoznawanie mowy jest procesem zamiany sygnału dźwiękowego na tekst. W większości rozwiązań proces ten realizowany jest poprzez oprogramowanie komputerowe, które korzysta z zaawansowanych algorytmów cyfrowego przetwarzania sygnałów, modeli językowych oraz procesów statystycznych. Na skuteczność rozpoznawania mowy wpływa wiele czynników. Nieustanny rozwój technik, metod i algorytmów komputerowych doprowadził do tego, że obecnie rozpoznawanie mowy może być przeprowadzane niemalże w czasie rzeczywistym z wysoką skutecznością, co czyni je użytecznym do wykorzystania w sterowaniu wszelkimi urządzeniami za pomocą ludzkiego głosu [1, 2, 3, 4, 5].

## 2. OPIS STANOWISKA BADAWCZEGO

W ramach pracy opracowano interfejs typu człowiek-maszyna, pozwalający sterować manipulatorem elektrohydraulicznym za pomocą komend głosowych. Manipulator posiada kinematykę równoległą i dwie sterowane osie. Napęd stanowią siłowniki hydrauliczne kontrolowane zaworami proporcjonalnymi. Zawory proporcjonalne zostały wysterowane za pomocą dedykowanych kart sterujących z wejściem analogowym w standardzie napięciowym  $\pm 10$  V. Układ pomiarowy położenia ramion stanowią enkodery inkrementalne o rozdzielczości 3600 impulsów/obrót, które zostały umiejscowione w przegubach. Sygnał z enkoderów pozwala określić położenie katowe ramion z dokładnością wynoszącą  $0,025^\circ$ .

Manipulator pracuje pod kontrolą sterownika PLC, który odpowiada za wyliczenie kinematyki odwrotnej i pozycji ramion. W układzie sterowania znajduje się również komputer z oprogramowaniem PVI Menager (ang. *Process Visualization Interface*), umożliwiającym wizualizację położenia ramion manipulatora. Komunikacja pomiędzy sterownikiem PLC a komputerem sterującym odbywa się za pomocą połączenia TCP/IP. Na rysunku 1 przedstawiono schemat stanowiska testowego.

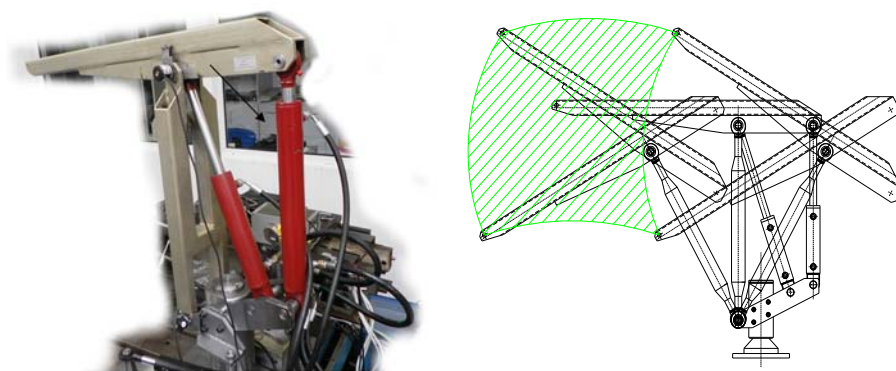


Rys. 1. Schemat stanowiska testowego

Podczas projektowania stanowiska do sterowania głosowego zdecydowano się nie ingerować w układ sterowania manipulatora. Interfejs typu człowiek-maszyna zaimplementowano na odrębnym komputerze PC oraz na smartfonie. Dzięki takiemu rozwiązaniu można było łatwiej wprowadzać zmiany w oprogramowaniu. Połączenie ze sterownikiem manipulatora w przypadku komputera odbywa się z wykorzystaniem portu COM w standardzie RS-232, a w przypadku smartfona z wykorzystaniem połączenia bezprzewodowego i adaptera Bluetooth, który w systemie jest rejestrowany jako port szeregowy. W obu połączeniach zdefiniowano taką samą ramkę danych, w której przesyłane są informacje.

### 3. STEROWANIE GŁOSOWE

Oprogramowanie do sterowania głosowego pozwala sterować pozycją manipulatora, poprzez podanie współrzędnych końca ramienia w jego przestrzeni roboczej. Operator wypowiada komendy głosowe, które są interpretowane w programie. Ze względu na to, że zdecydowano się opracować oprogramowanie na komputer PC oraz na smartfon z systemem Android, jako system rozpoznawania mowy wykorzystano Google Speech Recognition System [10]. Po przesłaniu próbki dźwięku do systemu, zwraca on odpowiedź w postaci rozpoznanego tekstu, która jest następnie interpretowana jako odpowiednia komenda.

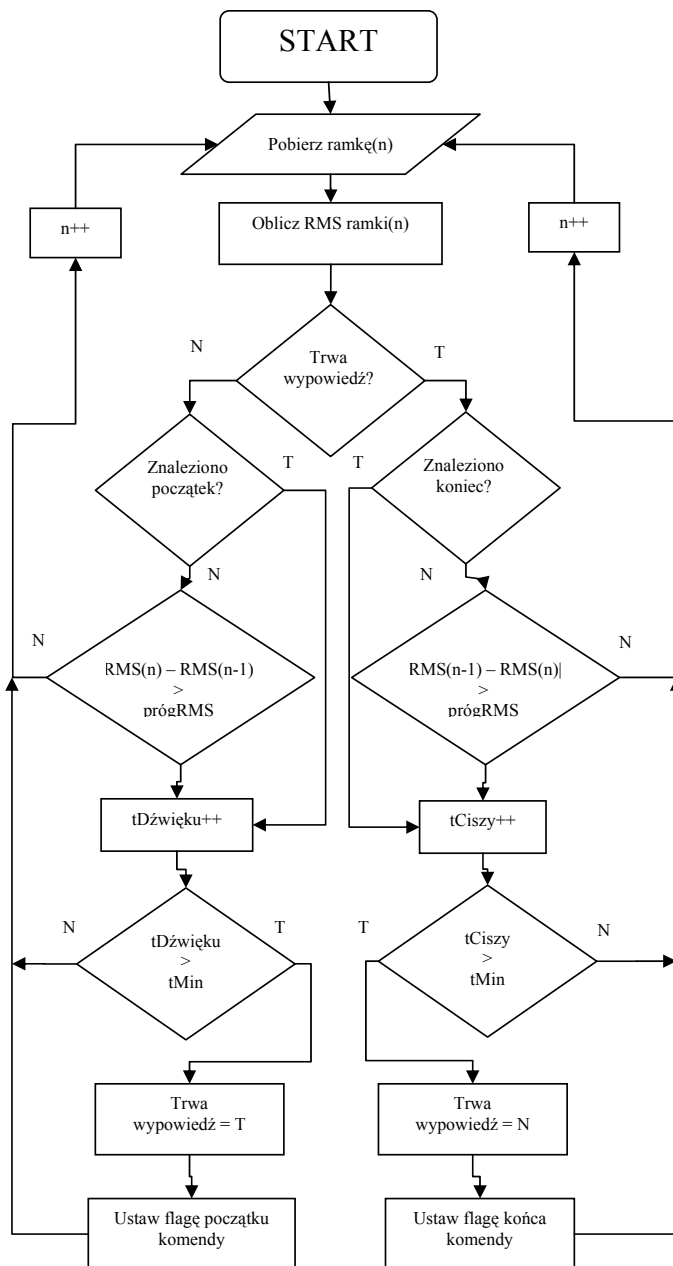


Rys. 2. Manipulator elektrohydrauliczny i jego pole robocze

Program na komputer klasy PC został napisany w środowisku Microsoft Visual Studio 2010 w języku programowania obiektowego C#. Dźwięk rejestrowany jest w czasie rzeczywistym z użyciem bibliotek NAudio, co pozwoliło w łatwy sposób zaimplementować algorytm detekcji początku i końca słowa.

Sposób wykrywania wypowiedzi polega na ciągłej analizie wartości RMS sygnału w czasie. Energia RMS wyliczana jest w ramach o długości 20 ms. Następnie badana jest pochodna jako przyrost lub spadek energii pomiędzy ramkami. Przekroczenie odpowiedniego progu determinuje początek wypowiedzi w przypadku przyrostu

energii. Podobnie zostaje wykryty jej koniec podczas spadku energii sygnału. Wartość progu wykrycia słowa i wzmacnienie sygnału mikrofonu dostosowywane jest za pomocą suwaków w celu uzyskania jak najlepszych rezultatów sterowania.



Rys. 3. Algorytm znajdowania początku i końca wypowiedzi

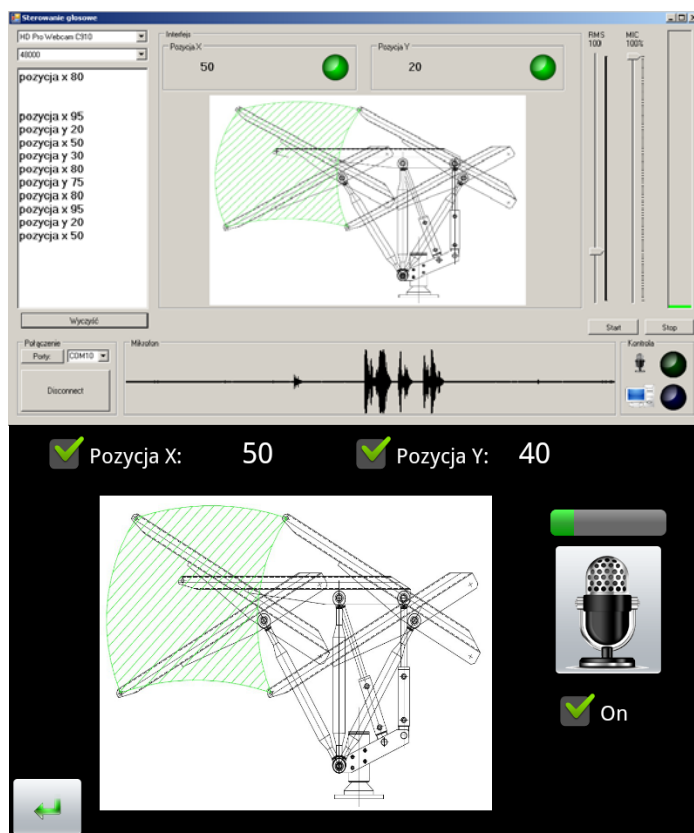
W algorytmie wprowadzono dodatkowe zabezpieczenia w postaci minimalnego czasu wypowiedzi oraz minimalnego czasu ciszy po zakończeniu wypowiedzi. Okres wypowiedzi i ciszy określany jest na podstawie średniej wartości energii RMS z 200 ms sygnału. Te zabiegi mają na celu wyeliminować błędną reakcję algorytmu na krótkie przerwy pomiędzy słowami komendy, które mogą zostać wykryte jako koniec wypowiedzi. Schemat blokowy algorytmu przedstawiono na rysunku 3. Przebieg rejestrowanego sygnału można obserwować na oscylogramie, podobnie jak poprawnie rozpoznane komendy, których historia widoczna jest na liście tekstowej. Aktualna pozycja manipulatora podana jest w głównym interfejsie. Na ekranie sygnalizowany jest również status wykrycia słowa i połączenia za pomocą kontroltek.

Aplikacja dedykowana na smartfon z systemem Android została napisana w języku Java w środowisku Eclipse. Dzięki temu, że język Java i C# są do siebie zbliżone, dużą część kodu z oprogramowania na komputer, można było zaimplementować w rozwiązaniu na urządzenie przenośne. Sposób działania aplikacji jest niemal identyczny, gdyż różnią się one głównie interfejsem graficznym oraz sposobem przesyłania danych. Ekran smartfona jest mniejszy, więc wyświetlane są na nim tylko informacje o aktualnym położeniu końcówki manipulatora. W obydwóch aplikacjach można w dowolnym momencie zatrzymać bądź uruchomić mikrofon i rejestrowanie sygnału za pomocą przycisków. Na rysunku 4 przedstawiono zrzuty ekranu programów podczas przeprowadzania testów. W obydwóch rozwiązaniach zaproponowano rozpoznawanie prostych komend, które są trój-członowe. Pierwsze słowo aktywujące to „pozycja”. Następnie użytkownik powinien podać oś manipulatora: X lub Y. Jako ostatni człon komendy należy podać wartość od 0 do 100, która odpowiada procentowej pozycji dla danej osi w stosunku do całego obszaru roboczego. Takie znormalizowanie przestrzeni roboczej ułatwia operatorowi sterowanie. Po poprawnym rozpoznaniu komendy zostaje ona przesłana do sterownika obsługującego pozycjonowanie manipulatora, który ustawia manipulator w żądanej pozycji. Pomimo konieczności wysłania próbek dźwięku na zewnętrzny serwer, aplikacja działa płynnie. Dzięki temu istniała możliwość bieżącego korygowania pozycji i zadawania następnych współrzędnych, nawet, gdy poprzednio zadana pozycja nie została jeszcze osiągnięta. Podczas testów, ze względu na duży hałas pompy hydraulicznej, zdecydowano się umiejscowić komputer i mikrofon za szybą w drugim pomieszczeniu, co znacząco wyeliminowało zakłócenia.

Przeprowadzono również badania skuteczności rozpoznawania komend przez system. W badaniu brało udział 5 mężczyzn w wieku od 24 do 28 lat. W celu ujednoczenia testów zdecydowano, że wszyscy operatorzy będą wypowiadali ten sam ciąg komend sterujących. Ustalono 6 różnych komend głosowych, które każdy badany mówca powtarzał 20 razy.

Ze względu na to, że opracowany interfejs wykorzystuje system rozpoznawania mowy Google Speech Recognition System, w ramach badań zdecydowano wydobycь trzy parametry:

- skuteczność rozpoznanych komend (stosunek liczby poprawnie rozpoznanych do wypowiedzianych komend, przy czym rozpoznana zdefiniowana jest jako zgodna tekstowo z wypowiedzianą),
- czas odpowiedzi serwera (czas od momentu zakończenia wypowiedzi do momentu rozpoznania komendy),
- ufność (ang. *confidence* – wartość zwracana przez system Google wraz odpowiedzią serwera, odpowiadająca ufności systemu dla rozpoznania danej wypowiedzi, przyjmuje wartość od 0 do 1).



Rys. 4. Aplikacja do sterowania głosowego przeznaczona na komputer i smartfon

W poniższych tabelach przedstawiono wyniki badań.

**TABELA 1**

Czas odpowiedzi serwera, ufnosć oraz skuteczność rozpoznawania komend

Mówca	Czas odpowiedzi [ms]	Ufnosć	Skuteczność [%]
# 1	1206,62	0,56	94,16
# 2	1256,77	0,65	84,17
# 3	1191,75	0,60	90,83
# 4	1300,84	0,54	80,00
# 5	1251,49	0,58	98,33
Max	1300,84	0,65	98,33
Min	1191,75	0,54	80,00
Średnia	1241,50	0,59	89,50

**TABELA 2**

Średnia ufnosć zwracana przez system dla poszczególnych komend

Mówca	pozycja x 80	pozycja x 95	pozycja y 20	pozycja x 50	pozycja y 30	pozycja y 75
# 1	0,61	0,61	0,41	0,44	0,68	0,72
# 2	0,66	0,68	0,71	0,61	0,67	0,71
# 3	0,66	0,68	0,41	0,49	0,71	0,72
# 4	0,59	0,61	0,37	0,45	0,66	0,64
# 5	0,68	0,62	0,40	0,46	0,68	0,73
Max	0,68	0,68	0,71	0,61	0,71	0,73
Min	0,59	0,61	0,37	0,44	0,66	0,64
Średnia	0,64	0,64	0,46	0,49	0,68	0,70

Osiągnięto średnią skuteczność rozpoznawania, sięgającą prawie 90%, przy czym średni czas odpowiedzi serwera nie przekraczał 1300 ms. Warto zauważyć, że skuteczność rozpoznawania komend w opracowanym systemie zależy od odpowiedzi serwera Google Speech Recognition System. Odpowiedź serwera w postaci tekstowej jest weryfikowana pod kątem zgodności z tekstem wypowiedzianej komendy. Jeśli odpowiedź tekstowa była różna od tekstu komendy, to klasyfikowano ją jako błędną. Nie zastosowano dodatkowych zabezpieczeń przed realizacją błędnych czynności przez urządzenie, gdyż uznano, że zastosowana metoda weryfikacji jest wystarczająca w tym przypadku.

#### 4. PODSUMOWANIE

Autorzy przedstawili w artykule zastosowanie technologii rozpoznawania mowy do sterowania manipulatorem elektrohydraulicznym. Problem optymalnej, naturalnej komunikacji człowieka z urządzeniami jest dziś szeroko poruszany. Dzięki rozwojowi technologii w zakresie rozpoznawania i syntezy mowy możliwe jest opracowanie interfejsów głosowych. W zaproponowanym rozwiązaniu przedstawiono oprogramowanie, które w prosty sposób realizuje koncepcje sterowania głosowego urządzeniem przemysłowym, jakim jest manipulator. Tego typu sterowanie może być dostosowane do różnego typu urządzeń mechatronicznych.

W ramach dalszych badań planuje się udoskonalenie systemu o możliwość sterowania większą ilością komend oraz wprowadzenie dodatkowych środków bezpieczeństwa w tym głosowych potwierdzeń zwrotnych. Stworzony system posłuży także jako szkielet do testowania różnych algorytmów wykrywania początku i końca wypowiedzi.

#### LITERATURA

1. Tadeusiewicz R.: Sygnał mowy. WKiŁ, Warszawa, 1988.
2. Basztura Cz.: Rozmawiać z komputerem. Wydawnictwo Prac Naukowych „FORMAT”, Wrocław, 1992.

3. Ziółko B., Ziółko M.: Przetwarzanie mowy. Wydawnictwa AGH, Kraków 2011.
4. Makowski R.: Automatyczne rozpoznawanie mowy – wybrane zagadnienia. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 2011.
5. Kasprzak W.: Przetwarzanie obrazów i sygnałów mowy. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2009.
6. Rogowski A.: Industrially oriented voice control system. Robotics and Computer-Integrated Manufacturing, nr 28, s. 303-315, 2012.
7. Pires J.N., Robot-by-voice: Experiments on commanding an industrial robot using the human voice, Industrial Robot, Vol. 32, 2005, No.6, Emerald Publishing.
8. Kacalak W., Majewski M.: Intelligent Layer of Two-Way Voice Communication of the Technological Device with the Operator. ICAISC s. 610-615, 2004.
9. Kacalak W., Majewski M.: I Intelligent human-machine speech communication system. International Journal of Information Technology, Vol. 12 No.4, 2006.
10. Schalkwyk J., Beeferman D., Beaufays F., Byrne B., Chelba C., Cohen M., Garret M., Strope B.: Google Search by Voice: A case study. Google Inc, <http://research.google.com/>

*Rękopis dostarczono dnia 21.08.2013 r.*

## VOICE CONTROL OF ELECTROHYDRAULIC MANIPULATOR

Roman REGULSKI, Piotr OWCZAREK,  
Dominik RYBARCZYK, Paweł BACHMAN,  
Jarosław GOŚLIŃSKI

**ABSTRACT** *Automatic speech recognition systems have more and more applications in field of mechatronics and robotics. They have become an essential element of intelligent human-machine interfaces. They also allow users to communicate with devices in the most natural way – using speech.*

*This paper presents an innovative solution of electrohydraulic manipulator control using voice commands. During the work, interfaces for PCs and mobile platform Android have been created. Both solutions allow to easily control the position of the manipulator using speech recognition technology. The main objective of this paper is to present the abilities and usefulness of speech recognition systems for device control.*

**Keywords:** *voice control, speech recognition, electrohydraulic manipulator*



**Mgr inż. Roman REGULSKI** – doktorant w Zakładzie Urządzeń Mechatronicznych, Wydziału Budowy Maszyn i Zarządzania Politechniki Poznańskiej. Absolwent kierunku Mechatronika w tej samej jednostce. Zainteresowania naukowe: mechatronika, sterowniki mikroprocesorowe, automatyczne systemy rozpoznawania mowy, sterowanie głosowe, elektrohydraulika, pneumatyka, programowanie, sterowniki PLC.



**Mgr inż. Piotr OWCZAREK** – asystent i doktorant w Zakładzie Urządzeń Mechatronicznych Politechniki Poznańskiej. Ukończył studia w 2011 r. na Wydziale Elektrycznym o kierunku Automatyka i Robotyka. Jest na studiach doktoranckich od 2011 r. na Wydziale Budowy Maszyn i Zarządzania. Jego zainteresowania obejmują nowoczesne metody cyfrowego przetwarzania obrazów, projektowanie urządzeń elektronicznych i mechatronicznych.

**Mgr inż. Dominik RYBARCZYK** – ukończył studia w 2010 r. na Wydziale Budowy Maszyn i Zarządzania Politechniki Poznańskiej. Od 2011 r. zatrudniony na Wydziale Budowy Maszyn i Zarządzania w Zakładzie Urządzeń Mechatronicznych Politechniki Poznańskiej (aktualnie na stanowisku asystenta). Jego zainteresowania obejmują urządzenia mechatroniczne, sterowniki mikroprocesorowe oraz metody sztucznej inteligencji.



**Dr inż. Paweł BACHMAN** – asystent w Instytucie Edukacji Techniczno Informatycznej, Wydział Mechaniczny, Uniwersytet Zielonogórski. Zainteresowania: mechatronika, sterowanie. Jest stypendystą w ramach Poddziałania 8.2.2 „Regionalne Strategie Innowacji”, Działania 8.2 „Transfer wiedzy”, Priorytetu VIII „Regionalne Kadry Gospodarki” Programu Operacyjnego Kapitał Ludzki współfinansowanego ze środków Europejskiego Funduszu Społecznego Unii Europejskiej i z budżetu państwa.

**Mgr inż. Jarosław GOŚLIŃSKI** – w 2011 roku ukończył studia na kierunku Automatyka i Robotyka. Podjął pracę w korporacji gdzie zajmował się wdrażaniem systemów pomiarowych, uzyskał certyfikat CLAD (Certified LabVIEW Associate Developer). Powrócił na Politechnikę Poznańską, gdzie zaczął studia doktoranckie. Zajmuje się estymacją stanu przy wykorzystaniu bezśladowych filtrów Kalmana. Pracuje nad robotami latającymi, modelowaniem dynamiki oraz identyfikacją parametrów modeli matematycznych.



