

Zastosowania systemów rozpoznawania mowy do sterowania i komunikacji głosowej z urządzeniami mechatronicznymi

Roman Regulski, Amadeusz Nowak

Zakład Urządzeń Mechatronicznych, Politechnika Poznańska

Streszczenie: Artykuł przedstawia przykłady wykorzystania systemów automatycznego rozpoznawania mowy do budowy głosowych interfejsów typu człowiek-maszyna. W artykule opisano sposób działania takich aplikacji pod kątem sterowania i komunikacji głosowej. W następnej części przedstawiono koncepcje i budowę systemu rozpoznawania mowy do komunikacji z 32-bitowym modułowym sterownikiem pralki.

Słowa kluczowe: automatyczne rozpoznawanie mowy, sterowanie głosowe, interfejs człowiek-maszyna, sterownik pralki

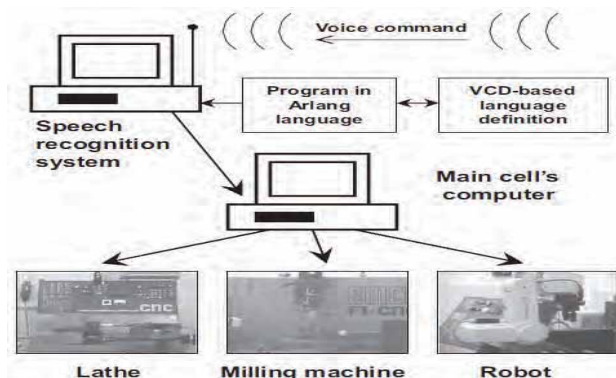
1. Wprowadzenie

Obecnie systemy rozpoznawania mowy najbardziej rozpowszechnione są jako programy komputerowe. Rozwiązania tego typu mogą np. uwolnić użytkownika komputera od posługiwania się myszką, a w wielu przypadkach nawet klawiaturą. Systemy mówiące można także wprowadzać do urządzeń sprzętu gospodarstwa domowego jak również w zastosowaniach przemysłowych. Urządzenia mechatroniczne, takie jak np.: robot, obrabiarka, transporter, mogłyby być w łatwy sposób sterowane głosem zamiast sterowania z użyciem skomplikowanych języków programowania. Istniejące rozwiązania w zakresie systemów rozpoznawania mowy oparte na zaawansowanych algorytmach zapewniają poprawność na poziomie ponad 90 %. Ich przykłady można znaleźć także w Polsce, gdzie system o nazwie „Primespeech” został zastosowany jako narzędzie w internetowym portalu głosowym, w którym stosuje się głosowe wydawanie poleceń, a generowane informacje są przekazywane użytkownikowi za pomocą syntezatora mowy. Wytrenowanie sieci neuronowych bardzo dużą liczbą próbek sygnałów dźwiękowych pozwoliło na osiągnięcie wysokiej skuteczności rozpoznawania mowy. System głosowy wdrożono również w Zarządzie Transportu Miejskiego w Warszawie oraz na Giełdzie Papierów Wartościowych [1].

Korzyści z zastosowania systemów ASR dostrzeżono także w branży motoryzacyjnej. Przykładem zastosowania tej technologii jest system głosowy *Blue&Me* koncernu Fiat. Kierowca auta może sterować radiem lub podłączonym telefonem komórkowym wypowiadając określone komendy. Może on wybrać numer telefonu, odebrać połączenie, czy też zmienić odtwarzany utwór muzyczny. Podobne rozwiązania wprowadzono w samochodach firmy Mercedes-Benz, w których możliwe jest dodatkowo sterowanie nawigacją satelitarną. Wystarczy głosowo podać nazwę miejscowości, a system sam wyznaczy trasę dojazdu. Dzięki temu wzrasta bezpieczeństwo oraz komfort użytkownika pojazdu. Wydawanie komend głosowych w telefonach komórkowych jest już standardem. Można głosowo wybrać kontakt z książki telefonicznej, aktywować połączenie albo podać słowo do wyszukiwarki internetowej. Wzrost skuteczności i szybkości działania systemów generacji oraz rozpoznawania mowy pozwolił na ich zastosowanie w takich aplikacjach i dziedzinach jak np.:

- militaria (polecenia głosowe w samolotach, helikopterach, np. F16 VISTA),
- automatyczna transkrypcja tekstu (Speech to text),
- interaktywne systemy informacyjne (PrimeSpeech, portale głosowe),
- urządzenia mobilne (telefony, smartfony, tablety),
- robotyka (Cleverbot – systemy ASR w połączeniu ze sztuczną inteligencją),
- motoryzacja (sterowanie podzespołami auta bez użycia rąk, np. Blue&Me),
- aplikacje domowe (systemy typu inteligentny dom).

Z punktu widzenia mechatroniki, robotyki, automatyki systemy głosowe nie są jeszcze popularne, lecz producenci urządzeń w tych branżach zaczynają dostrzegać zalety takich rozwiązań [2]. Przykładem może być technologia *Mazak Voice Adviser* firmy Mazak [3].



Rys. 1. Głosowe sterowanie gniazdem obróbkowym [4]
Fig. 1. Manufacturing cells voice control [4]

Urządzenie wyposażone w ten system informuje operatora komendami głosowymi o dokonanym wyborze oraz doradza podczas manualnej pracy maszyny. Funkcja wydawania poleceń głosowych przyczynia się do znacznej redukcji problemów związanych z błędami operatora.

W laboratoriach naukowych prowadzone są badania implementacji systemów rozpoznawania mowy do sterowania i komunikacji głosowej. Jednym z przykładów zastosowania komunikacji człowiek-maszyna jest gniazdo obróbkowe. Użytkownik może wydawać polecenia głosowe poszczególnym urządzeniom, podczas gdy jego ręce mogą być zajęte wykonywaniem innych czynności [4, 5]. Budowę systemu przedstawiono na rysunku 1.

W pracach [6, 7, 8] autor przedstawił koncepcję komunikacji głosowej robota przemysłowego, wykorzystanego do montażu elementów oraz spawania. Poruszył także zagadnienia programowania robotów z wykorzystaniem komend głosowych. W ramach projektu stworzono oprogramowanie komputerowe z wykorzystaniem języka C# do rozpoznawania i generacji mowy, używając bibliotek Microsoft Speech SDK 5.1.



Rys. 2. Robot przemysłowy sterowany komendami głosowymi [6]
Fig. 2. Industrial robot controlled by voice commands [6]

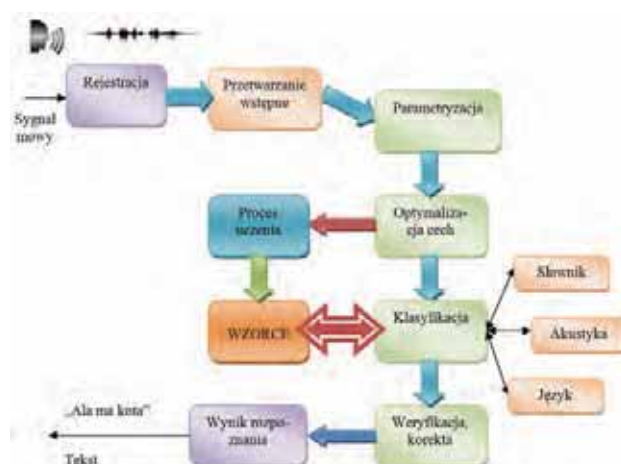
Na rysunku 2 przedstawiono robota przemysłowego sterowanego przez operatora za pomocą komend głosowych z wykorzystaniem mikrofonu.

Interesującymi przykładami są też próby sterowania robotami mobilnymi z wykorzystaniem ASR. Robot Godot [9, 10] wyposażony w system obustronnej komunikacji głosowej, który zapewnia porozumiewanie się robotem z wykorzystaniem naturalnego języka. Przykładowo robot podaje swoją pozycję na komendę „Where are you?”. W pracy [11] wykorzystano smartfon z systemem Android do sterowania robotem mobilnym. W urządzeniu zaimplementowano rozpoznawanie głosowe komend z wykorzystaniem systemu ASR Google.

2. Systemy ASR

2.1. Ogólna budowa systemów ASR

System automatycznego rozpoznawania mowy ASR (ang. *Automatic Speech Recognition*) jest rozbudowanym obiektem składającym się z kilku bloków (rys. 3), którego głównym celem jest przetworzenie sygnału mowy na tekst bądź inną użyteczną informację. Problemem, na jaki napotyka systemy ASR jest wydobycie z sygnału mowy użytecznych danych i skojarzenie ich z konkretnym słowem, które dla człowieka ma określone znaczenie. Wypowiedane dźwięki za każdym razem są inne. Zmienność sygnału mowy jest naturalną jego cechą, dzięki której na przykład rozpoznajemy akcent czy intonację. Zbudowanie systemu rozpoznającego mowę nie jest łatwe, lecz możliwe.



Rys. 3. Ogólny schemat systemu rozpoznawania mowy
Fig. 3. General scheme of speech recognition system

Każdy z elementów systemu rozpoznawania mowy ma określone zadanie. Należy mieć na uwadze, że poszczególne bloki są od siebie silnie zależne. Źle zarejestrowany lub błędnie przetworzony sygnał może skutkować dostarczeniem nieużytecznych parametrów. Ma to szczególne znaczenie w procesie uczenia systemu, kiedy to system ASR tworzy wzorce i modele, które wykorzystuje w następnym

etapie rozpoznawania. Na podstawie przetworzonego sygnału i jego parametrów powstaje model akustyczny, który będzie skojarzony z modelem języka. W ten sposób system uzyskuje informację o tym, jaką informację lingwistyczną niesie ze sobą sygnał mowy. Należy mieć na uwadze, że system rozpoznawania mowy może działać w oparciu o małą bądź dużą liczbę wzorców, co determinuje wielkość słownika, z jakim system współpracuje. Mogą to być małe słowniki w przypadku systemów rozpoznających izolowane słowa bądź komendy, jak również duże bazy danych zawierające odpowiednik zbioru językowego oraz uwzględniające model językowy (gramatykę). Na złożoność i wielkość systemu ASR wpływa wiele czynników, takich jak:

- liczba mówców (jeden, wielu),
- liczba rozpoznawanych słów,
- rodzaj mowy (izolowane komendy, mowa ciągła),
- warunki pracy systemu (zakłócenia, brak zakłóceń),
- wykorzystany sprzęt (komputer PC, mikrokontroler, mikrofon).

Wyżej wymieniono najważniejsze aspekty brane pod uwagę przy budowie systemu ASR. Oczywiście uwarunkowań jest więcej, istotnym czynnikiem może być również czas reakcji systemu.

2.2. Różnorodność systemów ASR

Różnorodność systemów ASR wynika z celów, jakie dany system ma spełnić, stąd coraz częściej systemy ASR optymalizuje się pod dane zastosowanie. Głównym zadaniem ASR jest zamiana mowy na tekst, którego znaczenie można łatwo zweryfikować. W systemach sterowania, w urządzeniach mechatronicznych i w systemach automatyki można zastosować system rozpoznawania, którego celem nadrzędnym będzie rozpoznanie komendy, a następnie wykonanie operacji z nią związanej. W tym przypadku wyjściowy tekst systemu ASR posłuży do klasyfikacji komendy i podjęcia decyzji, co ma wykonać.

Zastosowania mikrokontrolerów w takich systemach nie budzą zdziwienia. Powstaje wiele nowych rozwiązań systemów ASR dedykowanych do zastosowania w urządzeniach wyposażonych w mikroprocesor. Zbudowanie systemu rozpoznającego mowę z wykorzystaniem takiej jednostki jest trudniejsze niż np. napisanie programu na komputerze PC. Powodem tego jest konieczność uwzględnienia możliwości sprzętowych danego układu, oraz optymalizacji oprogramowania pod tym kątem. W przypadku programów pracujących np. na komputerach wyposażonych w system operacyjny, napisany i skompilowany program będzie mógł działać na różnych jednostkach sprzętowych.

Dzięki licznym badaniom prowadzonym na przestrzeni lat dokonano wielu optymalizacji algorytmów wykorzystywanych w systemach ASR. Uwzględniając przy tym rozwój w dziedzinach technik komputerowych oraz mikroproceso-

row, można zauważyć wzrost zastosowań systemów automatycznego rozpoznawania mowy typu wbudowanego.

Warto podkreślić, że takie aplikacje przystosowane są do konkretnego zadania potrafią osiągnąć skuteczność na poziomie ponad 95 %, przy czym rzeczywisty czas rozpoznania nie przekracza sekundy, co można uznać za rozpoznawanie mowy w czasie rzeczywistym.

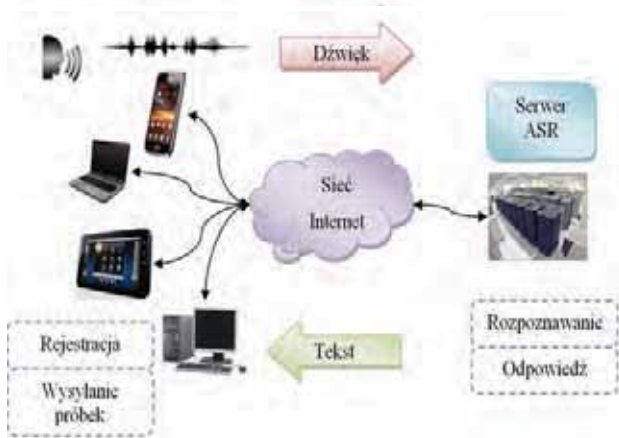
W systemach mikroprocesorowych wykorzystuje się te same algorytmy, które są implementowane w dużych programach komputerowych. Mikrokontrolery wyposażone są w wszelkie narzędzia sprzętowe, których wymagają systemy ASR. Standardem wyposażenia mikrokontrolerów są przetworniki ADC, które wykorzystuje się do rejestracji sygnału z mikrofonu oraz kontrolery PWM, które wykorzystuje się do generacji dźwięku. Istnieją również procesory sygnałowe DSP, które są zoptymalizowane pod kątem realizacji algorytmów przetwarzania sygnałów. Zastosowania systemów ASR typu wbudowanego można znaleźć w wielu dziedzinach. Takie rozwiązania są standardem np. w nawigacjach samochodowych oraz w sterowaniu podzespołami samochodu.

Przetwarzanie języka naturalnego (ang. *natural language processing*) to jedna z dziedzin mocno powiązanych z systemami rozpoznawania mowy. W dążeniu do porozumiewania się z urządzeniami wykorzystuje się naturalny język dla mowy potocznej człowieka, aby umożliwić jego rozmowę np. z komputerem. Takie podejście może być wykorzystane w nowych innowacyjnych systemach, w urządzeniach mechatronicznych, urządzeniach powszechnego użytku, czy też w szeroko rozumianym sterowaniu. Jednymi z ciekawszych, stworzonych do tego celu narzędzi są rozwiązania firmy Sensory Inc. Oferuje ona rodzinę naturalnych procesorów mowy (ang. *Natural Language Processor – NLP*). Są to procesory dedykowane do wbudowywania w różnorodnych urządzeniach, w których chcemy zastosować mowę jako sygnał sterujący. Procesory NLP są niewątpliwie warte zainteresowania, ze względu na możliwości i elastyczność ich struktury, co pozwala dostosować je do żądanej aplikacji. Po procesie weryfikacji oprogramowania otrzymujemy komponent elektroniczny z gotowym systemem ASR, który można wykorzystać w dowolnym urządzeniu. Dzięki zastosowaniu tego typu procesorów można stosunkowo szybko zbudować i zastosować zaawansowany system mówiący w dowolnym urządzeniu.

Google stworzyło system automatycznego rozpoznawania mowy na potrzeby wprowadzania tekstu w swoich wyszukiwarkach. Początkowo rozwiązanie było przeznaczone dla urządzeń mobilnych typu smartfon czy tablet, lecz szybko znalazło zastosowanie w komputerach klasy PC. System ma za zadanie ułatwić wprowadzanie fraz wyszukiwania z wykorzystaniem poleceń głosowych. Użytkownik telefonu może podać hasło wyszukiwania za pomocą mowy zamiast wprowadzać je z klawiatury, co jest uciążliwe w przypadku długich fraz.

Innowacyjne podejście do rozpoznawania mowy ze strony Google polega na tym, że urządzenie końcowe systemu (np. smartfon) posiada aplikację kliencką, której zadaniem polega jedynie na zebraniu próbek dźwięku i interpretacji wyniku rozpoznania. System automatyczne-

go rozpoznawania mowy znajduje się na zewnętrznych serwerach o dużej mocy obliczeniowej. Próbkę dźwięku zostają przesłane do serwera w celu rozpoznania, a serwer odsyła wyniki w postaci tekstu do urządzenia. Dzięki takiemu podejściu, system rozpoznawania mowy dostępny jest dla wielu urządzeń za pośrednictwem połączenia Internetowego. Urządzenia wykorzystujące ten system nie są obciążone obliczeniami związanymi z przetwarzaniem i rozpoznaniem mowy. Takie rozwiązanie umożliwia zbudowanie dużej bazy słownikowej dla systemu, która znajduje się tylko po stronie serwera. Wadą tego typu architektury jest konieczność zapewnienia stałego połączenia internetowego pomiędzy klientem a serwerem. Lecz w dzisiejszych czasach dostęp do Internetu nie jest problemem. Ideę rozpoznawania mowy z systemem ASR na zewnętrznym serwerze przedstawiono na rysunku 4.



Rys. 4. Idea zastosowania systemu ASR na zewnętrznym serwerze

Fig. 4. The idea of using ASR system on an external server

3. Projekt implementacji systemu ASR

3.1. Opis projektu

W ramach niniejszej pracy i realizacji projektu *Inteligentne sterowniki pralek* [12], zbudowano sterownik będący interfejsem dla urządzenia mechatronicznego, jakim jest pralka. Głównym zadaniem podczas budowy urządzenia było zaimplementowanie obsługi rozpoznawania komend głosowych do sterowania mową. Realizując ten cel, wykorzystano system ASR firmy Google, który zaimplementowano w oprogramowaniu przeznaczonym dla telefonu komórkowego typu smartfon oraz komputera PC. W ramach projektu wykonano również moduł w postaci płytki elektronicznej PCB.

Na rysunku 5 przedstawiono schemat blokowy interfejsu na tle innych modułów wchodzących w skład głównego sterownika pralki. Głównym elementem urządzenia jest sterownik zapewniający komunikację z płytą główną pralki po magistrali RS-485 oraz komunikację bezprzewodową z telefonem i komputerem PC w technologii Bluetooth. Takie rozwiązanie pozwala sterować pralką za pomocą wybranego urządzenia. Każde z urządzeń posiada dedykowaną aplikację, która umożliwia interpretację poleceń głosowych. Tego typu interfejsy są interesującą alternaty-

wą dla klasycznych rozwiązań z wykorzystaniem przycisków lub pokręteł w urządzeniach.



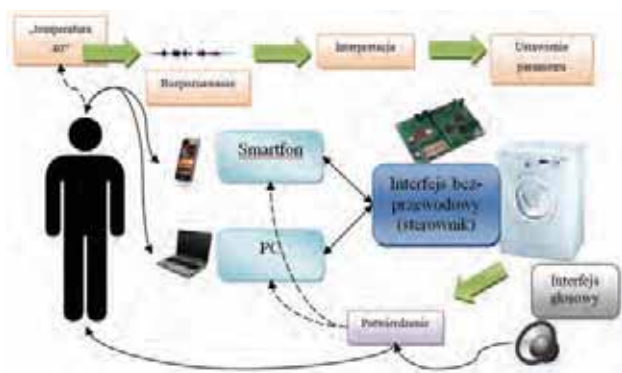
Rys. 5. Budowa modułowego sterownika pralki

Fig. 5. Structure of washing machine controller

Wykonany sterownik może równolegle współpracować z interfejsem LCD lub całkowicie go zastąpić. Innowacyjnym podejściem jest także zaimplementowana komunikacja bezprzewodowa umożliwiająca połączenie pralki z telefonem lub komputerem PC. Dzięki temu użytkownik może uruchomić i obserwować proces prania na odległość.

Za pomocą komend głosowych można wybrać program, ustawić parametry, uruchomić pranie oraz je zatrzymać. Wykonanie każdego polecenia można obserwować na wyświetlaczu lub monitorze. Dodatkowo użytkownik słyszy potwierdzenia głosowe, gdyż pralka wyposażona jest w interfejs głosowy, który odtwarza komendy za pomocą głośnika. Przykładowo, w odniesieniu do klasycznych sterowników pralek, aby uruchomić pranie należy wybrać parametry za pomocą oznaczonych przycisków, używając komunikacji głosowej wystarczy zaś powiedzieć pralce komendy, np.: „temperatura 40, obroty 600, program bawełna”. Sterownik rozpozna komendy i poinformuje użytkownika o wykonaniu poleceń za pomocą zwrotnych potwierdzeń głosowych. Taka obsługa jest bardzo intuicyjna. Użytkownik nie jest zmuszony do przeglądania instrukcji sterowania w celu zapoznania się ze znaczeniem np. ikon graficznych odpowiadającym danym programom prania. Ogromną część pracy w projekcie poświęcono oprogramowaniu. Moduł interfejsu zawiera implementację komunikacji RS-485 z protokołem MODBUS ASCII wykorzystywanej do przesyłu informacji między sterownikiem a płytą główną pralki. Płyta główna pralki jest urządzeniem typu Master, a interfejs typu Slave. Takie podejście pozwala na zarządzanie arbitrażem przesyłu danych. Komunikacja bezprzewodowa w technologii Bluetooth pozwala na połączenie pomiędzy dowolnymi urządzeniami posiadającymi taki adapter. Zdecydowano się na to rozwiązanie, ponieważ komputery i smartfony są standardowo wyposażone w moduły Bluetooth. Oprogramowanie na komputer PC napisano w języku C#, w którym zaimplementowano obsługę rejestracji głosu (z mikrofonu), komunikację internetową oraz komunikację szeregową z wykorzystaniem modułu Bluetooth.

Podobne rozwiązania zastosowano w aplikacji dla smartfona na platformę Android, która została napisana w języku Java. Moduł elektroniczny sterownika zawiera 32-bitowy mikrokontroler STM32, zaprogramowany w języku C. W ramach wyżej wymienionego oprogramowania zwrócono dużą uwagę na ich zgodność i niezawodność. Ramki komunikacyjne zawierają sumy kontrolne wykorzystywane do sprawdzenia poprawności przesyłu danych. Celem tego zabiegu jest wyeliminowanie błędnej pracy interfejsu.



Rys. 6. Komunikacja głosowa z wykorzystaniem sterownika
Fig. 6. Voice communication using driver

Na rysunku 6 przedstawiono zrealizowaną koncepcję wykorzystania komunikacji głosowej do sterowania i porozumiewania się z pralką za pomocą mowy. Interfejsem bezpośrednim jest smartfon albo komputer. W zależności od wyboru, dane urządzenie rozpoznaje komendę i przesyła polecenie do sterownika, który pośredniczy w wymianie danych między wybranym urządzeniem a pralką. Na podstawie rozpoznania dokonywana jest interpretacja komendy, a następnie podejmowana jest decyzja o wysłaniu instrukcji do sterownika głównego pralki, który ustawia dany parametr. Po tej operacji następuje potwierdzenie jej wykonania. Sprzężenie zwrotne polega na podaniu komendy głosowej z wykorzystaniem interfejsu głosowego, a w urządzeniu zostaje wyświetlona informacja o poprawności ustalenia danej wartości. Jest to najbardziej naturalna z punktu widzenia człowieka realizacja interfejsu człowiek–maszyna.

Podsumowując ogólny opis projektu, można w nim wyróżnić trzy główne elementy, które wchodzi w skład realizacji niniejszej pracy. Są to:

- sterownik w postaci modułu elektronicznego,
- aplikacja dla smartfona z systemem Android,
- aplikacja komputerowa dla środowiska MS Windows.

3.2. Sterownik

Pierwszym etapem było wykonanie projektu elektroniki modułu sterownika. Ze względu na ilość zadań, jakie ma spełniać, zdecydowano się wykorzystać mikrokontroler z 32-bitowy. Schemat oraz płytkę PCB zaprojektowano z wykorzystaniem programu EAGLE. Większość elementów przeznaczonych jest do montażu powierzchniowego, dlatego ze względu na dużą ilość połączeń i wysoki stopień skomplikowania rozmieszczenia elementów zdecydowano się na wykonanie płytki dwustronnej.

Na rysunku 7 przedstawiono wykonany sterownik. Podstawowymi elementami wchodzącymi w jego skład są:

- mikrokontroler STM32F103VBT6,
- moduł Bluetooth BTM-222,
- moduł komunikacji RS-485,
- układ zasilania ze stabilizatorem.

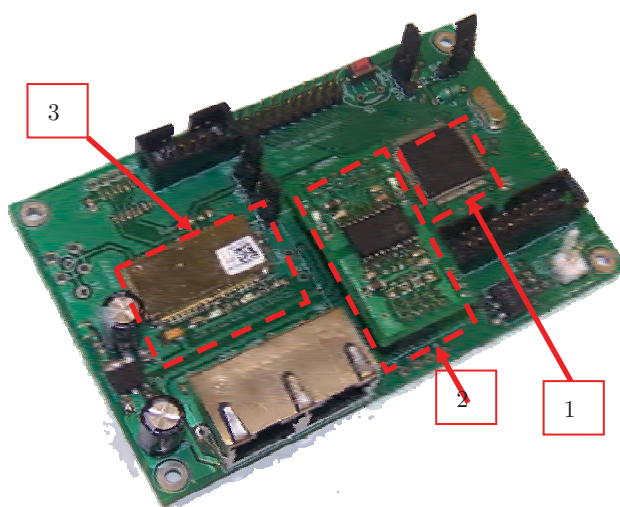
Układ STM32F103VBT6 jest 32-bitowym mikrokontrolerem z rdzeniem ARM Cortex™-M3. Główne parametry i peryferia, w które go wyposażono to:

- taktowanie do 72MHz,
- pamięć Flash 128 kB,
- interfejsy komunikacyjne (USART x3, I²C x2, CAN, USB)
- 7 timerów,
- przetworniki ADC, PWM, DMA,
- 80 wejść/wyjść generalnego przeznaczenia.

Układ BTM-222 jest modulem Bluetooth, który może pracować w trybie Master lub Slave. Konfigurację układu realizuje się z wykorzystaniem komend AT. Interfejsem komunikacyjnym do wymiany danych pomiędzy modulem a innym urządzeniem jest transmisja szeregową UART. Transmisja Bluetooth realizowana jest z protokołem SPP, co pozwala na wykorzystanie modułu jak adaptera szeregowego.

Moduł komunikacyjny RS-485 jest układem zaprojektowanym przez wykonawców głównego sterownika pralki. Oparty jest na układzie ADM2582. W sterowniku zaimplementowano jego obsługę.

Do realizacji obustronnej komunikacji z płytą główną pralki oraz komunikacji Bluetooth wykorzystano dwa oddzielne interfejsy UART mikrokontrolera. Pozwoliło to na bezkonfliktowe, niezależne przesyłanie danych. Przesył danych jest głównym zadaniem układu, który można traktować jako most komunikacyjny pomiędzy telefonem, komputerem a pralką.



Rys. 7. Wykonany sterownik do komunikacji głosowej 1 – mikroprocesor STM32, 2 – moduł komunikacyjny, 3 – moduł Bluetooth

Fig. 7. Voice communication using driver 1 – microcontroller STM32, 2 – communication module, 3 – Bluetooth module

Oprogramowanie sterujące mikrokontrolerem zostało napisane w języku C z wykorzystaniem środowiska programistycznego Ride 7. Do obsługi urządzeń peryferyjnych posłużono się bibliotekami *Standard Peripheral Library* przeznaczonymi dla mikrokontrolera STM32. Ułatwiło to i skróciło proces programowania urządzenia. Program główny ma za zadanie sterować przepływem danych pomiędzy telefonem, komputerem a płytą główną sterownika pralki poprzez moduł. Zaimplementowano w nim obsługę komunikacji RS-485 z protokołem MODBUS. W tym celu zaadaptowano biblioteki zawierające odpowiednie funkcje powstałe na potrzeby projektu *Inteligentne sterowniki pralek*. Komunikacja realizowana jest przy pomocy buforów nadawczo odbiorczych. Bufory odbiorcze znajdują się w przerwanach obsługujących dane interfejsu USART. Wysyłanie realizowane jest w pętli głównej programie. Dane są kontrolowane za pomocą sumy kontrolnej. Biblioteki MODBUS zawierają zaawansowaną obsługę błędów i sterują automatycznie transmisją pomiędzy układami Master-Slave.

3.3. Program na platformę Android

Po zapoznaniu się z możliwościami i sposobem programowania na platformę Android zdecydowano o jej wyborze do zrealizowania jednego z celów pracy, jakim było napisanie oprogramowania do współpracy telefonu ze sterownikiem pralki.

Stworzony program stanowi bezpośredni interfejs, który może być uruchomiony na dowolnym urządzeniu wyposażonym w system Android, takim jak smartfon lub tablet. Aplikację napisano w języku Java z wykorzystaniem programu Eclipse i zestawu bibliotek Android SDK. Graficznie program zgodny jest z interfejsem LCD wchodzącego w skład modułowego sterownika pralki. Co różnicą stanowi możliwość sterowania aplikacją za pomocą mowy. Struktura aplikacji jest typowa dla urządzeń wbudowanych czy mobilnych. Wykorzystując dotykowy ekran LCD, użytkownik ma możliwość wyboru danej opcji w postaci wirtualnego przycisku. Po wyborze następuje wyświetlenie odpowiedniego menu.

Podczas pisania aplikacji do jej testowania wykorzystano smartfon Galaxy S Plus I9001. Pozwoliło to na sprawdzenie wszystkich zaimplementowanych funkcji w aplikacji – takich jak rozpoznawanie mowy oraz komunikacja Bluetooth. Działanie aplikacji w dalszej części rozdziału zostanie przedstawione na zrzutach ekranu.

Po uruchomieniu aplikacji sprawdzany jest stan zasobów telefonu takich jak Bluetooth, czy możliwość wykorzystania rozpoznawania mowy Google. Jeśli moduł komunikacji bezprzewodowej jest wyłączony, aplikacja żąda jego włączenia, gdyż jest on wykorzystywany w programie do przesyłu danych. Po poprawnym załączeniu ukazuje się ekran powitalny (rys. 8).

W przypadku wyboru trybu głosowego użytkownik za pomocą dotyku użytkownik może jedynie aktywować mikrofon. Parametry wprowadza się za pomocą komend głosowych. Podczas gdy mikrofon jest aktywny urządzenie automatycznie wykrywa mowę co jest sygnalizowane na ekranie (rys. 9). W momencie wykrycia początku i końca wypowiedzi próbki dźwięku wysyłane są do serwera w celu

rozpoznania. Na podstawie odpowiedzi serwera weryfikowana jest zgodność rozpoznania komendy. Poprawność jest sygnalizowana poprzez zaznaczenie znaku obok danego parametru oraz ustalenie jego wartości. Rozpoczęcie prania następuje, jeśli wszystkie wartości parametrów zostaną ustalone. Po rozpoczęciu programu wyświetlany jest ekran końcowy, wspólny dla wszystkich trybów. Poprawność przykładowego rozpoznania komend przedstawiono na rysunku 10. Analizując rysunki 9 oraz 10 można dojść do wniosku, że ustalenie parametrów za pomocą interfejsu głosowego jest bardziej intuicyjne i mniej energochłonne niż w przypadku trybów dotykowych, w których użytkownik musi ustalić je na kolejnych ekranach. Warto podkreślić, że za pomocą mowy można wprowadzić kilka parametrów jednocześnie. Przykładowe komendy interpretowane przez sterownik to:

- Temperatura:
 - “40”, “50”, “60”,
- Obroty:
 - od “0” do “1200”,
- Zabrudzenie:
 - “małe”, “średnie”, “duże”,
- Płukanie:
 - “jeden”, “dwa”, “trzy”,
- Opóźnienie:
 - “zero”, “30”, “60”,
- Program:
 - “Bawełna”, “Jedwab”.

Podczas wykonywania programu prania sterownik cały czas komunikuje się z aplikacją z pomocą Bluetooth. Następuje cykliczna wymiana danych, co pozwala na sterowania pralką w czasie rzeczywistym, można w każdej chwili dokonać pauzy bądź zatrzymania programu.



Rys. 8. Ekran powitalny

Fig. 8. Welcome screen

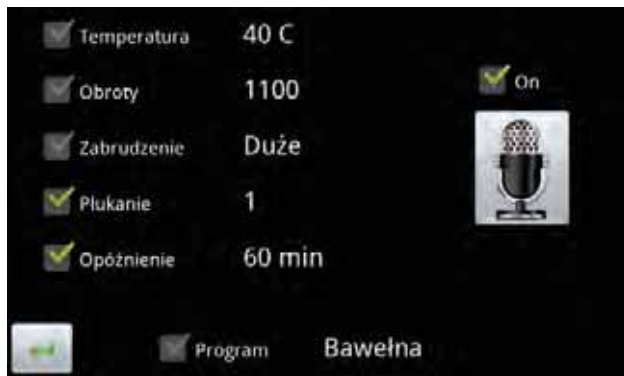
Aplikacja do sterowania głosowego wykazuje się dużą poprawnością rozpoznawania komend z względu na ich małą liczbę. Podawanie parametrów z jej wykorzystaniem jest bardzo efektywne. Dzięki bezprzewodowemu połączeniu można komunikować się ze sterownikiem na odległość, co zapewnia moduł Bluetooth. Jest to dobra alternatywa dla klasycznych interfejsów, w których żeby skontrolować proces prania należy się udać do pomieszczenia, gdzie znajduje się pralka. W przypadku zaproponowanego i zrealizowanego rozwiązania użytkownik może posiadać informację „w kieszeni”. Aplikacja na platformę Android

może być dostosowana do potrzeb i uruchomiona na różnej klasy urządzeniach. Jediną wadą interfejsu jest konieczność zapewnienia dostępu do Internetu, gdyż wykorzystany system ASR Google znajduje się na zewnętrznym serwerze. Jednak w dzisiejszych czasach nie jest to duży problem, gdyż prawie każde gospodarstwo domowe posiada dostęp do światowej sieci WWW.



Rys. 9. Wykrycie mowy

Fig. 9. Speech detection



Rys. 10. Rozpoznanie komend

Fig. 10. Command recognition

4. Podsumowanie

W artykule przedstawiono nowatorskie podejście do budowy interfejsów wykorzystujących systemy automatycznego rozpoznawania mowy oraz przykłady ich zastosowań. Podsumowując, można zauważyć, że systemy rozpoznawania mowy będą wkraczać w coraz to nowsze dziedziny, zarówno techniki jak i życia codziennego. Duży postęp technologiczny, jak i coraz skuteczniejsze algorytmy spowodują, że innowacyjne systemy wyposażone w interfejsy człowiek-maszyna będą pozwalały na kontakt przypominający rozmowę między ludźmi. Właśnie do takiej formy komunikacji ludzie są bardzo dobrze przygotowani. Ułatwi ona sterowanie, kontrolowanie i zarządzanie maszynami czy urządzeniami technicznymi. Dlatego warto podjąć prace nad wprowadzaniem techniki komunikacji głosowej człowieka z obrabiarką, robotem itp. Pozwoli to znacząco uprościć powszechnie dzisiaj stosowane interfejsy człowiek-maszyna typu panele i pulpity operatorskie. Wprowadzenie systemów mówiących do obrabiarek doprowadzi także zapewne do rewolucyjnych zmian w sposobie ich programowania i obsługi, powodując, że obsługa ta będzie bardziej intuicyjna i łatwiejsza dla człowieka

Praca naukowa finansowana ze środków na naukę ze środków Narodowego Centrum Badań i Rozwoju w latach 2010-2012 w ramach projektu rozwojowego pt.: „Inteligentne sterowniki pralek”.

Bibliografia

- [www.primespeech.pl] – Rozpoznawanie Mowy i Portale głosowe (10 listopada 2012).
- Lewis J., *Put on your human-machine interface*, “Design News” 8/2001.
- [www.mazak.eu] – Yamazaki Mazak Europe (10 listopada 2012).
- Rogowski A., *Głosowa komunikacja człowiek-maszyna w gniazdach obróbkowych o zróżnicowanym stopniu automatyzacji*, „Zeszyty Naukowe Politechniki Poznańskiej: Budowa Maszyn i Zarządzanie Produkcją”, 15/2011, 131–144.
- Rogowski A., *Industrially oriented voice control system*, “Robotics and Computer-Integrated Manufacturing”, 28/2012, 303–315.
- Pires J.N., *New challenges for industrial robotic cell programming*, “Industrial Robot”, 36/2008.
- Pires J.N., *Robot-by-voice: Experiments on commanding an industrial robot using the human voice*, “Industrial Robot”, 32/2005.
- Pires J.N., *Industrial Robots Programming: Building Applications for the Factories of the Future*, Springer 2007.
- Bos J., Klein E., Oka T., *Meaningful Conversation With A Mobile Robot*, “Annual Meeting of The European Chapter of The Association of Computational Linguistics”, Edinburgh 2003.
- [www.ltg.ed.ac.uk/godot/] – Godot – The Robot (10 listopada 2012).
- [http://www.agh.edu.pl/pl/aktualnosci/blog-naukowy.html,,0:vw:2928] – Smartfon steruje robotem (10 listopada 2012).
- Milecki A., Pittner G., *Modułowy, 32-bitowy sterownik pralki*, „Pomiary Automatyka Robotyka”, 2/2012, 357-361.

Applications of speech recognition systems to control and voice communication with mechatronic devices

Abstract: This paper presents examples of the use of automatic speech recognition systems to build human-machine voice interfaces. Also this paper briefly describes how these applications can work. The rest of the article shows the concept of usage speech recognition system based on own driver which cooperate with washing machine controller.

Keywords: automatic speech recognition, voice control, human-machine interface, washing machine controller

mgr inż. Roman Regulski

Doktorant w Zakładzie Urządzeń Mechatronicznych, Wydziału Budowy Maszyn i Zarządzania Politechniki Poznańskiej.
Zainteresowania naukowe: mechatronika, sterowniki mikroprocesorowe, automatyczne systemy rozpoznawania mowy, sterowanie głosowe.



e-mail: roman.regulski@doctorate.put.poznan.pl

mgr inż. Amadeusz Nowak

Doktorant Zakładu Urządzeń Mechatronicznych, Wydziału Budowy Maszyn i Zarządzania Politechniki Poznańskiej.
Zainteresowania naukowe: mechatronika, metody sztucznej inteligencji w sterowaniu.



e-mail: amadeusz.nowak@doctorate.put.poznan.pl