

Czy można rozmawiać z robotem spawalniczym?

Can we talk to a welding robot?

Streszczenie

W artykule przedstawiono zagadnienia związane ze sterowaniem głosowym robotami przemysłowymi, w tym robotami spawalniczymi. Omówiono celowość wykorzystania automatycznego rozpoznawania mowy w robotyce, potencjalny zakres zastosowań oraz specyficzne wymagania dotyczące aplikacji sterowania głosowego związanych z robotami przemysłowymi. W szczególności skoncentrowano się na głosowym wspomaganie programowania robotów przez uczenie. Poruszone zostało zagadnienie definiowania języka komend głosowych oraz różne aspekty integracji systemu rozpoznawania mowy z układem sterowania robota przemysłowego. Rozważania poparto przykładami ze zrealizowanej implementacji sterowania głosowego robotem Movemaster.

Słowa kluczowe: programowanie robotów, rozpoznawanie mowy

Abstract

Current paper deals with various aspects of voice control system that could be applied to industrial robots, particularly in welding applications. It discusses the usefulness of voice-based human-machine interfaces, potential areas of application, restrictions as well as specific requirements regarding these systems. In particular, it focuses on speech-aided teach-in robot programming. A separate chapter is dedicated to the issue of voice command language description. Integration of speech recognition system and robot controller is also broadly discussed. Description of these issues is illustrated by example of practically implemented voice control system applied to educational robot Movemaster.

Keywords: robot programming, speech recognition

Wstęp

Przełom 20 i 21 wieku, a w szczególności ostatnie lata, były okresem bardzo intensywnego rozwoju technik rozpoznawania mowy. Jak wiadomo, mowa jest najdogodniejszym, a zarazem najbardziej naturalnym sposobem komunikowania się ludzi pomiędzy sobą, stąd też coraz liczniejsze są próby zastosowania komunikacji głosowej jako interfejsu pomiędzy człowiekiem-operatorem a obsługiwaną przez niego maszyną. Pozytywne praktyczne doświadczenia z rozpoznawaniem mowy w takich dziedzinach jak np. systemy głosowego dostępu do informacji, zainspirowały naukowców do prób ich przeniesienia w dziedzinę robotyki. W szczególności dotyczy to sterowania głosowego robotami usługowymi [1÷5], jednak coraz częściej podejmuje się również próby przeniesienia tych doświadczeń do praktyki przemysłowej. Bardzo często ma to miejsce w kontekście coraz bardziej popularnej koncepcji współpracy ludzi i robotów, która miałyby się odbywać we wspólnej przestrzeni roboczej [6,7]. Odnosi się to zwłaszcza do szeroko rozumianych prac montażowych [7,8], choć powstały także inne opracowania np. dotyczące zastosowania rozpoznawania mowy do bezpośredniego [9] bądź zdalnego [10,11] sterowania zrobotyzowanym gniazdem obróbkowym. Funkcjonujące w praktyce opracowania ograniczają się na razie do testów laboratoryjnych [12÷14].

Jeśli chodzi o spawalnictwo, to na uwagę zasługuje fakt, że już w latach osiemdziesiątych ubiegłego wieku podejmowano próby wspomaganie pracy spawacza poprzez komunikację głosową pomiędzy nim a spawarką. Rezultatem tych prób był opracowany patent [15]. Scenariusz działania takiego systemu był następujący: spawacz, wyposażony w nadajnik audio, podaje głosem komendy, które po przeanalizowaniu przez komputer wpływają na sterowanie zasilaniem przez spawarkę. W ten sam sposób można także sterować prędkością podawania drutu elektrody. Autorowi niniejszego artykułu nie są znane dalsze losy tego patentu, jednak idea komunikacji głosowej człowiek – maszyna w dziedzinie spawalnictwa jest ciągle żywa. Również w znacznie nowszych opracowaniach dotyczących sterowania głosowego robotami, jako przykład zastosowania podano prowadzenie prac spawalniczych przez mieszany (ludzie – roboty) zespół [16]. Zawarte w tytule pytanie: „Czy można rozmawiać z robotem spawalniczym?” jest więc całkowicie zasadne. Należałoby jednak jeszcze je rozszerzyć o pytania kolejne: „O czym rozmawiać z robotem?” oraz „Jak rozmawiać z robotem?”. Innymi słowy, należy się zastanowić, w jakich sytuacjach komunikacja głosowa pomiędzy operatorem i robotem jest celowa, w jakim zakre-

sie i okolicznościach może ona ułatwić pracę człowiekowi i wreszcie – jak ją zrealizować. W artykule niniejszym omówiono potencjalne zastosowania rozpoznawania mowy robotyce, a także przedstawiono wybrane problemy z nimi związane oraz rozwiązania tych problemów.

Rozpoznawanie mowy w robotyce

Jak wspomniano wcześniej, głównym obszarem badań nad zastosowaniem rozpoznawania mowy w robotyce jest sterowanie głosowe robotami usługowymi. Stosunkowo nieliczne są prace dotyczące sterowania głosowego robotami przemysłowymi. Wynika to z wielu przyczyn. Badania przeprowadzone przez Profantera i in. [17] (badania te ukie-
runkowane były m.in. na różne metody komunikacji z robotami spawalniczymi) wykazały, że – wbrew pozorom – nie w każdej sytuacji wprowadzanie danych za pomocą głosu jest najbardziej efektywną i najłatwiejszą z punktu widzenia użytkownika metodą. W wielu przypadkach wygodniejsze było wprowadzanie danych za pomocą gestów, ekranów dotykowych itp. Zastosowanie mowy było konkurencyjne w stosunku do innych metod w dość ograniczonym zakresie – zwłaszcza przy wskazywaniu obiektów. Należy stąd wyciągnąć wniosek, że sposób i zakres funkcjonowania projektowanego systemu sterowania głosowego musi być najpierw bardzo dokładnie przemyślany. W szczególności trzeba rozważyć integrację takiego systemu z innymi metodami wprowadzania danych.

Ponadto, automatyczna analiza mowy w aplikacjach związanych z robotami przemysłowymi charakteryzuje się pewną specyfiką w stosunku do zastosowań w innych dziedzinach [18]:

- Bardzo ważne jest prawidłowe rozpoznanie wszystkich słów wypowiedzianej komendy głosowej. Niespełnienie tego warunku może być przyczyną awarii maszyny, a nawet zagrożenia życia ludzkiego.
- Prawidłowe rozpoznanie komendy może być wspomaganie przez uwzględnienie kontekstu. Dlatego pożądana jest współpraca systemu głosowego na przykład z systemem automatycznego rozpoznawania obrazu.
- Ponieważ na ogół reakcja maszyny na komendę głosową powinna być natychmiastowa, więc system rozpoznawania mowy powinien być zabezpieczony przed rezultatami np. takich zjawisk jak pauzy w wypowiedzi, będące przyczyną niejednoznacznej interpretacji intencji operatora.

Nie wchodząc w szczegóły, a zainteresowanego czytelnika odsyłając do literatury [18,19], można stwierdzić, że spełnienie wymienionych postulatów jest możliwe między innymi poprzez wykorzystanie reguł gramatyki formalnej np. gramatyki bezkontekstowej (CFG) do zdefiniowania języka komend głosowych. Ten fakt ma również pewien wpływ na wnioski dotyczące zakresu zastosowania rozpoznawania mowy przy interakcji operatora z robotem spawalniczym.

Kiedy więc może być pożądana komunikacja głosowa z robotem? Wydaje się, że w szczególności może to dotyczyć trzech przypadków. Najprostszy z nich, to reagowanie głosem w sytuacjach awaryjnych. Drugi, to zorientowane zadaniowo sterowanie złożonymi funkcjami realizowanymi przez robota (wydaje się, że szczegółowe sterowanie elementarnymi funkcjami robota w warunkach produkcyjnych nie miałyby większego sensu). Wreszcie trzeci obszar zastosowań, to głosowe wspomaganie programowania robota metodą uczenia. W wielu przypadkach może się okazać, że podawanie pewnych komend głosem będzie łatwiejsze, niż oddziaływanie na robota za pomocą programatora. Z tym właśnie obszarem związana jest tematyka niniejszego artykułu.

Język komend głosowych

Istotną rolę w systemie komunikacji głosowej z robotem przemysłowym odgrywa właściwie zdefiniowany język komend głosowych. Jak napisano w poprzednim rozdziale, w przypadku robotów spawalniczych (a właściwie ogólnie: robotów przemysłowych) najlepszym rozwiązaniem jest wykorzystanie reguł gramatyki formalnej do zdefiniowania tego języka. Narzędziem ułatwiającym to zadanie jest opracowany format VCD [14] (w niniejszym artykule będziemy się do tego formatu odwoływać, dlatego też w celu zrozumienia przykładów podawanych dalej, wskazane jest zapoznanie się z podaną pozycją literatury). Ważną cechą tego formatu jest możliwość wspólnego zapisu zarówno syntaktyki, jak i semantyki komend głosowych. Opis języka komend w tym formacie może być wykorzystany na wszystkich etapach przetwarzania komendy [20]:

- do wygenerowania reguł gramatyki służących do zwiększenia niezawodności rozpoznawania ciągów wyrazów przez aparat rozpoznawania mowy,
- podczas analizy syntaktycznej,
- podczas analizy semantycznej.

Dzięki takiemu rozwiązaniu, rozpoznawanie mowy ma charakter systemowy, a poszczególne jego etapy są ze sobą powiązane, co znacznie zwiększa niezawodność sterowania głosowego. Należy zauważyć, że rozwiązania bazujące nie na gramatyce formalnej, lecz na przetwarzaniu języka sensu stricte naturalnego, z reguły ograniczają się do analizy komend wprowadzanych z pomocą klawiatury [21].

Jaki zatem powinien być język komend głosowych stosowanych przy programowaniu przez uczenie robota spawalniczego? Podstawowe komendy dotyczyłyby przemieszczania efektora robota w przestrzeni oraz zapamiętywania osiągniętych pozycji (przykładowo: przypisywania im określonych numerów). Zapis takiej komendy w formacie VCD [14] mógłby mieć postać (ze względu na brak miejsca nie podano tu opisu frazy *liczba):

```
#com zapisz  
zapisz pozycję pod numerem *liczba: p5  
zapisz pozycję ?numer *liczba: p4
```

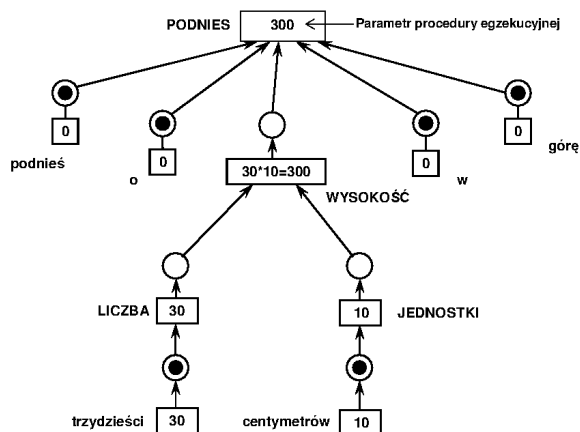
Słowa poprzedzone znakiem zapytania są opcjonalne. Wyrażenia poprzedzone gwiazdką są to nazwy fraz składowych komendy, których zapis też musi się znaleźć w opisie języka. Na końcu każdej linii może się znaleźć wyrażenie arytmetyczne określające wartość liczbową określającą semantykę (znaczenie) komendy. W powyższym przykładzie „wartość” komendy jest określona jako wartość parametru odpowiadającego odpowiednio piątemu i czwartemu słowu danego sformułowania komendy.

W równie prosty sposób można by też zapisać syntaktykę i semantykę komend dotyczących przemieszczania efektora robota. Przykładowy zapis ruchu w górę mógłby mieć postać:

```
#com podnies  
podnieś o *wysokość w górę: p3  
#def wysokość  
*liczba *jednostki: p1*p2  
#def jednostki  
milimetrów: 1  
centymetrów: 10  
metrów: 1000
```

Podany zapis uwzględnia bardzo dużą liczbę możliwych sformułowań komendy o nazwie „podnieś”. Przykładowe sformułowanie komendy przez operatora mogłoby brzmieć: „podnieś o trzydzieści centymetrów w górę”.

W rezultacie przetwarzania tej komendy następuje najpierw wygenerowanie jej drzewa składniowego (analiza syntaktyczna), a następnie proceduralnej sieci semantycznej, pokazanej na rysunku 1.



Rys. 1. Proceduralna sieć semantyczna reprezentująca wypowiedzianą komendę głosową

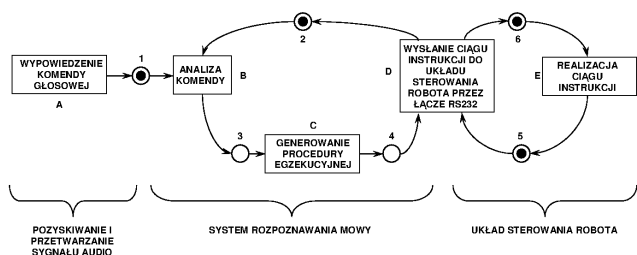
Fig. 1. Procedural semantic network representing voice command

Proceduralna sieć semantyczna wykorzystuje mechanizmy stosowane w sieciach Petriego [18]. W momencie uruchomienia algorytmu analizy semantycznej żetony umieszczane są w miejscach sieci, odpowiadających poszczególnym rozpoznany słowom komendy. Powoduje to aktywację kolejnych przejść, w wyniku czego realizowane są obliczenia zgodnie z opisem semantyki języka komend w formacie VCD. Ostatecznie cała komenda głosowa zostaje zredukowana do pojedynczej tzw. procedury egzekucyjnej („rozumiałej” dla sterowanej maszyny) oraz jej parametrów liczbowych. W podanym przypadku mamy pojedynczy parametr opisujący wysokość w milimetrach, o którą należy podnieść efektor robota. Przykładowa procedura egzekucyjna może mieć postać:

- MoveZ (300)

Integracja systemu rozpoznawania mowy z układem sterowania robota

System rozpoznawania mowy będzie z reguły uruchamiany na komputerze zewnętrznym w stosunku do układu sterowania robota. Zarazem jednak konieczna jest taka integracja wszystkich elementów, aby robot mógł na bieżąco i bez nadmiernych opóźnień reagować na komendy głosowe wypowiedziane przez operatora. Należy więc zapewnić odpowiednie mechanizmy komunikacji pomiędzy systemem rozpoznawania mowy a układem sterowania robota. Oznacza to konieczność takiego zaprogramowania robota, aby pracował on w synchronizacji z programem uruchomionym na zewnętrznym komputerze. Na Politechnice Warszawskiej przeprowadzone zostały badania mające na celu sprawdzenie możliwości języka Karel, służącego do programowania robota Fanuc, a także dokonano zakończonej sukcesem implementacji na robocie Movemaster. Schemat pokazujący ogólną koncepcję tej implementacji został pokazany na rysunku 2.



Rys. 2. System głosowego sterowania robotem przemysłowym

Fig. 2. Industrial robot voice control system

Działanie systemu głosowego sterowania robotem przemysłowym zostało przedstawione w postaci prostej sieci Petriego. Cyframi arabskimi zostały oznaczone tak zwane „miejsca” (places) tej sieci, zaś dużymi literami tak zwane „przejścia” (transitions). W wyniku przetworzenia komendy głosowej wygenerowana zostaje procedura egzekucyjna, składająca się z ciągu instrukcji sterujących robotem. Instrukcje te są wysyłane do układu sterowania robota przez złącze RS232. W zrealizowanej implementacji dla robota Movemaster, procedury egzekucyjne są zapisane w specjalnie w tym celu opracowanym języku Arlang [9÷11]. Najistotniejsze (z punktu widzenia tej implementacji) procedury języka Arlang są następujące:

- Load (nazwa instrukcji, parametry) – załadowanie do bufora instrukcji sterującej robotem
 - Execute() – wysłanie instrukcji zawartych w buforze do układu sterowania robota
 - ReadData(wykaz zmiennych) – odczytanie danych wysłanych z układu sterowania robota przez RS232
- Realizacja procedury ReadData jest możliwa po uprzednim uruchomieniu w układzie sterowania robota np. instrukcji PR (position read):

```
Load (PR)
Execute()
ReadData(x)
```

Poniżej znajduje się wykaz instrukcji robota Movemaster, istotnych dla zrozumienia podanych dalej przykładów:

- MP – ruch do pozycji opisanej współrzędnymi. Parametry: współrzędne x,y,z,a,b
- PD – zapamiętanie aktualnej pozycji. Parametr: numer przypisany pozycji
- PR – żądanie wysłania współrzędnych aktualnej pozycji

Przykładowa procedura egzekucyjna w języku Arlang, odpowiedzialna za podniesienie efektora robota o określoną liczbę milimetrów w górę, miałaby postać następującą:

```
procedure MoveZ
parameter h
Load (PR)
Execute()
ReadData (x,y,z,a,b)
Load (MP, x, y, z+h, a, b)
Execute()
endproc
```

Należy zwrócić uwagę, że przy programowaniu robotów przez uczenie operator nie zawsze potrafi określić konkretną wartość, o którą chciałby przesunąć efektor robota. Dlatego system sterowania głosowego powinien umożliwić definiowanie i realizowanie komend, w których przemieszczenia podawane byłyby w sposób „rozmyty” np. „Podnieś jeszcze trochę”, „Za daleko”, „Cofnij” itp. Zauważmy, że opis języka komend w formacie VCD, oraz podana architektura systemu, umożliwia dość łatwą realizację tego typu zadań. Przykładowo, podczas realizacji procedury egzekucyjnej związanej z przesunięciem efektora robota, system może zapamiętywać wartość ostatniego przesunięcia. Kolejne wywołanie procedury egzekucyjnej, tym razem jako efekt komendy określającej przemieszczenie w sposób „rozmyty”, może się odwoływać do zapamiętanej wartości. Jeśli ostatnio zrealizowane byłoby przesunięcie o wartość h_old, to odpowiednia procedura egzekucyjna mogłaby mieć następującą postać:

```
procedure JeszczeTroche
call MoveZ(h_old/5)
endproc
```

Podsumowanie

Przedstawione w artykule rozważania pokazują, że istnieje szerokie pole potencjalnych zastosowań automatycznego rozpoznawania mowy w robotyce, w tym również w obszarze zrobotyzowanego spawania. W szczególności pożądanymi byłyby wykorzystanie interfejsów głosowych przy programowaniu robotów przez uczenie. Aby jednak interfejsy te spełniały swoją funkcję efektywnie, muszą one być dobrze przemyślane i właściwie zaimplementowane. Bardzo istotną rolę odgrywa właściwe zdefiniowanie języka komend głosowych. Pomocne może być przy tym wykorzystanie formatu VCD.

Literatura

- [1] V. Kulyukin: Human-Robot Interaction Through Gesture-Free Spoken Dialogue, *Autonomous Robots*, Vol. 16, s. 239-257, 2004.
- [2] J. Bos, T. Oka: A spoken language interface with a mobile robot, *Artificial Life and Robotics*, Vol. 11(1), s. 42-47, 2007.
- [3] J. Savage i in.: VIRbot: a system for the operation of mobile robots, *Lecture Notes in Computer Science*, Vol. 5001, s. 512-519, 2008.
- [4] M. Yokota, K. Sugita, T. Oka: Natural language understanding based on mental image description language Lmd and its application to language-centered robot manipulation, *Artificial Life and Robotics*, Vol. 13(1), s. 84-88, 2008.
- [5] T. Oka i in.: RUNA: a multimodal command language for home robot users, *Artificial Life and Robotics*, Vol. 13(2), s. 455-459, 2009.
- [6] J. Pires: The Industrial robot as a human coworker: the role of the speech interfaces, materiały konferencyjne: International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion, Portugal, 2007.
- [7] M. Zaeh, W. Roesel: Safety aspects in a human-robot interaction scenario: a human worker is co-operating with an industrial robot, *Progress in Robotics*, Vol. 44(2), s. 53-62, 2009.
- [8] A. Weiss i in.: Exploring human-robot cooperation possibilities for semiconductor manufacturing, materiały konferencyjne: International Conference on Collaboration Technologies and Systems, Philadelphia, PA, USA, s. 173-177, 2011.
- [9] A. Rogowski: Głosowa komunikacja człowiek-maszyna w gniazdach obróbkowych o zróżnicowanym stopniu automatyzacji, *Zeszyty Naukowe Politechniki Poznańskiej* 15, s. 131-144, 2011.
- [10] A. Rogowski: Projekt i implementacja systemu programowania i sterowania gniazdem zrobotyzowanym przez Internet", *Prace Naukowe Politechniki Warszawskiej*, seria Elektronika 175, s. 311-320, 2010.
- [11] A. Rogowski: Remote programming and control of the flexible machining cell", *International Journal of Computer Integrated Manufacturing*, Vol. 28(6), s. 650-663, 2015.
- [12] J. Pires: Robot-by-voice: Experiments on commanding an industrial robot using the human voice, *Industrial Robot*, Vol. 32(6), s. 505-511, 2005.
- [13] G. Veiga, J. Pires, K. Nilsson: Experiments with service-oriented architectures for industrial robotic cells programming, *Robotics and Computer-Integrated Manufacturing*, Vol. 25(4-5), s. 746-755, 2009.
- [14] A. Rogowski: Industrially oriented voice control system, *Robotics and Computer-Integrated Manufacturing*, Vol. 28(3), s. 303-315, 2012.
- [15] G. Tunnell, Ch. Pomernacki, J. Gregg: Voice controlled welding system, Patent US 4641292A, 1987.
- [16] T. Fong, C. Kunz, L. Hiatt, M. Bugajska: The Human-Robot Interaction Operating System, materiały konferencyjne: 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Salt Lake City, Utah, USA s. 41-48, 2006.
- [17] S. Profanter, A. Perzylo, N. Somani, M. Rickert, A. Knoll: Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Hamburg, Germany*, 2015.
- [18] A. Rogowski: Analiza i synteza systemów sterowania głosowego w zrobotyzowanym wytwarzaniu, *Prace Naukowe Politechniki Warszawskiej, Mechanika* z. 244, 2012.
- [19] V. Kulyukin: Talk the Walk: robotic NLP vs. human sublanguage acquisition, *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, Palo Alto, CA, USA s. 33-35, 2007.
- [20] A. Rogowski: Web-based remote voice control of robotized cells, *Robotics and Computer-Integrated Manufacturing*, Vol. 29(4), s. 77-89, 2013.
- [21] M. Stenmark, P. Nugues: Natural Language Programming of Industrial Robots, *International Symposium on Robotics (ISR) Seoul, Korea*, 2013.