

RECENZJA

rozprawy doktorskiej mgr. inż. Bartłomieja Klina

pt. *Zwiększanie efektywności komunikacji człowiek-komputer
(human - to - machine) w języku naturalnym
w autonomicznych systemach robotycznych poprzez akwizycję i deskrypcję
zagregowanej informacji wielokanałowej*
opracowana na zlecenie Przewodniczącego Rady Dyscypliny
Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne
Politechniki Opolskiej – dr. hab. inż. Andrzeja Waindoka

1. Obszar problemowy

Analiza coraz bardziej zaawansowanych systemów komunikacji głosowej człowieka z autonomicznym układem robotycznym wymaga dalszego rozwoju metod i technik aby poszerzyć zakres ich bardziej efektywnego zastosowania w warunkach naturalnej interakcji głosowej. Autor podejmuje trafnie zidentyfikowany problem badawczy, wynikający zarówno z analizy literatury, jak i własnych doświadczeń wdrożeniowych. Opiniowana rozprawa realizowana w formule doktoratu wdrożeniowego jest udaną propozycją opracowania innowacyjnego systemu tzw. Smart Beamforming opartą na wykorzystaniu fuzji danych wielomodalnych, zaawansowanych technik adaptacyjnego beamformingu w oparciu o dane wizyjne i niezależne algorytmy akustyczne jak również mechanizmów zarządzania kontekstem wieloosobowym (warstwa sterowania nadzorczego). Sformułowana na str. 9 hipoteza badawcza „*Możliwe jest wykorzystanie autonomicznego systemu robotycznego do jednoczesnej obsługi konwersacji z wieloma rozmówcami w warunkach rzeczywistych*” wraz z podanymi mierzalnymi kamieniami milowymi pozwalającymi na empiryczną weryfikację hipotezy są poprawnie opracowane.

Na podstawie dobrze opracowanego przeglądu literatury i oceny stanu wiedzy dotyczącej systemów komunikacji głosowej człowieka z robotem wraz z podanymi ich ograniczeniami takimi jak między innymi ograniczenia schematu komunikacji REPL (ang. *Read-Eval-Print Loop*), niewystarczająca zdolność do rozróżniania mówców i separacji źródeł dźwięku, wpływu

niedoskonałości systemów percepcji (audio i wideo) na jakość interakcji, trudności w adaptacji do dynamicznych zmian w scenie interakcji czy brak skutecznych mechanizmów zarządzania kontekstem w rozmowach wieloosobowych zaproponowano system tzw. Smart Beamforming służący do akwizycji, synchronizacji i fuzji informacji wielokanałowej oraz wielokanałowego przetwarzania mowy, sprzęgniętego z warstwą sterowania nadzorczego. Zaproponowano adaptacyjny beamforming z równoległymi kanałami przypisanymi do aktywnych rozmówców, a także zestaw metryk i modułów diagnostycznych pełniących rolę czujników stanu percepcji, w tym pomiar SNR na wyjściu kanałów mowy interlokutorów, moduł Recovery Time Calculator mierzący czas powrotu jakości po rekonfiguracji wiązek oraz moduł oceny jakości rozpoznawania mowy. Opracowano też warstwę Virtual Context Factory (VCF) agregującą informację o rozmówcach, stabilizującą identyfikację kontekstów, wspierającą przypisanie wypowiedzi do mówców oraz estymującą adresowanie wypowiedzi do robota. Ważnym i oryginalnym rozwiązaniem jest zaprojektowanie algorytmów sterowania w postaci polityki nadzorczej, realizowanej przez moduły SuperManager, FallbackManager i FailsafeManager, które przełączają tryby pracy oraz generują komunikaty korygujące zachowanie interlokutorów, domykając pętlę sprzężenia zwrotnego pomiędzy jakością percepcji a przebiegiem interakcji. Przeprowadzono obszerne badania walidacyjne dla różnych scenariuszy obejmujących przypadki nominalne i krytyczne, przy czym badania obejmowały zarówno symulacje kontrolowane, jak i pomiary na stanowiskach badawczych z robotem Pepper oraz zewnętrzną macierzą mikrofonową własnej konstrukcji. Załączone liczne wyniki badań testowych w postaci zintegrowanych tabel i wykresów dobrze ilustrują funkcjonalności i charakterystyki zaproponowanego systemu komunikacji głosowej człowieka z robotem. Należy też odnotować, że zaproponowany system został wdrożony w firmie Weegree sp. z o.o. sp.k. Są załączone w rozdziale trzynastym opisy tego wdrożenia ale brak jest oficjalnego dokumentu potwierdzającego ten fakt wystawionego przez firmę np. w postaci załącznika.

2. Koncepcja oraz realizacja rozprawy

Rozprawa o ogólnej objętości 230 stron opracowana jest w języku polskim z jednostronicowymi streszczeniami w języku angielskim i polskim. Praca formalnie składa się z czternastu rozdziałów w tym z wstępu i podsumowania oraz wykazu 89 pozycji literaturowych będących podstawą opracowania stanu wiedzy w zakresie tematyki rozprawy dotyczących zarówno rozwiązań teoretycznych jak i częściowo technicznych.

Formalnie rozprawę można podzielić na dwie części, gdzie część pierwsza to rozdziały od drugiego do ósmego, które poświęcone są teoretycznemu wprowadzeniu do metod i algorytmów implementowanych w projektowanym systemie komunikacji głosowej człowieka z autonomicznym systemem robotycznym. Część druga to rozdziały od dziewiątego do trzynastego poświęcone omówieniu zaproponowanych stanowisk badawczych i wykonanych licznych eksperymentów symulacyjnych i technicznych.

Rozdział pierwszy jako wstęp jest ogólnym wprowadzeniem do tematyki rozprawy omawiającym motywację podjęcia badań w zakresie dalszego rozwoju systemów komunikacji głosowej człowieka z układem robotycznym. Sformułowano tutaj hipotezę badawczą wraz z określeniem zakresu badań wymaganych do potwierdzenia sformułowanej hipotezy. Stosunkowo krótki bo 3.5 stronicowy przegląd literatury dotyczy omówienia wybranych projektów badawczych i kierunków powiązanych z implementacją systemów typu multi-talker na platformach robotycznych. W kolejnych rozdziałach 3- 6, które trudno nazwać rozdziałami z uwagi na ich ogólną objętość 14 stron omawia się bardzo skrótowo zjawiska fizyczne z zakresu akustyki takie jak propagacja, odbicia dźwięku, pogłos czy tłumienie echa akustycznego. Kolejny rozdział siódmy (7 stron) dotyczy przetwarzania sygnałów akustycznych w systemach robotycznych a także zagadnień synchronizacji, weryfikacji oraz sterowania i układu architektury czasu rzeczywistego. Skrótowo omawia się podstawowe algorytmy i metody synchronizacji danych wieloczuJNIKOWYCH, synchronizację opartą na znacznikach czasu, na zdarzeniach czy na modelu filtru Kalmana. Oznaczony jako rozdział ósmy trudno z redakcyjnego punktu widzenia, podobnie jak poprzednie rozdziały 3-6, uznać jako rozdział bo liczy on zaledwie dwie strony i dotyczy wybranych metod lokalizacji dźwięku.

Część drugą otwiera rozdział dziewiąty, który omawia stanowiska badawcze wraz z bazą sprzętową laboratorium robotów humanoidalnych na Politechnice Opolskiej. Przedstawiono główne komponenty (moduły) opracowanego systemu jako zintegrowaną warstwę przetwarzania i ewaluacji sygnałów mowy w scenariuszach interakcji człowiek-robot. Przedstawiono wydruk pseudokodu opisującego zaproponowany system Smart Beamforming, oraz omówiono wraz wydrukami pseudokodów moduły obliczenia RIR (odpowiedzi impulsowej pomieszczenia), moduł diagnostyczny Recovery Time Calculator i WER Calculator, moduł redundancji audio-only, czy moduł Virtual Context Factory (VCF). Przedstawiono stanowisko pomiarowe (ver.1) zainstalowane w recepcji siedziby firmy Weegree Sp. z o.o. SK w Opolu, w którym robot Pepper pełnił rolę recepcjonisty. Drugą wersję stanowiska pomiarowego (ver.2)

zainstalowano w laboratorium Politechniki Opolskiej w celu opracowania oraz przetestowania zewnętrznego systemu beamformingu z fuzją danych audio-wideo, działającego niezależnie od wbudowanych modułów robota Pepper. Trzecia wersja stanowiska pomiarowego (ver.3) znajdującego się również na uczelni została zaprojektowana w celu empirycznej weryfikacji zależności pomiędzy parametrami strumienia wideo a wydajnością adaptacyjnego beamformera oraz systemu rozpoznawania mowy. Kolejny rozdział dziesiąty przedstawia ver.4 stanowiska pomiarowego będącego prototypem działającego systemu Smart Beamforming. Wersja ver.4 jest integracją macierzy 8-kanalowej, redundancji wideo i audio-only localization, modułu VCF oraz równoległej instancji STT/NLP, które stanowią finalny prototyp systemu zdolnego do jednoczesnej obsługi wielu rozmówców w warunkach rzeczywistych. Omawia się między innymi zagadnienia sterowania nadzorczego z automatem stanów, selekcję rozmówcy i sterowania komunikatami korygującymi warunki sceny, regulator jakości percepcji oraz mechanizmy fallback i failsafe jak również metryki jakości. Dokonuje się również porównania mechanizmów platformy Pepper (SoftBank/NAOqi) z opracowanym systemem Smart Beamforming, które pokazuje istotną innowacyjność opracowanego systemu. W zaproponowanym systemie wprowadzono warstwy percepcji i sterowania zaprojektowane pod scenariusze grupowe, dynamiczne i zaszumione. Obszerny rozdział jedenasty bo liczący 100 stron przedstawia finalne modele matematyczne komponentów systemu Smart Beamforming oraz ich powiązań w kompletny układ przetwarzania, monitoringu jakości i sterowania nadzorczego. Podane są obszerne tabele parametrów różnych scenariuszy, które z powodzeniem można było umieścić jako załączniki co w sposób istotny poprawiło by czytelność rozdziału. Podane są liczne wykresy i przebiegi sygnałów, które dobrze charakteryzują efektywność pracy zaproponowanego systemu. Opisane są przebiegi scenariusza testowego symulacji dwóch wariantów sterowania tzw. *baseline* i *full*, a wyniki podane są w tabeli oraz w postaci licznych wykresów i przebiegów sygnałów. W rozdziale dwunastym dokonano weryfikacji wskaźników KPI, których spełnienie potwierdzono eksperymentalnie w wariacie *full* a rozdział trzynasty dotyczy opisu wdrożonych wyników pracy doktorskiej. Ostatni rozdział to podsumowanie wyników zaprezentowanych w rozprawie.

Na podstawie skrótowego omówienia treści rozprawy stwierdzam, że mgr inż. Bartłomiej Klin w dobrym stopniu wykazał się umiejętnościami formułowania problemów naukowo-badawczych i technicznych na podstawie analizy stanu wiedzy z wykorzystaniem dostępnej

literatury międzynarodowej zawierającej 89 pozycji obejmujących artykuły opublikowane w czasopismach naukowych, materiałach konferencyjnych i monografiach. Sformułowany problem badawczy dotyczący zwiększenia efektywności systemów komunikacji człowiek-komputer w języku naturalnym w autonomicznych systemach robotycznych rozwiązał właściwymi metodami badawczymi dyscypliny automatyka, elektronika, elektrotechnika i technologie kosmiczne na akceptowalnym poziomie teoretycznym z zachowaniem rygorów matematycznych oraz na dobrym poziomie techniczno-aplikacyjnym. Dodatkowo wykonane liczne badania symulacyjne i eksperymentalne w warunkach rzeczywistych potwierdzają własności i zalety zaproponowanego systemu komunikacji człowiek-komputer. Rozprawa posiada charakter praktyczno-teoretyczny a do opisu metod i algorytmów zastosowano właściwy formalizm matematyczny. Ponadto doktorant wykazał się wiedzą z takich metod badawczych jak np. cyfrowe przetwarzanie sygnałów, cyfrowa komunikacja, projektowania systemów informatycznych, symulacji procesów i zdarzeń czy testowania złożonych systemów komunikacyjnych.

3. Osiągnięcia oryginalne

Realizując sformułowane cele badawcze i zadania mgr inż. Bartłomiej Klin uzyskał wiele ciekawych, ważnych i oryginalnych wyników naukowych potwierdzonych przede wszystkim na drodze badań eksperymentalnych i symulacyjnych. Do ważnych osiągnięć naukowych zaliczam między innymi następujące:

1. Opracowanie nowego paradygmatu percepcji w systemach robotycznych. Jednym z najważniejszych osiągnięć pracy jest odejście od klasycznego modelu, w którym tor percepcji (np. audio) pełni funkcję pasywnego źródła danych. Autor proponuje podejście, w którym percepcja mowy jest jawnie mierzona, diagnozowana i sterowana, metryki jakości (SNR, WER, RT) są integralnym elementem pętli decyzyjnej systemu oraz system adaptuje swoje zachowanie w zależności od aktualnej jakości percepcji. Takie ujęcie problemu stanowi istotną zmianę koncepcyjną względem tradycyjnych architektur typu REPL.
2. Zaprojektowanie i implementacja frameworka Smart Beamforming. Ważnym osiągnięciem rozprawy jest opracowanie autorskiego frameworka Smart Beamforming, który integruje akwizycję, synchronizację i fuzję danych wielokanałowych (audio i wideo), realizuje adaptacyjny beamforming z wieloma równoległymi kanałami dla różnych rozmówców oraz

współpracuje z warstwą sterowania nadzorczego. Framework ten stanowi kompleksową architekturę systemu percepcji i interakcji, zdolną do pracy w warunkach rzeczywistych, co zostało potwierdzone eksperymentalnie.

3. Opracowanie warstwy zarządzania kontekstem rozmówców (ang. Virtual Context Factory). Istotnym osiągnięciem jest zaprojektowanie warstwy Virtual Context Factory (VCF), która agreguje dane o rozmówcach z różnych sensorów, umożliwia stabilną identyfikację i śledzenie interlokutorów, wspiera przypisanie wypowiedzi do właściwego mówcy, oraz ocenia, czy wypowiedź jest skierowana do robota. Rozwiązanie to pokazuje jeden z kluczowych problemów interakcji wieloosobowej – zarządzanie kontekstem dialogowym.

4. Integracja percepcji z warstwą sterowania nadzorczego. Autor proponuje nowatorską architekturę sterowania, obejmującą trzy moduły: Super Manager, Fallback Manager oraz Failsafe Manager. Zaproponowane mechanizmy dynamicznie przełączają tryby pracy systemu, reagują na spadki jakości percepcji, generują komunikaty korygujące zachowanie użytkowników oraz zapewniają odporność systemu na sytuacje nienominalne. Stanowi to istotny wkład w obszar integracji percepcji i sterowania w systemach robotycznych.

5. Istotnym osiągnięciem jest szeroko zakrojona walidacja obejmująca symulacje kontrolowane, eksperymenty laboratoryjne, oraz testy w warunkach rzeczywistych. Badania uwzględniały scenariusze krytyczne, takie jak nakładanie się wypowiedzi, degradacja toru wizyjnego, zmiany pozycji rozmówców, czy przeciążenie liczby kontekstów. Przedstawione wyniki wykazały poprawę stabilności i skuteczności interakcji w porównaniu do wariantu bazowego.

6. Praca ma charakter wdrożeniowy – opracowane rozwiązania zostały zaimplementowane na robocie humanoidalnym Pepper, z wykorzystaniem autorskiej macierzy mikrofonowej oraz w środowisku rzeczywistym (firma Weegree). Te trzy opracowania potwierdzają praktyczną użyteczność proponowanych rozwiązań.

Przedstawione w rozprawie oryginalne wyniki naukowe potwierdzone przede wszystkim badaniami eksperymentalnymi i symulacyjnymi zostały opublikowane przez doktoranta w pięciu współautorskich artykułach. Są to czasopisma: *Sensors*, 2020, nr 8 i 15 (IF:3.5, 100 pkt), *IEEE Robotics and Automation Letters*, 2023 (IF:5.3, 200 pkt), oraz artykuły w pracy zbiorowej

Progress in Applied Electrical Engineering (PAEE), IEEE, 2018 i materiałach konferencyjnych *International Conference on Methods and Models in Automation and Robotics (MMAR)*, IEEE, 2024, Międzyzdroje.

Uwzględniając wymienione oryginalne wyniki badań naukowych oraz ich opublikowanie w pięciu artykułach i pracach zbiorowych, jak również implementację i wdrożenie opracowanego systemu uważam, że mgr inż. Bartłomiej Klin zrealizował cele badawcze, rozwiązał sformułowane zadania i problem praktyczny w stopniu dobrym i uzasadnił postawioną hipotezę na drodze badań symulacyjnych i eksperymentalnych.

4. Uwagi i komentarze

Ogólnie rozprawę pod względem merytorycznym oceniam dobrze, ale pod względem redakcyjnym tylko na poziomie dostatecznym i akceptowalnym. Pewne uwagi zamieszczone niżej dotyczą takich zagadnień i problemów jak.

1. Zaproponowany framework Smart Beamforming jest rozwiązaniem bardzo rozbudowanym, obejmującym wiele współpracujących modułów (m.in. VCF, SuperManager, FallbackManager, FailsafeManager). Choć świadczy to o kompleksowości podejścia, rodzi również istotne pytania. Odczuwalny jest brak pogłębionej analizy złożoności obliczeniowej systemu jak również jednoznacznego określenia wymagań sprzętowych (CPU/GPU, przepustowość, opóźnienia). Ponadto ograniczona jest dyskusja dotycząca skalowania systemu przy rosnącej liczbie rozmówców. W kontekście zastosowań praktycznych, szczególnie w systemach wbudowanych, zagadnienia te mają kluczowe znaczenie

2. Autor odnosi się do istniejących rozwiązań (np. HARK), jednak porównania mają często charakter jakościowy lub opisowy oraz brakuje jednoznacznych benchmarków i wspólnych zbiorów danych. Ponadto nie zawsze przedstawiono porównywalne metryki w identycznych warunkach eksperymentalnych. W rezultacie trudno jednoznacznie ocenić skalę przewagi proponowanego rozwiązania nad istniejącymi metodami.

3. W pracy przedstawiono modele matematyczne systemu, jednak ich poziom formalizacji jest miejscami nierównomierny, niektóre modele mają charakter bardziej opisowy niż analityczny, ponadto brakuje głębszej analizy ich własności (np. stabilności, zbieżności, optymalności). Przykładem jest model sterowania nadzorczego przedstawiony w rozdziale jedenastym.

Silniejsze powiązanie modeli matematycznych z wynikami eksperymentalnymi zwiększyłoby wartość naukową pracy.

4. Przeprowadzone eksperymenty i symulacje są rozbudowane i wartościowe, jednak ich prezentacja jest miejscami rozproszona, brakuje jednolitego schematu prezentacji wyników, jak również nie zawsze jasno pokazano wpływ poszczególnych komponentów systemu. Wprowadzenie bardziej sformalizowanej metodologii eksperymentalnej zwiększyłoby przejrzystość wyników. Ponadto brakuje pełnej informacji umożliwiającej odtworzenie wyników, przykładowo brak szczegółowych parametrów konfiguracji eksperymentów, brak informacji o dostępności kodu lub zbiorów danych testowych czy ograniczone są opisy procedur pomiarowych. W kontekście współczesnych standardów naukowych jest to istotny element.

5. Uwagi redakcyjne

a) Rozprawa jest bardzo obszerna (ponad 200 stron), co skutkuje miejscami nadmierną szczegółowością opisów technicznych, powtarzaniem niektórych koncepcji czy zaburzeniem proporcji między częścią teoretyczną a analityczną.

b) W pracy występuje mieszanie terminologii polskiej i angielskiej (np. „beamforming”, „framework”, „fallback”) czy skrótów bez konsekwentnego ich wprowadzenia (brakuje listy skrótów).

c) Struktura pracy jest logiczna, jednak niektóre rozdziały są bardzo rozbudowane i trudne w odbiorze, podrozdziały mają nierówną szczegółowość oraz brakuje syntetycznych podsumowań w części rozdziałów. Struktura pracy jest poprawna ale jej realizacja w ograniczonym stopniu spełnia wymagania stawiane przed rozprawami doktorskim z nauk inżynieryjno-technicznych: brak jest uzasadnienia przyjęcia aż 13 rozdziałów w sytuacji, gdy np. rozdział 8 to tylko 2 strony a np. rodz. 11 aż 101 stron. Prezentowane są obszerne tabele np. Tabela 4 na str. 129-161, które dla poprawienia czytelności rozprawy można było umieścić jako załączniki.

d) Praca zawiera liczne rysunki i tabele, jednak nie zawsze są one w pełni opisane lub interpretowane w tekście, brak jednolitego stylu prezentacji danych, niektóre ilustracje mają

charakter pogładowy, a nie analityczny. Struktura oznaczeń wzorów, tabel i rysunków jest różnorodna, np. większość wzorów nie jest numerowana, a te które są nie mają przypisania do rozdziałów, podobnie jak tabele, ale z kolei rysunki już posiadają właściwą numerację np. Rysunek 11.9. Ponadto są liczne tzw. dziury w tekście np. str. 22 oraz niepełne podpisy pod rysunkami na str. 44.

Podsumowanie

Mając na uwadze wszystkie aspekty rozprawy doktorskiej mgr. inż. Bartłomieja Klina a w szczególności odnotowane osiągnięcia naukowe oraz dobry wykaz publikacji współautorskich, jak również uwagi dyskusyjne, które nie podważają dobrej wartości pracy, lecz wskazują obszary, które mogłyby zostać rozwinięte lub doprecyzowane i nie mają decydującego wpływu na końcową pozytywną ocenę rozprawy:

- **Stwierdzam**, że rozprawa doktorska spełnia wszystkie wymagania Prawa o szkolnictwie wyższym i nauki z dnia 20 lipca 2018 roku (art. 168) w dziedzinie *inżynierijno-technicznej* w dyscyplinie *automatyka, elektronika, elektrotechnika i technologie kosmiczne*.
- **Wnoszę** o przyjęcie rozprawy przez Radę Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Opolskiej.

