

INNOWACJE W NAUKACH INŻYNIERYJNO-TECHNICZNYCH

W Y B R A N E A S P E K T Y



Redakcja:
Izabela Mołdoch-Mendoń
Kinga Kalbarczyk

Lublin 2023

Innowacje w naukach inżynieryjno- -technicznych – wybrane aspekty

Innowacje w naukach inżynieryjno- -technicznych – wybrane aspekty

Redakcja:
Izabela Mołdoch-Mendoń
Kinga Kalbarczyk

Lublin 2023

**Wydawnictwo Naukowe TYGIEL składa serdeczne podziękowania
zespółowi Recenzentów za zaangażowanie w dokonane recenzje
oraz merytoryczne wskazówki dla Autorów.**

Recenzentami niniejszej monografii byli:

- prof. dr hab. inż. Tadeusz Markowski
- prof. dr hab. inż. Beata Tryba
- dr hab. inż. Joanna Jankowska-Sandberg, prof. PK
- dr hab. inż. Donata Konopacka-Łyskawa
- dr hab. inż. Agnieszka Makara
- dr hab. Agnieszka Nowak-Brzezińska, prof. UŚ
- dr hab. inż. Zbigniew Paweł Omiotek
- dr hab. inż. Jarogniew Rykowski, prof. UEP
- dr hab. n. med. Agnieszka Woźniak-Kosek
- dr n. o zdr. Mariola Janiszewska
- dr inż. Radosław Patyk
- dr inż. Anna Skarbak-Żabkin

Wszystkie opublikowane rozdziały otrzymały pozytywne recenzje.

Skład i łamanie:
Monika Maciąg

Projekt okładki:
Marcin Szklarczyk

Korekta:
Ewelina Chodźko

© Copyright by Wydawnictwo Naukowe TYGIEL sp. z o.o.

ISBN 978-83-67104-79-1

Wydawca:
Wydawnictwo Naukowe TYGIEL sp. z o.o.
ul. Głowackiego 35/341, 20-060 Lublin
www.wydawnictwo-tygiel.pl

Spis treści

Leszek Majkut, Ryszard Olszewski Modelowanie i analiza drgań zazębienia w przekładni walcowej.....	7
Stanisław Pietrzyk, Piotr Palimąka Wychwytywanie dwutlenku węgla w solach stopionych – prototyp reaktora CCMS	18
Justyna Jońca Czujniki gazu na bazie heterozłączy tlenków metali zaprojektowanych w nanoskali	40
Regina Przywara, Mateusz Przywara, Wojciech Zapała Właściwości adsorpcyjne wybranych polarnych faz stacjonarnych.....	61
Marcin Krajewski Badania wpływu imperfekcji na rozkład obciążeń stabilizujących stężoną kratownicę	74
Andrzej Opaliński, Krzysztof Regulski, Jakub Swadźba, Piotr Sitkowski, Paweł Wąsowicz, Agnieszka Kwietniewska-Śmietana Predykcja badań laboratoryjnych w optymalizacji kontroli jakości i obciążenia pracy analizatorów	85
Mirosław Wieczorek Wpływ zmiany sztywności płyty nad słupem na wartość sił wewnętrznych w płycie żelbetowej konstrukcji płytowo-słupowej	95
Olena Stryhuniwska, Oksana Dzyundzja, Ramis Rasulov Harmonizacja podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ	107
Paweł Urzenitzok Czy sztuczna inteligencja powinna mieć formę jak najbardziej zbliżoną do człowieka?	120
Ewelina Jurczak Wybrane kierunki rozwoju Sztucznej Inteligencji okiem socjologa i... z perspektywy biznesu.....	131
Mariusz Wiśniewski, Stanisław Deniziak, Daniel Kaczmarski, Paweł Pięta, Piotr Picheta, Aleksandra Królak, Mirosław Borkowski System ekspercki wspomaganie decyzji lekarza oparty o zbiór reguł medycznych	157

Karolina Kuligowska, Maciej Stanusch, Marek Koniew Zastosowanie automatycznego rozpoznawania mowy w transkrypcji wywiadów medycznych – porównanie silników ASR dla języka polskiego	173
Karolina Pianowska, Joanna Kluczka, Grzegorz Benke, Karolina Goc, Joanna Malarz, Michał Ochmański, Katarzyna Leszczyńska-Sejda Zastosowanie ekstrakcji cieczowej w technologii odzysku i rozdziału metali z grupy platynowców	184
Indeks Autorów.....	201

Modelowanie i analiza drgań ząbienia w przekładni walcowej

1. Wprowadzenie

Przekładnie zębate stanowią powszechny element łańcucha kinematycznego układu przeniesienia mocy od silnika do odbiornika energii. Powinny się więc charakteryzować wysoką trwałością i niezawodnością działania. Podstawową ich funkcją jest zmiana przenoszonego momentu i prędkości obrotowej przy zachowaniu minimalizacji zaburzeń dynamicznych i maksymalnej sprawności. Zatem z punktu widzenia eksploatacji istotna jest ocena aktywności drganiowej przekładni, która między innymi pozwala ocenić jej bieżący stan techniczny.

Ze względu na kluczową rolę w łańcuchu napędowym, uszkodzenie w obrębie współpracujących ze sobą zębów przekładni prowadzić może do uszkodzenia kolejnych par lub oddziaływanie na nie, którego następstwem może być w szczególności trudna do przewidzenia w skutkach awaria. Wskazaniem jest więc zastosowanie modelu dynamicznego, który pozwoli na analizę pracy przekładni w różnych warunkach obciążenia. Zgodnie z danymi, które zawarto w pracy [1] około 60% awarii przekładni zębatych wywołane jest uszkodzeniem kół zębatych. W literaturze [2-4] autorzy wyróżniają następujące ich rodzaje:

- wykruszenie wierzchołka zęba;
- pęknięcie zęba u podstawy;
- jego częściowe lub całkowite wyłamanie;
- pęknięcie wieńca zębatego;
- zmęczeniowe wykruszenia warstwy wierzchniej współpracujących powierzchni roboczych zębów (*pitting*);
- zatarcie współpracujących powierzchni;
- inne.

Występujące uszkodzenia i ich skutki przyczyniły się do podjęcia prac, których celem jest opracowanie metod pozwalających na wykrywanie uszkodzeń kół zębatych takich jak między innymi pęknięcie u podstawy zęba w możliwie jego wczesnej fazie [5-8]. Autorzy stawiają sobie za cel możliwie szybkie wykrycie rozwijającego się uszkodzenia, aby zapobiec następstwom awarii. W tym celu korzystają z zaawansowanych metod numerycznych zaimplementowanych w obliczeniowych pakietach inżynierskich, które oparte są między innymi na:

- metodzie elementów skończonych;
- metodzie sztywnych elementów skończonych;
- metodzie różnic skończonych;
- metodzie objętości skończonych.

¹ majkut@agh.edu.pl, Katedra Mechaniki i Wibroakustyki, Wydział Inżynierii Mechanicznej i Robotyki, AGH Akademia Górniczo-Hutnicza.

² olszewsk@agh.edu.pl, Katedra Mechaniki i Wibroakustyki, Wydział Inżynierii Mechanicznej i Robotyki, AGH Akademia Górniczo-Hutnicza.

Niewątpliwie metody te dają bardzo dobre rezultaty z uwzględnieniem kinematyki zazębienia, ale sposób implementacji w zamkniętych pakietach obliczeniowych utrudnia zastosowanie ich do identyfikacji rozwijającego się uszkodzenia i oceny przydatności przekładni do dalszej eksploatacji z określeniem czasu warunkowo bezpiecznego jej użytkowania do chwili wyłączenia. W przypadku konieczności prowadzenia obserwacji przekładni w celu identyfikacji wystąpienia rodzaju uszkodzenia wraz z określeniem jego zakresu, celowe staje się skorzystanie z modelu analitycznego, który uzupełniony o parametry możliwe do wyznaczenia np. we współpracy z metodą elementów skończonych, pozwoli możliwie szybko dokonać oceny stanu zazębienia nadzorowanej przekładni.

Problem diagnostyki przekładni jest problemem aktualnym i nadal otwartym. W pracy [9] dotyczącej zagadnienia uwzględnienia wielokrotnych pęknięć autorzy zastosowali dynamiczny model zazębienia przekładni, w którym sztywność zazębienia wyznaczana jest między innymi metodą elementów skończonych z uwzględnieniem scenariuszy zazębienia zębów uszkodzonych z zębami nieuszkodzonymi, wykorzystując do tego metody dekompozycji.

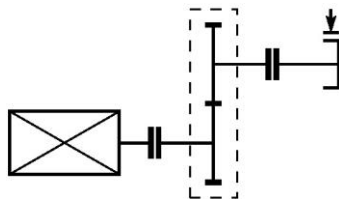
Praca dotyczy możliwości uwzględnienia w modelu dynamicznym zmiany sztywności zęba w zależności od stopnia jego uszkodzenia (pęknięcia u podstawy). Skutkiem tego rodzaju uszkodzenia jest zmiana współczynnika sprężystości uszkodzonego zęba. Sztywność zazębienia została oszacowana z wykorzystaniem Metody Elementów Skończonych zaimplementowanej w pakiecie *Salome-Meca*, które jako jądro obliczeniowe bazuje na *Code_Aster* na licencji GNU GPL. Do oceny stanu technicznego zazębienia przekładni wykorzystano metodę dekompozycji empirycznej pozwalającą na wczesną detekcję i kwantyfikację uszkodzenia zęba u jego podstawy.

2. Model przekładni

Jako obiekt badań wykorzystano stanowisko badawczo-demonstracyjne DMG-1A wyposażone w przekładnię zębatą walcową o przełożeniu $i = 1$, w którego skład wchodzi:

- silnik komutatorowy z reduktorem zębatym CELMA Cieszyn;
- badana przekładnia zębata o prostych zębach;
- pompa zębata wraz z zaworem przelewowym stanowiących obciążenie.

Schemat kinematyczny stanowiska badawczego pokazano na rysunku 1.



Rysunek 1. Schemat kinematyczny stanowiska badawczego DMG-1A, opracowanie własne

Parametry kół zębatych, które poddano badaniu:

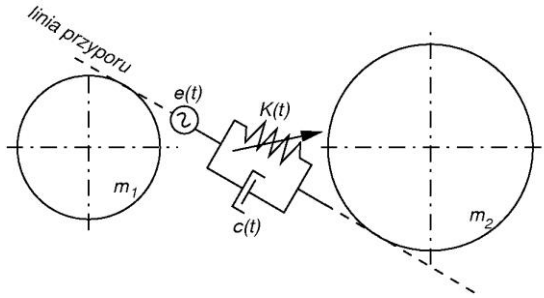
- liczba zębów $z = 29$;
- moduł $m_n = 3,175$ mm ($DP = 8$);
- czynna wysokość zęba $h = 6,16$ mm;
- zęby proste wiórkowane;
- klasa dokładności wykonania 7.

2.1. Model zazębienia

Model matematyczny zazębienia współpracujących ze sobą kół zębatych oparto na modelu fizycznym [10] układu drgającego o jednym stopniu swobody przedstawionego na rysunku 2, wyrażony jest następującą zależnością:

$$m \frac{d^2 x(t)}{dt^2} + c(t) \frac{dx(t)}{dt} + F_d = F_s \quad (1)$$

gdzie: $x(t)$ wyraża przemieszczenie wzdłuż linii przyporu, m jest proporcjonalne do masy zazębionych kół (model zredukowany jest do dyskretnego układu drgającego o jednym stopniu swobody), $c(t)$ jest zmiennym tłumieniem, F_d jest siłą dynamiczną, zaś F_s wyraża obciążenie statyczne.



Rysunek 2. Model fizyczny zazębienia przekładni, opracowanie własne na podstawie [10]

Siła dynamiczna F_d może być modelowana z wykorzystaniem nieliniowej funkcji luzu BL , która wyrażona jest iloczynem współczynnika sztywności zazębienia $K(t)$ i wartości odkształcenia zębów w sposób następujący:

$$F_d = \begin{cases} K(t)[x(t) + e(t)] & \text{dla } x + e \geq 0 \\ 0 & \text{dla } -BL \leq x + e < 0 \\ K(t)[x(t) + e(t) + BL] & \text{dla } x + e \leq -BL \end{cases} \quad (2)$$

W modelu siły dynamicznej uwzględniono funkcję przenoszenia błędów $e(t)$, która opisuje błąd powstający przy obrocie koła zębatego jako niedokładność wykonania każdego z zębów wchodzących w przypór. Może ona być opisana szeregiem różnych zależności. Jako jeden z możliwych sposobów jej określenia przyjęto postać sumy szeregu Fouriera:

$$e = a_1 \cos(2\pi f_{01} t) + a_2 \cos(2\pi f_{02} t) + \sum_{i=1}^n b_i \cos(2\pi f_m t) \quad (3)$$

gdzie: f_m jest częstotliwością zazębienia, zaś f_{01} i f_{02} odpowiednio częstotliwością związaną z obrotem koła zębatego i częstotliwością obrotu zębownika.

Zmienna sztywność w funkcji czasu, a tym samym zazębienia, stanowi podstawę modelu i ma bezpośredni wpływ na jego jakość. Sztywność zazębienia może być modelowana w funkcji obrotu i wyrażona jako:

$$K(t) = \begin{cases} K_a(1+\alpha) & \text{dla } (k-1)T \leq t < (C_r - 2 + k)T \\ K_a(1-\alpha) & \text{dla } (C_r - 2 + k)T \leq t < kT \end{cases} \quad (4)$$

dla $k = 1, 2, \dots, n_t$, gdzie n_t oznacza liczbę zębów. K_a wyraża średnią sztywność, α określa zmianę średniej wartości sztywności, zaś $T = 1/f_m$ jest okresem zazębienia, natomiast C_r oznacza wskaźnik przyporu czołowego współpracujących kół.

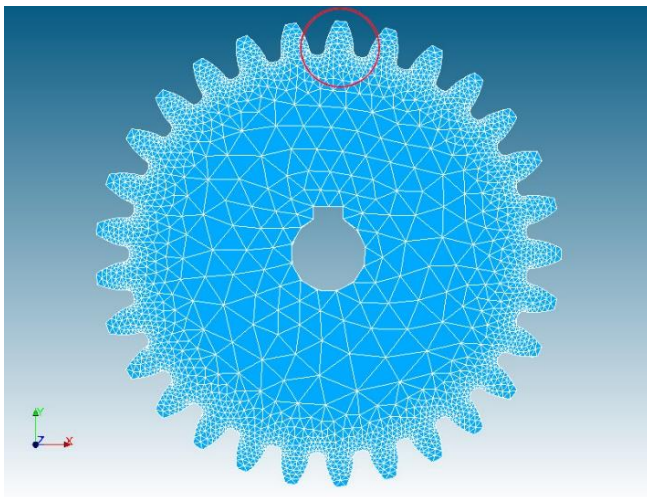
Współczynnik tłumienia $c(t)$ może być wyrażony przez tłumienie krytyczne zależne od zmiennej sztywności $K(t)$:

$$c(t) = C_l \sqrt{K(t)} \quad (5)$$

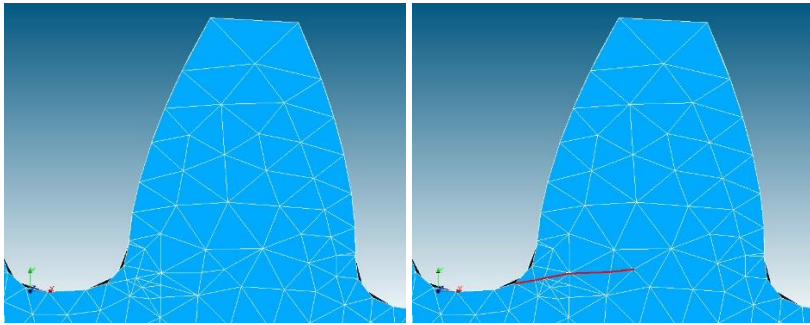
gdzie: C_l jest współczynnikiem tłumienia krytycznym zależnym od momentu bezwładności i sztywności $K(t)$ określonej równaniem (4).

2.2. Sztywność zazębienia z uwzględnieniem pęknięcia u podstawy zęba

Ogromnym problemem modelu opisanego równaniem (1) jest poprawne określenie sztywności zazębienia w funkcji kąta obrotu wału. Wskaźnik przyporu C_r , pozwala na oszacowanie sztywności zazębienia $K(t)$ w funkcji czasu na podstawie jej wartości średniej w przypadku braku uszkodzenia. Problem ulega skomplikowaniu w przypadku pojawienia się pęknięcia. Wówczas zmiana punktu przyłożenia siły przyporu do boku zęba silnie wpływa na zmianę funkcji sztywności zazębienia [11, 12]. Na potrzeby modelu obliczeniowego skorzystano z możliwości analizy Metodą Elementów Skończonych współpracujących ze sobą kół zębatach w przypadku braku uszkodzenia, jak i uszkodzenia polegającego na pęknięciu podstawy zęba w zależności od jego głębokości. Na rysunku 3 przedstawiono koło modelowane zębate, fragment ujęty w okręgu pokazano na rysunku 4, stopa zęba bez uszkodzenia (a) oraz z uszkodzeniem (b) wraz z linią modelowanego uszkodzenia.



Rysunek 3. Siatka węzłów i elementów modelowanych kół zębatach, opracowanie własne



rys. a

rys. b

Rysunek 4. Fragment w obrębie zęba, bez uszkodzenia – rys. a, z modelowanym uszkodzeniem – rys. b, opracowanie własne

Zarys pęknięcia podstawy zęba uzyskano, korzystając z możliwości techniki powielania (dublowania) węzłów bezpośrednio na siatce elementów skończonych i w ten sposób rozłączenia sąsiadujących ze sobą elementów. Na podstawie obliczeń MES w zależności od długości modelowanej szczeliny pęknięcia przyjęto procentowe zmniejszenie wartości Ka występującej w zależności (4) jako 1%, 3%, 5%, 10%, 15%, 20% względem sztywności zęba nieuszkodzonego.

3. Metoda dekompozycji empirycznej

Metoda dekompozycji empirycznej została opracowana przez Huanga i innych w 1998 roku [13] i z powodzeniem stosowana jest w zagadnieniach diagnostyki przekładni zębatych [14-16]. Opiera się ona na prostym założeniu, że każdy sygnał zawiera różne proste wewnętrzne oscylacje (mody). Każdy liniowy bądź nieliniowy mod będzie miał taką samą ilość ekstremów i przejść przez zero. Pomiedzy kolejnymi przejściami przez zero występuje tylko jedno ekstremum. Każdy mod powinien być niezależny od pozostałych. W ten sposób każdy sygnał może być rozłożony na pewną ilość niezależnych funkcji IMF (*Intrinsic Mode Functions*) spełniających warunki:

1. w całym zbiorze danych, ilość ekstremów i ilość przejść przez zero musi być sobie równa bądź różnić się tylko o 1,
2. w każdym punkcie średnia wartość obwiedni zdefiniowanej przez maksima lokalne i obwiedni zdefiniowanej przez minima lokalne jest równa 0.

Funkcja IMF reprezentuje tzw. prosty mod oscylacyjny, który odpowiada prostej funkcji harmonicznej. Na podstawie tej definicji, każdy sygnał może być rozłożony w następujący sposób:

1. Wyznaczenie lokalnych ekstremów i aproksymacja lokalnych maksimów za pomocą krzywej sklepanej 3 stopnia, uzyskując górną obwiednię sygnału.
2. Powtórzenie powyższej operacji dla lokalnych minimów pozwoli uzyskać dolną obwiednię sygnału. Wszystkie dane powinny się zawierać pomiędzy dolną a górną obwiednią.
3. Wyznaczenie średniej górnej i dolnej obwiedni oznaczonej jako m_1 oraz różnicy pomiędzy sygnałem $x(t)$ i m_1 jest pierwszym komponentem h_1

$$x(t) - m_1 = h_1 \tag{6}$$

idealnie byłoby, gdyby h_1 był funkcją IMF, wtedy byłby to pierwszy komponent sygnału $x(t)$.

4. Jeśli h_1 nie jest funkcją IMF, jest traktowane jako oryginalny sygnał i kroki 1., 2. i 3. są powtarzane

$$h_1 - m_{11} = h_{11} \quad (7)$$

Powtarzając operację tzw. przesiewania k -razy otrzymujemy h_{1k} , który to komponent jest funkcją IMF, czyli:

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (8)$$

Co jest oznaczone jako:

$$c_1 = h_{1k} \quad (9)$$

5. Komponent c_1 jest usuwany z sygnału $x(t)$. Otrzymujemy:

$$r_1 = x(t) - c_1 \quad (10)$$

gdzie r_1 jest traktowany jako oryginalny sygnał i cały proces jest powtarzany. Otrzymujemy drugą funkcję IMF c_2 sygnału $x(t)$. Cały proces powtarza się n -krotnie, otrzymując n funkcji IMF sygnału $x(t)$. Wtedy:

$$\begin{aligned} r_1 - c_2 &= r_2 \\ &\vdots \\ r_{n-1} - c_n &= r_n \end{aligned} \quad (11)$$

Proces dekompozycji kończy się, kiedy r_n będzie funkcją monotoniczną, z której nie da się już uzyskać funkcji IMF. Na podstawie równań (6) i (11) można napisać:

$$x(t) = \sum_{j=1}^n c_j + r_n \quad (12)$$

Funkcje IMF c_1, c_2, \dots, c_n zawierają różne zakresy pasm częstotliwości od wysokiej do niskiej. Składowe częstotliwościowe zawarte w każdym paśmie są różne i zmieniają się wraz ze zmianą w czasie sygnału $x(t)$ a r_n reprezentuje główny trend sygnału.

4. Lokalne widmo Hilberta

Dla każdej funkcji IMF można wyznaczyć jej transformatę Hilberta (13):

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(t')}{t-t'} dt' \quad (13)$$

oraz tzw. sygnał analityczny (14):

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t)e^{j\Phi_i(t)} \quad (14)$$

w którym:

$$a_i(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]} \quad (15)$$

$$\Phi_i(t) = \arctg \frac{H[c_i(t)]}{c_i(t)} \quad (16)$$

Na podstawie równania (16) wyznaczyć można pewną częstość, zwaną częstością chwilową:

$$\omega_i(t) = \frac{d\Phi_i(t)}{dt} \quad (17)$$

Po wyznaczeniu transformaty Hilberta każdej funkcji IMF oryginalny sygnał można opisać jako część rzeczywistą w poniższej postaci:

$$x(t) = \operatorname{Re} \sum_{i=1}^n a_i(t) e^{j\Phi_i(t)} = \operatorname{Re} \sum_{i=1}^n a_i(t) e^{j \int \omega_i(t) dt} \quad (18)$$

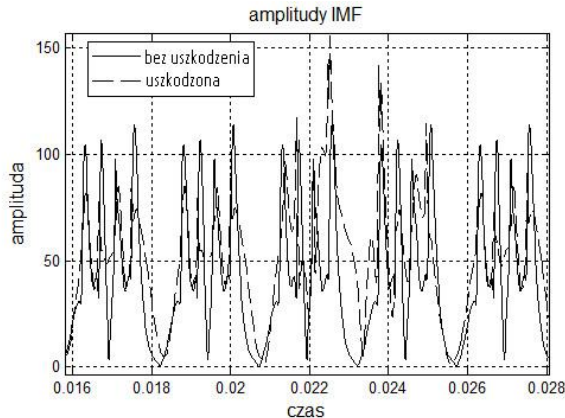
Równanie (18) przedstawia zarówno amplitudę, jak i częstotliwość w funkcji czasu. Ten czasowo-częstotliwościowy rozkład amplitudy jest określanany jako widmo Hilberta $H(\omega, t)$:

$$H(\omega, t) = \operatorname{Re} \sum_{i=1}^n a_i(t) e^{j \int \omega_i(t) dt} \quad (19)$$

5. Diagnostyka zazębienia przekładni

5.1. Dekompozycja empiryczna

Zarejestrowane sygnały drgań dla przekładni z nieuszkodzonym zazębieniem i z wprowadzonym uszkodzeniem poddano dekompozycji empirycznej EMD.



Rysunek 5. Przebieg funkcji IMF dla przekładni bez uszkodzenia i uszkodzonej, opracowanie własne

Na rysunku 5 pokazano zmiany amplitud funkcji IMF z dekompozycji sygnałów zarejestrowanych dla przekładni zdrowej (linia ciągła) i uszkodzonej (linia przerywana). Jak można zauważyć, zmiany wartości amplitud funkcji IMF są tak małe, że tylko na tej podstawie nie jest możliwe stwierdzenie wystąpienia uszkodzenia przekładni. Zmiany tych częstości są niezauważalne przy analizie sygnału obejmującego jedynie jeden obrót wału przekładni. Stąd do analizy należy wybrać sygnał, zarejestrowany co najmniej dla kilkunastu obrotów wału przekładni, dzięki czemu możliwe jest badanie zarówno okresowości, jak i intensywności zmian sygnału spowodowanych uszkodzeniem zęba.

5.2. Analiza cepstralna

Jednym z narzędzi do badania okresowości widma sygnału jest analiza cepstralna. Cepstrum rzeczywiście definiuje się jako (20):

$$C(\tau) = F^{-1}[\ln(F[x(t)])] \quad (20)$$

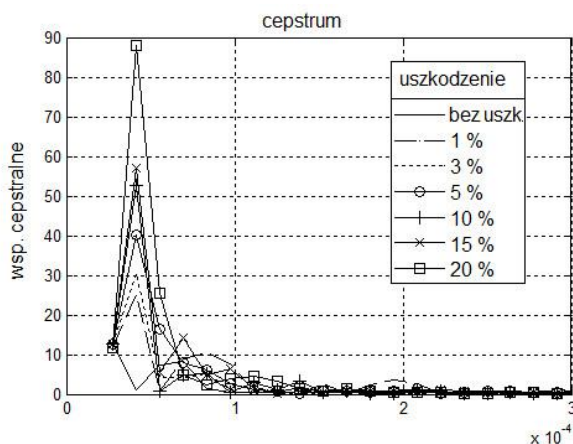
gdzie:

F – oznacza transformatę Fouriera sygnału $x(t)$,

F^{-1} – odwrotna transformata Fouriera.

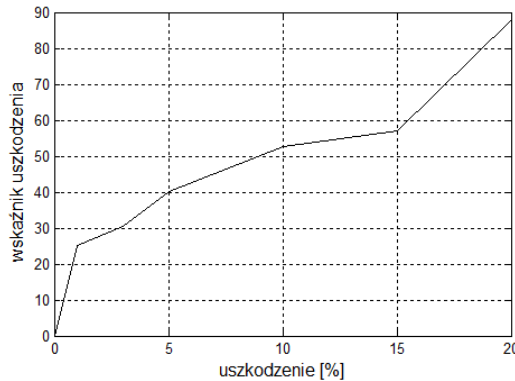
Czyli cepstrum wyznacza się jako odwrotną transformatę Fouriera logarytmu widma sygnału $x(t)$.

Zastosowanie analizy cepstralnej pozwala na uzyskanie informacji czy zazębienie w przekładni jest uszkodzone, natomiast „intensywność” tzn. wielkość uszkodzenia badać można, analizując wartość współczynników cepstralnych. Przebiegi transformaty cepstralnej częstości chwilowej (17) wyznaczonych dla zazębienia przekładni o różnym stopniu uszkodzenia pokazano na rysunku 6.



Rysunek 6. Współczynniki cepstralne częstości chwilowych, opracowanie własne

Autorzy proponują, by jako wskaźnik uszkodzenia przyjąć wartość współczynnika cepstralnego wyznaczonego dla drugiej próbki transformaty cepstralnej. Wg autorów ten współczynnik odpowiada okresowości związanej z zazębieniem z uszkodzonym zębem. Z tego powodu współczynnik jest równy zero dla przekładni nieuszkodzonej.



Rysunek 7. Proponowany wskaźnik uszkodzenia, opracowanie własne

Na rysunku 7 pokazano przebieg proponowanego skalarnego wskaźnika uszkodzenia w funkcji stopnia uszkodzenia zęba (głębokości pęknięcia u podstawy zęba) wyrażonego przez zmianę współczynnika sprężystości zęba, który ma wpływ na sztywność zazębienia.

6. Podsumowanie

Jak wspomniano we wstępie, przekładnie zębate są istotnym elementem napędowego łańcucha kinematycznego. Ich uszkodzenie prowadzić może do uszkodzenia kolejnych elementów łańcucha kinematycznego, a co za tym idzie do wypadku. W pracy zastosowano model dynamiczny zazębienia, w którym sztywność oszacowano w oparciu o Metodę Elementów Skończonych w zależności od długości szczeliny pęknięcia u podstawy zęba. Zaproponowano diagnostykę przekładni zębatych opartą o metodę dekompozycji empirycznej oraz analizę cepstralną tzw. częstości chwilowych sygnału drganiowego, który jest sygnałem rejestrowanym.

Bezpośrednia analiza amplitud funkcji IMF, będących elementami dekompozycji empirycznej nie pozwala na stwierdzenie, czy dana przekładnia jest uszkodzona. Dopiero zastosowanie analizy cepstralnej pozwala na uzyskanie informacji o uszkodzeniu przekładni, natomiast „intensywność” tzn. wielkość uszkodzenia badać można, analizując wartość współczynników cepstralnych.

Analizując przebiegi współczynników cepstralnych, na rysunku 6 zauważyć można wyraźne ich zróżnicowanie w zależności od wielkości uszkodzenia zęba zarówno w porównaniu z przebiegami o innym stopniu uszkodzenia, jak i w porównaniu z przebiegiem dla przekładni nieuszkodzonej. Zauważalny jest również wyraźny trend proponowanego wskaźnika w funkcji wzrostu uszkodzenia.

Współczynnik cepstralny wyznaczony dla drugiej próbki transformaty cepstralnej ze względu na korelację z okresowością związaną zazębieniem z uszkodzonym zębem, jest bardzo dobrym wskaźnikiem uszkodzenia.

Zastosowanie nowoczesnego narzędzia analizy sygnałów, jakim jest dekompozycja empiryczna w powiązaniu z transformacją cepstralną daje dobre efekty w diagnozowaniu stanu przekładni zębatej. Współczynnik cepstralny ze względu na swoje własności diagnostyczne w powiązaniu z zastosowaniem techniki obliczeniowej opartej np. na metodzie elementów skończonych dodatkowo pozwala oszacować fizyczną wielkość uszkodzenia bez odwoływania się do tzw. uszkodzeń małych, średnich i znacznych.

Praca finansowana z subwencji badawczej nr 16.16.130.942.

Literatura

1. James Li C., Limmer J.D., *Model – Based Condition Index for Tracking Gear Wear and Fatigue Damage*, Wear, 241(1), 2000, s. 26-32.
2. Dąbrowski Z., Radkowski S., Wilk A., *Dynamika przekładni zębatych – Badania i symulacja w projektowaniu eksploatacyjnie zorientowanym*, Wydawnictwo i Zakład Poligrafii Instytutu Technologii Eksploatacji w Radomiu, Warszawa-Katowice-Radom 2000.
3. Müller L., *Przekładnie zębate – obliczenia wytrzymałościowe*, WNT, Warszawa 1972.
4. Staszewski W.J., Tomlinson G.R., *Local Tooth Fault Detection in Gearboxes Using a Moving Window*, MSSP, 1997, 11(3), s. 331-350.
5. Wojnar G., Łazarz B., *Wykrywanie różnych uszkodzeń kół przekładni zębatych*, XXXIII Ogólnopolskie Sympozjum DIAGNOSTYKA MASZYN, Węgierska Górka 6.03-11.03.2006, s. 41-58.
6. Lohrengel A., Kramarczyk W., Kruk R., Twardoch K., Wieczorek A., *Modelowanie zjawisk dynamicznych w przekładniach zębatych z wykorzystaniem Metody Elementów Szybywnych*, Górnictwo i Geologia Wydawnictwo Politechniki Śląskiej, 6(11), 2011.
7. Cioch W., Dąbrowski D., *Methods synchronized by operating cycle for condition monitoring of a planetary gearbox in non-stationary operations*, 5th International congress on Technical diagnostics 2012. Kraków, 3rd-5th September 2012.
8. Guo Y., Keller J., Parker R., *Dynamic Analysis of Wind Turbine Planetary Gears Using an Extended Harmonic Balance Approach*, International Conference on Noise and vibration Engineering Leuven, Belgium 17-19.09.2012.
9. Yang X., Wei D., Zuo M.J., Tian Z., *Analysis of vibration signals and detection for multiple tooth cracks in spur gearboxes*, Elsevier, Mechanical Systems and Signal Processing, 185, 2023, article 109780.
10. Bosch M., *Das dynamische Verhalten von Stirnradgetriebe unter besonderer Berücksichtigung der Verzahnungsgenauigkeit*, Industrie-Anzeiger, Teil 1, No. 102, 1965, Teil 2, No. 88, 1966.
11. Jaśkiewicz Z., Wąsiewski A., *Przekładnie walcowe. Geometria, wytrzymałość, dokładność wykonania*, Wydawnictwa Komunikacji i Łączności, Warszawa 1992.
12. Bartelmus W., *Computer-aided Multistage Gearbox Diagnostics Inference by Computer Simulation*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2002.
13. Huang N.E., Shen Z., Long S.R., Wu N.C., Shih H.H., Zheng Q., Yen N.C., Tung C.C., Liu H.H., *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, Proc. of the Royal Society of London A 454, 1998, s. 903-995.
14. Lei Y., He Z., Zi, Y., *Application of the EEMD method to rotor fault diagnosis*, Mechanical Systems and Signal Processing, 23, 2009, s. 1327-1338.
15. Liu B., Riemenschneider S., Xu Y., *Gearbox fault detection using empirical mode decomposition and Hilbert spectrum*, Mech. Sys. and Sig. Proc., 20, 2006, s. 718-734.
16. Yang W., Tavner P.J., *Empirical mode decomposition, an adaptive approach for interpreting shaft vibratory signals of large rotating machinery*, Journal of Sound and Vibration, 321, 2009, s. 1144-1170.

Modelowanie i analiza drgań zażębenia w przekładni walcowej

Streszczenie

W pracy przedstawiono detekcję uszkodzeń zębów przekładni zębatej z wykorzystaniem dynamicznego modelu zażębenia, w którym sztywność zęba wyznaczana jest Metodą Elementów Skończonych oraz zastosowanie dekompozycji empirycznej i transformacji cepstralnej. Dekompozycja empiryczna jest metodą analizy sy-

gnału poprzez rozłożenie go na zbiór ortogonalnych tzw. *Intrinsic Mode Functions*. Pozwala również na wyznaczenie tzw. częstości chwilowych. Zastosowanie analizy cepstralnej do „sygnału” częstości chwilowych pozwoliło na uzyskanie informacji czy przekładnia jest uszkodzona, natomiast „intensywność” tzn. wielkość uszkodzenia badano, analizując wartość współczynników cepstralnych.

Słowa kluczowe: diagnostyka, koło zębate, dekompozycja empiryczna

Modeling and vibration analysis of spur gearbox meshing

Abstract

This paper presents detection of damage to the gear teeth using a dynamic meshing model, in which the stiffness of the tooth is determined by the Finite Element Method, and the use of Empirical Mode Decomposition and cepstral analysis. EMD is a way to decompose a signal into so-called Intrinsic Mode Functions (IMF), and obtain instantaneous frequency data. Cepstral analysis of instantaneous frequency was used for obtaining information about gear health, cepstral coefficients allowed to determine the fault intensity.

Keywords: diagnostics, spur gear, empirical decomposition

Wychwytywanie dwutlenku węgla w solach stopionych – prototyp reaktora CCMS

1. Wprowadzenie

Dwutlenek węgla jest jednym z gazów odpowiedzialnych za efekt cieplarniany i zwiększenie średniej temperatury na Ziemi. Jego stężenie w atmosferze jest najwyższe od co najmniej 2 mln lat. Głównymi źródłami emisji CO₂ do atmosfery są sektory: energetyczne, transportowe oraz przemysłowe, bazujące na spalaniu paliw kopalnych. Na chwilę obecną wiadomo, iż paliwa kopalne pozostaną wiodącym źródłem energii zarówno dla produkcji energii, jak i transportu. Stąd też bardzo istotne jest rozwinięcie efektywnych metod wychwytu oraz sekwestracji CO₂ ze strumienia spalin. Ograniczenie jego emisji do atmosfery jest kluczowe w najbliższych latach, w celu zapobieżenia nieodwracalnym zmianom klimatu na Ziemi. Nowy cel Unii Europejskiej to obniżenie emisji CO₂ o 55% do 2030 roku, co oznacza jego redukcję w elektroenergetyce o 71-76% oraz w przemyśle o 22-25% [1].

Główne kierunki związane z ograniczeniem emisji CO₂ polegają na:

- zmniejszeniu emisji poprzez znajdowanie zamienników energetycznych, które nie opierają się na węglowodorach, a tym samym nie emitują CO₂ (tzw. odnawialne źródła energii – OZE);
- wychwytywaniu CO₂ u źródła jego powstawania, kompresji i geologicznej sekwestracji (ang. *Carbon Capture and Sequestration* – CCS);
- wychwytywaniu i utylizacji CO₂ (ang. *Carbon Capture and Utylization* – CCU), czyli wyłapywaniu z gazów spalinowych, a następnie wykorzystywaniu w miejsce węglowodorów w procesach energetycznych i przemysłowych.

Niekwestionowaną zaletą technologii CCS/CCU jest to, iż mogą zostać wdrożone do już istniejących instalacji – nie wymagają specjalnej przebudowy działających układów.

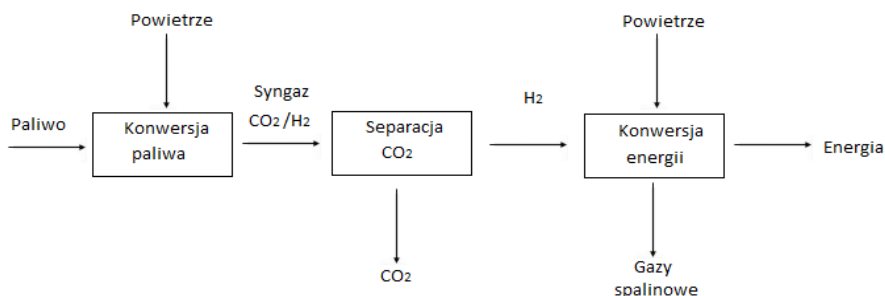
2. Technologie wychwytu dwutlenku węgla

Metody ograniczania emisji CO₂ przez jego wychwytywanie ze spalin energetycznych można podzielić na następujące technologie [2]: przed procesem spalania (ang. *pre-combustion*), po procesie spalania (ang. *post-combustion*) spalanie w atmosferze tlenowej (ang. *oxy-fuel combustion*), a także spalanie w pętli chemicznej (ang. *chemical looping combustion* – CLC).

W technologii wychwytu przed procesem spalania, paliwo jest częściowo utleniane i w wyniku tego procesu powstaje gaz syntezowy (tlenek węgla i wodór), tzw. syngaz, który jest przekształcany w CO₂ i wodór. W rezultacie CO₂ jest łatwy do separacji ze strumienia gazu syntezowego, a wodór może być użyty jako paliwo w turbinie gazowej. Schemat blokowy wychwytu metodą *pre-combustion* przedstawia rysunek 1.

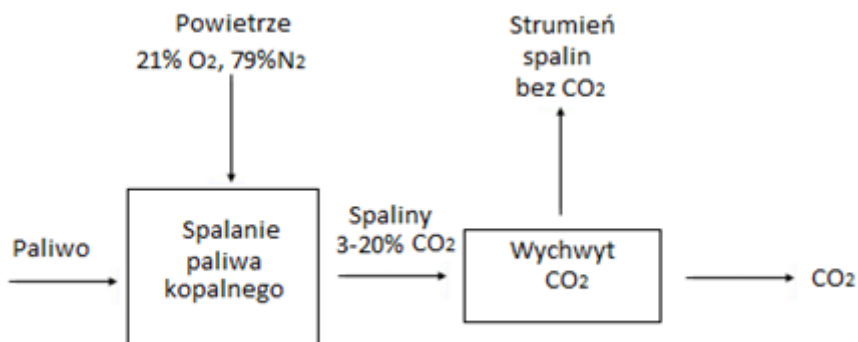
¹ pietstan@agh.edu.pl, Katedra Fizykochemii i Metalurgii Metali Nieżelaznych, Wydział Metali Nieżelaznych, Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie, www.agh.edu.pl.

² palimaka@agh.edu.pl, Katedra Fizykochemii i Metalurgii Metali Nieżelaznych, Wydział Metali Nieżelaznych, Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie, www.agh.edu.pl.



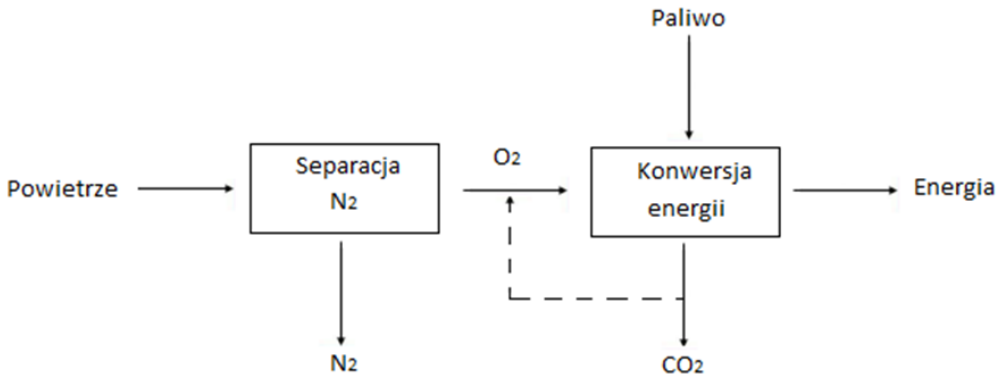
Rysunek 1. Schemat wychwytywania CO₂ przed procesem spalania, opracowanie własne na podstawie [3]

Z kolei wychwytywanie CO₂ po procesie spalania, polega na jego usuwaniu ze spalin wychodzących z kotła lub turbiny gazowej (rys. 2). Jest on następnie desorbowany, osuszany, sprężany i transportowany do miejsca magazynowania. Do zalet tej technologii zalicza się fakt, że najłatwiej jest ją dołączyć do istniejącej instalacji, bez istotnej zmiany konfiguracji technologii spalania w danym zakładzie energetycznym. Technologia *post-combustion* może być dołączona do źródeł emisji CO₂, pochodzących z przemysłu cementowego czy metalurgicznego.

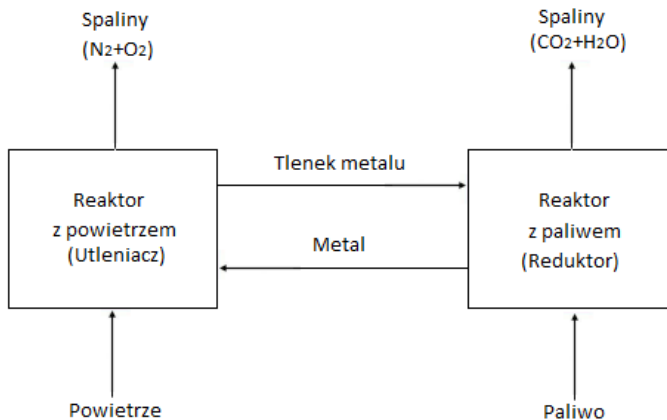


Rysunek 2. Schemat wychwytywania CO₂ po procesie spalania, opracowanie własne na podstawie [4]

W przypadku spalania tlenowo-paliwowego paliwo jest spalane w czystym tlenie zamiast w powietrzu. Podstawowe elementy technologii *oxy-combustion* to separacja azotu z powietrza, spalanie paliwa, a następnie zawracanie CO₂ i jego doczyszczanie oraz wychwytywanie (rys. 3). Aby ograniczyć temperaturę płomienia do poziomu typowego dla tradycyjnego spalania, schłodzone spaliny są zawracane do obiegu i wtryskiwane do komory spalania. Gazy spalinowe składają się głównie z CO₂ i pary wodnej, która skrapla się podczas chłodzenia. Rezultatem jest prawie czyste CO₂, który nadaje się do sprężania, transportu i magazynowania.

Rysunek 3. Schemat technologii *oxy-combustion*, opracowanie własne na podstawie [3]

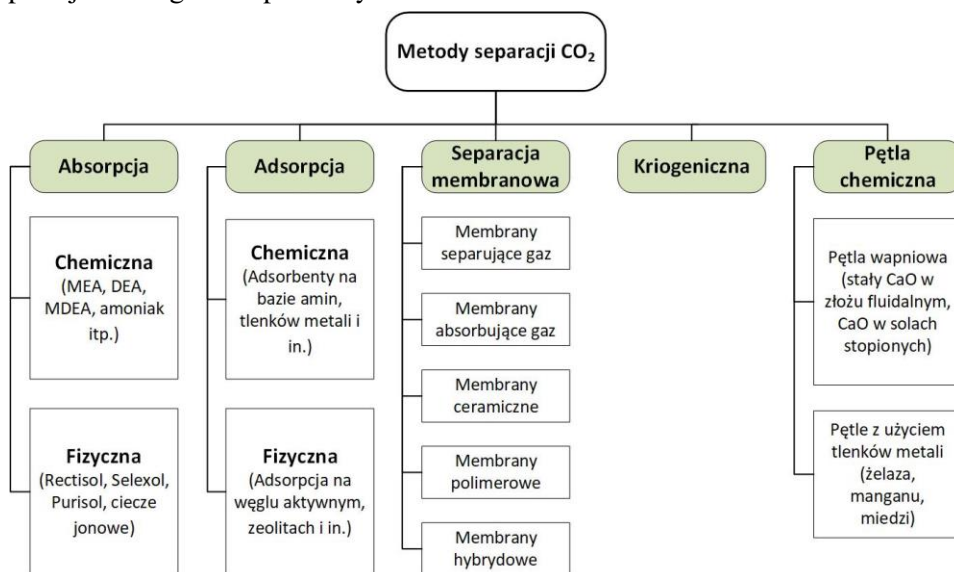
Technologia spalania w pętli chemicznej CLC opiera się na dwóch reaktorach ze złożem fluidalnym, których wypełnieniem jest tlenek metalu i metal (lub tlenki metalu na różnych stopniach jego utlenienia) jako nośniki tlenu. Jego zadaniem jest przenoszenie tlenu w celu prowadzenia reakcji spalania, po czym jest on przenoszony z powietrzem do reaktora, gdzie ponownie zachodzi utlenianie przed skierowaniem nośnika do reaktora z paliwem, w celu zamknięcia pętli. Schemat technologii CLC przedstawia rysunek 4. Dzięki temu, że powietrze i paliwo znajdują się w osobnych reaktorach, CO_2 pochodzący ze spalania paliwa nie zawiera N_2 , co znacznie ułatwia jego wychwyty. W reaktorze z paliwem zachodzi reakcja z tlenkiem metalu, w wyniku której powstają spaliny zawierające głównie CO_2 i parę wodną. Po skropleniu pary wodnej, CO_2 jest sprężany do postaci ciekłej. Zredukowany tlenek metalu jest transportowany do reaktora z powietrzem, w celu ponownego utlenienia. W wyniku reakcji z tlenem, powstaje tlenek metalu. Pozostały po tym etapie gaz, zawiera azot i pozostały O_2 , jest uwalniany do atmosfery. Tlenek metalu jest kierowany do reaktora z paliwem, gdzie oddaje tlen do procesu spalania, a sam ulega redukcji do postaci metalicznej. W procesie CLC są stosowane metale, takie jak: Fe, Ni, Cu, Co i Cd.



Rysunek 4. Schemat technologii CLC, opracowanie własne na podstawie [3]

3. Metody i techniki separacji dwutlenku węgla ze spalin

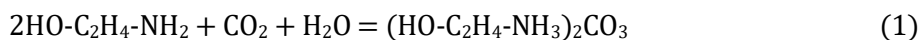
Spośród technologii omówionych w podrozdziale 2, technologia *post-combustion* jest najlepiej poznana, najszerzej stosowana i jest obecnie wykorzystywana również w innych (poza produkcją energii z paliw kopalnych) zastosowaniach przemysłowych, chociaż na mniejszą skalę aniżeli są oczekiwania w skali komercyjnej. Wychwytywanie po spalaniu jest najbardziej popularne w badaniach i innowacjach, ponieważ elektrownie na paliwa kopalne można zmodernizować tak, aby obejmowały technologię CCS w tej konfiguracji. Na rysunku 5 zestawiono wybrane, najczęściej stosowane metody separacji CO₂ z gazów spalinowych.



Rysunek 5. Metody separacji CO₂ stosowane w technologii wychwytu *post-combustion*, opracowanie własne na podstawie [5]

Absorpcja chemiczna jest metodą wychwytywania CO₂, która obecnie jest najbardziej znana i najczęściej stosowana. Zasada procesu opiera się na reakcji między dwutlenkiem węgla a rozpuszczalnikiem chemicznym. Technologia mycia aminowego powstała ponad 60 lat temu w przemyśle naftowym i chemicznym, m.in. do usuwania siarkowodoru i CO₂ ze strumieni gazu. Z komercyjnego punktu widzenia jest to najbardziej ugruntowana metoda z dostępnych technik, chociaż praktyczne doświadczenia dotyczą głównie strumieni gazu, które są chemicznie redukujące (przeciwieństwo utleniającego środowiska strumienia gazów spalinowych).

Do najpopularniejszych rozpuszczalników zalicza się roztwory wodne alkanoloamin, takie jak: monoetanolamina (MEA), dietanolamina (DEA) lub metyldietanolamina (MDEA) [5]. Proces realizowany jest w dwóch etapach. W pierwszym, gazy będące składnikami spalin reagują z rozpuszczalnikiem w absorberze, wychwytyjąc CO₂. Następnie, nasycony roztwór kierowany jest do drugiej kolumny, gdzie w podwyższonych temperaturach (383K), następuje uwolnienie dwutlenku węgla. Reakcja towarzysząca absorpcji CO₂ jest odwracalna i może być opisana równaniem (1) [2]:



Zregenerowany roztwór jest zawracany do kolumny absorbera i rozpoczyna się kolejny cykl procesu. Strumień dwutlenku węgla uwolniony w desorberze jest przekazywany do sprężania i składowania lub wykorzystania (utyliczacji). Dzięki zastosowaniu 30% roztworów MEA i MDEA, można uzyskać zarówno wysoką wydajność procesu, jak i wysoki stopień czystości dwutlenku węgla. Szacuje się, że metodą aminową można wychwycić ze spalin ok. 85-95% dwutlenku węgla o czystości 99,95%. Obecnie oprócz konwencjonalnych rozpuszczalników (aminy-MEA, DEA, amoniak, piperazyna), stosuje się mieszanki rozpuszczalników, które posiadają wyższe właściwości absorpcyjne niż pojedyncze składniki. Do wad metody aminowej, zalicza się najczęściej korozyjność środowiska reakcji oraz degradację aminy spowodowaną obecnością tlenu w gazach [2].

Alternatywą dla amin są ciecze jonowe, czyli niskotopliwe ($T_{\text{top}} < 373 \text{ K}$) sole składające się tylko z kationów i anionów. Sole takie można łączyć (mieszać), uzyskując zdecydowanie lepsze właściwości absorpcyjne. Ciecze jonowe mogą absorbować CO_2 fizycznie lub chemicznie, w zależności od zastosowanego w układzie ciśnienia.

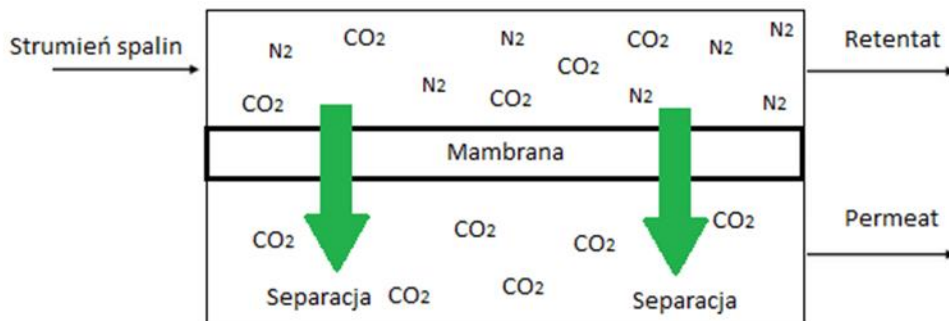
Metoda absorpcji fizycznej, opiera się na zastosowaniu chemicznie obojętnego rozpuszczalnika, który wychwytuje CO_2 fizycznie. Absorpcja zachodzi w wodzie lub absorbentach organicznych (metanol, N-metylo-2-pirolidon, eter dimetylowy). Metoda osiąga najlepsze wyniki dla niskich temperatur (poniżej temperatury otoczenia) i wysokich ciśnień wydzielanego gazu, dlatego służy głównie do wychwytywania CO_2 w procesie zgazowania węgla. Wyróżnia się w niej procesy, z zastosowaniem rozpuszczalników, takich jak SelexolTM, RectisolTM, IfpexolTM, FluorTM, PurisolTM, SulfinolTM i MorphysorbTM [2, 6, 7].

Adsorpcja to proces wykorzystujący powierzchnię ciała stałego, na której CO_2 może być zaadsorbowany dzięki międzycząsteczkowym siłom adhezji lub chemisorpcji. Adsorpcja fizyczna zachodzi na powierzchni materiałów silnie porowatych, takich jak węgiel aktywny, tlenek glinu, tlenki metali czy zeolity [8]. Węgiel aktywny zawierający węgiel amorficzny jest względnie tanim materiałem, którego największą zaletą jest duża powierzchnia właściwa i możliwości modyfikacji porowatej struktury, jednak słaba energia wiązania z dwutlenkiem węgla powoduje, że materiał ten musi być wysoce mikroporowaty, aby nadawał się do wychwytywania CO_2 [9]. Z kolei zeolity (krystaliczne glinokrzemiany), mają dobre właściwości adsorpcyjne, lecz największą ich wadą jest hydrofilowość, co powoduje, że obecność wody silnie osłabia oddziaływanie ich powierzchni z CO_2 .

Innym rodzajem adsorbentów są hybrydowe porowate ciała, tzw. szkielety metaloorganiczne (MOF). Ich przestrzenna struktura sprawia, iż charakteryzują się wieloma zaletami, takimi jak: łatwość projektowania i syntezy, wysoka porowatość oraz możliwość dostosowania porów właściwościami do adsorbowanego gazu [10]. Innym niewęglowym materiałem posiadającym dobre zdolności adsorpcji CO_2 jest krzemionka. Do jej największych zalet należą: duża powierzchnia właściwa, jednorodny rozkład porów, stabilność mechaniczna i możliwość modyfikacji. Mezoporowate materiały krzemionkowe do wychwytywania CO_2 wykorzystują najczęściej aminy [9, 11].

Metody adsorpcyjne ze względu na sposób prowadzenia cykli adsorpcyjno-desorpcyjnych, dzieli się m.in. na: ciśnieniową (PSA, ang. *Pressure Swing Adsorption*), temperaturową (TSA, ang. *Temperature Swing Adsorption*), próżniową (VSA, ang. *Vacuum Swing Adsorption*), ciśnieniowo-temperaturową (PTSA) czy zmiennoelektryczną (ESA, ang. *Electrical Swing Adsorption*).

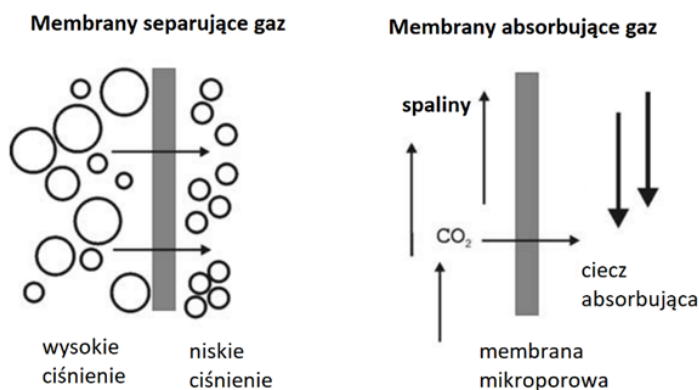
Separacja gazu na membranach polega na wykorzystaniu różnicy w oddziaływaniu materiału membrany z gazami. Zadaniem membrany jest rozdzielenie strumienia gazu na dwa: przenikający (*permeat*) i zatrzymany (*retentat*) (rys. 6). Siłą napędową w procesie permeacji gazów jest różnica ciśnień cząstkowych usuwanych zanieczyszczeń po obu stronach membrany [12].



Rysunek 6. Schemat procesu wychwytywania CO₂ przy zastosowaniu membrany, opracowanie własne na podstawie [13]

Membrany absorpcyjne są mikroporowatymi ciałami stałymi, dzięki którym możliwe jest selektywne przeniesienie danego składnika i zaabsorbowanie go np. w cieczy. Oznacza to, iż po jednej stronie membrany musi znajdować się ciecz absorbująca dany składnik, ze strumienia gazów znajdujących się po jej drugiej stronie (w przypadku CO₂ jako ciecz absorbującą można zastosować np. aminy). Membrany dobiera się tak, by charakteryzowały się wysoką selektywnością separowanych składników.

W przypadku wychwytywania dwutlenku węgla stosuje się membrany separujące gaz (ceramiczne, polimerowe oraz hybrydowe) oraz absorbujące gaz (rys. 7). Najczęściej, zastosowanie pojedynczej membrany, nie daje oczekiwanego efektu wysokiego stopnia separacji danego gazu ze strumienia głównego, dlatego też stosuje się układy wielostopniowe, zarówno ceramiczne, jak i polimerowe [14].



Rysunek 7. Schemat działania membran separujących i adsorbujących CO₂ [14]

Metoda kriogeniczna polega na sprężaniu i schładzaniu gazu do temperatury około -56°C , a następnie wydzielaniu CO₂ w postaci ciekłej. Selektywność procesu frakcjonowania kriogenicznego wynika z różnych warunków kondensacji poszczególnych

składników spalin. Separacja kriogeniczna jest szeroko rozpowszechniona i stosowana komercyjnie dla strumieni spalin, które już mają wysokie stężenie CO₂ (zwykle >90%) i wysokie ciśnienie (wychwytywanie przed spalaniem lub spalanie paliwa w tlenie), natomiast nie jest używana do bardziej rozcieńczonych strumieni CO₂ [14].

Separacja węgla metodą kriogeniczną należy do metod kosztownych, ale może być uzasadniona w połączeniu z innymi technikami, szczególnie jeżeli skroplony CO₂ miałby być kierowany do składowania w głębokich formacjach skalnych, a więc kiedy musiałby być sprężony. Metoda pozwala na uzyskanie CO₂ o czystości powyżej 99% [15].

Na potrzeby separacji CO₂ metodą kriogeniczną stosuje się dwa procesy. W jednym, gazowy CO₂ jest poddawany resublimacji, a następnie rozmrożeniu na powierzchni wymiennika ciepła, dzięki czemu uzyskiwany jest w postaci ciekłej [16]. W drugim procesie, dwutlenek węgla jest poddawany resublimacji na powierzchni odpowiednich materiałów, po czym jest uwalniany w postaci gazowej [17].

Spośród wielu dotychczas opracowanych technik absorpcji CO₂ z rozcieńczonych gazów, szczególnie jedna z najnowszych, wyróżnia się wysoką wydajnością i cyklicznością. Jest to metoda wychwytywania na sorbencie w postaci tlenku metalu rozpuszczonego w solach stopionych (ang. *Carbon Capture in Molten Salts* – CCMS). Metoda ta została opisana w następnym rozdziale.

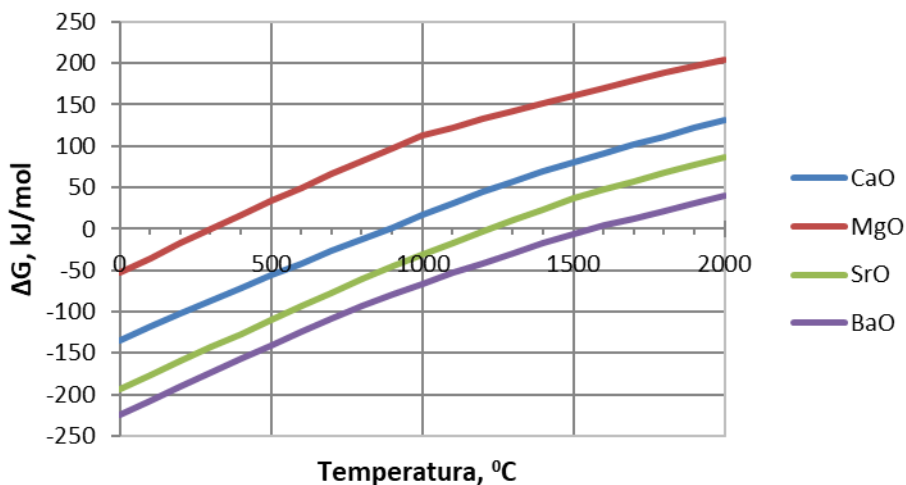
W Polsce w Akademii Górniczo Hutniczej od października 2019 do grudnia 2023 roku realizowany jest projekt o tytule: „Carbon Capture in Molten Salts – Prototype”, o akronimie „CCMS-P”. Projekt finansowany jest z Funduszy Norweskich w ramach programu Polsko-Norweskiej Współpracy Badawczej – POLNOR CCS 2019, za pośrednictwem Narodowego Centrum Badań i Rozwoju. Koordynatorem projektu jest Akademia Górniczo-Hutnicza, a partnerem zagranicznym jest Norwegian University of Life Science (NMBU) z Norwegii. Głównym celem projektu jest budowa prototypowego reaktora do sorpcji i desorpcji CO₂ w środowisku soli stopionych. W niniejszej pracy omówiono podstawy teoretyczne i zasadę działania tej metody oraz aktualne postępy prac nad zbudowaniem prototypowego reaktora do wychwytywania CO₂, realizowanych w ramach ww. projektu.

4. Separacja CO₂ w solach stopionych – CCMS

Metoda CCMS powstała na bazie metody pętli wapniowej. Pętla (ang. *calcium looping* – CaL) lub regeneracyjny cykl wapniowy (ang. *regenerative calcium cycle* – RCC) to technologia wychwytywania dwutlenku węgla drugiej generacji. Jest to najbardziej rozwinięta forma pętli węglanowej, w której metal (M) podlega odwracalnej reakcji między jego formą węglanową (MCO₃) a formą tlenkową (MO), w celu oddzielenia dwutlenku węgla od innych gazów pochodzących z elektrowni lub zakładu przemysłowego [18-21]. Metoda opiera się na reakcji (2), która przebiega odwracalnie:



Rysunek 8 przedstawia wykresy przebiegu zmian entalpii swobodnej Gibbsa (ΔG) dla reakcji tworzenia kilku węglanów.



Rysunek 8. Entalpia swobodna tworzenia węglanów, opracowanie własne na podstawie [22]

Jak widać dla ujemnych wartości entalpii swobodnej ($\Delta G < 0$), reakcja (2) przebiega w kierunku na prawo:



który jest procesem tworzenia węglanu, a z punktu widzenia CO_2 jest to jego wychwytywanie w procesie karbonizacji. W przypadku gdy $\Delta G > 0$, reakcja (2) przebiega w kierunku na lewo:



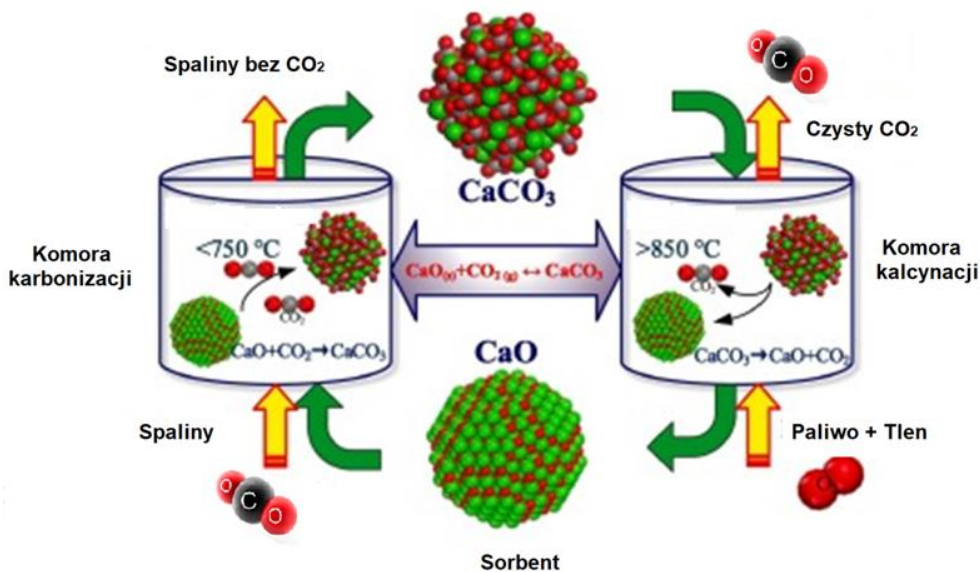
który z kolei jest procesem rozkładu węglanu, czyli kalcynacji, a dla CO_2 stanowi proces jego uwalniania do postaci gazowej.

Tlenek metalu MO pełni rolę sorbentu CO_2 . Reakcja (2) to proces sorpcji CO_2 , czyli wychwytywania dwutlenku węgla z gazów spalinowych. Natomiast reakcja (3) to proces desorpcji CO_2 , czyli uwalniania czystego, gazowego CO_2 , a zarazem regeneracji sorbentu. Wychwycony dwutlenek węgla można następnie przetransportować na miejsce składowania (CCS), zastosować do ulepszanego wydobycia ropy naftowej lub wykorzystać jako surowiec chemiczny (CCU).

Procesy karbonizacji tlenku/sorpcji CO_2 oraz kalcynacji węglanu/desorpcji CO_2 , przebiegają w oddzielnych reaktorach w tzw. warstwie fluidalnej. Jak wynika z rysunku 9, rolę sorbentu mogą pełnić różne tlenki metali, ale o praktycznym zastosowaniu decydują jak zwykle niskie koszty otrzymywania i łatwa dostępność na rynku, dlatego największe zastosowanie przypadło tlenkowi wapnia CaO (4):



a proces separacji dwutlenku CO_2 nazwano pętlą wapniową – CaL .

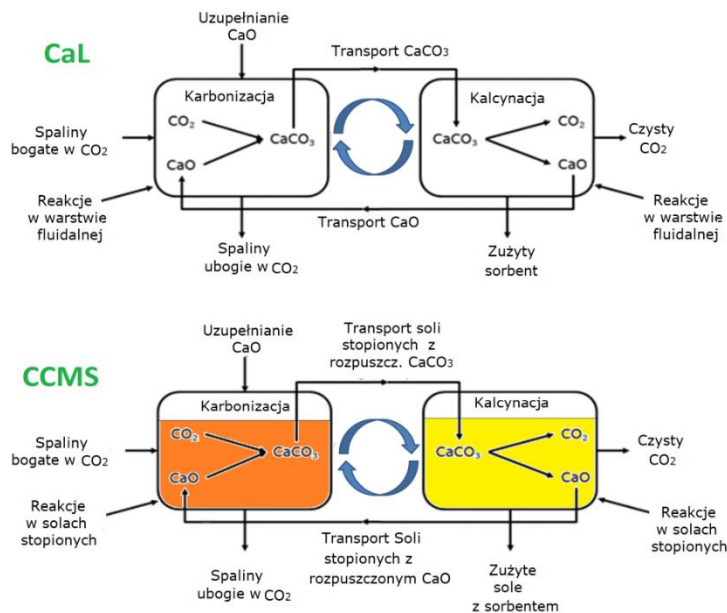


Rysunek 9. Zasada pętli wapniowej do wychwytu i separacji CO_2 ze spalin, opracowanie własne na podstawie [23]

CaL to proces cykliczny, składający się z dwóch etapów:

- karbonizacji: stały CaO jest usuwany z kalcynatora i podawany do karbonatora, gdzie jest schładzany do około 750°C i kontaktowany z gazami spalinowymi zawierającymi CO_2 i reagując tworzy CaCO_3 ;
- kalcynacji: stały CaCO_3 jest wprowadzany do kalcynatora, gdzie jest podgrzewany powyżej 850°C w celu termicznego rozkładu na gazowy CO_2 i stały CaO. Powstały prawie czysty strumień gazowego CO_2 jest następnie usuwany do przechowywania lub wykorzystania.

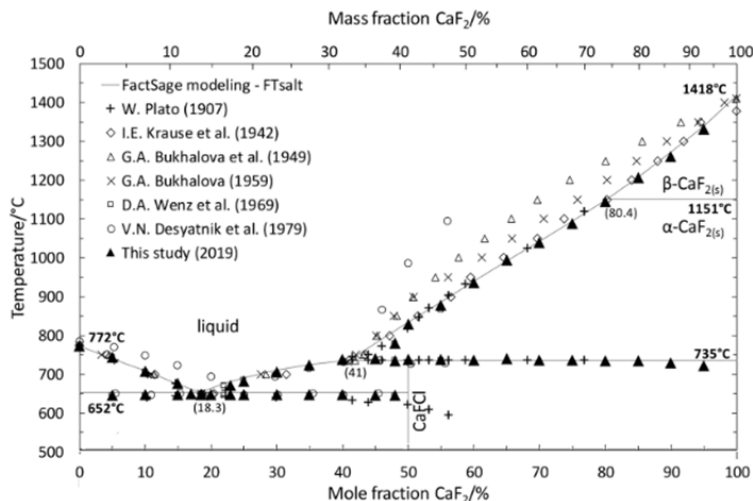
Niestety sorbent CaO ulega degradacji podczas cyklicznej pracy, maleje pojemność absorpcyjna CO_2 i wydajność procesu. Z tego powodu konieczne jest okresowe usunięcie części sorbentu z układu i zastąpienie go świeżym. W XXI wieku powstała nowa metoda wychwytywania dwutlenku węgla na podobnej jak w pętli wapniowej zasadzie sorpcyjno-desorpcyjnej, jednak przebiegająca w środowisku soli stopionych, stąd nazwa CCMS. W metodzie tej sorbent nieprzereagowany jako CaO i po przereagowaniu jako CaCO_3 , są rozpuszczone w roztworach soli stopionych (halogenki metali alkalicznych – Ca, Na, Li, głównie chlorki i fluorki). Graficzne porównanie metody pętli wapniowej i wychwytywania dwutlenku węgla w solach stopionych przedstawiono na rysunku 10.



Rysunek 10. Porównanie metody CaL z CCMS [opracowanie własne]

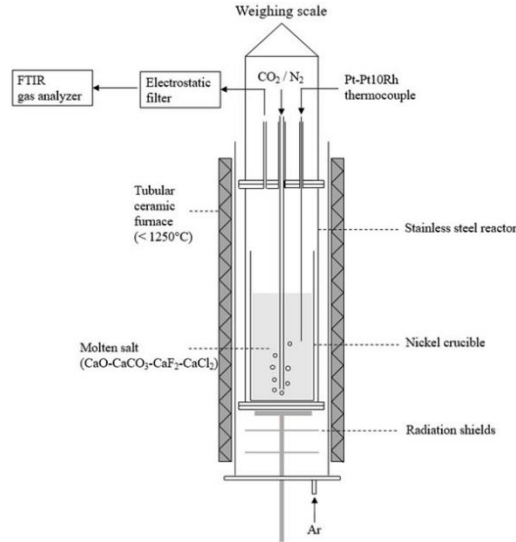
5. Opracowanie prototypu reaktora CCMS – założenie projektowe

Dotychczas metoda CCMS jest na etapie zaawansowanych badań w skali laboratoryjnej [24-29]. Roztwory soli, w których głównie prowadzono wychwytywanie, oparte były na układzie fazowym $\text{CaF}_2\text{-CaCl}_2$ o składach okołoeutektycznych (przemiana eutektyczna w 652°C dla 18,3% mol. CaF_2 , rys. 11) z dodatkami CaO/CaCO_3 [27].



Rysunek 11. Diagram fazowy $\text{CaCl}_2\text{-CaF}_2$ [30]

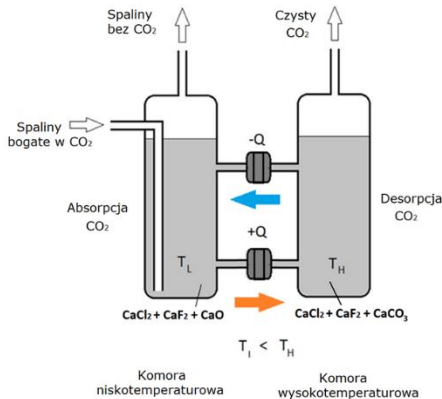
Badania cykli sorpcji i desorpcji przeprowadzono jedynie w reaktorze jednokomorowym (rys. 12), co odbiega od idei patentu [31] i podobieństwa do metody CaL, w której stosuje się dwie komory (karbonatora i kalcynatora).



Rysunek 12. Jednokomorowy reaktor do badania sorpcji i desorpcji CO_2 w solach stopionych [24-29]

Po zakończeniu etapu sorpcji, zamykano dopływ mieszanki gazowej z CO_2 , podgrzewano reaktor z zaadsorbowanym CO_2 pod postacią CaCO_3 do temperatury, w której termodynamicznie trwałą postacią jest CO_2 i przeprowadzano etap desorpcji, polegający na kalcynacji węglanu wapnia. Po usunięciu CO_2 z soli stopionych, ich roztwór zawierał rozpuszczony CaO , zregenerowany do kolejnego etapu wychwytu, który rozpoczynano, przepuszczając spaliny, po wcześniejszym ostudzeniu komory do niższej temperatury, w której fazą termodynamicznie trwałą jest CaCO_3 .

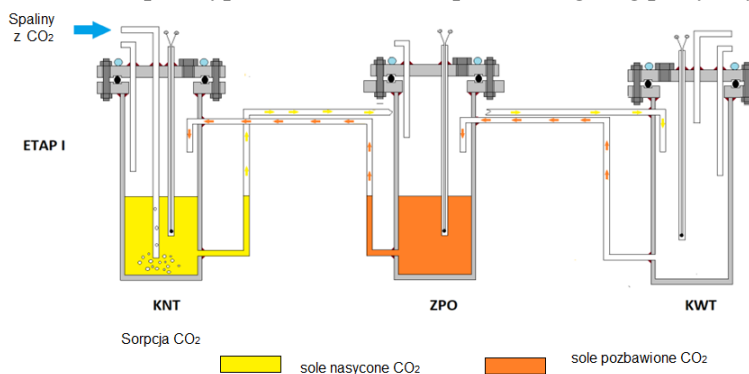
Idea pracy reaktora zaprezentowana w patencie [31], zakłada pracę reaktora w układzie dwukomorowym, który pozwala rozdzielić od siebie etapy sorpcji i desorpcji CO_2 . Dlatego w ramach realizacji prototypu reaktora założono, że reaktor dla sprawnego i wydajnego działania, powinien pracować w systemie dwukomorowym. Jednak przedstawiony w rozwiązaniu patentowym reaktor (rys. 13) przewiduje, że wymiana soli stopionych z rozpuszczonym CaO lub CaCO_3 , będzie się odbywać na zasadzie dyfuzji sześceniowej i temperaturowej, poprzez system dwóch kanałów łączących obie komory.



Rysunek 13. Zasada działania reaktora CCMS wg patentu [31]

Jeden z nich będzie schładzał do temperatury około 750°C przepływające sole stopione po procesie kalcynacji do komory niskotemperaturowej (sorpcyjnej). Natomiast drugi, będzie podgrzewał do temperatury około 950°C przepływające sole stopione po procesie karbonizacji do komory wysokotemperaturowej (desorpcyjnej). W patencie brak szczegółowych wyjaśnień, ale takie rozwiązanie ma umożliwić działanie reaktora w sposób ciągły. Istnieje jednak obawa, że utrudni to przeprowadzenie do końca sorpcji i desorpcji CO₂ oraz nie pozwoli w pełni wykorzystać zdolności sorpcyjnych sorbentu.

Rozważania nad prowadzeniem procesu i działaniem reaktora doprowadziły autorów niniejszej pracy do pomysłu, że reaktor powinien składać się trzech podstawowych elementów tj. komory sorpcyjnej (niskotemperaturowej – KNT), komory desorpcyjnej (wysokotemperaturowej – KWT) oraz tzw. zbiornika pośredniego (ZPO), który będzie pełnił funkcję rezerwuaru soli stopionych do procesu sorpcji. Takie rozwiązanie pozwoli skrócić czas trwania pojedynczego cyklu. Wyeliminowane zostanie oczekiwanie na zakończenie procesów sorpcji (wyczerpanie zdolności pochłaniania CO₂ w KNT) lub procesu desorpcji (brak wydzielania CO₂ z rozkładu CaCO₃ w KWT). Na rysunku 14 przedstawiono schemat prototypu reaktora CCMS opracowanego wg powyższych założeń.



Rysunek 14. Schemat prototypu reaktora CCMS [opracowanie własne]

Reaktor będzie pracował cyklicznie. Każdy cykl będzie składał się z trzech etapów, przedstawionych w tabeli 1.

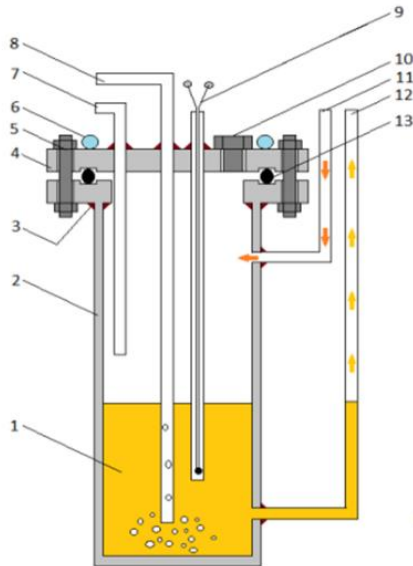
Tabela 1. Etapy pracy reaktora dwukomorowego CCMS ze zbiornikiem pośrednim przy wstępnym napełnieniu zbiorników ZWT i ZNT [opracowanie własne]

Zbiornik	Etap 0 (start)	Cykl 1			Cykl 2
		Etap I	Etap II	Etap III	Etap I
ZWT					
		Opróżnianie do ZPO	Napełnianie z KNT	Desorpcja CO ₂	Opróżnienie do ZPO
ZPO					
		Napełnianie z KWT		Opróżnianie do KNT	Napełnianie z KWT
ZNT					
		Sorpcja CO ₂	Opróżnianie do KWT	Napełnianie z ZPO	Sorpcja CO ₂
		sole nasycone CO ₂	sole pozbawione CO ₂		

Cykle i etapy przebiegają zatem w następujący sposób:

- Etap 0: początek procesu – napełnione KWT i KNT solami stopionymi bez CO₂;
- Etap I: napełnienie ZPO z KWT solami stopionymi bez CO₂, sorpcja CO₂ w KNT, start Etapu I, CYKLU 1;
- Etap II: napełnienie KWT z KNT solami stopionymi nasyconymi CO₂, zbiornik ZPO wypełniony solami stopionymi bez CO₂;
- Etap III: desorpcja CO₂ z KWT, napełnienie KNT z ZPO solami stopionymi bez CO₂, koniec CYKLU I;
- Etap I: napełnienie ZPO z KWT solami stopionymi bez CO₂, sorpcja CO₂ w KNT, start Etapu I CYKLU 2 itd.

Na rysunku 15 przedstawiono schemat komory sorpcyjnej KNT. Transport soli stopionych pomiędzy elementami reaktora będzie wymuszony ciśnieniem gazu obojętnego (N₂ lub Ar), który będzie wtłaczany do komory, ponad poziom soli stopionych, co wywoła przepompowanie soli do kolejnej komory czy zbiornika pomocniczego. Tego typu system działa analogicznie jak w przypadku windy gazowej. Komory KNT i KWT oraz zbiornik ZPO będą częściowo usytuowane we wnętrzu modułów grzewczych, pozwalających utrzymywać temperaturę soli stopionych na odpowiednim poziomie.



Rysunek 15. Koncepcja budowy komory niskotemperaturowej: 1 – Sole stopione, 2 – Komora reaktora, 3 – Miejsca spawania elementów reaktora, 4 – Pokrywa reaktora, 5 – Śruby, 6 – Chłodnica pokrywy reaktora, 7 – Rura do transportu soli pomiędzy komorami, 8 – Rura zasilająca komorę reaktora w mieszankę spalin do usuwania CO₂, 9 – Termoelement, 10 – Otwór do zasypu soli w postaci stałej, 11 – Rura zasilająca reaktor w sole stopione po desorpcji CO₂, 12 – Rura do usuwania z komory soli stopionej nasyconej CO₂, 13 – uszczelka [opracowanie własne]

6. Założenia konstrukcyjne prototypu reaktora CCMS

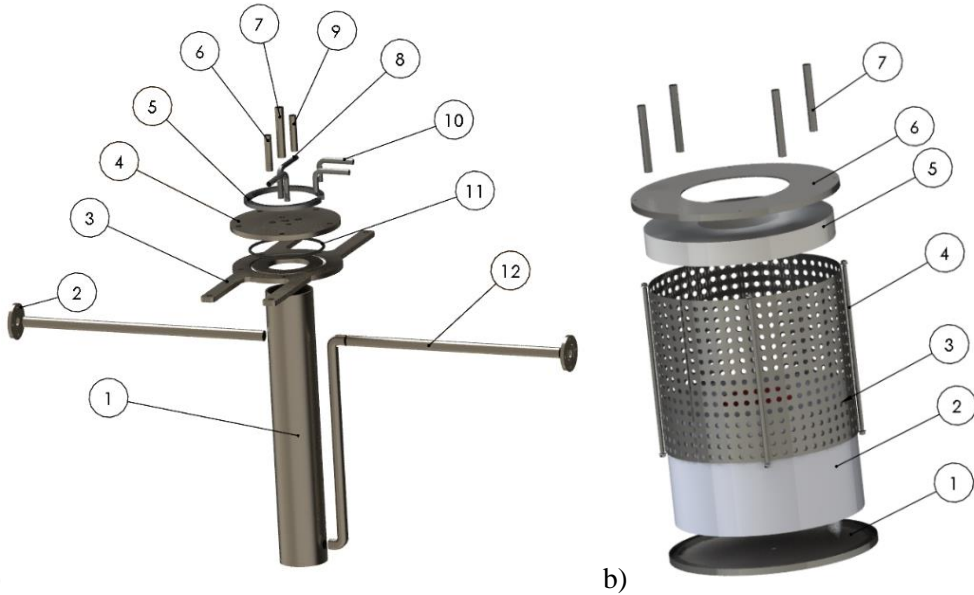
Na bazie wieloletniej praktyki autorów w badaniach soli stopionych oraz opracowanych koncepcji i założeń, zaprojektowano prototyp reaktora przy wykorzystaniu tzw. technicznego projektowania wspomaganego komputerowo CAD przy użyciu oprogramowania SOLIDWORKS. Dzięki temu możliwe było:

- przygotowanie projektu 3D za pomocą podzielenia na części i złożenia;
- utworzenie wizualizacji projektu;
- sprawdzenie kompatybilności poszczególnych elementów złożenia.

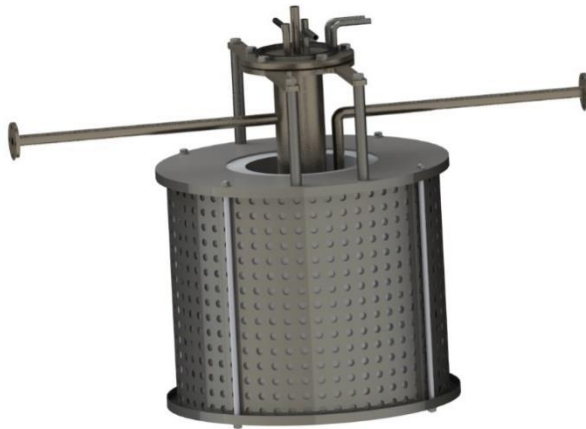
Projektowanie oparte było na cyfrowym modelowaniu geometrycznym, które miało na celu opracowanie zapisu dokumentacji konstrukcyjnej. Dzięki temu możliwe było utworzenie postaci konstrukcyjnej elementów zawierającej zbiór trzech informacji: geometrycznej, dynamicznej oraz niektórych cech technologicznych na przykład rodzaju materiału.

W oparciu o wytyczne i założenia koncepcyjne dokonano stworzenia modeli bryłowych poszczególnych elementów reaktora. Pierwszym elementem, który opracowano, była komora reaktora. Średnicę komory oraz planowany stopień jej wypełnienia solami stopionymi (wysokość słupa soli) dobrano, biorąc pod uwagę warunki eksperymentów w reaktorze jednokomorowym, przedstawionym w pracach [24-29]. Wysoki stopień wychwytu dwutlenku węgla osiągnięto w solach umieszczonych w tyglu pomiarowym o średnicy 50 mm i wysokości 300 mm, w którym po stopieniu mieszaniny stałych składników, wysokość słupa cieczy wynosiła 100 mm. Stwierdzono, że dwukrotnie wyższa wysokość słupa cieczy w projektowanym prototypie, będzie wystarczająca, aby zapewnić równie wysoki stopień absorpcji CO₂, jak w pracach [24-29]. Średnicę zewnętrzną komory reaktora ustalono wstępnie na 65 mm przy grubości ścianki 3 mm. Wewnętrzna średnica komory musi pozwolić na bezkolizyjne umieszczenie w niej: co najmniej jednego termoelementu, rury wprowadzającej do stopionych soli spaliny zawierające CO₂, otworu do dozowania i uzupełniania soli stopionych w komorze oraz wylotu gazów do analizatora. Wysokość reaktora ustalono na 350 mm, co zapewni możliwość uzyskania 200 mm słupa stopionej soli oraz wyjście komory reaktorowej ponad moduł grzejny i umieszczenie rur do transportu soli. System transportu stopionych soli pomiędzy KNT, KWT i ZPO, będzie zrealizowany rurkami o średnicy zewnętrznej 15 mm i grubości ścianki 1,5 mm. Komora reaktorowa będzie spawana do kołnierza, spoczywającego na czterech wspornikach umieszczonych w obudowie elementu grzejnego. Dzięki temu zamocowanie komory w piecu grzewczym będzie stabilne.

Na rysunku 16 przedstawiono model 3D niskotemperaturowej komory reaktora oraz jej elementów składowych. Rurka transportowa odprowadzająca sole stopione, powinna być umieszczona powyżej dna rury reaktorowej, z uwagi na możliwość kumulowania się w dolnej części komory stałych produktów reakcji karbonizacji w wyniku sedymentacji i niebezpieczeństwo jej zatkania.



a) b)
 Rysunek 16. Modele 3D elementów składowych: a) niskotemperaturowej komory reaktora: 1 – rura reaktora, 2 – kołnierz łączący, 3 – kołnierz reaktora, 4 – pokrywa reaktora, 5 – chłodzenie pokrywy, 6-9 – rurki wprowadzające/wyprowadzające gazy i tempopary, 10 – rurki wprowadzające wodę chłodzącą, 11 – uszczelka, 12 – rura transportowa, b) modułu grzewczego z obudową: 1, 6 – pokrywy modułu, 2 – moduł grzewczy z izolacją, 3 – osłona z blachy perforowanej, 4 – pręty ściągające, 5 – warstwa izolacji, 7 – pręty do mocowania reaktora [opracowanie własne]



Rysunek 17. Model 3D złożenia elementów konstrukcyjnych komory reaktora do wychwytywania dwutlenku węgla metoda CCMS [opracowanie własne]

Utrzymanie temperatury powyżej temperatury krzepnięcia soli stopionej jest ważne dla zapewnienia ciągłej pracy reaktora. Element grzewczy reaktora powinien posiadać odpowiednią izolację pozwalającą na możliwość utrzymania temperatury wewnątrz modułu do 750°C w przypadku komory niskotemperaturowej i zbiornika pośredniego oraz 950°C dla zbiornika wysokotemperaturowego, w sposób ciągły, z możliwością regulacji do 1°C.

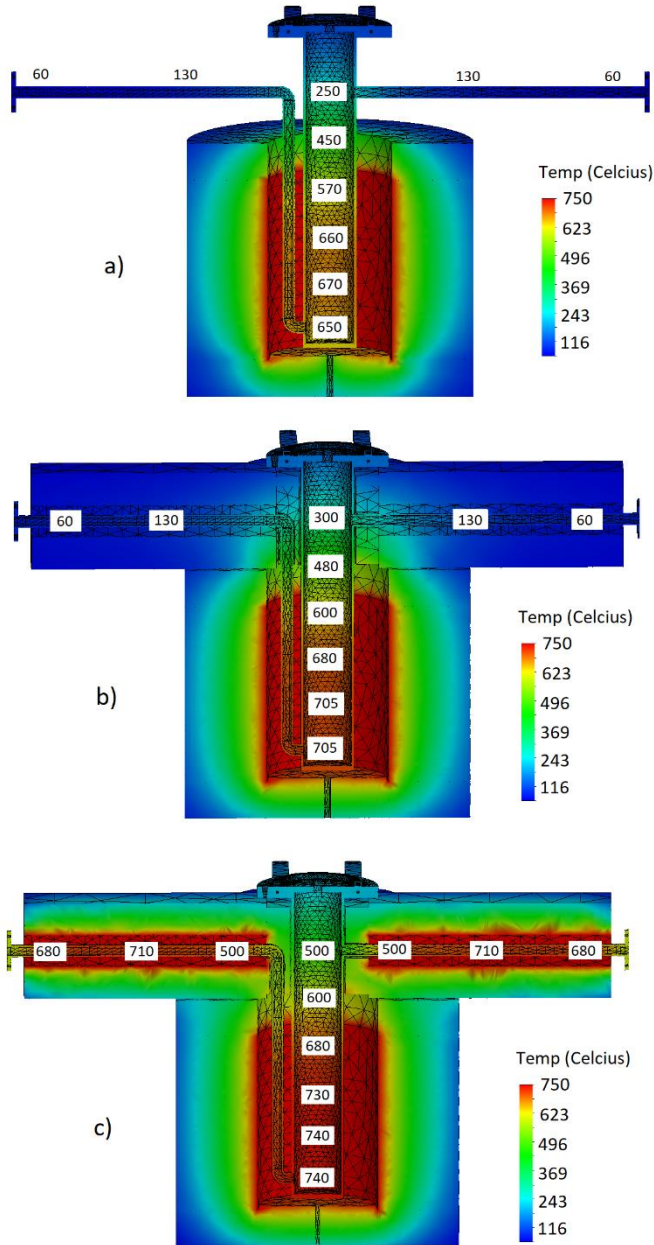
Połączenia rur transportowych z kolejnymi elementami konstrukcji reaktora powinny być zrealizowane za pomocą połączeń kołnierzowych, z zastosowaniem uszczeltek ceramicznych pomiędzy płaszczyznami przyłgowymi. Połączenie pomiędzy kołnierzem a rurką transportową powinno być połączeniem spawanym. Rura reaktorowa wraz z elementami opisanymi powyżej, powinna być zamocowana w obudowie elementu grzejnego, na wspornikach z możliwością skręcenia tych elementów śrubami.

Wewnętrzna średnica modułu grzejnego powinna wynosić 150 mm, a zewnętrzna 350 mm, natomiast wysokość 250 mm. Średnica zewnętrzna wynika ze stopnia zaizolowania elementu grzejnego, gdzie grubość warstwy izolacyjnej (próżniowo formowane włókna ceramiczne $\text{Al}_2\text{O}_3 + \text{SiO}_2$ z zawartością min. 40% Al_2O_3) wynosząca 100 mm, jest wystarczająca do zabezpieczenia ciągłej pracy pieca w temperaturze nawet 1100°C.

Jak widać na rysunku 17, część elementów rurociągów do transportu soli stopionych będzie znajdowała się poza strefą grzewczą. Istotnym zagadnieniem do zbadania było zatem określenie rozkładu pola temperatur elementów transportujących sole stopione, znajdujących się poza strefą grzewczą modułu głównego. W tym celu przeprowadzono badanie termiczne (analizę termiczną) w środowisku oprogramowania SOLIDWORKS. W oparciu o warunki generowania, przewodzenia, konwekcji i promieniowania ciepła, możliwe było uzyskanie mapy temperaturowej obiektów oraz gradientu temperatury w wybranych komponentach reaktora. Z uwagi na wysoki stopień skomplikowania elementów bryłowych modelu i ograniczoną moc obliczeniową komputerów, na których realizowano badania, do analizy uwzględniono tylko najważniejsze elementy komory reaktora. Badania realizowano przy następujących założeniach i warunkach brzegowych:

- siatka bryłowa z trójwymiarowych elementów czworościennych dla wszystkich elementów złożenia (108000 węzłów);
- solver FFEPlus (iteracyjny);
- typ rozwiązania: stan ustalony;
- brak oporów kontaktowych;
- obciążenia cieplne: temperatura, konwekcja, promieniowanie;
- temperatura elementów grzejnych modułów grzewczych 750°C.

Na rysunku 18 przedstawiono wyniki analiz termicznych dla trzech konfiguracji komory niskotemperaturowej. Analizując wyniki dla pierwszego wariantu (rys. 18a), można stwierdzić, iż reaktor zamocowany w module grzewczym osiągnie w stanie ustalonym temperaturę nie wyższą niż 670°C. Problemem jest również niska temperatura poziomych odcinków, spadająca do wartości nawet 130°C. Taka sytuacja w trakcie pracy prototypu reaktora, grozi wykrystalizowaniem soli podczas ich przetłaczania pomiędzy zbiornikami i zatkaniem rur transportowych. W wariantcie drugim (rys. 18b) doizolowano górną część reaktora oraz boczne rury transportowe. Do symulacji zastosowano formowane próżniowo włókno ceramiczne o współczynniku przewodzenia ciepła 0,21 W/m·K, takie samo jak w przypadku głównego modułu grzejnego.



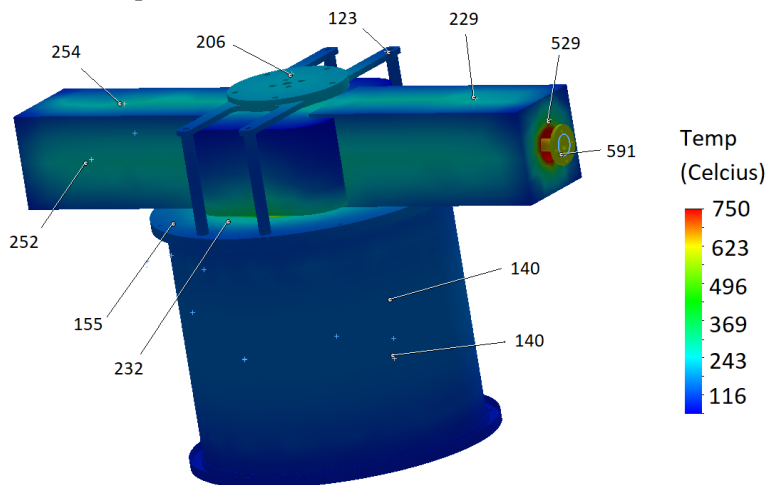
Rysunek 18. Wyniki analiz termicznych zilustrowane na przekroju poprzecznym komory niskotemperaturowej reaktora w konfiguracji: a) bez izolacji górnej części reaktora oraz rurek transportowych, b) z izolacją górnej części reaktora oraz rurek transportowych, c) z dodatkowymi poziomymi modułami grzewczymi na rurekach transportowych reaktora. Temperatura wszystkich modułów grzewczych: 750°C [opracowanie własne]

O ile warunki temperaturowe w środku reaktora uległy poprawie (temperatura wzrosła do 705°C), to na bocznych rurach transportowych nie zaszła zmiana i nadal grozi zamarznięciem soli w czasie ich transportu. Dlatego też, w celu zapewnienia odpo-

wiedniej temperatury, gwarantującej przepływ stopionej soli do reaktora i z reaktora, konieczne jest umieszczenie na poziomych rurach dodatkowych, niewielkich elementów grzejnych. Po uwzględnieniu ich w analizie, w postaci dodatkowego źródła ciepła (rys. 18c), warunki temperaturowe wyglądają korzystniej, aczkolwiek uzyskanie jednolitej temperatury na całej długości całej rury transportowej, może być trudne. Najniższe wartości temperatury (500°C) mogą budzić obawę (temperatura krzepnięcia mieszaniny o składzie eutektycznym $\text{CaCl}_2\text{-CaF}_2$ wynosi 652°C). Niemniej jednak należy wziąć pod uwagę wysoką pojemność cieplną soli, która prawdopodobnie nie pozwoli na ich krystalizację podczas szybkiego przepływu przez rury transportowe, nawet jeżeli temperatura na niektórych odcinkach spadnie do 500°C . Rozwiązaniem może być oczywiście podniesienie na czas przetłaczania temperatury rur transportowych, poprzez odpowiednie sterowanie dodatkowymi, poziomymi modułami grzewczymi.

Przeprowadzone analizy wykazały więc konieczność zainstalowania dodatkowych modułów grzejnych na poziomych rurach transportujących sole stopione, pomiędzy zbiornikami prototypu reaktora. Powinny one być rozkładane, w przeciwnym wypadku nie będzie możliwości wymiany rur transportowych.

Z kolei na rysunku 19, przedstawiono rozkład temperatury na zewnętrznych powierzchniach elementów zbiornika niskotemperaturowego, w którym moduły grzejne ustawione są na temperaturę 750°C .



Rysunek 19. Rozkład temperatury na zewnętrznych elementach komory niskotemperaturowej prototypu reaktora do wychwytywania CO_2 metodą CCMS [opracowanie własne]

Wyniki analizy termicznej wskazują, iż niektóre powierzchnie, szczególnie poziomych modułów grzewczych, mogą mieć dość wysoką temperaturę (do ok. 250°C). Stąd też wskazane jest zastosowanie dodatkowych, perforowanych blach osłonowych, mocowanych na dystansach, uniemożliwiających bezpośredni kontakt z modułem i bezpieczną obsługę prototypu reaktora.

Podobne symulacje, obliczenia konstrukcyjne i analizy cieplne przeprowadzono dla wszystkich podstawowych elementów reaktora (KNT, KWT i ZPO) wraz z rurociągami, elementami grzejnymi i chłodzącymi (pokryw górnych) oraz izolacjami. Finalny model całości prototypu wraz z konstrukcją nośną przedstawiono na rysunku 20.



Rysunek 20. Modelu 3D prototypu reaktora CCMS wraz z konstrukcją nośną [opracowanie własne]

7. Wnioski

Analizując opracowany na podstawie założeń koncepcyjnych projekt niskotemperaturowej komory reaktora do wychwytywania dwutlenku węgla w solach stopionych, można stwierdzić, iż:

- dobrane średnice komory reaktorowej oraz modułu grzewczego zapewniają bezproblemowe umieszczenie reaktora w module grzewczym, bez zagrożenia jego mechanicznego uszkodzenia (elementy grzejne modułów są zwykle nieosłonięte i niezabezpieczone);
- dobrana długość komory reaktorowej i jej umiejscowienie zapewnia utrzymanie odpowiednio wysokiej temperatury w reaktorze wypełnionym solami stopionymi, niezbędnej do prawidłowej pracy reaktora;
- w celu zapewnienia odpowiedniej temperatury, gwarantującej przepływ stopionych soli do i z komory reaktora, konieczne jest umieszczenie dodatkowych elementów grzejnych na poziomych odcinkach rurek transportowych, które znajdują się powyżej modułu grzejnego komory reaktora;
- pokrywa górna reaktora, przez którą wprowadzane są elementy kontrolno-pomiarowe oraz spaliny i gaz do windy gazowej powinna być chłodzona. Bez układu chłodzącego osiąga ona temperaturę powyżej 200°C;
- wszystkie elementy konstrukcyjne, mające bezpośredni kontakt z komorą reaktora muszą być łączone z nią za pomocą szczelnych i stabilnych połączeń: spawanych i wciskanych;
- w przypadku zamknięcia pieca pokrywą montowaną do kołnierza reaktora za pomocą śrub konieczne jest zastosowanie uszczelki, najlepiej odpornej na działanie wysokich temperatur.

Prototyp reaktora CCMS jest obecnie w fazie wykonawstwa na bazie wyników uzyskanych w niniejszej pracy. Rezultaty testów prototypu zostaną opublikowane.

Podziękowania

Niniejsze badania zostały zrealizowane przy i dzięki wsparciu finansowym Narodowego Centrum Badań i Rozwoju (NCBiR) w ramach konkursu POLNOR CCS 2019 (grant nr NOR/POLNORCCS/CCMS-P/0004/2019-00).

Literatura

1. Masson-Delmotte V., Zhai P., Pirani A., Connors S.L., Péan C., Berger S., Caud N., Chen Y., Goldfarb L., Gomis M.I., Huang M., Leitzell K., Lonnoy E., Matthews J.B.R., Maycock T.K., Waterfield T., Yelekci O., Yu R., Zhou B. (red.) *IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge 2020, doi:10.1017/9781009157896.
2. Kuroпка J., *Możliwości Ograniczania Emisji Dwutlenku Węgla ze Spalin Energetycznych*, s. 179-178, http://www.pzits.not.pl/docs/ksiazki/Pol_%202012/Kuroпка%20179-188.pdf [data dostępu: 21.12.2022].
3. Spigarelli B.P., Kawatara S.K., Li S.H., *Opportunities and challenges in carbon capture*, Journal of CO₂ Utilisation, 2013, s. 69-87.
4. Elwell L.C., Grant W.S., *Technology options for capturing CO₂*, Power, 150, 2006, s. 60-65.
5. Madejski P., Chmiel K., Subramanian N., Kus T., *Methods and Techniques for CO₂ Capture: Review of Potential Solutions and Applications in Modern Energy Technologies*, Energies, 15, 2022, s. 88-892.
6. Vega F., Cano M., Camino S., Fernandez L.M.G., Portillo E., Navarrete B., *Solvents for Carbon Dioxide Capture. In Carbon Dioxide Chemistry, Capture and Oil Recovery*, Intech Open, London 2018, ISBN 978-1-78923-575-3.
7. Sifat N.S., Haseli Y., *A critical review of CO₂ Capture Technologies and Prospects for Clean Power Generation*, Energies, 12, 2019.
8. Madejski P., Janda T., Modliński N., Nabagło D.A., *Combustion Process Optimization and Numerical Analysis for the Low Emission Operation of Pulverized Coal-Fired Boiler*, [w:] Kyprianidis K., Skvaril J. (red.), *Developments in Combustion Technology*, Intech Open Limited, London 2016.
9. Nord L., Bolland O., *Carbon Dioxide Emission Management in Power Generation*, Wiley-VCH Verlag GmbH & Co, Weinheim, Germany 2020.
10. Li J.-R., Ma Y., McCarthy M.C., Sculley J., Yu J., Jeong H.-K., Balbuena P.B., Zhou H.-C., *Carbon dioxide capture-related gas adsorption and separation in metal-organic frameworks*, Coordination Chemistry Reviews, 255, 2011, s. 1791-1823.
11. Osman A.I., Hefny M., Maksoud M.I.A.A., Elgarahy A.M., Rooney D.W., *Recent advances in carbon capture storage and utilization technologies: A review*, Environmental Chemistry Letters, 19, 2020, 19, 797-849.
12. Kotowicz J., Janusz K., *Sposoby redukcji emisji CO₂ z procesów energetycznych*, Rynek Energii, 1, 2007, s. 1-15.
13. Pires J.C.M., Martins F.G., Akvim-Ferrez M.C.M., Simones M., *Recent developments on carbon capture and storage: an overview*, Chemical Engineering Research and design, 89, 2011, s. 1446-1460.
14. Rautenbach R., *Procesy membranowe. Podstawy projektowani modułów i instalacji*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.
15. Mazurkiewicz M., Uliasz-Bocheńczyk A., Mokrzycki E., Piotrowski Z., Pomykała R., *Metody separacji i wychwytywania CO₂*, Polityka energetyczna, 8, 2005, s. 527-538.
16. Clodic D., Younes M., *A new Method for CO₂ capture: frosting CO₂ at atmospheric pressure*, [w:] Proc. 6th. Conf. On Greenhouse Gas Control Technologies, GHGT6, Kyoto, 2002, s. 155-160.

17. Tuinier M.J., Annaland S.M., Kuipers J.A.M., *A Novel process for cryogenic CO₂ Capture using dynamically operatedn packed beds- an experimental and numerical study*, International Journal of Greenhouse Gas Control 5, 2011, s. 694-701.
18. Dean C.C., Blamey J., Florin N.H., Al-Jeboori M.J., Fennell P.S., *The calcium looping cycle for CO₂ capture from power generation, cement manufacture and hydrogen production*, Chemical Engineering Research and Design, 89(6), 2011, s. 836-855.
19. Blamey J., Anthony E.J., Wang J., Fennell P.S., *The calcium looping cycle for large-scale CO₂ capture*, Progress in Energy and Combustion Science, 36(2), 2010, s. 260-279.
20. Shimizu T., Hiramata T., Hosoda H., Kitano K., Inagaki M., Tejima K., *A Twin Fluid-Bed Reactor for Removal of CO₂ from Combustion Processes*, Chemical Engineering Research and Design, 77(1), 1999, s. 62-68.
21. Blamey J., Anthony E.J., Wang J., Fennell P.S., *The Calcium Looping Cycle for large-scale CO₂ capture*, Progress in Energy and Combustion Science, 36(2), 2010, s. 260-279.
22. Program Komputerowy HSC Chemistry Ver.5.11, Outokumpu Research Oy, Pori Finland.
23. Wang Y. Feng X. Guo, *Atomistic mechanisms study of the carbonation reaction of CaO for high-temperature CO₂ capture*, Applied Surface Science Volume, 532(1), 2020, 147425.
24. Olsen E., Tomkute V., *Carbon capture in molten salts*, Energy Science and Engineering, 1, 2013, s. 144-150.
25. Tomkute V., Solheim A., Olsen E., *Investigation of high-temperature CO₂ capture by CaO in CaCl₂ molten salt*, Energy Fuels, 27, 2013, s. 5373-5379.
26. Tomkute V., Solheim A., Olsen E., *CO₂ capture by CaO in molten CaF₂-CaCl₂: Optimization of the process and cyclability of CO₂ capture*, Energy Fuels, 28m, 2014, s. 5345-5353.
27. Tomkute V., Solheim A., Sakirzanovas S., *Reactivity of CaO with CO₂ in molten CaF₂-NaF: formation and decomposition of carbonates*, Energy Science and Engineering, 4, 2016, s. 205-216.
28. Nygård H.S., Tomkute V., Olsen E., *Kinetics of CO₂ Absorption by Calcium Looping in Molten Halide Salts*, Energy Procedia, 114. 2017, s. 250-258.
29. Nygard H.S., Hansen M., Alhaj-Saleh Y., Olsen E., Palimąka P., Pietrzyk S., *Experimental evaluation of chemical systems for CO₂ capture by CaO in eutectic CaF₂-CaCl₂*, AIMS Energy, 7(5), 2019, s. 619-633.
30. Claquesin J., Gibilaro M., Massot L., Lemoine O., Bourges G., Chamelot P., *Thermodynamic Properties of CaCl₂-CaF₂-CaO System: Phase Diagram Investigation*, Materials Sciences and Applications, 12, 2021, s. 139-151.
31. Olsen E., inventor; 2013 Sep. 24. CO₂-Capture in Molten Salts, Patent No.: US 8,540,954 B2.

Wychwytywanie dwutlenku węgla w solach stopionych – prototyp reaktora CCMS

Streszczenie

W pracy dokonano przeglądu metod wychwytu i sorpcji dwutlenku węgla z gazów po procesie spalania w instalacjach bazujących na paliwach kopalnych, ze szczególnym uwzględnieniem metody pętli wapieniowej i jej ostatniej wersji z zastosowaniem soli stopionych jako rozpuszczalnika dla sorbentu tlenkowego. Metoda ta (ang. *carbon capture in molten salts* – CCMS), stosowana dotychczas w skali laboratoryjnej w środowisku soli stopionych na bazie CaCl₂ i CaF₂ zapewnia uzyskanie wysokiej wydajności sorpcji CO₂, a sorbent (CaO) nie ulega degradacji nawet po kilkunastu cyklach. Niniejsza praca dotyczy opracowania prototypu reaktora CCMS. Przedstawia nowatorskie rozwiązanie reaktora, składającego się z trzech podstawowych elementów, tj. komory niskotemperaturowej, komory wysokotemperaturowej oraz zbiornika pomocniczego. Zadanie dwóch pierwszych elementów to sorpcja i desorpcja CO₂, natomiast zbiornik pomocniczy ma za zadanie zwiększyć efektywność wykorzystania zdolności absorpcyjnej CaO i wydajność procesu. W pracy przedstawiono zagadnienia związane z opracowaniem koncepcji działania i budowy prototypu reaktora CCMS. Przeprowadzone analizy termiczne reaktora pozwoliły potwierdzić trafność koncepcji i zaprojektować jego konstrukcję wykonawczą. Prototyp reaktora CCMS jest obecnie w fazie wykonawstwa na bazie wyników uzyskanych w niniejszej pracy.

Słowa kluczowe: wychwytywanie CO₂, dekarbonizacja, sole stopione

Carbon capture in molten salts – prototype of the CCMS reactor

Abstract

This paper reviews the methods of CO₂ capture from gases after the combustion process in installations based on fossil fuels with a particularly focus on the CaL method and its latest version using molten salts as a solvent for the calcium oxide sorbent. In this method, which was developed on a laboratory scale so far, the molten salts mixture (CaCl₂+CaF₂) ensures high efficiency of CO₂ sorption, and the sorbent (CaO) does not degrade even after a several cycles. This work concerns the development of a CCMS reactor prototype. It presents an innovative reactor solution consisting of three basic elements, i.e. a low-temperature chamber, a high-temperature chamber and an reservoir tank. The task of the first two chambers is sorption and desorption of CO₂, while the reservoir tank is designed to increase the efficiency of the process and the sorption capacity of CaO. The paper presents issues related to the development of the concept of operation and construction of the CCMS reactor prototype. The thermal analysis conducted in this work allowed to confirm the accuracy of the concept, design and its execution plan. Based on the results obtained in this work, the prototype of the CCMS reactor is currently in the realization phase.

Keywords: CO₂ capture, decarbonization, molten salts

Czujniki gazu na bazie heterozłączy tlenków metali zaprojektowanych w nanoskali

1. Wprowadzenie

Czujniki gazu na bazie tlenków metali (MOS) to urządzenia półprzewodnikowe, które znalazły zastosowanie w analizie jakości powietrza, diagnozie chorób, czy też wykrywaniu materiałów wybuchowych, substancji odurzających i gazów toksycznych [1]. Mimo swoich licznych zalet (np. wysoka czułość, długi czas życia, krótki czas odpowiedzi), czujniki tego typu posiadają również liczne ograniczenia (np. niska selektywność oraz niestabilność odpowiedzi w czasie). Wykazano, że zastosowanie dwóch (lub więcej) tlenków metali w celu utworzenia heterozłącza może mieć pozytywny wpływ na wydajność detekcji czujnika, zwłaszcza na jego selektywność. W ostatnich latach efekt ten został wzmocniony poprzez projektowanie heterozłączy w nanoskali [2, 3]. Badania początkowo obejmowały np. mieszanie nanocząstek wybranych tlenków metali i wykorzystanie tak powstałej mieszaniny jako warstwy czulej. Inna strategia zakładała wytworzenie tzw. podwójnej warstwy czulej, gdzie warstwa jednego rodzaju tlenku metalu była nakładana na drugą. Z czasem zaczęto jednak stosować bardziej zaawansowane systemy, m.in. nanostruktury typu rdzeń-powłoka, czy też nanocząstki tlenku metalu udekorowane mniejszymi nanocząstkami drugiej fazy. W niniejszym przeglądzie wyszczególniono różne klasy morfologiczne dostępnych heterozłączy opartych na tlenkach metali, a następnie przedstawiono dominujące elektroniczne i chemiczne mechanizmy wpływające na działanie tych materiałów jako warstw czułych. Przedstawiono liczne tabele podsumowujące przykłady zastosowań heterozłączy w konstrukcji czujników gazu. W ostatnim rozdziale omówiono wyzwania związane z zastosowaniem heterozłączy jako warstw czułych, które będą napędzały badania w nadchodzących latach.

2. Zasada działania czujników MOS

W czujnikach typu MOS porcja tlenku metalu nakładana jest na elektrody, za pomocą których mierzy się zmiany oporności elektrycznej tak powstałej warstwy czulej wywołane jej interakcjami z oznaczaną mieszaniną gazową. Kinetyka tych reakcji zależy od temperatury, więc większość czujników MOS jest wyposażona w grzałkę pozwalającą na pracę z zadaną temperaturą, optymalną dla danego zastosowania [4]. O przydatności danego czujnika do konkretnego zastosowania decydują jego parametry, takie jak: czułość, selektywność, granica wykrywalności i oznaczalności, powtarzalność, stabilność odpowiedzi w czasie oraz szybkość odpowiedzi i regeneracji [5].

Mechanizm działania czujnika MOS typu n można pokrótce opisać następująco. W warunkach normalnych tlen atmosferyczny ulega adsorpcji na powierzchni warstwy czulej, gdzie jest redukowany do jednej ze swoich postaci zjonizowanych (tj. O_2^- , O^{2-} , O^-). W wyniku tego procesu warstwa czuła pozbawiana jest swoich nośników ładunku,

¹ justyna.jonca@pwr.edu.pl, Katedra Inżynierii Ochrony Środowiska, Wydział Inżynierii Środowiska, Politechnika Wroclawska, <https://wis.pwr.edu.pl/>.

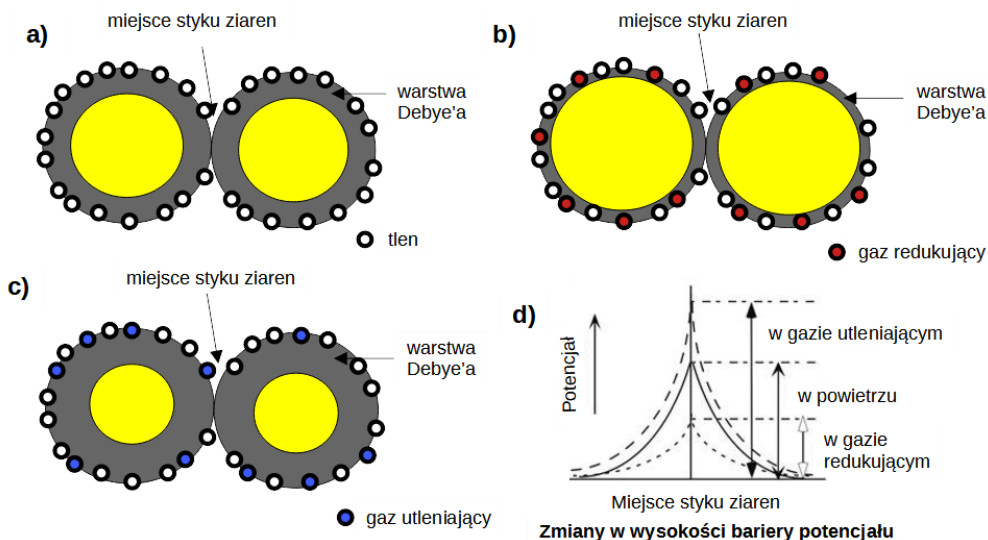
czyli elektronów, co powoduje wzrost oporności czujnika. W przypadku gdy warstwa czuła zaprojektowana jest w mikroskali, tlen adsorbuje się tylko na powierzchni ziaren tlenku metalu i co za tym idzie – elektrony są z nich usuwane tylko do określonej głębokości, zwanej długością Debye’a, która zwykle wynosi 2-100 nm. Region ten może się zmniejszać bądź zwiększać, gdy oprócz tlenu, adsorpcji na warstwie czułej ulegną inne gazy (rys. 1). Oporność czujnika może wówczas wzrastać lub maleć w obecności analizowanego gazu w zależności od tego czy jest to, odpowiednio, gaz utleniający, czy redukujący [4, 5]. Dla półprzewodnika typu *p* obserwuje się odwrotne zachowanie, czyli wzrost oporności w atmosferze redukującej oraz spadek – w utleniającej (tab. 1).

Wykazano, że zaprojektowanie warstwy czułej w nanoskali (tj. wykorzystanie nanocząstek tlenków metali zamiast ich odpowiedników w mikroskali) skutkuje zwiększeniem wydajności detekcji czujnika [6]. W przypadku nanocząstek, zwłaszcza o średnicy mniejszej niż 20 nm, warstwa zubożona w elektrony (dla półprzewodników typu *n*) obejmuje zarówno powierzchnię, jak i wnętrze ziaren (rys. 2). Skutkuje to zwiększeniem oporności w porównaniu do warstw czułych zbudowanych z większych ziaren, w których w procesie adsorpcji tlenu nie bierze udziału znaczna część materiału. W przypadku detekcji gazu redukującego materiał czuły zaprojektowany w nanoskali oferuje zatem więcej zjonizowanych form tlenu, które mogą utlenić analit. W procesie tym do warstwy czułej zwrócona zostaje duża porcja elektronów, co prowadzi do drastycznego spadku jej oporności, a to z kolei przekłada się na większą czułość czujnika. Wydajność detekcji czujników na bazie nanocząstek tlenków metali może być dodatkowo zwiększona poprzez dekorowanie warstwy czułej metalami szlachetnymi (np. Ag, Au, Pt, Pd, itp.), domieszkowanie nanomateriałami węglowymi (nanorurki węglowe, grafen, itp.) czy też tworzenie heterozłączy [5, 7]. Ostatnie podejście zostanie szczegółowo opisane w niniejszym przeglądzie.

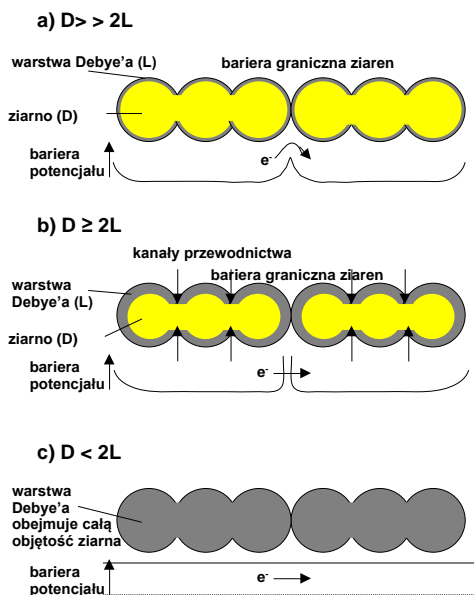
Tabela 1. Reakcja warstw czułych na bazie półprzewodników w zależności od typu półprzewodnika i gazu z nim wchodzącego w interakcję

	Półprzewodnik typu <i>n</i>	Półprzewodnik typu <i>p</i>
Gaz redukujący	Wzrost oporności	Spadek oporności
Gaz utleniający	Spadek oporności	Wzrost oporności
Nośniki ładunku	Elektrony	Dziury elektronowe

Źródło: opracowanie własne.



Rysunek 1. Mechanizm detekcji gazów utleniających i redukujących za pomocą czujników na bazie półprzewodników typu n : a) adsorpcja zjonizowanych form tlenu i wytworzenie warstwy zubożonej w nośniki ładunku, b) adsorpcja gazu redukującego i zmniejszenie warstwy zubożonej w nośniki ładunku, c) adsorpcja gazu utleniającego i zwiększenie warstwy zubożonej w nośniki ładunku, d) zmiany wysokości bariery potencjału podczas detekcji poszczególnych rodzajów gazów dla półprzewodnika typu n .
 Źródło: opracowanie własne



Gdy $D \gg 2L$, na większość wnętrza ziaren nie mają wpływu oddziaływania powierzchniowe z mieszaniną gazową. Przewodność czujnika zależy od wysokości bariery granicznej ziaren. W tym przypadku czułość czujnika jest niezależna od wielkości ziarna.

Gdy $D \geq 2L$, w warstwie zubożonej tworzą się kanały przewodnictwa. Przewodność czujników zależy zarówno od tych kanałów, jak i od wysokości bariery granicznej ziaren. W tym przypadku czułość czujnika staje się zależna od wielkości ziarna i wzrasta wraz z jej zmniejszaniem.

Gdy $D < 2L$, warstwa zubożona rozciąga się na całe ziarno, a kanały transferowe między ziarnami zanikają. Prowadzi to do drastycznego spadku przewodnictwa czujnika, a czułość jest całkowicie zależna od wielkości ziarna.

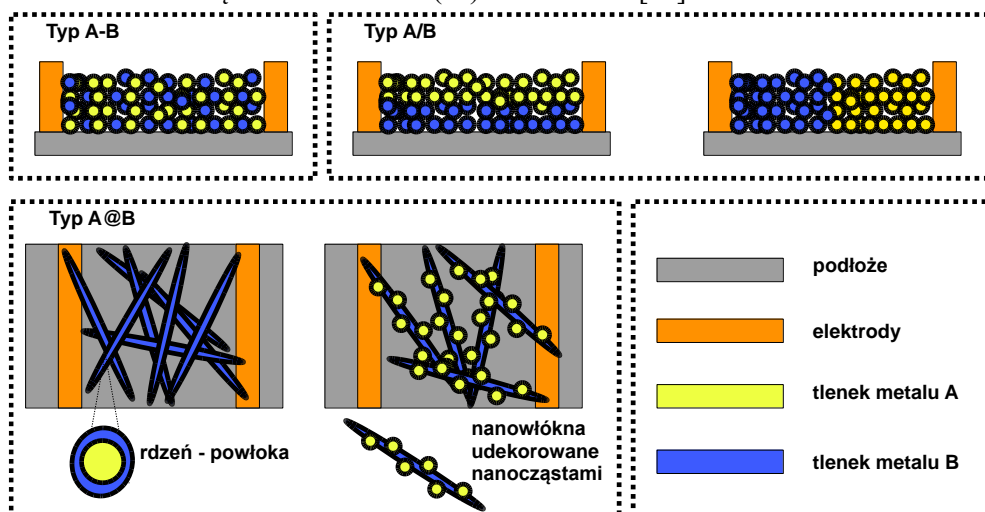
Rysunek 2. Schematyczny model wpływu wielkości ziaren na czułość półprzewodnikowych czujników na bazie tlenków metali: (a) $D \gg 2L$; (b) $D \geq 2L$; (c) $D < 2L$. Źródło: opracowanie własne

3. Morfologia heterozłączeń

Heterozłączenie odnosi się do obszaru na granicy styku między dwoma półprzewodnikami o różnych szerokościach przerwy energetycznej. Ze względu na różne właściwości obu materiałów zjawisko niedopasowania nadaje heterozłączeniom nowe właściwości, które znalazły zastosowanie m.in. w konstrukcji czujników gazu. W zależności od tego, jakie typy półprzewodników zostaną ze sobą połączone wyróżnia się następujące rodzaje heterozłączeń, tj. *n-n*, *p-p* oraz *p-n* (lub *n-p*). Bardzo istotna jest także morfologia heterozłączenia. Wyróżniamy trzy jej rodzaje (rys. 3):

1. Heterozłączenia **typu A-B**, które powstały w wyniku zmieszania dwóch (lub więcej) tlenków metali, a następnie wykorzystanie powstałej mieszaniny jako warstwy czulej. Ten rodzaj heterozłączenia w niniejszej pracy oznaczać będziemy symbolem „-” stawianym pomiędzy dwoma tlenkami metali (np. ZnO-CuO [8]).
2. Heterozłączenia **typu A/B**, które powstały w wyniku nałożenia jednej warstwy czulej na drugą oznaczać będziemy symbolem „/” stawianym pomiędzy dwoma tlenkami metali, np. SnO₂/ZnO, gdzie ZnO stanowi warstwę dolną (tj. tę bezpośrednio nałożoną na elektrody pomiarowe) [9].
3. Heterozłączenia **typu A@B**, które powstały w wyniku nałożenia nanocząstek tlenku metalu na materiał bazowy, który sam jest tlenkiem metalu o odmiennym składzie chemicznym. Ten rodzaj heterozłączenia oznaczać będziemy symbolem „@” stawianym pomiędzy dwoma tlenkami metali. W tej kategorii znajdują się np. nanoarkusze NiO udekorowane nanocząstkami Fe₂O₃ (Fe₂O₃@NiO, gdzie NiO stanowi tutaj materiał bazowy [10]), czy struktury typu rdzeń-powłoka (np. CuO@SnO₂, gdzie SnO₂ stanowi rdzeń, czyli materiał bazowy [11]).

Sytuacje, w których zastosowano więcej niż jedno rozwiązanie, należy rozpatrywać indywidualnie, np. poprzez zapis (γ -Fe₂O₃-In₂O₃)/In₂O₃ należy rozumieć system złożony z dwóch warstw, z czego dolna składa się z tlenku indy, a górna powstała w wyniku zmieszania nanocząstek tlenku żelaza(III) i tlenku indy [12].



Rysunek 3. Schematyczne przedstawienie morfologii heterozłączeń. Źródło: opracowanie własne

3.1. Heterozłącza typu A-B

Najłatwiejszy sposób przygotowania heterozłączy polega na mieszanii dostępnych komercyjnie pudrów tlenków metali. W pracy de Lacy Costello i in. [13] opracowano w ten sposób serię czujników na bazie takich tlenków, jak ZnO, SnO₂, In₂O₃ oraz ich mieszanin (przygotowanych w stosunku masowym 1 : 1). Ze wszystkich testowanych warstw czułych, mieszanina SnO₂-ZnO okazała się być najbardziej czuła na wybrane lotne związki organiczne (z grupy alkoholi, ketonów, aldehydów i węglowodorów). Parę lat później, ten sam zespół badał wpływ stosunku masowego SnO₂ do ZnO na wydajność detekcji etanolu [14]. Najlepsze parametry uzyskano dla warstwy czułej składającej się w 25% SnO₂ oraz w 75% z ZnO. Ponadto, czujnik ten charakteryzował się wyższą czułością na etanol niż dostępne komercyjnie czujniki Figaro (TGS 800, TGS 822, TGS 842).

Obecnie, w procesie tworzenia heterozłączy typu A-B coraz częściej wykorzystuje się tlenki metali opracowane w nanoskali za pomocą technik na bazie koloidów, które pozwalają na lepszą kontrolę składu chemicznego oraz wielkości i kształtu ziaren w porównaniu z proszkami dostępnymi komercyjnie [15]. W pracy Jońca i in. przygotowano anizotropowe nanocząstki ZnO i CuO poprzez rozkład odpowiednich prekursorów metaloorganicznych w obecności alkioloamin [9]. Produkty syntezy połączono ze sobą, uzyskując mieszaniny o różnych stosunkach masowych ZnO i CuO (tj. 1 : 3, 1 : 1 oraz 3 : 1), które wykorzystano następnie jako warstwy czułe. Najlepsze właściwości zaobserwowano dla mieszaniny 1 : 3, zwłaszcza w przypadku detekcji tlenku węgla(II). W temperaturze pracy czujnika wynoszącej 165°C odpowiedź na 100 ppm CO wynosiła 624%. Poza tym czujnik ten wykazał niską czułość na inne testowane gazy ($K_{CO/C_3H_8} = 14,5$ i $K_{CO/NH_3} = 26$), co pozwala na wykorzystanie urządzenia do detekcji tlenku węgla w obecności amoniaku i propanu.

Ciekawym sposobem na opracowanie heterozłączy typu A-B jest prowadzenie syntezy w obecności dwóch prekursorów jednocześnie. W pracy Mataung i in., wykorzystując podejście hydrotermalne, przygotowano nanocząstki CeO₂ i SnO₂ o średnicy około odpowiednio 14 i 11 nm [16]. Po przeprowadzeniu syntezy w obecności obu prekursorów jednocześnie uzyskano mieszaninę CeO₂-SnO₂ o średnicy nanocząstek odpowiednio 46 i 6 nm. Obecność prekursora ceru hamowała zatem wzrost nanocząstek tlenku cyny(IV). Parametry czujnika na bazie CeO₂-SnO₂ odznaczały się na tle czujników przygotowanych na bazie czystego CeO₂ i SnO₂. Czujnik ten okazał się być bowiem wyjątkowo czuły na H₂. Czujnik CeO₂-SnO₂ wykazywał też znacznie niższe odpowiedzi na inne testowane gazy tj. CO czy NO₂ oraz nie był w ogóle czuły na C₂H₅OH, NH₃ oraz CH₄, co pozwala na jego selektywne zastosowanie w detekcji wodoru.

W pracy Shaposhnik i in. porównywano natomiast właściwości czujników SnO₂-TiO₂, w których warstwy czułe zostały przygotowane poprzez zmieszanie nanocząstek tlenków metali oraz dzięki technice współstrącania z odpowiednich prekursorów [17]. Oba podejścia pozwoliły na uzyskanie wyższej czułości detekcji H₂ w porównaniu z czujnikami na bazie SnO₂, jednak optymalne parametry uzyskano, stosując inne stosunki masowe SnO₂ : TiO₂, tj. 4 : 1 w podejściu wykorzystującym zmieszane tlenki oraz 9 : 1 dla heterozłącza powstałego w procesie współstrącania.

3.2. Heterozłącza typu A/B

Heterozłącza typu A/B przygotować można za pomocą kilku metod, w tym zol-żel, rozpylanie, czy osadzanie elektrochemiczne [17-23]. Dwuwarstwowe heterostrukтуры z dużym, zdefiniowanym złączem pozwalają na łatwiejszą charakterystykę jego właściwości ze względu na dobrze zdefiniowane wymiary i obszar kontaktu. Struktury te mogą być jednak mniej pożądane ze względu na małą dostępność gazu do złącza w porównaniu do systemu typu A-B. W dwuwarstwach horyzontalnych gaz musi bowiem dyfundować przez górny materiał, aby dotrzeć do heterozłącza. W pracy Dandeneau i in. zminimalizowano ten problem, stosując pirolizę w celu zwiększenia porowatości górnej warstwy CuO [20], co poskutkowało m.in. zmniejszeniem czasu odpowiedzi czujnika CuO/ZnO na H₂.

Stosowanie nanocząstek tlenków metali zamiast ich odpowiedników w mikroskali również może ułatwić dyfuzję gazów do heterozłącza i pozytywnie wpływać na wydajność detekcji. Maziarz przygotował czujniki TiO₂/SnO₂ oraz TiO₂/CuO o grubości warstwy dolnej (tj. SnO₂ lub CuO) wynoszącej około 200 nm [24]. Grubość warstwy górnej (tj. TiO₂) wynosiła natomiast zaledwie 40 nm. Czujnik TiO₂/SnO₂ wykazał dwukrotnie wyższą czułość na NO₂ w porównaniu z czujnikiem SnO₂. Co ciekawe, poprawy wydajności detekcji tego gazu nie zaobserwowano w przypadku kombinacji TiO₂/CuO. Ekspozycja warstwy TiO₂ (półprzewodnik typu *n*) na gaz utleniający skutkuje wzrostem oporności czujnika, a warstwy CuO (półprzewodnik typu *p*) – jej spadkiem. W odpowiednich warunkach efekty te mogą się kompensować, co było prawdopodobnie przyczyną niższej wydajności detekcji w tym przypadku.

W pracy Sendi i in. natomiast badano właściwości złącza CuO/SnO₂, które porównano z odpowiedziami czujników CuO i SnO₂ na HCHO, CO i NO₂ [25]. Badania przeprowadzono dla dwóch różnych temperatur pracy, tj. 100°C oraz 500°C. Czujnik CuO wykazał się dobrą czułością na HCHO, podczas gdy czujnik SnO₂ – na NO₂. W obu przypadkach uzyskano wyższą odpowiedź w niższej temperaturze pracy. Czujnik CuO/SnO₂ z kolei okazał się bardzo czuły na CO przy wyższej temperaturze pracy. Jednak to w przypadku HCHO uzyskano najciekawsze wyniki dla tego czujnika. W zależności od temperatury pracy czujnik CuO/SnO₂ w obecności tego gazu wykazywał odpowiedzi charakterystyczne dla półprzewodników typu *n* (100°C) lub *p* (500°C). Fenomen ten zwany inwersją sygnału jest często obserwowany w przypadku heterozłączeń typu *p-n* i może być indukowany temperaturą, światłem UV, czy też obecnością konkretnych analitów [8, 25, 26].

Heterozłącza typu A/B mogą być realizowane zarówno horyzontalnie (przykłady opisane powyżej), jak i wertykalnie. W tym drugim podejściu (znacznie rzadziej stosowanym) tlenek metalu A pozostaje w kontakcie z jedną elektrodą pomiarową, podczas gdy tlenek metalu B – z drugą. Tutaj również możemy mieć do czynienia z fenomenem kompensacji odpowiedzi, czy też inwersją sygnału w przypadku heterozłączy typu *p-n*. Bardzo ciekawe wyniki uzyskano w pracy Kwon i in. [26], w której czujnik na bazie NiO/SnO₂ wykazał taką samą tendencję odpowiedzi zarówno wobec gazu utleniającego (NO₂), jak i redukującego (H₂). Stwierdzono, że takie nieoczekiwane właściwości wynikały z możliwości modulacji wysokości bariery potencjału na złączach *p-n*.

3.3. Heterozłącza typu A@B

3.3.1. Dekoracja materiału bazowego nanocząstkami drugiej fazy

Heterozłącza typu A@B można przygotować zarówno za pomocą prostego podejścia hydrotermalnego, jak i bardziej zaawansowanych technik. W pracy Kim i in. za pomocą metody hydrotermalnej, nanocząstkami NiO udekorowano hierarchiczne sfery SnO₂ w celu minimalizacji wpływu wilgoci na wydajność detekcji CO [27]. W suchej atmosferze czujniki na bazie niedomieszkowanych struktur SnO₂ wykazują wysoką wydajność detekcji CO. Jednak czułość, czas odpowiedzi oraz rezystancja czujnika ulegają pogorszeniu lub znacznie zmieniają się w wilgotnej atmosferze. Gdy hierarchiczne nanostruktury SnO₂ są domieszkowane NiO, wszystkie właściwości wykrywania gazu pozostają podobne zarówno w atmosferze suchej, jak i wilgotnej. Badania spektroskopowe wykazały, że większość wilgoci jest absorbowana przez NiO, a zatem oddziaływania między wodą a powierzchnią SnO₂ są zablokowane. Ponadto, czujniki NiO@SnO₂ charakteryzowały się wyjątkowo krótkim czasem odpowiedzi (1,6 s) i regeneracji (2,8 s) oraz wysoką czułością na CO (50 ppm, R_a/R_g = 2,8).

Z hydrotermalnego podejścia korzystano także w pracy Lou i in. [28]. W pracy przygotowano hierarchiczne struktury ZnO przypominające kwiaty, które następnie udekorowano nanocząstkami PdO. Czujnik PdO@ZnO wykazywał wyższą czułość na toluen i etanol niż czujnik ZnO. Co więcej, przy temperaturze pracy wynoszącej 240°C czujnik PdO@ZnO był wyjątkowo czuły na toluen, podczas gdy odpowiedź na etanol była bardzo niska. Odwrotną tendencję zaobserwowano przy temperaturze 320°C. Dzięki modulacji temperatury udało się zatem osiągnąć selektywność detekcji wybranych gazów.

W pracy Suh i in. wykorzystano bardziej zaawansowaną technikę wytwarzania heterozłączy [29]. W pracy przygotowano heterozłącze NiO@Co₃O₄ poprzez dekorację nanoprętów Co₃O₄ nanocząstkami NiO. Warstwa czuła została przygotowana przez metodę znaną jako *glancing-angle deposition* (ang.), będącą modyfikacją techniki fizycznego osadzania z fazy gazowej. W metodzie tej podłoże jest obracane pod różnym kątem w stosunku do padającego nań strumienia gazu, co zwiększa porowatość końcowego produktu. Równomierne pokrycie nanoprętów na całej ich długości skutkowało wytworzeniem licznych złączy typu *p-p*. Czujniki na bazie heterozłącza wykazywały kilkunastokrotnie wyższą czułość na benzen niż czujniki Co₃O₄. Autorzy pracy podkreślili, że zwiększona wydajność detekcji czujnika została uzyskana nie tylko dzięki występowaniu złącza, ale także dzięki katalitycznym właściwościom NiO oraz zwiększonej ilości miejsc aktywnych w warstwie czułej, co było wynikiem zastosowanej techniki jej przygotowania.

3.3.2. Struktury typu rdzeń-powłoka

Struktury typu rdzeń-powłoka należą do najbardziej obiecujących typów heterostruktur do konstrukcji czujników gazowych. Ta morfologia zapewnia maksymalizację powierzchni złącza między dwoma tlenkami, jednocześnie minimalizując ilość tlenku służącego jako materiał bazowy. Istnieje kilka sposobów wytwarzania struktur typu rdzeń-powłoka. Materiał bazowy może być syntetyzowany za pomocą metody hydrotermalnej lub zolowo-żelowej, fizycznego i chemicznego osadzania z fazy gazowej (PVD, CVD), elektroprzędzenia oraz wielu innych. Następnie na materiał bazowy osadzana jest cienka warstwa (zwykle <20 nm) drugiej fazy, wykorzystując podejście hydrotermalne, elektroprzędzenie, napylenie katodowe czy też osadzanie warstw atomo-

wych (ang. *atomic-layer deposition*, ALD) [30]. Techniki napyłania katodowego i ALD, w odróżnieniu od pozostałych metod, pozwalają na precyzyjną kontrolę grubości powłoki, która uzależniona jest od czasu trwania procesu. Liczne badania wykazały, że grubość powłoki znacząco wpływa na wydajność detekcji gazu. W pracy Xu i in. [31] przygotowano nanoprepy $\text{NiO}@\alpha\text{-Fe}_2\text{O}_3$ oraz $\text{CuO}@\alpha\text{-Fe}_2\text{O}_3$ za pomocą podejścia hydrotermalnego. Obie struktury wykorzystano jako warstwy czułe i porównano ich możliwości detekcji trietyloaminy z czujnikiem na bazie $\alpha\text{-Fe}_2\text{O}_3$. W obu przypadkach, tj. $\text{NiO}@\alpha\text{-Fe}_2\text{O}_3$ oraz $\text{CuO}@\alpha\text{-Fe}_2\text{O}_3$ uzyskano wyższą czułość na analizowany gaz niż dla czujnika $\alpha\text{-Fe}_2\text{O}_3$, przy zachowaniu częściowej selektywności w obecności etanolu, propan-2-olu, acetonu, benzenu i ksyleny. Dodatkowo, w pracy wykazano, że najwyższą odpowiedź na analizowane gazy uzyskuje się, gdy grubość warstwy NiO (lub CuO) jest zbliżona do długości Debye'a.

W pracy Chang i in. [32] przygotowano trzy warstwy czułe $\text{TiO}_2@\text{SnO}_2$ za pomocą techniki ALD, w których grubość powłoki TiO_2 wynosiła ok. 2, 4 oraz 6 nm. Celem pracy było głównie zmniejszenie temperatury pracy czujnika do pokojowej. Osiągnięto to za pomocą fotodiody UV. Najlepsze parametry uzyskano dla czujnika, w którym wykorzystano nanocząstki o grubości powłoki wynoszącej ok. 4 nm. Czujnik ten był bardzo czuły na HCHO i wykazywał częściową selektywność w obecności NH_3 , CO, $\text{CH}_3(\text{CO})\text{CH}_3$. Zastosowanie fotodiod w oznaczaniu gazów za pomocą czujników MOS cieszy się coraz większym zainteresowaniem. Zastosowanie diody generuje w materiale fotoczułym (TiO_2) pary dziura-elektron, co sprzyja adsorpcji tlenu i tworzeniu zjonizowanych jego form na powierzchni półprzewodnika (analogicznie jak ma to miejsce podczas ogrzewania materiału czułego za pomocą grzałki do wysokiej temperatury). Podczas detekcji HCHO, gaz ten reaguje z zaadsorbowanym tlenem, w wyniku czego do warstwy czułej dostarczane są elektrony. Zamiast jednak rekombinować się z dziurami, elektrony te odpływają w kierunku SnO_2 , co powoduje drastyczny spadek rezystancji czujnika i co za tym idzie jego zwiększoną czułość.

Struktury typu $A@B$ również mogą wykazywać inwersję sygnału jak opisano w Huang i in. [33]. W pracy przygotowano warstwy czułe SnO_2 oraz $\text{ZnO}@\text{SnO}_2$ za pomocą techniki CVD, a następnie zbadano ich właściwości detekcji wybranych gazów (H_2 , CH_4 , CO, NH_3). Zarówno $\text{ZnO}@\text{SnO}_2$, jak i SnO_2 odznaczały się wysoką selektywnością wobec H_2 , ale czułość tego pierwszego na analizowany gaz była znacznie wyższa. Ponadto, zaobserwowano, że czujnik $\text{ZnO}@\text{SnO}_2$ wykazuje zależne od stężenia przejścia $n\text{-}p\text{-}n$ podczas detekcji wodoru. Czujnik ten w niskich oraz bardzo wysokich stężeniach wodoru reaguje spadkiem rezystancji (co jest charakterystyczne dla półprzewodnika typu n), a dla średnich stężeń tego gazu – wzrostem (co jest charakterystyczne dla półprzewodnika typu p). To nietypowe zachowanie czujnika w czasie detekcji H_2 wytłumaczono możliwością tworzenia się heterozłącza $n\text{-}\text{ZnO}/p\text{-}\text{ZnO}\text{-}\text{Sn}/n\text{-}\text{SnO}_2$ w warstwie czułej.

4. Mechanizmy odpowiedzialne za zwiększoną wydajność detekcji czujników na bazie heterozłączeń

4.1. Rodzaj heterozłącza

Rodzaj heterozłącza ($n\text{-}n$, $p\text{-}p$, $p\text{-}n$) jest prawdopodobnie najważniejszym czynnikiem podczas analizy mechanizmów detekcji czujników opartych na mieszaninach tlenków metali. Bardzo powszechnym złączeniem jest złącze typu $p\text{-}n$. Czujniki gazu typu p są rzadziej stosowane niż czujniki gazu typu n ze względu na ich mniejszą stabilność

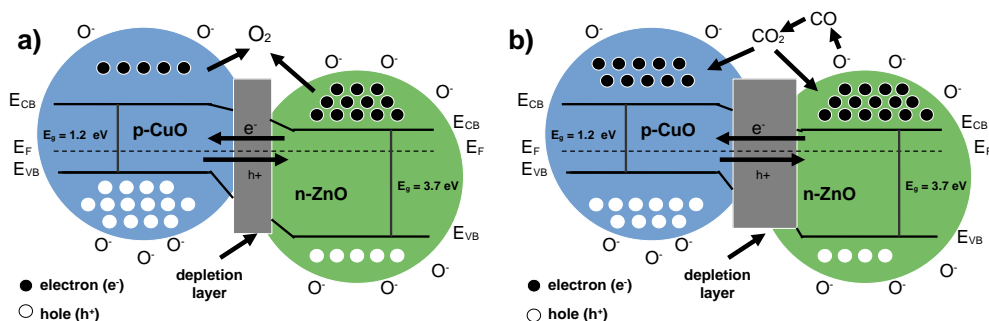
w środowiskach o niskiej zawartości tlenu i kompatybilność z urządzeniami pomiarowymi. Jednak materiały te można łączyć na różne sposoby z półprzewodnikami typu n , aby wykorzystać efekty złącza $p-n$ i poprawić wydajność detekcji. Poniżej przedstawiono mechanizmy działania złącz $p-n$, $n-n$ oraz $p-p$ w czujnikach gazowych.

4.1.1. Heterozłącze typu $p-n$

Dla lepszego zrozumienia mechanizm detekcji czujnika na bazie heterozłącza $p-n$ zostanie przedstawiony na konkretnym przykładzie, tj. czujnika CuO-ZnO [8] (gdzie CuO to półprzewodnik typu p , a ZnO – typu n). Zarówno czujnik CuO, jak i CuO-ZnO wykazują wzrost rezystancji w obecności gazów redukujących, co jest charakterystyczne dla półprzewodników typu p . Gdy warstwa czuła na bazie CuO jest wystawiona na działanie czystego powietrza, cząsteczki tlenu adsorbują się na jej powierzchni, gdzie ulegają redukcji za pomocą wolnych elektronów CuO. Sprzyja to tworzeniu tzw. warstwy akumulacyjnej dziur (ang. *hole accumulation layer*, HAL). Po wystawieniu na działanie gazu redukcyjnego (tj. CO) cząsteczki gazu reagują ze wstępnie zaadsorbowanymi formami tlenu na powierzchni CuO i zwracają przechwycone elektrony z powrotem do pasma przewodnictwa (ang. *conductivity band*, CB) CuO. Tymczasem grubość HAL staje się cieńsza i wzrasta rezystancja CuO.

Poprawa właściwości detekcji gazu czujnika na bazie półprzewodnika typu p w wyniku jego połączenia z półprzewodnikiem typu n można opisać następująco. Kiedy CuO łączy się z ZnO dochodzi do wytworzenia wewnętrznego pola elektrycznego wskutek dyfuzji nośników ładunku wywołanej różnymi właściwościami chemicznymi i fizycznymi obu tlenków (tj. różna struktura pasmowa, stała dielektryczna, stała sieci krystalograficznej, powinowactwo elektronowe, itd.). Elektrony poruszają się od ZnO do CuO, a dziury w przeciwnym kierunku do momentu, w którym ustali się równowaga (tj. zostanie osiągnięty poziom Fermiego, EF) jak przedstawiono na rysunku 4a. W miejscu styku dwóch półprzewodników następuje rekombinacja dziura-elektron, a w okolicy złącza tworzy się warstwa zubożona w nośniku ładunku (ang. *depletion region*) tj. po stronie CuO tworzy się obszar zubożony w dziury, a po stronie ZnO – w elektrony. W związku z powyższym powstaje bariera potencjału, a oporność czujnika CuO-ZnO w powietrzu rośnie w porównaniu z czujnikami CuO i ZnO.

Kiedy oprócz tlenu w mieszaninie gazowej znajduje się gaz redukujący (tj. CO), jego cząsteczki reagują ze wstępnie zaadsorbowanymi formami tlenu i elektrony wracają do CuO i ZnO (rys. 4b). Elektrony wracające do pasma przewodnictwa CuO łączą się z dziurami w paśmie walencyjnym (VB), co zmniejsza stężenie dziur w warstwie akumulacyjnej dziur (HAL) i zwiększa rezystancję CuO. Z drugiej strony, elektrony, które powróciły do ZnO, odpływają do CuO aż do ustalenia się nowego stanu równowagi. W efekcie jeszcze większa ilość dziur w paśmie walencyjnym CuO ulega rekombinacji z napływającymi elektronami, co dodatkowo zwiększa rezystancję CuO. W końcowym efekcie uzyskuje się znaczną poprawę odpowiedzi czujnika CuO/ZnO na CO w porównaniu do czujników CuO i ZnO.



Rysunek 4. Mechanizm działania czujnika na bazie heterozłącza *p-n* (CuO-ZnO) w powietrzu (a) oraz w atmosferze gazu redukującego (b) [8]

4.1.2. Heterozłącze typu *n-n* oraz *p-p*

Heterozłącze typu *n-n* badano np. w pracy Zeng i in. [34], w której wykazano, że heterostruktura SnO₂-Sn₃O₄ oferuje lepszą wydajność detekcji NO₂ niż czujnik SnO₂. Mechanizm działania czujnika na bazie heterozłącza można opisać następująco. Z powodu różnej pracy wyjścia Sn₃O₄ (3,9eV) i SnO₂ (4,9eV), elektrony płyną z Sn₃O₄ do SnO₂ aż do osiągnięcia poziomu Fermiego (EF). Warstwa akumulacji elektronów (ang. *electron accumulation layer*) i warstwa zubożona w te nośniki ładunku (ang. *electron depletion layer*) tworzą się na granicy faz odpowiednio po stronie SnO₂ i Sn₃O₄. Kiedy system ten wystawiony jest na działanie NO₂, elektrony uwięzione w warstwie akumulacyjnej mają tendencję do reagowania z tym gazem utleniającym, ponieważ zapewnia ona więcej aktywnych miejsc do adsorpcji cząsteczek docelowych niż jakiegokolwiek inne części heterostruktury. Spora część elektronów jest wówczas zużywana w tym procesie, zmniejszając ich ilość w warstwie akumulacyjnej i prowadząc do gwałtownego wzrostu rezystancji czujnika.

Heterozłącze typu *p-p* było badane m.in. przez Wang i in. [35]. W wyniku wytworzenia złącza, na granicy faz powstają warstwa akumulacji dziur oraz warstwa zubożona w te nośniki po stronie odpowiednio NiO oraz CuO. Adsorpcja tlenu na powierzchni warstwy czulej powoduje w tym przypadku zwiększenie liczby dziur elektronowych, zmniejszając szerokość warstwy zubożonej i obniżając rezystancję czujnika. Kiedy heterozłącze jest wystawione na działanie gazu redukującego (tj. H₂S), jego molekule zostaną utlenione przez wcześniej zaadorbowane formy tlenu. Do warstwy czulej dostają się wówczas elektrony, które rekombinują się z obecnymi w warstwie akumulacyjnej dziurami, zwiększając rezystancję czujnika.

4.2. Charakter kwasowo-zasadowy tlenków

Przy doborze tlenków metali tworzących warstwę czułą powinno brać się pod uwagę nie tylko rodzaj nośników ładunku w nich przeważających, ale również ich charakter kwasowo-zasadowy. Liczne badania wykazały, że detekcja alkoholi przebiega z wytworzeniem produktu pośredniego, którego skład zależy właśnie od składu chemicznego warstwy czulej [36-38]. Alkohol może ulec dehydratacji lub dehydrogenacji, jeżeli jego detekcja odbywa się z wykorzystaniem warstw czulej zrobionej z tlenku, odpowiednio kwasowego lub zasadowego. W czasie detekcji np. etanolu produktami pośrednimi mogą być zatem eten lub etanal. Dodatkowo wykazano, że dalszy rozkład produktów pośred-

nich do CO_2 i H_2O przebiega wydajniej, gdy produktem tej reakcji jest aldehyd [37]. W pracy Rumyantseva i in. [38] wykorzystano ten fakt, dodając zasadowego Fe_2O_3 (tj. o przeważającym charakterze zasadowym) do kwasowego SnO_2 (tj. o przeważającym charakterze kwasowym), co zwiększyło rolę dehydrogenacji w procesie detekcji etanolu.

W pracy de Lacy Costello i in. [13] przygotowano czujniki na bazie ZnO-SnO_2 i obserwowano produkty procesu rozkładu 1-butanolu. W badaniach wykazano, że SnO_2 było w dużej mierze odpowiedzialne za utlenianie butanolu do butanal, ale to ZnO odgrywało główną rolę w dalszych procesach rozkładu produktu pośredniego. W ten sposób system $\text{SnO}_2\text{-ZnO}$ działał synergistycznie, aby całkowicie rozłożyć butanol i wykazywał lepszą wydajność detekcji niż czujniki ZnO i SnO_2 . Otrzymane wyniki pokazują analogię pomiędzy wykrywaniem gazu a katalizą. Literatura dotycząca katalizy często przedstawia szczegółową analizę ścieżek reakcji zachodzących na powierzchni czujników MOS. Warto zatem po nią sięgnąć, konstruując czujniki na bazie heterozłącz tlenków metali [39, 40].

4.3. Efekt *spill-over*

Z efektem *spill-over* mamy do czynienia, gdy oznaczana molekula reaguje w specyficzny sposób z tlenkiem warstwy czulej, tworząc produkt pośredni, który pozostaje z nią związany (tj. zaadsorbowany), wpływając w sposób znaczący na wydajność detekcji czujnika. Najbardziej znanym przykładem jest detekcja siarkowodoru z wykorzystaniem warstwy czulej na bazie CuO . H_2S reaguje wówczas z nanocząstkami CuO , przekształcając je w CuS [41–44]. Wodór, który powstaje w tej reakcji, zostaje utleniony na powierzchni warstwy czulej, zmniejszając jej oporność.

Efekt *spill-over* pozwala na zwiększenie wydajności detekcji H_2S głównie poprzez całkowite usunięcie warstwy zubożonej w nośniki ładunku. W pracy Shao i in. przygotowano system CuO@SnO_2 poprzez dekorację nanoprętów SnO_2 nanocząstkami CuO [45]. CuO , tworząc heterozłącze *p-n* z SnO_2 zwiększa drastycznie oporność czujnika (patrz wyżej). CuS jest umiarkowanym przewodnikiem, a zatem po jego wytworzeniu powstaje złącze omowe, które eliminuje obszar zubożenia i umożliwia łatwe przewodzenie przez złącze i wzdłuż nanoprętów. W pracy Chowdhuri i in. [46] przygotowano strukturę CuO@SnO_2 typu rdzeń-powłoka oraz wykazano, że grubość powłoki CuO miała znaczący wpływ na wydajność detekcji H_2S . Gdy nie przekraczała 8 nm całkowita konwersja CuO do CuS była możliwa, a oporność czujnika spadała gwałtownie. Opisane systemy wykorzystują zatem wiele mechanizmów jednocześnie, aby zmaksymalizować wydajność detekcji H_2S .

5. Przykłady czujników opracowanych na bazie heterozłącz

Przykłady zastosowań czujników na bazie heterozłącz przedstawiono w tabelach 2–4, gdzie tabela 2 podsumowuje parametry czujników na bazie heterozłącz *n-n*, tabela 3 – *p-n*, a tabela 4 – *p-p*.

Tabela 2. Przykłady czujników na bazie heterozłącz *n-n*

Morfologia heterozłącza	Analizowany gaz i jego stężenie	Odpowiedź czujnika	Czas odpowiedzi/regeneracji	LOD	Selektywność	Ref.
ZnO@SnO_2	etanol (100 ppm)	13,3 ^b	-/-	5 ppm	CO , NH_3 , CH_4	[47]
ZnO-SnO_2	siarkowodór (10 ppb)	2,7 ^b	150 s/40 s	10 ppb	etanol, aceton, NH_3 , H_2	[49]

Czujniki gazu na bazie heterozłącz tlenków metali zaprojektowanych w nanoskali

ZnO/SnO ₂	etanol (5 ppm)	200 ^a	72 s/-	-	aceton, propan, CO, H ₂ , NO ₂	[9]
ZnO@SnO ₂	formaldehyd (100 ppm)	76 ^c	36 s/73 s	1,91 ppb	NH ₃ , etanol, aceton, metanol, benzen, toluen	[49]
SnO ₂ @ZnO	tlenek azotu(IV) (5ppb)	1,2 ^d	60 s/45 s	5 ppb	aceton, toluen, etanol, H ₂ S, CHCl ₃ , NH ₃	[50]
ZnO-SnO ₂	ozon (20 ppb)	8 ^f	13 s/90 s	20 ppb	CO, NO ₂ , NH ₃	[51]
TiO ₂ @SnO ₂	trietyloamina (100 ppm)	52,3 ^a	12 s/22 s	2 ppm	glikol, etanol, benzen, izopropanol, aceton, paraksylen	[52]
SnO ₂ @TiO ₂	aceton (100 ppm)	13,7 ^a	2 s/60 s	10 ppm	formaldehyd, metanol, etanol, toluen, NH ₃	[53]
SnO ₂ -CeO ₂	wodór (60 ppm)	1323 ^a	17 s/24 s	0,5 ppm	CO, NH ₃ , CH ₄ , NO ₂ , etanol	[54]
ZnO-CdO	tlenek azotu(II) (33 ppm)	22,6 ^a	35 s/630 s	1,2 ppm	CO, NH ₃ , CH ₄ , NO ₂ , H ₂ S, SO ₂ , CHCl ₃ , etanol, metanol, aceton	[55]
TiO ₂ @MoO ₃	etanol (10 ppm)	4,8 ^a	40 s/40 s	10 ppm	H ₂ , NH ₃ , CH ₄	[53]
In ₂ O ₃ @ZnO	tlenek azotu(IV) (3 ppm)	1419,9 ^e	13 s/189 s	500 ppb	H ₂ S, NH ₃ , SO ₂ , aceton, etanol	[57]
Nb ₂ O ₅ -TiO ₂	etanol (500 ppm)	21,64 ^a	-/-	17 ppb	NH ₃ , NO ₂ , kwas octowy, toluen, izopropanol	[58]

Źródło: opracowanie własne.

Tabela 3. Przykłady czujników na bazie heterozłącz *p-n*

Morfologia heterozłącza	Analizowany gaz i jego stężenie	Odpowiedź czujnika	Czas odpowiedzi/regeneracji	LOD	Selektywność	Ref.
CuO/ZnO	tlenek węgla(II) (100 ppm)	624 ^b	50 s/-	-	NH ₃ , propan	[8]
ZnO@CuO	etanol (50 ppm)	3,5 ^e	5 s/18 s	5 ppm	CO, CH ₄ , metanol, aceton, izopropanol	[59]
NiO@ZnO	trietyloamina (100 ppm)	185 ^a	7 s/33 s	2 ppm	etanol, aceton, benzen, izopropanol, p-ksylen, cykloheksan, heksan, metanol	[60]
ZnO@Co ₃ O ₄	formaldehyd (10 ppm)	5,65 ^c	27 s/15 s	1 ppm	NO ₂ , NH ₃ , etanol, aceton	[61]
NiO@SnO ₂	trietyloamina (10 ppm)	46,5 ^a	11 s/34 s	2 ppm	etanol, aceton, benzen, ksylen	[62]
SnO ₂ @NiO	H ₂ S (50 ppm)	183,5 ^g	45 s/60 s	1,5 ppb	NO ₂ , NH ₃ , toluen, formaldehyd	[68]
Cu ₂ O@SnO ₂	tlenek węgla(II) (1 ppm)	3 ^g	17 s/12 s	-	NH ₃ , aceton, benzen, formaldehyd	[11]

CuO/TiO ₂	wodór (1000 ppm)	2 ^f	7,4 min/6,8 min	100 ppm	NO ₂ , CHCl ₃ , aceton, etanol	[63]
Fe ₂ O ₃ @NiO	etanol (100 ppm)	170,7 ^d	0,5 s/14,6 s	200 ppb	NH ₃ , metanol, toluen, DMA, DMF, formaldehyd	[10]
SnO-SnO ₂	tlenek azotu(IV) (0,2 ppm)	2,5 ^d	57 s/5 min	0,1 ppm	NH ₃ , etanol, aceton, benzen, toluen	[64]
SnO-SnO ₂	formaldehyd (50 ppm)	80,9 ^f	7/27 s	8,15 ppb	aceton, etanol, metanol, trietyloamina	[65]
Cr ₂ O ₃ @Nb ₂ O ₅	wodór (200 ppm)	5,24 ^a	40 s/-	-	NH ₃ , H ₂ S, etanol	[66]
NiO-In ₂ O ₃	siarkowodór (3 ppm)	6 ^a	14 s/22 s	3 ppm	NH ₃ , H ₂ , CH ₄ , C ₂ H ₂ , metanol, benzen, toluen	[67]

Źródło: opracowanie własne.

Tabela 4. Przykłady czujników na bazie heterozłączeń *p-p*

Morfologia heterozłącza	Analizowany gaz i jego stężenie	Odpowiedź czujnika	Czas odpowiedzi/regeneracji	LOD	Selektywność	Ref.
CuO@NiO	tlenek azotu(IV) (100 ppm)	77,16 ^b	2 s/-	1 ppm	CO, CH ₄ , H ₂	[69]
CuO-NiO	glikol (100 ppm)	10,35 ^f	45 s/45 s	78 ppb	NO ₂ , etanol, propanol, izopropanol	[70]
Cu ₃ Mo ₂ O ₉ @CuO	tlenek azotu(IV) (5 ppm)	1,6 ^b	49 s/241s	2,3 ppb	NH ₃	[71]
NiO@Co ₃ O ₄	benzen (50 ppm)	16,78 ^a	-/-	13,9 ppb	CO, aceton, etanol, toluen, acetaldehyd	[29]
CoWO ₄ -Co ₃ O ₄	ksylen (100 ppm)	50 ^e	236 s/254 s	300 ppb	NH ₃ , NO ₂ , etanol, metanol, benzen, toluen, formaldehyd	[72]
NiWO ₄ @NiO (240°C)	trietyloamina (50 ppm)	65 ^d	48 s/15 s	1 ppm	ksylen, toluen,	[73]
NiWO ₄ @NiO (270°C)	ksylen (50 ppm)	28 ^d	40 s/21 s	-	anilina, metanol, aceton trietyloamina, toluen, anilina, metanol, aceton	
CuO-Co ₃ O ₄	tlenek azotu(IV) (10 ppm)	37,86 ^b	158 s/738 s	200 ppb	CO, NH ₃ , H ₂ S, etanol, toluen, ksylen	[74]
Co ₃ O ₄ -CoMoO ₄	aceton (5 ppm)	8,5 ^d	-/-	10 ppb	CO, NH ₃ , CH ₄ , NO ₂ , etanol, acetaldehyd, toluen	[75]
PdO@NiO	wodór (50 ppm)	82 ^f	53 s/78 s	500 ppb	CO, NO ₂ , CH ₄ , etanol, aceton, metanol,	[76]

PdO-CuO	siarkowodór (50 ppm)	6,8 ^a	1,8 s/4,1 s	5 ppm	CH ₄ , etanol, metanol, izopropanol, aceton, toluen, trietyloamina	[77]
PdO/CuWO ₄ / CuO	wodór (10000 ppm)	3,4 ^d	-/-	-	-	[78]

Źródło: opracowanie własne.

Uwaga do tabel 2-4. Sposób obliczania odpowiedzi czujnika różni się pomiędzy poszczególnymi pracami:

^aS = R_a/R_g , ^bS = $(R_a-R_g)/R_g \times 100$, ^cS = $(I_g-I_a)/I_a \times 100$, ^dS = R_g/R_a , ^eS = $(R_g-R_a)/R_a \times 100$, ^fbrak danych,
^gS = $(R_a-R_g)/R_g$.

6. Aktualne wyzwania

6.1. Stabilność odpowiedzi w czasie

W miarę zmniejszania rozmiarów ziaren tlenków metali i ich wykorzystania w heterozłączach pojawił się problem z termiczną i długoterminową stabilnością odpowiedzi w czasie czujników na ich bazie. Nanocząstki o mniejszych wymiarach są bardziej podatne na aglomerację w wysokich temperaturach niż ich większe i bardziej stabilne odpowiedniki. Proces aglomeracji zmienia właściwości warstwy czulej i co za tym idzie parametry czujnika w czasie. Dodatkowo, w wysokich temperaturach procesy dyfuzji w heterozłączach mogą doprowadzić do powstania tlenków mieszanych lub faz intermetalicznych, które mogłyby poważnie wpłynąć na właściwości elektroniczne bariery potencjału. Tworzeniu się tych faz, jeśli jest niepożądane, można zapobiec, stosując techniki syntez niskotemperaturowych oraz poprzez obniżenie temperatury pracy czujnika [79]. Dlatego też trwają intensywne prace nad konstrukcją czujników, które mogłyby operować w niskich temperaturach pracy [23, 33, 56, 65]. Wyższą stabilność można również osiągnąć za pomocą struktur 1D (nanopręty, nanowłókna, itd.), które generalnie mają wyższy stopień krystaliczności i większe powierzchnie płaskie niż sferyczne nanocząstki [80].

6.2. Powtarzalność produkcji

Powtarzalność wytwarzania czujników nie jest często opisywana w literaturze, ale jest istotnie ważna, jeżeli urządzenia te mają zostać w przyszłości skomercjalizowane. Metody produkcji powinny pozwalać na precyzyjną kontrolę grubości i średnicy warstwy czulej, czy też rozmiarów nanocząstek. Jak stwierdzono powyżej, długość Debye'a nanocząstek typu rdzeń-powłoka jest na ogół tego samego rzędu co grubość samej powłoki. Dlatego też, jeśli grubość powłoki będzie kilka nanometrów większa niż długość Debye'a, odpowiedź czujnika może być znacznie niższa niż powinna. Najwyższy stopień kontroli syntezy systemów rdzeń-powłoka zapewnia osadzanie plazmowe (np. rozpylanie katodowe, PLD i ALD) [2]. Jednak dużo tańsze metody, takie jak *spin-coating* oraz *dip-coating*, mogą być wystarczająco dobre dla niektórych zastosowań. Rozwój nowych sposobów przygotowania systemów rdzeń-powłoka pozwalających na kontrolę rozmiarów nanocząstek i grubości ich powłok będzie miał zatem znaczenie w nadchodzących latach.

6.3. Integracja z platformami MEMS

Coraz częściej czujniki gazowe konstruowane są przy wykorzystaniu platform opracowanych w technologii MEMS (ang. *microelectromechanical system*), z którą nie wszystkie techniki przygotowania warstw czułych są kompatybilne [2, 81]. Na przykład, popularna metoda sitodruku (ang. *screen-printing*) jest bardzo trudna do zastosowania w przypadku układów MEMS, w których rejon depozycji jest mniejszy niż $100\ \mu\text{m} \times 100\ \mu\text{m}$. Nanocząstki tlenków metali mogą być jednak precyzyjnie osadzone w postaci cienkich warstw metodami, takimi jak druk atramentowy (ang. *ink-jet printing*). Należy jednak pamiętać, że aglomeraty większe niż $5\ \mu\text{m}$ mogą blokować głowicę drukującą. Inne metody, tj. CVD czy PLD są również ciekawym rozwiązaniem pod warunkiem zastosowania odpowiednich technik maskowania sąsiadujących z warstwą czułą elementów czujnika.

6.4. Zrozumienie mechanizmów detekcji

Liczne przykłady zebrane w rozdziale 5 niniejszej pracy wskazują jasno, że odpowiedni dobór morfologii i składu chemicznego heterozłącza pozwala na kontrolę właściwości warstwy czułej i co za tym idzie na projektowanie czujników do ściśle określonych zastosowań. Dlatego też w miarę pojawiania się nowych technik wytwarzania złożonych nanostruktur oraz projektowania czujników gazowych na ich bazie zrozumienie podstawowych mechanizmów detekcji stało się niezwykle istotne. Jednak zbyt duża liczba możliwości nie pozwala na zastosowanie czysto empirycznego podejścia do rozwiązania tego problemu. Zaobserwować można obecnie dwa trendy w badaniach mechanizmów detekcji gazów z wykorzystaniem heterozłącz. Przede wszystkim kładzie się nacisk na badania z wykorzystaniem pojedynczych nanowłókien w celu zmniejszenia ilości zmiennych oraz ograniczenia wpływu różnego rodzaju czynników na zachowanie czujnika [80]. Poza tym, mechanizmy detekcji z wykorzystaniem heterozłącz badane są także intensywnie przez społeczność naukową zajmującą się katalizą [2]. W najbliższych latach praca interdyscyplinarna powinna przynieść odpowiedzi na wiele nurtujących nas obecnie pytań.

7. Podsumowanie

W pracy przedstawiono sposoby wytwarzania heterozłącz w nanoskali oraz ich zastosowania jako warstwy czułe w czujnikach gazowych. W pracy wykazano, że na wydajność detekcji wpływa nie tylko rodzaj użytych półprzewodników do konstrukcji heterozłącza (n - n , p - n , p - p), czy też jego morfologia (A/B, A-B, A@B), ale także inne aspekty, tj. charakter kwasowo-zasadowy warstwy czułej, obecność efektu *spill-over*, inwersja sygnału (n - p , czy też n - p - n). Nie jest to oczywiście wyczerpująca lista możliwości, jakie dostarczają nam heterozłącza w projektowaniu czujników gazu. Warto zwrócić uwagę, że omówiono tutaj jedynie możliwości, jakie dają systemy oparte wyłącznie na tlenkach metali. Celowo pominięto systemy mieszane, w których wykorzystuje się, obok tlenków metali, nanostruktury typu grafen, $g\text{-C}_3\text{H}_4$, czy też MoS_2 itp., gdyż jest to temat zbyt obszerny, by mógł zostać jednocześnie zawarty w tej monografii.

Niniejszy przegląd ukazuje także jak skomplikowane mechanizmy interakcji między warstwą czułą a analizowanym gazem mają miejsce w czasie detekcji. Badania w najbliższym czasie będą napędzane m.in. chęcią dokładnego zrozumienia tychże mechanizmów oraz dalszą poprawą wydajności detekcji czujników na bazie heterozłącz. Szczególny nacisk kładziony będzie na uzyskanie stabilności odpowiedzi czujników w czasie oraz powtarzalność ich produkcji.

Podziękowania

Autorka pracy otrzymała dofinansowanie w unijnego programu badań i innowacji Horyzont 2020 w ramach umowy o grant Marie Skłodowska-Curie nr 101033564.

Literatura

1. Kumar P., Morawska L., Martani C., Martini C., Bisk G., Neophyou M., Di Sabatino S., Bell M., Norford L., Britter R., *The rise of low-cost sensing for managing air pollution in cities*, Environmental International, 75, 2015, s. 199-205.
2. Miller D.R., Akbar S.A., Morris P.A., *Nanoscale metal oxide-based heterojunctions for gas sensing: A review*, Sensors and Actuators B: Chemical, 204, 2014, s. 250-272.
3. Meng F.-J., Xin R.-F., Li S.-X., *Metal Oxide Heterostructures for Improving Gas Sensing Properties: A Review*, Materials, 16(1), 2023, s. 263.
4. Sun Y.F., Liu S.B., Meng F.L., Liu J.Y., Jin Z., Kong L.T., Liu J.H., *Metal oxide nanostructures and their gas sensing properties: a review*, Sensors (Basel), 12, 2012, s. 2610-2631.
5. Jońca J., *Nanocząstki tlenków metali i ich zastosowanie w konstrukcji czujników gazów*, [w:] *Rozwiązania technologiczne XXI wieku – skutki i perspektywy rozwoju. Tom 1*, Kamil Maciąg, Joanna Kozłowska (red.), Wydawnictwo Naukowe TYGIEL Sp. z o.o., Lublin 2022, s. 104-127.
6. Barsan N., Schweizer-Berberich M., Göpel W., *Fundamental and practical aspects in the design of nanoscaled SnO₂ gas sensors: a status report*, Fresenius' Journal of Analytical Chemistry, 365, 1999, s. 287-304.
7. Choi S.J., Kim I.D., *Recent developments in 2D nanomaterials for chemiresistive-type gas sensors*, Electronic Materials Letters, 14, 2018, s. 221-260.
8. Jońca J., Fajerweg K., Kahn M., Menini Ph., Sówka I., Fau P., *Towards selective and sensitive detection of carbon monoxide with CuO/ZnO heterojunction nanocomposite prepared by an organometallic approach*, Sensors & Transducers, 2022, 259(5), s. 99-108.
9. Kim K.-W., Cho P.-S., Kim S.-J., Lee J.-H., Kang C.-Y., Kim J.-S., Yoon S.-J., *The selective detection of C₂H₅OH using SnO₂-ZnO thin film gas sensors prepared by combinatorial solution deposition*, Sensors and Actuators B: Chemical, 123(1), 2007, s. 318-324.
10. Tan W., Tan J., Fan L., Yu Z., Qian J., Huang X., *Fe₂O₃-loaded NiO nanosheets for fast response/recovery and high response gas sensor*, Sensoras and Actuators B: Chemical, 256, 2018, s. 282-293.
11. Kim J.H., Lee J.H., Kim J.Y., Mirzaei A., Kim H.W., Kim S.S., *Enhancement of CO and NO₂ sensing in n-SnO₂-p-Cu₂O core-shell nanofibers by shell optimization*, Journal of Hazardous Materials, 376, 2019, s. 68-82.
12. Ivanovskaya M., Kotsikau D., Faglia G., Nelli P., Irkaev S., *Gas-sensitive properties of thin film heterojunction structures based on Fe₂O₃-In₂O₃ nanocomposites*, Sensors and Actuators B: Chemical 93, 2003, s. 422-430.
13. de Lacy Costello B.P.J., Ewen R., Jones P.R., Ratcliffe N., Wat R.K., *A study of the catalytic and vapour-sensing properties of zinc oxide and tin dioxide in relation to 1-butanol and dimethyldisulphide*, Sensors and Actuators B: Chemical, 61, 1999, s. 199-207.
14. de Lacy Costello B.P., Ewen R., Guernion N., Ratcliffe N., *Highly sensitive mixed oxide sensors for the detection of ethanol*, Sensors and Actuators B: Chemical, 87, 2002, s. 207-210.
15. Kahn M.L., Monge M., Colliere V., Senocq F., Maisonnat B., Chaudret B., *Size- and Shape-Control of Crystalline Zinc Oxide Nanoparticles: A New Organometallic Synthetic Method*, Advanced Functional Materials, 15(3), 2005, s. 458-468.
16. Motaung D.E., Mhlongo G.H., Makgwane P.R., Dhonge B.P., Cummings F.R., Swart H.C., Ray S.S., *Ultra-high sensitive and selective H₂ gas sensor manifested by interface of n-n*

- heterostructure of CeO₂-SnO₂ nanoparticles*, Sensors and Actuators B: Chemical, 254, 2018, s. 984-995.
17. Shaposhnik D., Pavelko D.R., Llobet E., Gispert-Guirado F., Vilanova X., *Hydrogen sensors on the basis of SnO₂-TiO₂ systems*, Engineering Proceedings, 25, 2011, s. 1133-1136.
 18. Kim K.-W., Cho P.S., Kim S.J., Lee J.H., Kang C.Y., Kim J.S., Yoon S.K., *The selective detection of C₂H₅OH using SnO₂-ZnO thin film gas sensors prepared by combinatorial solution deposition*, Sensors and Actuators B: Chemical, 123, 2007, s. 318-324.
 19. Ivanovskaya M., Kotsikau D., Faglia G., Nelli P., *Influence of chemical composition and structural factors of Fe₂O₃/In₂O₃ sensors on their selectivity and sensitivity to ethanol*, Sensors and Actuators B: Chemical, 96, 2003, s. 498-503.
 20. Dandeneau C.S., Jeon Y.H., Shelton C.T., Plant T.K., Cann D.P., Gibbons B.J., *Thin film chemical sensors based on p-CuO/n-ZnO heterocontacts*, Thin Solid Films, 517, 2009, s. 4448-4454.
 21. Kosci I., Hotovy I., Rehacek V., Griesseler R., Predanocny M., Wilke M., Spiess L., *Sputtered TiO₂ thin films with NiO additives for hydrogen detection*, Applied Surface Science, 269, 2013, s. 110-115.
 22. Cui G., Zhang M., Zou G., *Resonant tunneling modulation in quasi-2D Cu₂O/SnO₂ p-n horizontal-multi-layer-heterostructure for room temperature H₂S sensor application*, Scientific Reports, 3, 2013, s. 1-8.
 23. Cui G., Gao L., Yao B., Wang S., Zhang P., Zhang M., *Electrochemistry of CuO/In₂O₃ p-n heterojunction nano/microstructure array with sensitivity to H₂ at and below room-temperature*, Electrochemical Communications 30, 2013, s. 42-45.
 24. Maziarz, W. *TiO₂/SnO₂ and TiO₂/CuO thin film nano-heterostructures as gas sensors*, Applied Surface Science 480, 2019, s. 361-370.
 25. Sendi A., Fau P., Fajerweg P., Kahn M.L., Menini Ph., *Detection and Discrimination of Formaldehyde with CuO/SnO₂ dual layers MOS Gas Sensors Operated with a Pulsed Temperature Modulation*, Conference: 6th International Conference on Sensors Engineering and Electronics Instrumentation Advances (SEIA' 2020), Porto, Portugal.
 26. Kwon H., Yoon J.-S., Dong Y.L., Kim Y., Beak C.-K., Kim J.Y., *An array of metal oxides nanoscale hetero p-n junctions toward designable and highly-selective gas sensors*, Sensors and Actuators B: Chemical, 255(2), 2018, s. 1663-1670.
 27. Kim H.-R., Haensch A., Kim I.D., Barsan N., Weimar U., Lee J.-H., *The Role of NiO Doping in Reducing the Impact of Humidity on the Performance of SnO₂-Based Gas Sensors: Synthesis Strategies, and Phenomenological and Spectroscopic Studies*, Advanced Functional Materials, 21(23), 2011, s. 4456-4463.
 28. Jianan Z.L., Wang D.L., Teng L.W., Zhang F.T., *Toluene and ethanol sensing performances of pristine and PdO-decorated flower-like ZnO structures*, Sensors and Actuators B: Chemical, 176, 2013, s. 323-329.
 29. Suh J.M., Sohn W., Shim Y.S., Choi J.S., Song J.G., Kim T.L., Jeon J.M., Kwon K.C., Choi K.S., Kang C.Y., Byun H.G., Jang H.W., *p-p Heterojunction of Nickel Oxide Decorated Cobalt Oxide Nanorods for Enhanced Sensitivity and Selectivity toward Volatile Organic Compounds*, ACS Applied Materials & Interfaces, 10(1), 2018, s. 1050-1058.
 30. Choi S.-W., Park J.Y., Kim S.S., *Synthesis of SnO₂-ZnO core-shell nanofibers via a novel two-step process and their gas sensing properties*, Nanotechnology, 20, 2009, 465603.
 31. Zichao Q.X., Xiaopan Z., Shuai S., Zhiwen Y., Hongyan Q., Cao X.B., *Improving the triethylamine sensing performance based on debye length: A case study on α -Fe₂O₃@NiO(CuO) core-shell nanorods sensor working at near room-temperature*, Sensors and Actuators B: Chemical, 245, 2017, s. 375-385.
 32. Chang H.K., Ko D.-S., Cho D.H., Kim S., Lee H.-N., Lee H.S., Kim H.-J., Park T.J., Park Y.M., *Enhanced response of the photoactive gas sensor on formaldehyde using porous*

- SnO₂@TiO₂ heterostructure driven by gas-flow thermal evaporation and atomic layer deposition*, *Ceramics International*, 47(5), 2021, s. 5985-5992.
33. Huang H., Gong H., Chow C.L., Guo J., White T.J., Tse M.S., Tan O.K., *Low-Temperature Growth of SnO₂ Nanorod Arrays and Tunable n-p-n Sensing Response of a ZnO/SnO₂ Heterojunction for Exclusive Hydrogen Sensors*, *Advances Functional Materials*, 21(14), 2011, s. 2680-2686.
 34. Zeng W., Liu Y., Mei J., Tang C., Luo K., Li S., Zhan H., He Z., *Hierarchical SnO₂-Sn₃O₄ heterostructural gas sensor with high sensitivity and selectivity to NO₂*, *Sensors and Actuators B: Chemical*, 301, 2019, 127010.
 35. Wang Y., Qu F., Liu J., Wang Y., Zhou J., Ruan S., *Enhanced H₂S sensing characteristics of CuO-NiO core-shell microspheres sensors*, *Sensors and Actuators B: Chemical*, 209, 2015, s. 515-523.
 36. Jinkawa T., Sakai G., Tamaki J., Miura N., Yamazoe N., *Relationship between ethanol gas sensitivity and surface catalytic property of tin oxide sensors modified with acidic or basic oxides*, *Journal of Molecular Catalysis A: Chemical*, 155, 2000, s. 193-200.
 37. Maekawa T., Tamaki J., Miura N., Yamazoe N., *Development of SnO₂-based ethanol gas sensor*, *Sensors Actuators B: Chemical*, 9, 1992, s. 63-69.
 38. Rumyantseva M., Kovalenko V., Gaskov A., Makshina E., Yuschenko V., Ivanova I., Ponzoni A., Faglia G., Comini E., *Nanocomposites SnO₂/Fe₂O₃: sensor and catalytic properties*, *Sensors Actuators B: Chemical*, 118, 2006, s. 208-214.
 39. Maldotti A., Molinari A., *Design of heterogeneous photocatalysts based on metal oxides to control the selectivity of chemical reactions*, *Photocatalysis*, 303, 2011, s. 185-216.
 40. Gao P.-X., Shimpi P., Gao H., Liu C., Guo Y., Cai W., Liao K.-T., Wrobel G., Zhang Z., Ren Z., Lin H.-J., *Hierarchical assembly of multifunctional oxide-based composite nanostructures for energy and environmental applications*, *International Journal of Molecular Sciences*, 13, 2012, s. 7393-7423.
 41. Verma M.K., Gupta V., *SnO₂-CuO nanocomposite thin film sensor for fast detection of H₂S gas*, *Journal of Experimental Nanoscience*, 8, 2013, s. 326-331.
 42. Chowdhuri A., Gupta V., Sreenivas K., Kumar R., Mozumdar S., Patanjali P.K., *Response speed of SnO₂-based H₂S gas sensors with CuO nanoparticles*, *Applied Physical Letters*, 84, 2004, 1180.
 43. Kim S.-J., Na C.W., Hwang I.-S., Lee J.-H., *One-pot hydrothermal synthesis of CuO-ZnO composite hollow spheres for selective H₂S detection*, *Sensors and Actuators B: Chemical*, 168, 2012, s. 83-89.
 44. Kim S.S., Na H.G., Choi S.-W., Kwak D.S., Kim H.W., *Novel growth of CuO functionalized, branched SnO₂ nanowires and their application to H₂S sensors*, *Journal of Physics D: Applied Physics*, 45, 2012, 205301.
 45. Shao F., Hoffmann M.W.G., Prades J.D., Zamani R., Arbiol J., Morante J.R., Varechkina E., Rumyantseva M., Gaskov A., Giebelhaus I., Fischer T., Mathur S., Hernández-Ramírez F., *Heterostructured p-CuO (nanoparticle)/n-SnO₂ (nanowire) devices for selective H₂S detection*, *Sensors and Actuators B: Chemical*, 181, 2013, s. 130-135.
 46. Chowdhuri A., Gupta V., Sreenivas K., Kumar R., Mozumdar S., Patanjali P.K., *Response speed of SnO₂-based H₂S gas sensors with CuO nanoparticles*, *Applied Physical Letters*, 84, 2004, 1180.
 47. Gong H., Zhao C., Wang F., *On-chip growth of SnO₂/ZnO core-shell nanosheet arrays for ethanol detection*, *IEEE Electron Device Letters*, 39, 2018, s. 1065-1068.
 48. Fu D., Zhu C., Zhang X., Li C., Chen Y., *Two-dimensional net-like SnO₂/ZnO heteronanostructures for high-performance H₂S gas sensor*, *Journal of Materials Chemistry A*, 4, 2016, s. 1390-1398.

49. Zhao L., Chen Y., Li X., Lia X., Lin S., Lib T., Rumyantseva M.N., Gaskov A.M., *Room temperature formaldehyde sensing of hollow SnO₂/ZnO heterojunctions under UV-LED activation*, IEEE Sensors Journal, 19, 2019, s. 7207-7214.
50. Zhang Z., Xu M., Liu L., Ruan X., Yan J., Zhao W., Yun J., Wang Y., Qin S., Zhang T., *Novel SnO₂@ZnO hierarchical nanostructures for highly sensitive and selective NO₂ gas sensing*, Sensors and Actuators B: Chemical, 257, 2018, s. 714-727.
51. da Silva L.F.F., M'Peco J.-C., Catto A.C., Bernardini S., Mastelaro V.R., Aguir K., Ribeiro C., Longo E., *UV-enhanced ozone gas sensing response of ZnO-SnO₂ heterojunctions at room temperature*, Sensors and Actuators B: Chemical, 240, 2017, s. 573-579.
52. Xu H., Ju J., Li W., Zhang J., Wang J., Cao B., *Superior triethylamine-sensing properties based on TiO₂/SnO₂ n-n heterojunction nanosheets directly grown on ceramic tubes*, Sensors and Actuators B: Chemical, 228, 2016, s. 634-642.
53. Li F., Gao X., Wang R., Zhang T., Lu G., *Study on TiO₂-SnO₂ core-shell heterostructure nanofibers with different work function and its application in gas sensor*, Sensors and Actuators B: Chemical, 248, 2017, s. 812-819.
54. Motaung D.E., Mhlongo G.H., Makgwane P.R., Dhonge B.P., Cummings F.R., Swart H.C., Ray S.S., *Ultra-high sensitive and selective H₂ gas sensor manifested by interface of n-n heterostructure of CeO₂-SnO₂ nanoparticles*, Sensors and Actuators B: Chemical, 254, 2018, s. 984-995.
55. Naderi H., Hajati S., Ghaedi M., Espinos J.P., *Highly selective few-ppm NO gas-sensing based on necklace-like nanofibers of ZnO/CdO n-n type I heterojunction*, Sensors and Actuators B: Chemical, 297, 2019, 126774.
56. Chen Y.J., Xiao G., Wang T.S., Zhang F., Ma Y., Gao P., Zhu C.L., Zhang E., Xu Z., Li Q.H., *α -MoO₃/TiO₂ core/shell nanorods: controlled-synthesis and low-temperature gas sensing properties*, Sensors and Actuators B: Chemical, 155, 2011, s. 270-277.
57. Vishnuraj R., Karuppanan K.K., Aleem M., Pullithadathil B., *Boosting the performance of NO₂ gas sensors based on n-n type mesoporous ZnO@In₂O₃ heterojunction nanowires: in situ conducting probe atomic force microscopic elucidation of room temperature local electron transport*, Nanoscale Advances, 2, 2020, s. 4785-4797.
58. Li G., Zhang X., Lu H., Yan C., Chen K., Lu H., Gao T., Yang Z., Zhu G., Wang C., He Z., *Ethanol sensing properties and reduced sensor resistance using porous Nb₂O₅-TiO₂ n-n junction nanofibers*, Sensors and Actuators B: Chemical, 283, 2019, s. 602-612.
59. Yin M., Wang F., Fan H., Xu L., Liu S., *Heterojunction CuO@ZnO microcubes for superior p-type gas sensor application*, Journal of Alloys and Compounds, 672, 2016, s. 374-379.
60. Ju D., Xu H., Qiu Z., Guo J., Zhang J., Cao B., *Highly sensitive and selective triethylamine-sensing properties of nanosheets directly grown on ceramic tube by forming NiO/ZnO PN heterojunction*, Sensors and Actuators B: Chemical, 200, 2014, s. 288-296.
61. Kim J.H., Lee J.H., Mirzaei A., Kim H.W., Kim S.S., *SnO₂(n)-NiO(p) composite nanowires: gas sensing properties and sensing mechanisms*, Sensors and Actuators B: Chemical, 258, 2018, s. 204-214.
62. Ju D., Xu H., Xu Q., Gong H., Qiu Z., Guo J., Zhang J., Cao B., *High triethylamine-sensing properties of NiO/SnO₂ hollow sphere P-N heterojunction sensors*, Sensors and Actuators B: Chemical, 215, 2015, s. 39-44.
63. Alev O., Sennik E., Öztürk Z.Z., *Improved gas sensing performance of p-copper oxide thin film/n-TiO₂ nanotubes heterostructure*, Journal of Alloys and Compounds, 749, 2018, s. 221-228.
64. Yu H., Yang T., Wang Z., Li Z., Zhao Q., Zhang M., *p-N heterostructural sensor with SnO-SnO₂ for fast NO₂ sensing response properties at room temperature*, Sensors and Actuators B: Chemical, 258, 2018, s. 517-526.

65. Li N., Fan Y., Shi Y., Xiang Q., Wang X., Xu J., *A low temperature formaldehyde gas sensor based on hierarchical SnO/SnO₂ nano-flowers assembled from ultrathin nanosheets: synthesis, sensing performance and mechanism*, Sensors and Actuators B: Chemical, 294, 2019, s. 106-115.
66. Park S., Kheel H., Sun G.J., Kim H.W., Ko T., Lee C., *Room-temperature hydrogen gas sensing properties of the networked Cr₂O₃-functionalized Nb₂O₅ nanostructured sensor*, Metals and Materials International, 22, 2016, s. 730-736.
67. Yue X.J., Hong T.S., Yang Z., Huang S.P., *Room temperature H₂S micro-sensors with anti-humidity properties fabricated from NiO-In₂O₃ composite nanofibers*, Chinese Science Bulletin, 58, 2013, s. 821-826.
68. Liu L., Wang Y., Dai Y., Li G., Wang S., Li T., Zhang T., Qin S., *In situ growth of NiO@SnO₂ hierarchical nanostructures for high performance H₂S sensing*, ACS Applied Materials and Interfaces, 11, 2019, s. 44829-44836.
69. Xu H., Zhang J., Rehman A.U., Gong L., Kan K., Li L., Shi K., *Synthesis of NiO@CuO nanocomposite as high-performance gas sensing material for NO₂ at room temperature*, Applied Surface Sciences, 412, 2017, s. 230-237.
70. Su C., Zhang L., Han Y., Ren C., Zeng M., Zhou Z., Su Y., Hu N., Wei H., Yang Z., *Controllable synthesis of heterostructured CuO-NiO nanotubes and their synergistic effect for glycol gas sensing*, Sensors and Actuators B: Chemical, 304, 2020, 127347.
71. Adamu B.I., Falak A., Tian Y., Tan X., Meng X., Chen P., Wang H., Chu W., *p-p Heterojunction Sensors of p-Cu₃Mo₂O₉ Micro/Nanorods Vertically Grown on p-CuO Layers for Room-Temperature Ultrasensitive and Fast Recoverable Detection of NO₂*, ACS Applied Materials & Interfaces, 2020, 12, 7, s. 8411-8421.
72. Wu H., Zhou Y., Liupeng J.G., Wang T., Yan X., Wang C., Liu F., Sun P., Lu G., *Highly sensitive and selective xylene sensor based on p-p heterojunctions composites derived from off-stoichiometric cobalt tungstate*, Sensors and Actuators B: Chemical, 351, 2022, 130973.
73. Wang D., Gu K., Zhang J., Xing D., Zhang M., *MOF-derived NiWO₄@NiO p-p heterostructure for distinguish detection of TEA and xylene by temperature regulation*, Journal of Alloys and Compounds, 875, 2021, 160015.
74. Fang H., Li S., Zhao H., Deng J., Wang D., Li J., *Enhanced NO₂ gas sensing performance by hierarchical CuO-Co₃O₄ spheres*, Sensors and Actuators B, 352(2) 2022, 131068.
75. Song L., Ahn J., Xu L., Beak W.J., Shin E., Kim I.-D., *Facile Synthesis of Co₃O₄/CoMoO₄ Heterostructure Nanosheets for Enhanced Acetone Detection*, ACS Sensors, 2022, 7, 11, s. 3540-3550.
76. Song Z., Zhang L., Zhou Q., Zhang Z., Dong Z., Nie L., Liu Q., Pan G., *In-situ synthesis of needle-like PdO-decorated NiO thin films on Al₂O₃ substrates for high-performance H₂ sensors*, Ceramics International, 48(21), 2022, s. 31746-31754.
77. Dun M., Tang M., Zhao D., Li X., Huang X., *Synergistic effect of PdO and parallel nanowires assembled CuO microspheres enables high performance room-temperature H₂S sensor*, Sensors and Actuators B: Chemical, 358, 2022, 131520.
78. Kumar N., Haviar S., Zeman P., *Three-Layer PdO/CuWO₄/CuO System for Hydrogen Gas Sensing with Reduced Humidity Interference*, Nanomaterials, 11(12), 2021, 3456.
79. Eranna G., Joshi B.C., Runthala D.P., Gupta R.P., *Oxide materials for development of integrated gas sensors – a comprehensive review*, Critical Reviews in Solid State and Materials Sciences, 29, 2004, s. 111-188.
80. Arafat M.M., Dinan B., Akbar S.A., Haseeb A.S.M.A., *Gas sensors based on one dimensional nanostructured metal-oxides: a review*, Sensors 12, 2012, s. 7207-7258.
81. Vander Wal R.L., Berger G.M., Kulis M.J., Hunter G.W., Xu J.C., Evans L., *Synthesis methods, microscopy characterization and device integration of nanoscale metal oxide semiconductors for gas sensing*, Sensors (Basel), 9, 2009, s. 7866-7902.

Czujniki gazu na bazie heterozłączy tlenków metali zaprojektowanych w nanoskali

Streszczenie

Zastosowanie zminiaturyzowanych czujników gazu wzbudza coraz większe zainteresowanie w takich dziedzinach, jak monitorowanie środowiska, bezpieczeństwo, diagnostyka medyczna i rolnictwo. Detektory oparte na półprzewodnikowych tlenkach metali (MOS) są najpopularniejszym wyborem w zastosowaniach komercyjnych ze względu na swoje liczne zalety w postaci długiego czasu życia, niskiej ceny, małych rozmiarów i wysokiej czułości. Na działanie czujników mają wpływ różne parametry, w tym temperatura pracy urządzenia, fizyczne i chemiczne właściwości materiałów czujnikowych (np. wielkość ziarna, gęstość defektów i obecność luk tlenowych), czy konstrukcja samego detektora (np. geometria elektrod). Jednym z rozwiązań badanych intensywnie w ostatnim czasie jest wykorzystanie osiągnięć nanotechnologii w celu zaprojektowania warstw czułych w nanoskali. Co więcej, tworzenie heterozłączy na bazie nanomateriałów MOS (tzw. heteronanostruktur) może wywierać dodatkowy wpływ na właściwości czujników i poprawiać ich parametry, co zostało szczegółowo przeanalizowane. W pracy przedstawiono zatem kilka koncepcji w projektowaniu czujników gazu na bazie heteronanostruktur. Do każdej z nich przypisano po kilka przykładów oraz szczegółowo omówiono mechanizm detekcji. Ponadto omówiono zarówno wpływ wzajemnego połączenia nanomateriałów, jak i sposobu ich umieszczenia na elektrodach na wydajność czujnika. W pracy skupiono się na wynikach badań własnych i porównano je z osiągnięciami opisanymi przez inne grupy badawcze.

Słowa kluczowe: półprzewodnikowe czujniki gazu, nanocząstki tlenków metali, heterozłącza

Metal oxide nanoparticles and theirs applications in gas sensors construction

Abstract

The use of miniaturized gas sensors is driving more and more applications, including the creation of standards for environmental protection, safety, medical diagnostics and agriculture. Metal Oxide Semiconductor (MOS) detectors are the most popular effect in application due to their many advantages in lifetime operation, effective pricing, small size and high efficiency. Sensor performance is influenced by various parameters, including the operating temperature of the devices, the physical and chemical properties of the sensor materials (e.g. grain size, appearance of defects and sources of oxygen vacancies), and the design of the electrode itself (e.g. electrode geometry). One of the most advanced scientists recently is using nanotechnology to design nanoscale sensitive layers. Moreover, the creation of a heterojunction on the device of MOS nanomaterials (so-called heteronanostructures) can have an additional impact on the properties of the sensors and improve their parameters, which allows for the analyzed results. Therefore, the paper presents several concepts within gas sensors based on the principle of heteronanostructures. Several events are assigned to each of them and the detection mechanism is discussed. In addition, the influence of the interaction of nanomaterials as well as the way they are placed on the electrodes on the efficiency is also discussed. The work focused on the results of scientific and comparative research with the results achieved by other research groups.

Keywords: semiconductor gas sensors, metal oxide nanoparticles, heterojunctions

Właściwości adsorpcyjne wybranych polarnych faz stacjonarnych

1. Wprowadzenie

Do rozdzielania i oczyszczania wielu substancji chemicznych, takich jak np.: farmaceutyki, peptydy, aminokwasy wykorzystywane są procesy adsorpcyjne. Adsorpcja zachodzi w wyniku fizycznych lub chemicznych oddziaływań pomiędzy składnikami płynu a adsorbentem (czyli porowatym ośrodkiem stałym lub inaczej fazą stacjonarną). Od właściwości adsorbentu zależy w głównej mierze mechanizm wpływający na przebieg procesu adsorpcji oraz jego efektywność zarówno w rozdzielaniu, jak i oczyszczaniu substancji chemicznych [1].

Jednym z ważniejszych zastosowań adsorbentów jest chromatografia cieczowa. Najczęściej stosowane są w tym procesie fazy stacjonarne na bazie czystej krzemionki. Wykorzystuje się je w wielu rodzajach chromatografii cieczowej, jak np. chromatografia w normalnym układzie faz (NPLC, ang. *Normal Phase Liquid Chromatography*) oraz odwróconym układzie faz (RPLC, ang. *Reverse Phase Liquid Chromatography*) [2]. Adsorbenty krzemionkowe stosowane są także do separacji substancji w chromatografii oddziaływań hydrofilowych (HILIC, ang. *Hydrophilic Interaction Liquid Chromatography*). Czysta krzemionka ma pewne szczególne zalety, takie jak np.: odporność mechaniczna, odporność na wysokie ciśnienie czy na działanie rozpuszczalników organicznych, brak pęcznienia oraz wysoką wydajność, co czyni ją obiecującą substancją, będącą bazą do otrzymywania faz stacjonarnych do chromatografii HPLC (ang. *High Performance Liquid Chromatography*) [3].

Wielkościami charakteryzującymi proces chromatograficzny są m.in. selektywność oraz czas retencji. Aby poprawić retencję i selektywność różnych typów próbek można także stosować adsorbenty krzemionkowe modyfikowane chemicznie grupami np.: diolowymi, aminowymi, amidowymi, cyklodekstrynowymi, itp. Występują także adsorbenty modyfikowane grupami zwitterjonowymi (obojnaczymi) [1, 4].

Jedną z ważniejszych obecnie technik chromatograficznych jest chromatografia HILIC. Jest ona powszechnie stosowana do rozdzielania związków polarnych i hydrofilowych. Technika ta stanowi alternatywę dla chromatografii cieczowej w odwróconym układzie faz [1, 4, 5]. Mimo że HILIC jest szeroko stosowana, mechanizmy retencji substancji chromatografowanych są złożone i wciąż nie do końca poznane. Alpert jako pierwszy wprowadził termin HILIC i zasugerował, że główny mechanizm retencji wynika z różnych oddziaływań rozpuszczalnik-rozpuszczalnik, które przyczyniają się do podziału rozpuszczalnika pomiędzy hydroorganiczną fazę ruchomą i bogatą w wodę

¹ d555@stud.prz.edu.pl, Szkoła Doktorska Politechniki Rzeszowskiej, Wydział Chemiczny, Politechnika Rzeszowska im. Ignacego Łukasiewicza, w.prz.edu.pl

² m.przywara@prz.edu.pl, Katedra Inżynierii Chemicznej i Procesowej, Wydział Chemiczny, Politechnika Rzeszowska im. Ignacego Łukasiewicza, w.prz.edu.pl

³ ichwz@prz.edu.pl, Katedra Inżynierii Chemicznej i Procesowej, Wydział Chemiczny, Politechnika Rzeszowska im. Ignacego Łukasiewicza, w.prz.edu.pl

warstwę częściowo unieruchomioną na fazie stacjonarnej. Dokładna siła różnych oddziaływań zależy od zastosowanej fazy stacjonarnej i ruchomej, ale także od właściwości analitu, czyli substancji badanej [1, 4]. Wybór fazy stacjonarnej i fazy ruchomej ma duży wpływ na retencję, co znacznie komplikuje opracowanie/optymalizację samej metody rozdzielania w HILIC [6]. Zastosowanie polarnych faz stacjonarnych w połączeniu z mieszaninami wodno-organicznymi jako eluentami, zapewnia możliwość oznaczania polarnych związków obojętnych i naładowanych, a także zapewnia lepszą selektywność w porównaniu z tradycyjnie stosowaną techniką RPLC, której możliwości w rozdzielaniu substancji polarnych są ograniczone [7]. W HILIC stosowane są fazy ruchome zawierające wysoki procent organicznego modyfikatora. Tworzą je najczęściej roztwory acetonitrylu z niewielkim udziałem procentowym (<40%) wody. Jak wspomniano, dokładny mechanizm retencji jest nadal badany [8]. Wyniki badań wskazują, że istotną rolę w tym mechanizmie odgrywają takie czynniki, jak obecność wiązań wodorowych, oddziaływania dipolowo-dipolowe, elektrostatyczne, hydrofobowe oraz (co bardzo istotne) budowa i właściwości (w tym kształt i rozmiar cząsteczek) stosowanego adsorbentu [9].

Poprzez uwzględnienie danych równowagowych procesu adsorpcji oraz właściwości adsorpcyjnych zarówno adsorbentu, jak i adsorbentu, za pomocą różnych modeli izoterm adsorpcji, można opisywać mechanizmy interakcji pomiędzy adsorbentem a adsorbentem w stałej temperaturze. Modelowanie matematyczne wykorzystujące dane równowagowe jest bardzo istotnym sposobem przewidywania mechanizmów adsorpcji różnych układów adsorpcyjnych [10]. Do opisu pracy kolumny chromatograficznej stosuje się modele, wymagające uwzględnienia izoterm adsorpcji dla danego układu. Do wyznaczania izoterm adsorpcji stosuje się metody statyczne lub dynamiczne [11]. W niniejszej pracy zastosowano metodę statyczną. Za pomocą statycznych metod wyznaczania izoterm adsorpcji mierzona jest przede wszystkim pojemność każdego adsorbentu [12]. Statyczna metoda pomiaru izoterm adsorpcji nie analizuje krzywych czasu stężenia, lecz wykorzystuje jedynie informacje o stanach równowagi. Metoda ta opiera się na określeniu różnicy stężeń składnika adsorbowanego w roztworze wyjściowym i roztworze będącym w stanie równowagi z adsorbentem. Prowadząc badania adsorpcyjne, należy pamiętać o zachowaniu stałych parametrów prowadzenia procesu. W klasycznej metodzie statycznej znane objętości lub masy adsorbentu miesza się w zamkniętym naczyniu z roztworem o objętości V , posiadającym znane stężenia początkowe roztworów. Po wystarczająco długim czasie kontaktu między tymi dwiema fazami ustala się równowaga między stężeniem badanego związku w roztworze i na powierzchni adsorbentu. Odpowiednie równowagowe ilości substancji badanej adsorbowane na jednostkę objętości lub masy adsorbentu oblicza się za pomocą równania nr 1. W celu wyznaczenia izoterm adsorpcji należy zmierzyć jedynie końcowe stężenie równowagowe.

$$q = \frac{V_0 \cdot (C_0 - C)}{V_a} \quad (1)$$

gdzie: q – stężenie substancji adsorbowanej na powierzchni adsorbentu, V_0 – objętość roztworu, V_a – objętość adsorbentu, C_0 – stężenie początkowe badanego związku, C – stężenie badanego związku po adsorpcji.

Wykonując badania dla różnych stężeń początkowych C_0 badanego związku, otrzymuje się doświadczalny przebieg izotermi adsorpcji $q = f(C)$. Swobodnymi parametrami, które mogą być zmieniane są stężenia początkowe i objętość adsorbentu. Parametry te muszą być starannie dobrane, aby pokryć szeroki obszar izotermi [13, 14].

Na proces adsorpcji, a zwłaszcza na oddziaływania pomiędzy powierzchnią fazy stałej i cząsteczkami zawartymi w cieczy, istotnie wpływa geometryczny nieporządek i chemicznie niejednorodna struktura rzeczywistych powierzchni stałych proszkowych adsorbentów. Energetycznie jednorodna powierzchnia to powierzchnia, na której każdy punkt jest równoważny pozostałym nawet na molekularnym poziomie rozdzielczości. Jednak w skali molekularnej większość powierzchni stałych charakteryzuje się różnymi układami atomów lub cząsteczek na powierzchni. Wykazują one różne składy chemiczne. Ta dyspersja energii adsorpcji jest powszechnie nazywana heterogenicznością powierzchni [11, 12].

W celu wyznaczenia rozkładu energii adsorpcji na powierzchni adsorbentów stosuje się różne metody matematyczne bazujące na danych empirycznych. Programy matematyczne są niezbędne w każdej dziedzinie nauki w celu wdrożenia modeli i wyprowadzenia użytecznych informacji fizycznych z danych eksperymentalnych. Metoda oczekiwania-maksymalizacja (EM, ang. *Expectation – Maximization*) pozwala rozwiązać podstawowe równanie, które odnosi się do rozkładu energii adsorpcji (AED, ang. *Adsorption Energy Distribution*) powierzchni i ogólnej eksperymentalnej izotermi adsorpcji [13, 14]. Metoda ta może ilościowo oszacować stopień heterogeniczności procesu adsorpcji, pod warunkiem, że dostępne są empiryczne dane adsorpcyjne, takie jak izoterma adsorpcji. Wyznaczenie AED jest możliwe z bezpośredniej analizy matematycznej danych adsorpcyjnych q^*C (równanie nr 2). Ogólny problem polega na inwersji fundamentalnego równania adsorpcji na powierzchniach heterogenicznych:

$$q^*C = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} F(\varepsilon)\Theta(\varepsilon, C)d\varepsilon \quad (2)$$

gdzie: $F(\varepsilon)$ – przedstawia funkcję AED, $\theta(\varepsilon, C)$ – homogeniczny model adsorpcji, energia adsorpcji, a \min, \max – są minimalnymi i maksymalnymi możliwymi energiami adsorpcji [15-17].

W badaniach równowagi kinetycznej najczęściej przedstawia się dane w postaci wykresu Scatcharda. Wykres Scatcharda jest najbardziej odpowiedni do jakościowej oceny informacji eksperymentalnej. Nieliniowa, wklęsła zależność Scatcharda odpowiada energetycznie heterogenicznym oddziaływaniom substancji adsorbującej z fazą stacjonarną, natomiast oddziaływania homogeniczne odpowiadają liniowemu przebiegowi tej zależności [18, 19].

Na wydajność/selektywność kolumny chromatograficznej wpływa wiele właściwości fazy stacjonarnej. Jedną z ważniejszych jest wielkość cząstek, a bardziej precyzyjnie rozkład wielkości cząstek (PSD, ang. *Particle Size Distribution*). Rozkład wielkości cząstek w kolumnach z wypełnieniem, stosowanych w HPLC zawsze był uznawany za istotny czynnik wpływający na ich wydajność. Intuicyjnie wydaje się, że pożądane jest stosowanie wąskich rozkładów ziaren materiałów wypełniających HPLC, co jest poparte badaniami pojawiającymi się w literaturze naukowej [20]. Cząstki o różnych rozmiarach różnią się powierzchnią, a także mogą różnić się strukturą porów. Proces migracji cząstek analitu do porów we wnętrzu cząstki adsorbentu może być znacznie dłuższy

dla większych cząstek. W związku z tym rozkład wielkości cząstek może wpływać na proces adsorpcji. Na proces adsorpcji może mieć wpływ również szerokość lub rozpiętość rozkładu wielkości cząstek. Szeroki rozkład wielkości cząstek ma znaczący wpływ na efektywność rozdzielania substancji [21]. Stwierdzono empirycznie, że wydajność kolumny poprawia się poprzez zawężenie rozkładu wielkości cząstek [22]. Dane literaturowe wskazują, że istotną rolę odgrywa również obecność drobnych cząstek, które w znacznym stopniu determinują wydajność kolumn chromatograficznych [23]. Rozkład wielkości cząstek może również wpływać na upakowanie cząstek w kolumnie chromatograficznej. Małe cząstki mają tendencję do osadzania się w porach otaczających większe cząstki, co prowadzi do bardzo dużej gęstości upakowania, a także negatywnie wpływa na przepuszczalność kolumny oraz proces rozdzielania substancji [24]. Analiza rozkładu wielkości cząstek w niniejszej pracy została wykonana za pomocą metody dyfrakcji laserowej (DLS, ang. *Dynamic Light Scattering*). Dyfrakcja laserowa jest skuteczną metodą pomiaru cząstek o wielkościach mikrometrycznych i nanometrycznych. Jest ona szeroko stosowana w wielu dziedzinach, w tym w materiałoznawstwie, chemii, biologii i medycynie [23, 25].

Celem pracy była analiza rozkładu wielkości cząstek adsorbentów oraz na podstawie uzyskanych wyników analiza i ocena równowagi adsorpcyjnej oraz rozkładu energii adsorpcji na badanych fazach stacjonarnych. Badania przeprowadzono z wykorzystaniem dwóch polarnych faz stacjonarnych. Równowagę adsorpcji i rozkład energii adsorpcji analizowano z użyciem dwóch związków testowych (fenolu i kofeiny) w układach HILIC, w których jako związki organiczne eluentu zastosowano acetonitryl i/lub metanol.

2. Materiały zastosowane w eksperymencie

Do badań eksperymentalnych jako substancje testowe zastosowano fenol (wyprodukowany przez SIGMA – ALDRICH, czystość >99,5%) oraz kofeinę (Fluka Analytical, czystość $\geq 99,0\%$). Badane związki należą do różnych klas chemicznych dzięki temu mogły w różny sposób oddziaływać z badanymi adsorbentami co umożliwiło szerszą i bardziej obiektywną ocenę mechanizmu retencji. Jako fazy ruchome zastosowano układy MeOH-woda i ACN-woda zawierające 95% objętości odpowiedniego rozpuszczalnika organicznego. Wszystkie rozpuszczalniki organiczne klasy chromatograficznej zakupiono w firmie Merck (czystość $\geq 99,9\%$). Jako adsorbenty polarne zastosowano LiChrospher 60 Si (12 μm) oraz Silica gel 60 (0,2-0,5 mm) firmy Merck. Do odgazowania faz ruchomych zastosowano łaźnię ultradźwiękową przez 5 minut bezpośrednio po wymieszaniu. Woda destylowana była przygotowywana na dejonizatorze wody SolPure-78Z (ELKAR). Jako ciecze dyspergujące stosowane do pomiaru rozkładu wielkości cząstek adsorbentów zastosowano wodę dejonizowaną oraz 2-propanol firmy Merck.

3. Metodyka badawcza

3.1. Pomiar rozkładu wielkości cząstek

W celu scharakteryzowania adsorbentów przeprowadzono pomiar rozkładu wielkości cząstek. Analizę przeprowadzono za pomocą urządzenia Mastersizer (Malvern Instruments, Malvern, Worcestershire, UK). Urządzenie to pozwala określać wielkość cząstek z wykorzystaniem metody dynamicznego rozpraszania światła. Do pomiarów wykorzystano tzw. metodę „na mokro”, tzn. pomiar prowadzono, dyspergując próbkę materiału proszkowego w cieczy. Jako ciecz dyspergującą zastosowano wodę dejonizowaną oraz

2-propanol. Procedura pomiarowa składała się z kilku etapów. Pierwszym było przygotowanie i przeczyszczenie komory pomiarowej, następnie uruchomienie jednostki optycznej i przystawki dyspergującej. W kolejnym etapie włączono pompę perystaltyczną po uprzednim umieszczeniu w zlewce dyspergant. Po wyosiowaniu części optycznych wykonano pomiar tzw. tła, tj. rozpraszania światła laserowego dla czystego dyspergant. W kolejnym kroku w zlewce z cieczą umieszczono odpowiednią ilość fazy stacjonarnej, tak aby wskaźnik przysłonięcia wiązki mieścił się w zakresie od 10 do 20%. Rozpoczęto analizę w momencie uzyskania żądanej wartości wskaźnika. Pomiary prowadzono przy ciągłym, intensywnym mieszaniu i stosując co najmniej 4 serie ultradźwięków po 30 s w celu rozbicia aglomeratów cząstek. Dodatkowo wyznaczono następujące parametry: $d(0,5)$ – mediana średnicy, średnica objętościowa, od której 50% objętości próbki posiada mniejsze ziarna i 50% większe, $d(0,1)$ – średnica objętościowa, od której 10% objętości próbki posiada mniejsze ziarna, $d(0,9)$ – średnica objętościowa, od której 90% objętości próbki posiada mniejsze ziarna, $d(4,3)$ – średnia średnica objętościowa, $d(3,2)$ – średnia średnica powierzchniowa, SSA – powierzchnia właściwa zewnętrzna, jednorodność rozkładu – całkowite odchylenie od mediany, span – szerokość rozkładu, obliczana następująco (równanie nr 3):

$$span = \frac{d_{90\%} - d_{10\%}}{d_{50\%}} \quad (3)$$

3.2. Wyznaczanie izoterm adsorpcji

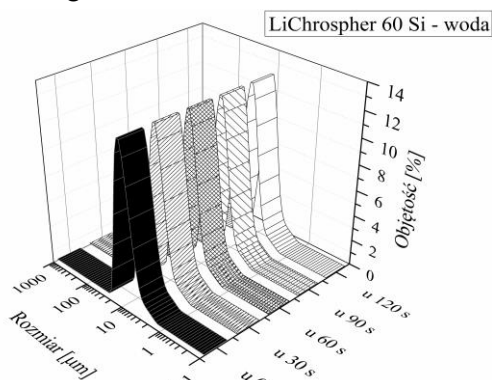
Izoterm adsorpcji wyznaczono metodą statyczną. Przygotowano różne stężenia roztworów po 3 cm³ dla fenolu i kofeiny w metanolu i acetonitrylu: dla fenolu 1; 6; 10; 20; 30; 40; 50; 60; 70; 80; 90; 100; 120 g/dm³ oraz dla kofeiny 0,1; 0,5; 0,8 ; 1; 2; 3,5; 5; 6,75; 8,25; 10,5; 12; 13,5; 15 g/dm³. W kolejnym etapie do kolb odważono adsorbenty LiChrospher 60 Si oraz Silica gel 60 w ilości 0,3 g. Następnie wspomnianymi roztworami zalewano zważone adsorbenty. W celu uzyskania odpowiedniej dynamiki procesu adsorpcji oraz stałych warunków temperatury zastosowano wycząsarkę Orbital Shaker-Incubator ES-20/60. Urządzenie to wyposażone jest w mechanizm napędu bezpośredniego umożliwiający ruch platformy, co zapewnia stabilne warunki pomiarowe dla długotrwałych eksperymentów. Prędkość obrotową ustawiono na 100 obr/min, próbki pobierano po jednej godzinie za pomocą strzykawki z filtrem o średnicy porów 0,2 μm, a następnie umieszczano w probówkach. Pomiary wykonano za pomocą wysokosprawnego chromatografu cieczonego Primaide firmy Merck-Hitachi (Niemcy), który składał się z: pompy eluentu (model 1110), termostatu kolumnowego (model 1310) oraz detektora UV (model 1410). Wszystkie pomiary przeprowadzono w warunkach izokratycznych przy przepływie eluentu 0,5 ml/min dla danego składu eluentu oraz przy stałej temperaturze 20°C. Objętość iniekcji (próbki rozpuszczonej w fazie ruchomej) wynosiła 0,02 cm³ roboczych roztworów wzorcowych. Czasy retencji odczytywano dla fenolu przy długości fali 294 nm, natomiast dla kofeiny przy 303 nm.

4. Analiza wyników

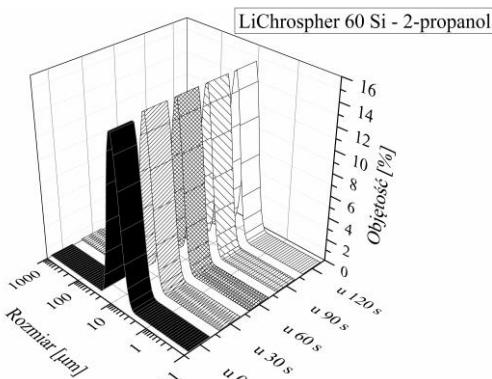
4.1. Wyniki pomiarów rozkładu wielkości ziaren

W pierwszym etapie prac przeprowadzono analizę rozkładu wielkości ziaren badanych adsorbentów. Do pomiarów wykorzystano tzw. metodę „na mokro”, zawieszając

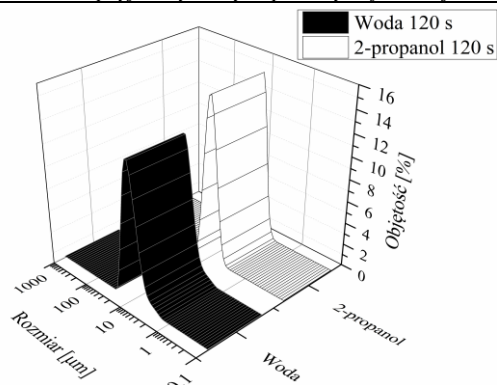
cząstki materiału proszkowego w cieczy dyspergującej. Ciecz dyspergująca miała za zadanie odpowiednio zwilżenie ziaren oraz uzyskanie zawiesiny niezawierającej aglomeratów. W związku z zaobserwowaną tendencją materiału sypkiego do aglomeracji zastosowano dodatkowe ekspozycje zawiesin próbek na działanie ultradźwięków. Ultradźwięki pomogły w rozbiciu aglomeratów cząstek i uzyskaniu równomiernej zawiesiny. Rysunki 1 oraz 2 przedstawiają wpływ ekspozycji próbki LiChrospher60 Si na działanie ultradźwięków. Zastosowano cztery serie ultradźwięków oznaczanych jako u , po 30 s każda. Zmiany rozkładu wielkości ziaren dla zawiesiny w wodzie oraz 2-propanolu w czasie były niewielkie, po czasie ekspozycji na ultradźwięki równym 60 s nie obserwowano dalszych zmian. Zmiany zaobserwowano natomiast, porównując zawiesinę w wodzie oraz 2-propanolu (rys. 3). W wodzie zaobserwowano większą ilość cząstek o większych ziarnach, co świadczy o możliwym występowaniu aglomeratów cząstek. W związku z tym jako ostateczny rozkład należy traktować ten uzyskany w 2-propanolu, w którym nie obserwowano aglomeratów.



Rysunek 1. Wpływ czasu ekspozycji na ultradźwięki (u) dla próbki LiChrospher 60 Si zawieszonyj w wodzie na rozkład wielkości ziaren

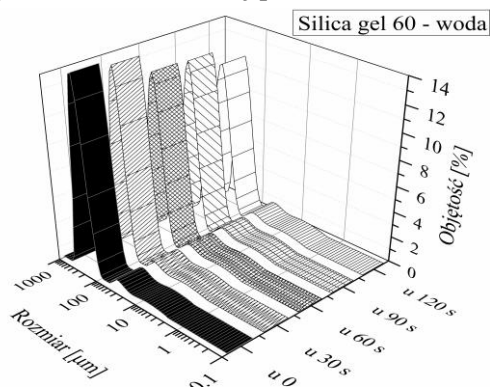


Rysunek 2. Wpływ czasu ekspozycji na ultradźwięki (u) dla próbki LiChrospher 60 Si zawieszonyj w 2-propanolu na rozkład wielkości ziaren



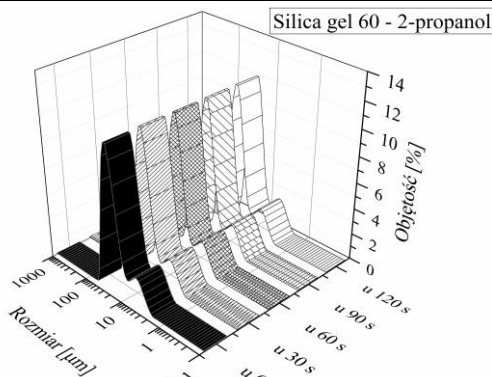
Rysunek 3. Porównanie rozkładu wielkości ziaren próbki LiChrospher 60 Si zawieszonyj w wodzie i 2-propanolu

Podobną analizę wpływu ekspozycji na ultradźwięki na rozkład wielkości ziaren przeprowadzono dla adsorbentu Silica gel 60 zawieszonyj w wodzie oraz 2-propanolu. Na rysunku 4 są widoczne zmiany rozkładu w czasie ekspozycji na ultradźwięki. Są one wyraźne i wraz ze wzrostem czasu ekspozycji wskazują na zmniejszanie rozmiaru ziaren, tzn. rozbijanie aglomeratów w badanej próbce.



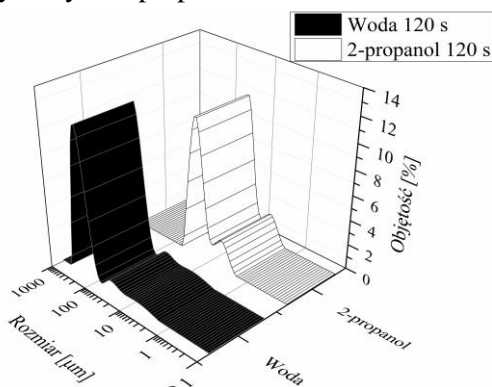
Rysunek 4. Wpływ czasu ekspozycji na ultradźwięki (u) dla próbki Silica gel 60 zawieszonyj w wodzie na rozkład wielkości ziaren

W przypadku 2-propanolu (rys. 5) zaobserwowano niewielkie zmiany rozkładu wielkości ziaren w czasie ekspozycji na ultradźwięki, co świadczy o braku aglomeratów w próbce.



Rysunek 5. Wpływ czasu ekspozycji na ultradźwięki (u) dla próbki Silica gel 60 zawieszonyj w 2-propanolu na rozkład wielkości ziaren

Rysunek 6 przedstawia porównanie rozkładu wielkości ziaren dla próbki Silica gel 60. W przypadku tego adsorbentu zaobserwowano wyraźne różnice pomiędzy próbkami zawieszonymi w wodzie i 2-propanolu. Próbki zawieszony w wodzie charakteryzowały się obecnością większych ziaren, co świadczy o występowaniu aglomeratów cząstek. W przypadku 2-propanolu nie zaobserwowano aglomeratów. Można zatem stwierdzić, że podobnie jak w przypadku adsorbentu LiChrospher 60 Si jako ostateczny rozkład należy traktować ten uzyskany w 2-propanolu.



Rysunek 6. Porównanie rozkładu wielkości ziaren próbki Silica gel 60 zawieszonyj w wodzie i 2-propanolu

W tabeli 1 przedstawiono porównanie wybranych wielkości charakteryzujących rozkład wielkości ziaren badanych próbek. Wielkości zestawiono dla wyników ostatecznych, tzn. uzyskanych po czasie 120 s ekspozycji na ultradźwięki. Badane adsorbenty wykazują znaczne różnice we właściwościach związanych z rozkładem wielkości ziaren. LiChrospher 60 Si charakteryzuje się znacznie mniejszymi rozmiarami ziaren z medianą średnicy na poziomie 8,61-8,80 μm w zależności od zastosowanej cieczy dyspergującej. Adsorbent Silica gel 60 wykazuje znacznie większe wartości mediany średnicy 247,5-233,9 μm . Również pozostałe średnie średnice są znacznie większe w przypadku Silica gel 60. Silica gel 60 ma znacznie większą wartość szerokości rozdziału, obecność cząstek znacznie różniących się rozmiarem może mieć wpływ na proces adsorpcji. Ponadto Silica gel 60 ma znacznie mniejszą powierzchnię zewnętrzną, przez co proces adsorpcji może przebiegać mniej wydajnie. W przypadku zastosowania róż-

nych cieczy dyspergujących zaobserwowano niewielkie różnice w wartościach wyznaczonych właściwości dla LiChrospher 60 Si, natomiast dla Silica gel 60 różnice były wyraźniejsze. Zaobserwowano znacznie większe wartości uzyskanych parametrów dla pomiarów w wodzie, co może świadczyć o występowaniu aglomeratów cząstek.

Tabela 1. Wybrane właściwości badanych adsorbentów

Adsorbent	d(0,1) [μm]	d(0,5) [μm]	d(0,9) [μm]	Span [-]	d(4,3) [μm]	d(3,2) [μm]	SSA [m ² /g]	Jedno- rodność [-]
LiChrospher 60 Si – woda	5,72	8,61	12,82	0,82	9,00	8,18	0,73	0,26
LiChrospher 60 Si – 2-propanol	6,34	8,80	12,27	0,67	9,08	8,51	0,71	0,21
Silica gel 60 – woda	51,91	247,57	420,94	1,49	252,12	76,51	0,078	0,41
Silica gel 60 – 2-propanol	41,82	233,94	386,17	1,47	233,48	83,30	0,072	0,40

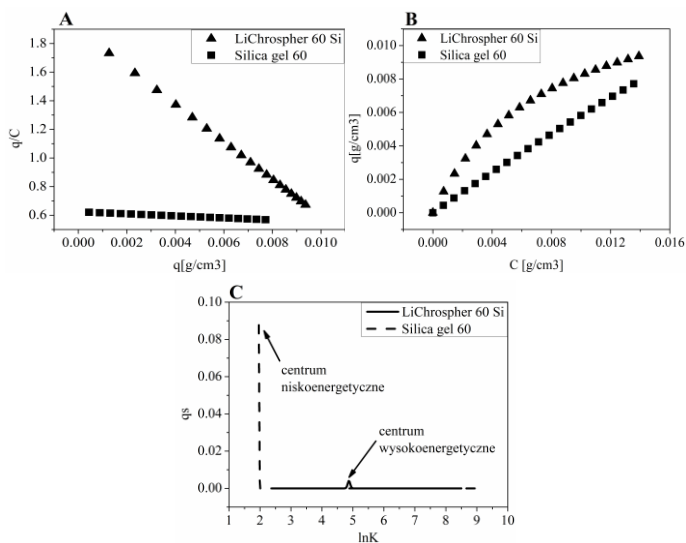
Źródło: Opracowanie własne.

4.2. Wyniki pomiarów równowagi adsorpcji

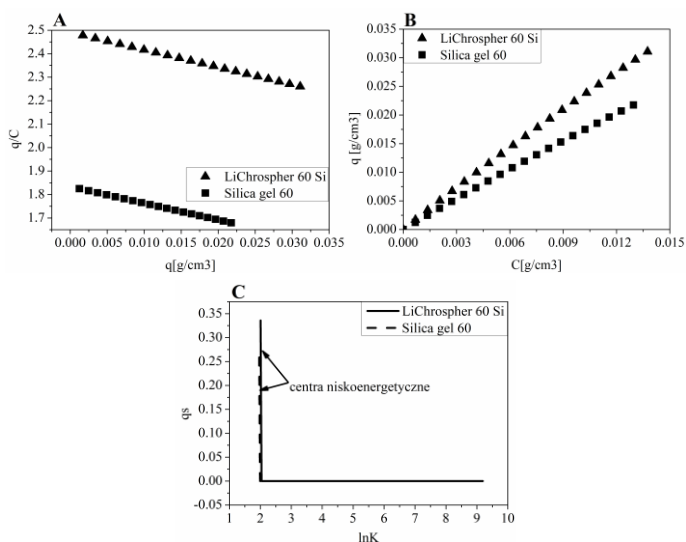
Przykładowe wyniki badań (tj. wykresy Scatcharda, izotermy adsorpcji oraz obliczenia AED) przedstawiono na rysunkach 7, 8 i 9. Analizując otrzymane izotermy adsorpcji, można zauważyć, że rodzaj zastosowanego adsorbentu oraz skład fazy ruchomej są ściśle powiązane z energią adsorpcji badanych substancji. Wyniki dla układów z eluentem zawierającym 95% v/v MeOH i 5% v/v H₂O dla kofeiny przedstawiono na rysunku nr 7. Zarówno dla adsorbentu LiChrospher 60 Si, jak i Silica gel 60 obserwuje się liniowe przebiegi zależności Scatcharda mogące świadczyć o homogenicznych rozkładach energii adsorpcji w tych układach. Potwierdzają to wyniki obliczeń AED. Uzyskane, homogeniczne rozkłady energii adsorpcji różnią się jednak zarówno pod względem pojemności chłonnych, *q_s*, jak i wartości energii adsorpcji, *lnK*. Wyraźnie wyższa energia adsorpcji jest potrzebna w układzie z adsorbentem LiChrospher 60 Si. Jednocześnie, w układzie z tym adsorbentem pojemność chłonna jest wyraźnie mniejsza w porównaniu z Silica gel 60. Na rysunku 8 przedstawiono wyniki uzyskane dla układów z eluentem zawierającym 95% v/v ACN i 5% v/v H₂O dla kofeiny. Liniowa zależność Scatcharda obserwowana jest dla fazy stacjonarnej Silica gel 60 oraz LiChrospher 60 Si. Liniowe zależności Scatcharda odpowiadają homogenicznym rozkładom energii adsorpcji na wykresach AED, widoczna jest obecność jednego centrum aktywnego. Warto zauważyć że w obu układach zarówno pojemności chłonne, jak i energia adsorpcji są do siebie zbliżone. Na rysunku 9 przedstawiono wyniki zebrane dla układów z eluentem zawierającym 95% v/v MeOH i 5% v/v H₂O dla fenolu. Nieliniowa zależność Scatcharda jest obserwowana dla fazy stacjonarnej Silica gel 60, natomiast liniowa dla układu z LiChrospher 60 Si. Obserwacje te potwierdzają wyniki obliczeń AED. W przypadku fazy stacjonarnej Silica gel 60 można zaobserwować obecność dwóch, dobrze rozdzielonych pasm: jednego nisko- i jednego wysokoenergetycznego. Co istotne, pojemność chłonna centrum wysokoenergetycznego jest stosunkowo niewielka, co wskazuje na

niewielki wpływ tego centrum na mechanizm adsorpcji. W układzie z LiChrospher 60 Si rozkład energii adsorpcji jest homogeniczny – występuje tylko jedno pasmo. Porównując oba układy, warto także zauważyć, że zarówno pojemności chłonne, jak i energia adsorpcji centrów niskoenergetycznych są do siebie zbliżone.

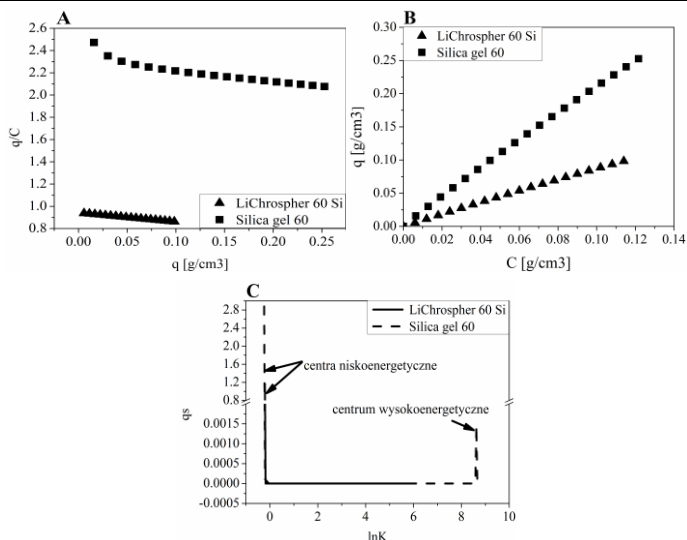
Przedstawione wyniki badań wskazują, że w testowanych układach największy wpływ na mechanizm procesu adsorpcji mają niskoenergetyczne centra aktywne. Jednak w celu potwierdzenia tej tezy potrzebne są dalsze badania w różnorodnych układach adsorpcyjnych.



Rysunek 7. Wykres Scatcharda (a), izotermi adsorpcji (b) i wykres AED (c) dla kofeiny w układzie 95% v/v MeOH i 5% v/v H₂O



Rysunek 8. Wykres Scatcharda (a), izotermi adsorpcji (b) i wykres AED (c) dla kofeiny w układzie 95% v/v ACN i 5% v/v H₂O



Rysunek 9. Wykres Scatcharda (a), izotermę adsorpcji (b) i wykres AED (c) dla fenolu w układzie 95% v/v MeOH i 5% v/v H₂O

5. Wnioski

W niniejszej pracy wyznaczono wybrane właściwości adsorpcyjne polarnych faz stacjonarnych. Do tego celu zastosowano fenol i kofeinę jako substancje testowe analizów w różnych układach HILIC. Ponadto, w celu dokładnej oceny zjawisk zachodzących w badanych układach, określono rozkłady wielkości cząstek adsorbentów zastosowanych w badanych układach. Wnioski wynikające z przeprowadzonych badań można podsumować następująco:

- uzyskane wyniki potwierdzają możliwość występowania zróżnicowanych miejsc aktywnych na powierzchni badanych adsorbentów;
- niejednorodność powierzchni adsorbentu może w istotny sposób wpływać na mechanizm sorpcji HILIC w analizowanych układach;
- w analizowanych układach główną rolę w mechanizmie sorpcji odgrywają miejsca niskoenergetyczne;
- pojemność adsorpcyjna miejsc wysokoenergetycznych na badanych adsorbentach jest stosunkowo niska;
- rozkład wielkości cząstek w badanych adsorbentach jest niejednorodny (polidispersyjny), co może wpływać zarówno na mechanizm adsorpcji, jak i przenoszenia masy.

Literatura

1. Guiochon G., Felinger A., Shirazi D.G., Katti A.M., *Fundamentals of Preparative and Nonlinear Chromatography*, Elsevier, Amsterdam 2006.
2. Giraldo L.F., López B.L., Pérez L., Urrego S., Sierra L., Mesa M., *Mesoporous silica applications*, In *Macromolecularsymposia*, 258, 2007, s. 129-141.
3. Qiao L., Lva W., Changa M., Shi X., Xua G., *Surface-bonded amide-functionalized imidazolium ionic liquid as stationary phase for hydrophilic interaction liquid chromatography*, *Journal of Chromatography A*, 1559, 2018, s. 141-148.

4. Yu J., Wey M., Firooz S.K., Armstrong D.W., *Ionizable Cyclofructan 6 Based Stationary Phases for Hydrophilic Interaction Liquid Chromatography Using Superficially Porous Particles*, *Chromatographia*, 84, 2021, s. 821-832.
5. Alpert A.J., *Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds*, *Journal of Chromatography A*, 499, 1990, s. 177-196.
6. Zahirakhaa A.V., Uzhela A.S., Loshina A.A., Chernobrovkina A.V., Smolenkova A.D., Shpiguna O.A., *Novel Stationary Phases for Ion Chromatography and Hydrophilic Interaction Liquid Chromatography*, *Journal of Analytical Chemistry*, 74, 2019, s. 12-20.
7. Roca L.S., Schoemaker S.E., Piroka B.W.J., Gargano A.F.G., Schoenmakers P.J., *Accurate modelling of the retention behaviour of peptides in gradient-elution hydrophilic interaction liquid chromatography*, *Journal of Chromatography A*, 1614, 2019, s. 460650.
8. McCalley D.V., *Understanding and manipulating the separation in hydrophilic interaction liquid chromatography-a review*, *Journal of Chromatography A*, 1523, 2017, s. 49-71.
9. Popova A.S., Maksimova G.S., Shpiguna O.A., Chernobrovkina A.V., *Adsorbents with a Covalently Bonded Polymer Layer for Hydrophilic Interaction Liquid Chromatography*, *Journal of Analytical Chemistry*, 77, 2022, s. 1173-1183.
10. Ayawei N., Ebelegi A.N., Wankasi D., *Modelling and Interpretation of Adsorption Isotherms*, *Journal of Chemistry*, 2017, 2017, s. 11.
11. Qiu X., Armstrong D.W., Berthod A., *Thermodynamic studies of a zwitterionic stationary phase in hydrophilic interaction liquid chromatography*, *Journal of Chromatography A*, 1272, 2013, s. 81-89.
12. Torkia Y.B., Khalfaoui M., Lamine A.B., *On the use of the Hill's model as local isotherm in the interpretation of the behaviour of the adsorption energy distributions*, *Journal of Applied Physics*, 6, 2014, s. 2278-4861.
13. Rudzinski W., Steel W.A., Zgrablich G., *Equilibria and dynamics of gas adsorption on heterogeneous solid surfaces*, Elsevier, Washington 1997, s. 883.
14. Gritti F., Guiochon G., *Limits of the numerical estimation of the adsorption energy distribution from adsorption isotherm data using the expectation-maximization method*, *Journal of Chromatography A*, 1144, 2007, s. 208-220.
15. Chu B., *Laser light scattering*, *Spectroscopy in Biochemistry*, 21, 1995, s. 145-174.
16. Stanley B.J., Bialkowski S.E., Marshall D.B., *Analysis of 1st-order rate-constant spectra with regularized least-squares and expectation maximization. Theory and numerical characterization*, *Analytical Chemistry*, 65, 1993, s. 259-267.
17. Stanley B.J., Guiochon G., *Analysis of the surface heterogeneity of unmodified and modified silica by capillary inverse gas-solid chromatography at finite dilution*, *Langmuir*, 10, 1994, s. 1573.
18. Stanley B.J., Guiochon G., *Importance of the Accuracy of Experimental Data in the Nonlinear Chromatographic Determination of Adsorption Energy Distributions*, *Langmuir*, 10, 1994, s. 4278.
19. Kalbitzer H.R., Stehlik D., *On the Analysis of Competitive Binding of Various Ligands to Cooperative and Independent Binding Sites of Macromolecules*, *Journal of Biosciences*, 9-10, 1979, s. 757-769.
20. Nakanishi K., Yamamoto S., Matsuno R., Kamikubo T., *Diffusion of Saccharides and Amino Acids in Cross-linked Polymers*, *Agricultural and Biological Chemistry*, 41(12), 1977, s. 2455-2462.
21. Horváth K., Lukács D., Sepsey A., Felinger A., *Effect of particle size distribution on the separation efficiency in liquid chromatography*, *Journal of Chromatography A*, 1361, 2014, s. 203-208.
22. Stanley B.J., Guiochon G., *Numerical estimation of adsorption energy distributions from adsorption isotherm data with the expectation-maximization method*, *Journal of Physical Chemistry A.*, 97, 1993, s. 8098.

23. Billen J., Guillarme D., Rudaz S., Veuthey J.L., Ritchie H., Grady B., Desmet G., *Relation between the particle size distribution and the kinetic performance of packed columns Application to a commercial sub-2m particle material*, Journal of Chromatography A, 1161, 2007, s. 224-233.
24. Liekens A., Billena J., Sherant R., Ritchie H., Denayer J., Desmet G., *High performance liquid chromatography column packings with deliberately broadened particle size distribution: Relation between column performance and packing structure*, Journal of Chromatography A, 1218, 2011, s. 6654-6662.
25. Zhang W., Shen J., Thomas J.C., Mu T., Xu Y., Xiu W., Xu M., Zhu M., *Particle size distribution recovery in dynamic light scattering by optimized multi-parameter regularization based on the singular value distribution*, Powder Technology, 353, 2019, s. 320-329.

Właściwości adsorpcyjne wybranych polarnych faz stacjonarnych

Streszczenie

Chromatografia cieczowa, w tym także chromatografia oddziaływań hydrofilowych, jest obecnie szeroko stosowana do rozdzielania związków polarnych. Wybór fazy stacjonarnej odgrywa w tej metodzie kluczową rolę. Powierzchnie współczesnych adsorbentów są energetycznie heterogeniczne z dwóch głównych powodów. Po pierwsze, wszystkie zanieczyszczenia istniejące w materiale sypkim użytym do przygotowania adsorbentu mają tendencję do segregacji na jego powierzchni. Po drugie, niejednorodność/konformacja związanych z matrycą krzemionkową łańcuchów może również pełnić funkcję oddzielnych miejsc aktywnych. W niniejszej pracy przeanalizowano mechanizm sorpcji kofeiny i fenolu jako substancji testowych w różnych układach adsorpcyjnych. Badania przeprowadzono z wykorzystaniem dwóch różnych polarnych adsorbentów. Jako fazy ruchome zastosowano układy metanol-woda i acetonitryl-woda o różnej zawartości rozpuszczalnika organicznego. Na podstawie izoterm adsorpcji uzyskanych dla badanych układów wyznaczono wykresy Scatcharda oraz rozkłady energii adsorpcji. Wyniki te uzyskano przy użyciu metody matematycznej opracowanej przez Stanleya i współpracowników. W celu lepszego zrozumienia uzyskanych wyników uzupełniono je o wyniki badań pomiarów granulometrycznych (tj. rozkładów wielkości cząstek badanych adsorbentów). Wyniki badań wskazują na złożony mechanizm adsorpcji wynikający z różnorodnych oddziaływań analitów z badanymi fazami stacjonarnymi, a także na niejednorodne rozkłady wielkości cząstek w badanych adsorbentach.

Słowa kluczowe: adsorbent, izoterma adsorpcji, chromatografia cieczowa, rozkład wielkości cząstek, heterogeniczność energetyczna

Adsorption properties of selected polar stationary phases

Abstract

Liquid chromatography, including hydrophilic interaction chromatography, is now widely used to separate polar compounds. The choice of stationary phase plays a key role in this method. The surfaces of modern adsorbents are energetically heterogeneous for two main reasons. First, any impurities existing in the bulk material used to prepare the adsorbent tend to segregate on its surface. Second, the heterogeneity/conformation of the silica matrix-bound chains can also act as separate active sites. The present study analyzed the sorption mechanism of caffeine and phenol as test substances in different adsorption systems. The study was carried out using two different polar adsorbents. Methanol-water and acetonitrile-water systems with different organic solvent contents were used as mobile phases. Based on the adsorption isotherms obtained for the studied systems, Scatchard plots and adsorption energy distributions were determined. These results were obtained using the mathematical method developed by Stanley et al. In order to better understand the results obtained, they were supplemented with the results of granulometric measurements (i.e. particle size distributions of the studied adsorbents). The results of the study indicate a complex mechanism of adsorption resulting from various interactions of analytes with the studied stationary phases, as well as heterogeneous particle size distributions in the studied adsorbents.

Keywords: adsorbent, adsorption isotherm, liquid chromatography, particle size distributions, energetic heterogeneity

Badania wpływu imperfekcji na rozkład obciążeń stabilizujących stężoną kratownicę

1. Wprowadzenie

Kratownice płaskie często stosowane są jako elementy nośne zadaszeń wielkogabarytowych obiektów budowlanych. Charakteryzują się one dużą nośnością i sztywnością w swojej płaszczyźnie. W celu stabilizacji wiązarów kratowych w płaszczyźnie połączy dachu stosuje się stężenia np. tężniki połączeniowe poprzeczne, tężniki pionowe. Ich rolą, w wielu przypadkach, jest nie tylko przeniesienie sił poziomych od wiatru, ale również zapewnienie geometrycznej niezmienności konstrukcji dachowym [1].

W trakcie procesu projektowania często zakłada się, że stężenia są sztywne. Na tej podstawie do obliczeń przyjmuje się, że długość wyboczeniowa ściskanego pasa kratownicy (z płaszczyzny) równa jest odległości pomiędzy punktami bocznego podparcia. Analizy numeryczne stateczności kratownic wykonane przy założeniu, że podpory boczne są podatne, były przedmiotem prac [2, 3], a rezultaty badań doświadczalnych przedstawiono między innymi w pracach [4-6]. Badania wpływu sztywności rotacyjnej podpór bocznych, usytuowanych w pasie górnym, na nośność kratownicy poddanej obciążeniu od wiatru zaprezentowano w pracach [7, 8]. Sztywność rotacyjna podpór bocznych wynika ze sztywności giętej elementów stężających i ze sztywności połączenia pomiędzy tymi elementami a pasem górnym kratownicy (1). W wyżej wymienionych pracach wykazano, że uwzględnienie w obliczeniach tej sztywności podpór bocznych miało istotny wpływ na wzrost nośności wyboczeniowej ściskanego pasa dolnego kratownicy.

$$\frac{1}{k_{rot}} = \frac{1}{K_{roof}} + \frac{1}{K_{con}} \quad (1)$$

gdzie: k_{rot} – sztywność rotacyjna sprężystej podpory bocznej, K_{roof} – sztywność gięta elementów stężających, K_{con} – sztywność połączenia między elementem stężającym a pasem górnym kratownicy.

Obciążenie stabilizujące wiązary dachowe, powstałe na skutek działania obciążeń pionowych na kratownice posiadające wstępną imperfekcją łukową, wyznaczane jest na podstawie aktualnych przepisów normowych [9]. Należy jednak podkreślić, że zależności przedstawione w normie odnoszą się do elementów ściskanych stałą siłą na całej długości i podpartych przegubowo. W pracy [10] przedstawiono rozwiązanie umożliwiające wyznaczenie tego obciążenia na podstawie normy [9], ale bez konieczności przeprowadzania pracochłonnego procesu iteracyjnego. W tym celu należało wyznaczyć podatność tężnika w środku rozpiętości. W pracy [6] zmodyfikowano wyżej opisane rozwiązanie poprzez wprowadzenie parametru k , tj. sztywności bocznego podparcia (sztywność na przesuw w płaszczyźnie prostopadłej do płaszczyzny kraty) w miejsce opisanej wyżej podatności. Na tej podstawie otrzymano równie (2).

¹ markraje@pg.edu.pl, Katedra Mechaniki Budowli, Wydział Inżynierii Lądowej i Środowiska, Politechnika Gdańska, www.pg.edu.pl.

$$q_d = \frac{8(e_o + \frac{L_{st}}{k} q_z) \sum_{i=1}^m N_{ed}}{L^2 - 8 \frac{L_{st}}{k} \sum_{i=1}^m N_{ed}}, R_d = q_d L_{st} \quad (2)$$

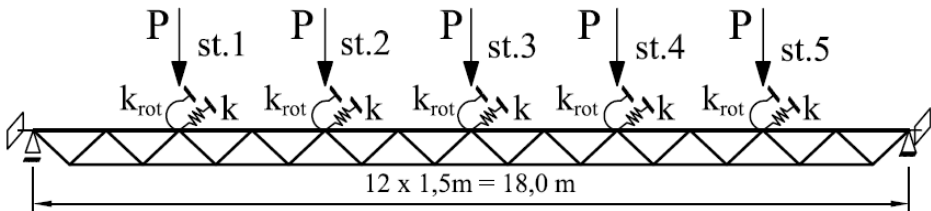
gdzie: q_d – obciążenie stabilizujące, L_{st} – rozstaw podpór bocznych, e_o – amplituda imperfekcji, k – sztywność podpory sprężystej, q_z – poziome obciążenie zewnętrzne, N_{ed} – siła normalna w pasie górnym kratownicy, m – liczba elementów stężanych, R_d – siła stabilizująca.

Niniejsza praca poświęcona jest badaniom stateczności i nośności stężonej kratownicy. Badana kratownica była podparta sprężystymi podporami, przy czym rozkład sztywności podpór (na długości pasa) był stały lub zmienny. Analizowany był wpływ sztywności pośrednich podpór bocznych (sztywność na przesuw k oraz na obrót k_{rot}) na obciążenie krytyczne konstrukcji (LBA). W pracy przedstawione zostały również rezultaty analiz statycznych nieliniowych dla konstrukcji posiadającej wstępne imperfekcje geometryczne. Ponadto, porównano siły stabilizujące kratownicę otrzymane na podstawie opisanego wyżej równania (2), do reakcji w podporach sprężystych, otrzymanych na podstawie analiz GMNIA.

2. Opis badanej konstrukcji

Badania przeprowadzone zostały dla kratownicy wykonanej ze stali S275. Rozpiętość konstrukcji wynosiła $L = 18,0$ m, a wysokość $h = 0,7$ m. Pas górny wykonany został z elementów o przekroju RHS 120 x 100 x 4, a pas dolny z RHS 100 x 100 x 4. Zastosowano wykratowanie typu V [1], przy czym krzyżulce wykonane zostały z rur kwadratowych RHS 50 x 50 x 3. Badania nośności wyboeczeniowej konstrukcji o podobnej geometrii, poddanej obciążeniu od wiatru, były poprzednio tematem prac [7, 8].

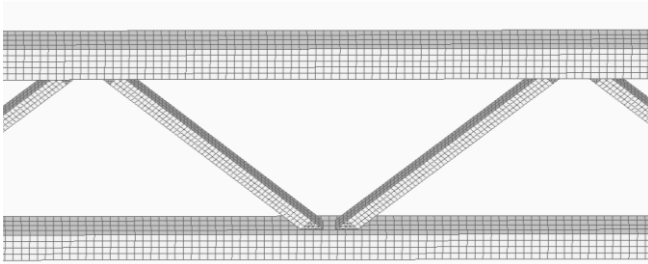
Skrajne węzły kratownicy zastały podparte przegubowo. Na jednej ze skrajnych podpór zablokowana została możliwość translacji we wszystkich trzech kierunkach, a na drugiej możliwy był przesuw jedynie wzdłuż osi podłużnej badanej konstrukcji. Na podporach skrajnych zablokowana została możliwość obrotu w płaszczyźnie prostopadłej do płaszczyzny kraty (blokada skręcania pasa górnego). W miejscach usytuowania stężeń (stężenia 1-5, rys. 1) zastosowane zostały podpory sprężyste o sztywności na przesuw i i na obrót (w płaszczyźnie prostopadłej do płaszczyzny kraty) w rozstawie $L_{st} = 3,0$ m.



Rysunek 1. Schemat statyczny kratownicy

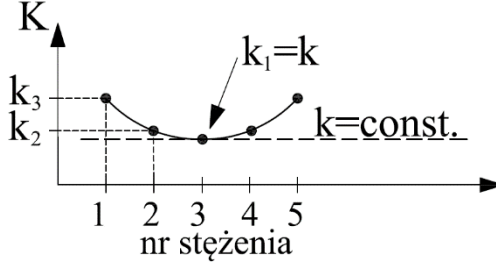
Analizy numeryczne stateczności i nośności wykonano w programie [11] dla modelu powłokowego konstrukcji. Do obliczeń wykorzystano metodę elementów skończonych [12]. Zastosowano około 42 tys. elementów typu Quad [11, 13] (4-węzłowych i 3-węzłowych, o sześciu stopniach swobody w węźle). Wymiar elementów w przekroju pasa górnego i dolnego wynosił około 2,5 cm x 2,5 cm, natomiast w przypadku krzyżulców 1,5 cm x 1,5 cm (rys. 2). W miejscach usytuowania spoin, łączących pasy

z krzyżulcami, zastosowane zostały elementy sztywne (typu Rigid [11]). Kratownica została obciążona w stężonych węzłach za pomocą sił skupionych (obciążenie grawitacyjne). W celu przeprowadzenia analiz zastosowano idealnie sprężysto-plastyczny model ciała ($E = 210 \text{ GPa}$, $\nu = 0,3$, $f_y = 275 \text{ MPa}$) oraz sterowanie metodą *arc-length* [11, 12].



Rysunek 2. Model powłokowy kratownicy

Stężenia wykonane zostały za pomocą elementów typu *Dof-spring* [11] o zadanej sztywności w kierunku przesuwu i obrotu (z płaszczyzny kraty). Zostały one połączone z wybranymi węzłami pasa górnego (połączenie z półką górną przekroju rurowego). W pracy przedstawione zostały rezultaty badań przeprowadzonych dla kratownicy z pięcioma podporami sprężystymi o jednakowej sztywności k (3). Ponadto wykonano również analizy w przypadku gdy rozkład sztywności (na przesuw) był nierównomierny (4) (rys. 3).



Rysunek 3. Rozkład sztywności podpór bocznych na długości kratownicy

$$K(x) = k \tag{3}$$

gdzie: $K(x)$ – funkcja opisująca rozkład sztywności podpór bocznych na długości kratownicy, k – sztywność (na przesuw) podpory sprężystej, x – odległość od przegubowo-przesuwnej podpory skrajnej do punktu usytuowania podpory sprężystej w pasie górnym.

$$K(x) = \frac{1}{\delta(x)} \tag{4}$$

gdzie: $\delta(x)$ – funkcja opisująca podatność tężnika połączeniowego poprzecznego.

W tym przypadku sztywność podpory sprężystej wyznaczona została na podstawie podatności danej równaniem (5). Wobec powyższego do obliczeń przyjęto, że sztywność podpór nr 2 i 4 (k_2) była o 12,5%, a podpór nr 1 i 5 (k_3) o 80% większa w porównaniu do sztywności podpory nr 3 (k_1), usytuowanej w środku rozpiętości kratownicy (6),

(rys. 1 i 3). W każdym przypadku przyjęto, że podpory boczne charakteryzowała jednakowa sztywność rotacyjna k_{rot} (niezależnie od miejsca usytuowania stężenia).

$$\delta(x) = \frac{-4\delta_o}{L^2}x^2 + \frac{4\delta_o}{L}x \quad (5)$$

gdzie: δ_o – podatność tężnika połączowego w środku rozpiętości, L – rozpiętość kratownicy.

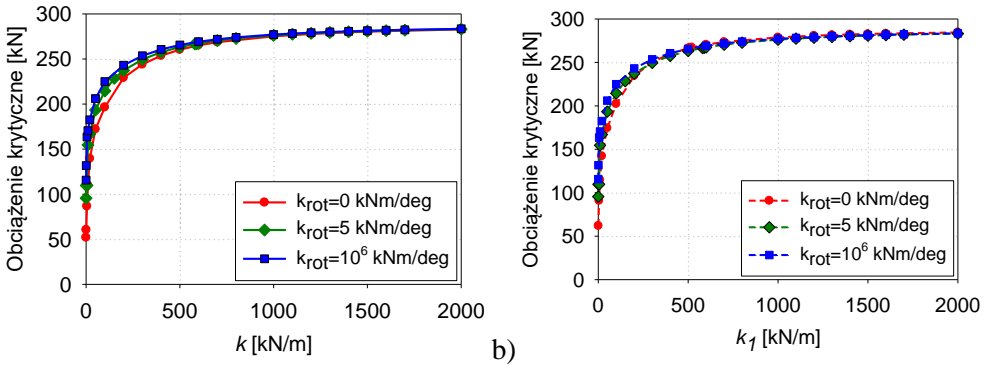
$$k_1 = k, k_2 = 1,125k, k_3 = 1,8k \quad (6)$$

gdzie: k_1 – sztywność podpory sprężystej usytuowanej w środku rozpiętości kratownicy, k_2 – sztywność podpór sprężystych usytuowanych w odległości 3,0 m od środka rozpiętości, k_3 – sztywność podpór sprężystych usytuowanych w odległości 6,0 m od środka rozpiętości.

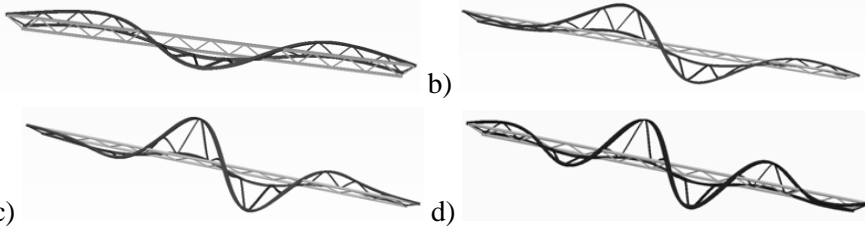
3. Rezultaty analiz numerycznych

W pierwszym etapie badań przeprowadzone zostały liniowe analizy stateczności (LBA), na podstawie których wyznaczono wpływ sztywności bocznego podparcia na nośność wyboczeniową kratownicy. Na podstawie rezultatów przedstawionych na rysunku 4 można stwierdzić, że w każdym przypadku wzrost sztywności sprężystych podpór powodował wzrost obciążenia krytycznego. Badania wykazały również, że istniała pewna progowa (minimalna) sztywność podpór bocznych, przy której nośność kratownicy była zbliżona do maksymalnej nośności uzyskanej przy sztywnym podparciu. W przypadku, gdy zastosowano podpory boczne o jednakowej sztywności ($K(x) = \text{const}$). Rysunek 4a) $k = 650$ kN/m obciążenie krytyczne było o 10% mniejsze w porównaniu do obciążenia uzyskanego dla $k = 10^6$ kN/m. Tę samą wartość obciążenia ($P_{cr} = 267$ kN) uzyskano również, gdy sztywność podpór wynosiła $k = 590$ kN/m i $k_{rot} = 5$ kN/m. W przypadku, gdy zastosowano zmienny rozkład sztywności podpór na długości kratownicy ($K(x)$ wg równania (4), rysunek 4b) wyżej podana wartość obciążenia została uzyskana dla $k_1 = 510$ kN/m ($k_2 = 573,75$ kN/m, $k_3 = 918$ kN/m, $k_{rot} = 0$ kNm/rad). Na podstawie przeprowadzonych badań można stwierdzić, że uwzględnienie w obliczeniach sztywności rotacyjnej (k_{rot}) stężenia powodowało wzrost nośności wyboczeniowej konstrukcji, przy czym wzrost ten był tym mniejszy im większa była wartość k (sztywność na przesuw). Ponadto uwzględnienie wartości k_{rot} powodowało w każdym przypadku redukcję wartości sztywności progowych. Obciążenie krytyczne kratownicy wzrosło (maksymalnie o 5%) w przypadku, gdy zastosowano podpory sprężyste o sztywnościach k_1, k_2, k_3 (wg rys. 3) w porównaniu do przypadku, gdy sztywność każdej punktowej podpory wynosiła k .

Wybrane postacie wyboczeniowe kratownicy zostały przedstawione na rysunku 5. W przypadku zastosowania stężeń o sztywności $k = 100$ kN/m i $k_{rot} = 5$ kN/m, postać wyboczeniowa przyjmowała formę przedstawioną na rysunku 5b. Przy zastosowaniu podpór sprężystych sztywnych na przesuw ($k = 10^6$ kN/m), niezależnie od wartości k_{rot} , kratownica wybaczala się pomiędzy stężeniami (rys. 5d).



a) b) Rysunek 4. Wykres zależności obciążenia krytycznego kratownicy od sztywności podpór sprężystych: a) w przypadku zastosowania podpór o sztywności k , b) w przypadku zastosowania podpór o sztywności k_1, k_2, k_3 , dla wybranych sztywności rotacyjnych k_{rot}



Rysunek 5. Postać wyboczeniowa kratownicy z podporami sprężystymi o sztywnościach: a) $k = 100$ kN/m lub $k_1 = 100$ kN/m, $k_2 = 112,5$ kN/m, $k_3 = 180$ kN/m, b) $k = 200$ kN/m lub $k_1 = 200$ kN/m, $k_2 = 225$ kN/m, $k_3 = 360$ kN/m, c) $k = 650$ kN/m lub $k_1 = 510$ kN/m, $k_2 = 573,75$ kN/m, $k_3 = 918$ kN/m dla $k_{rot} = 0$ kNm/deg lub $k_{rot} = 10^6$ kNm/deg d) $k = 10^6$ kN/m lub $k = 10^6$ kN/m i $k_{rot} = 10^6$ kNm/deg

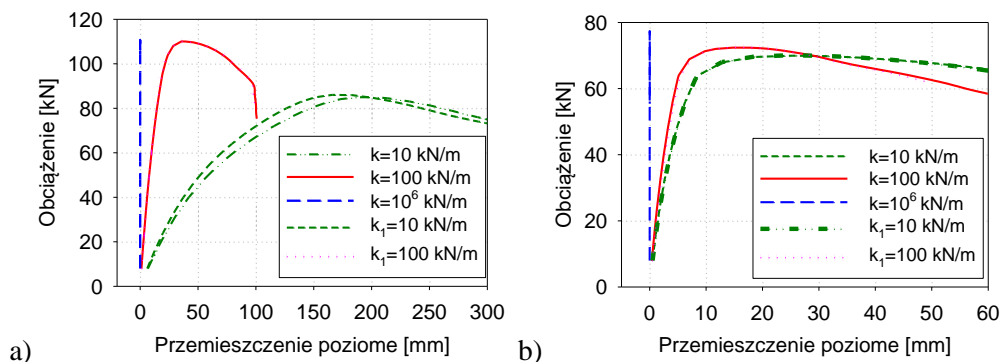
Analizy statyczne, fizycznie i geometrycznie nieliniowe (GMNIA) zostały przeprowadzone zarówno w przypadku, gdy badana konstrukcja posiadała wstępną imperfekcję geometryczną w formie łukowej krzywizny pasa górnego (wg [9], rys. 6), jak i w przypadku gdy zastosowana została imperfekcja, której kształt odpowiadał pierwszej postaci wyboczenia (na podstawie analiz LBA, przy sztywnych podporach bocznych, rys. 5d). W obydwu przypadkach przyjęto do obliczeń amplitudę imperfekcji (e_0) równą $L/500$ (maksymalne przemieszczenie poziome z płaszczyzny kraty, wyznaczone przy założeniu, że liczba elementów stężanych $m = 1$, na podstawie wytycznych normy [9]).



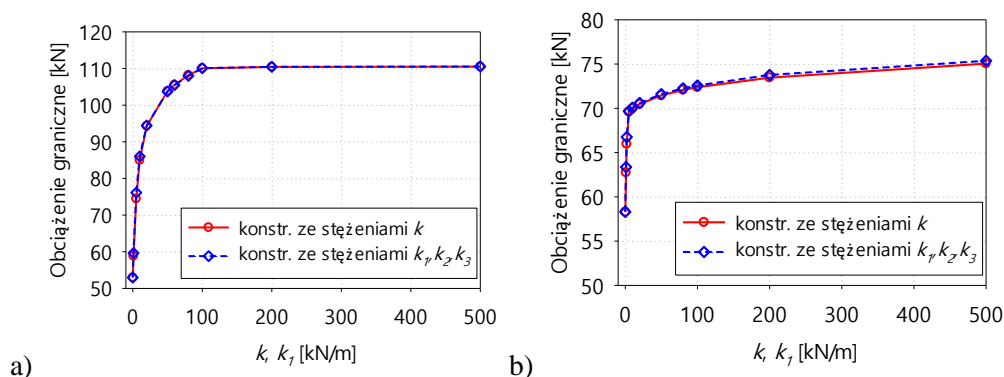
Rysunek 6. Wstępna imperfekcja geometryczna kratownicy

Ścieżki równowagi otrzymane dla kratownicy podpartej stężeniami o sztywności k oraz k_1 (wraz z odpowiadającymi k_2, k_3) zostały przedstawione na rysunku 7. Badania przeprowadzono przy założeniu, że sztywność rotacyjna podpór bocznych wynosiła $k_{rot} = 0$ kNm/rad. W większości analizowanych przypadków maksymalne obciążenie konstrukcji (obciążenie graniczne) zależało od sztywności (na przesuw) podpór bocznych.

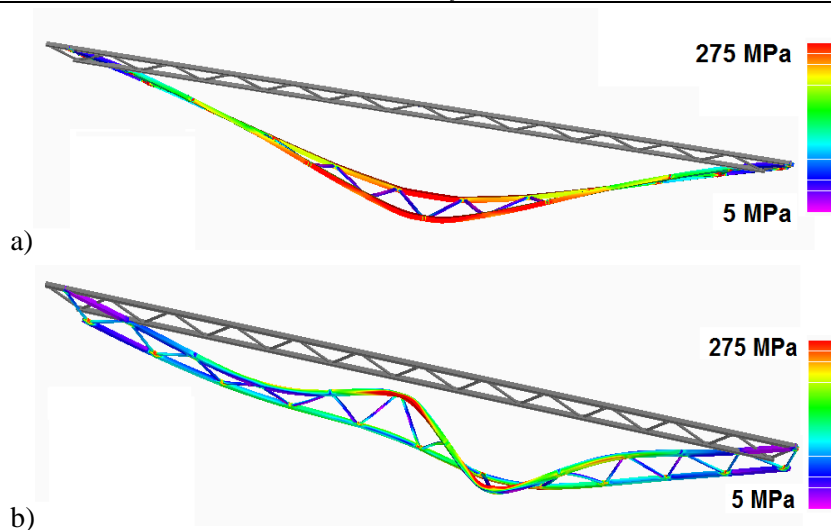
Podobnie jak w przypadku analiz LBA, wprowadzenie do modelu obliczeniowego podpór sprężystych o nierównomiernym rozkładzie sztywności (k_1, k_2, k_3) miało niewielki wpływ (do 2%) na wzrost nośności kratownicy w porównaniu do przypadku, gdy sztywność każdego stężenia wynosiła k . Badania wykazały, że kształt wstępnej imperfekcji geometrycznej miał istotny wpływ zarówno na nośność, jak i na przemieszczenia boczne kratownicy. Obciążenie graniczne konstrukcji z imperfekcją w formie postaci wybozczenia było mniejsze (do 35%) w porównaniu do rezultatów otrzymanych przy zastosowaniu imperfekcji łukowej. W przypadku imperfekcji zadanej wg wytycznych normy [9] zwiększenie sztywności podpór powyżej 100 kN/m (dla k lub k_1) miało niewielki wpływ na wzrost nośności kratownicy (poniżej 1%, rys. 8). Należy jednak podkreślić, że przy wyżej wymienionej sztywności stężeń, przemieszczenia boczne (poziome) w środku rozpiętości konstrukcji, były wielokrotnie większe (sięgały 34 mm, rys. 7a) od przemieszczeń otrzymanych przy sztywnym podparciu bocznym (tj. dla $k = 10^6$ kN/m, rys. 7a). Kształt wstępnej imperfekcji miał również istotny wpływ na deformację konstrukcji uzyskaną w momencie największego obciążenia (rys. 9). W zależności od poziomu obciążeń, strefy plastyczne pojawiały się zarówno w węzłach badanej konstrukcji, jak i w pasie górnym oraz dolnym.



Rysunek 7. Wykres zależności obciążenia od przemieszczenia poziomego pasa górnego kratownicy: a) z imperfekcją łukową (w środku rozpiętości), b) z imperfekcją w formie postaci wybozczenia (w odl. 3,0 m od środka rozpiętości), dla wybranych sztywności k i k_1 (oraz odpowiadające k_2 i k_3)



Rysunek 8. Wykres zależności obciążenia granicznego od sztywności stężeń dla konstrukcji: a) z imperfekcją łukową, b) z imperfekcją w formie postaci wybozczenia

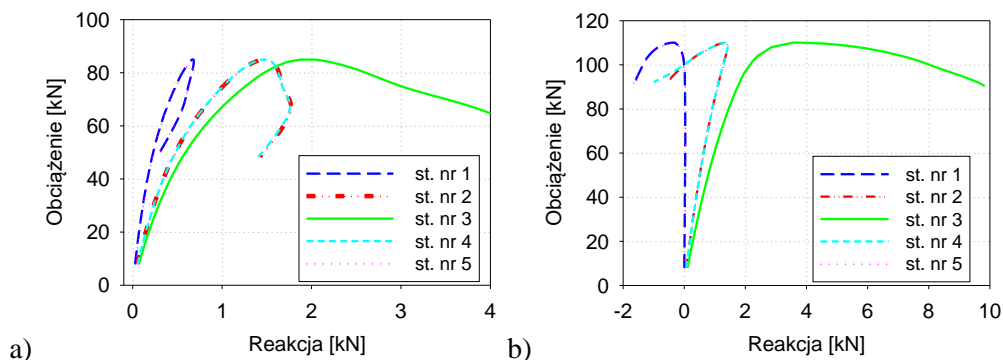


Rysunek 9. Deformacja oraz mapa naprężeń (wg HMH, na powierzchni górnej elementów) przy maksymalnym obciążeniu kratownicy: a) z imperfekcją łukową, b) z imperfekcją w formie postaci wyboczenia, w przypadku zastosowania podpór sprężystych o sztywności $k = 100 \text{ kN/m}$

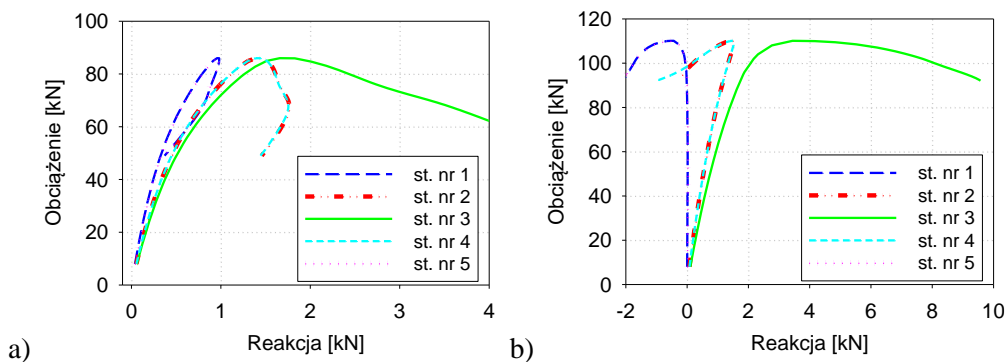
Rezultatem przeprowadzonych analiz GMNIA było również uzyskanie reakcji w podporach sprężystych, które zależały od poziomu obciążenia kratownicy (rys. 10-13). Niezależnie od rozkładu sztywności stężeń (k lub k_1 , k_2 , k_3) największą siłę uzyskano w podporze usytuowanej w środku rozpiętości, w przypadku gdy konstrukcja posiadała wstępną imperfekcję łukową. Badania wykazały jednak, że siła w tej podporze może być najmniejsza (co do wartości bezwzględnej) czego przyczyną był kształt wstępnej deformacji odpowiadający postaci wyboczenia. Na podstawie przeprowadzonych analiz można stwierdzić, że w momencie największego obciążenia reakcje w poszczególnych podporach (st. 1, st. 2, st. 3, st. 4, st. 5, rys. 1) mogły być znakozmienne, na co istotny wpływ miała zarówno sztywność podpór bocznych, jak i kształt imperfekcji. Dla wybranych przypadków, stężenia usytuowane w pobliżu podpór skrajnych były do pewnego poziomu obciążenia ściskane, a następnie rozciągane, na co istotny wpływ miał sposób deformacji konstrukcji w trakcie procesu obciążania (np. dla $k = 100 \text{ kN/m}$ lub $k_1 = 100 \text{ kN/m}$, $k_2 = 112,5 \text{ kN/m}$, $k_1 = 180 \text{ kN/m}$ – konstrukcja z imperfekcją łukową). Różnice pomiędzy siłami uzyskanymi w najbardziej wyężonym stężeniu o sztywności k lub k_1 , dla konstrukcji z imperfekcją łukową, sięgały 14% (np. dla $k = 10 \text{ kN/m}$). Różnice te wynosiły około 13%, gdy kratownica posiadała wstępną imperfekcję w formie postaci wyboczenia.

Siły w podporach sprężystych otrzymane na podstawie analiz numerycznych zostały porównane do sił stabilizujących kratownicę obliczonych na podstawie wytycznych normy [9]. Do wyznaczenia obciążenia stabilizującego wykorzystane zostało równanie (2) przedstawione w drugim podrozdziale niniejszej pracy. W obliczeniach pominięty został wpływ poziomych obciążeń zewnętrznych ($q_z = 0 \text{ kN/m}$). Reakcje wyznaczone zostały, przy obciążeniu pionowym kratownicy równym 70 kN . Było to maksymalne obciążenie otrzymane na podstawie analiz GMNIA, jakie może przenieść badana konstrukcja z imperfekcją w formie postaci wyboczenia, w przypadku zastosowania podpór o sztywności $k \geq 10 \text{ kN/m}$ (lub $k_1 \geq 10 \text{ kN/m}$ i odpowiadające k_2 , k_3). Do obliczeń

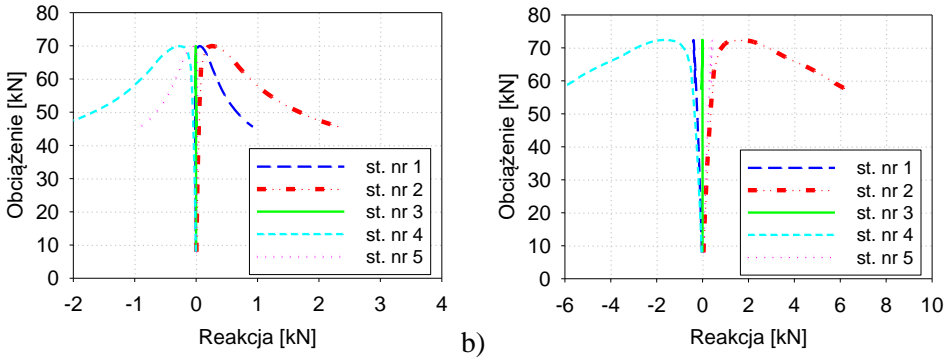
przyjęto siłę N_{ed} równą maksymalnej sile normalnej w ściskanym pasie górnym kratownicy (tj. w środku rozpiętości). Rezultaty dla kratownicy podparłej stężeniami o sztywności $k = 100$ kN/m zostały zaprezentowane na rysunku 14. Na podstawie otrzymanych wyników można stwierdzić, że siła $R_d = 0,85$ kN (2) wyznaczona na podstawie normy [9] była w większości przypadków mniejsza od sił otrzymanych na podstawie analiz GMNIA, powstałych w najbardziej wyężonym stężeniu. Dla kratownicy z imperfekcją łukową (wg [9]), podparłej stężeniami o sztywności $k = 100$ kN/m, różnice pomiędzy wartościami sił uzyskanymi na podstawie analiz numerycznych i obliczeń normowych sięgały 30% (w środku rozpiętości). Ponadto należy podkreślić, że wyznaczone na podstawie normy obciążenie stabilizujące nie było znakozmienne, co stanowiło istotną różnicę w porównaniu do rezultatów GMNIA otrzymanych dla konstrukcji ze wstępną imperfekcją w formie postaci wyboczenia.



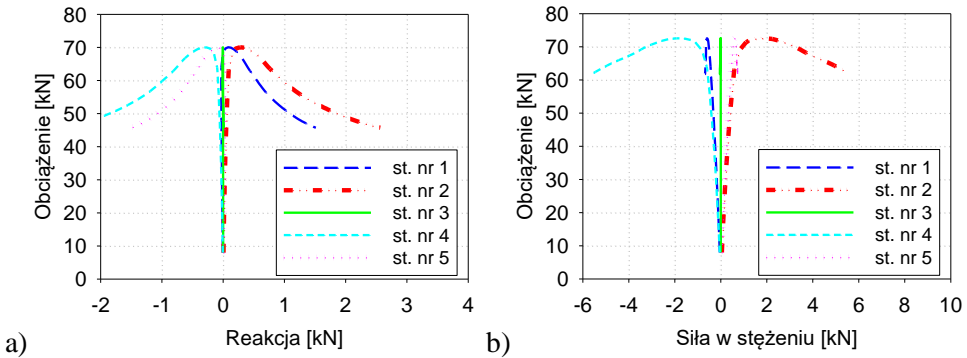
Rysunek 10. Wykres zależności obciążenia kratownicy (z imperfekcją łukową) od reakcji w podporach sprężystych o sztywności: a) $k = 10$ kN/m, b) $k = 100$ kN/m



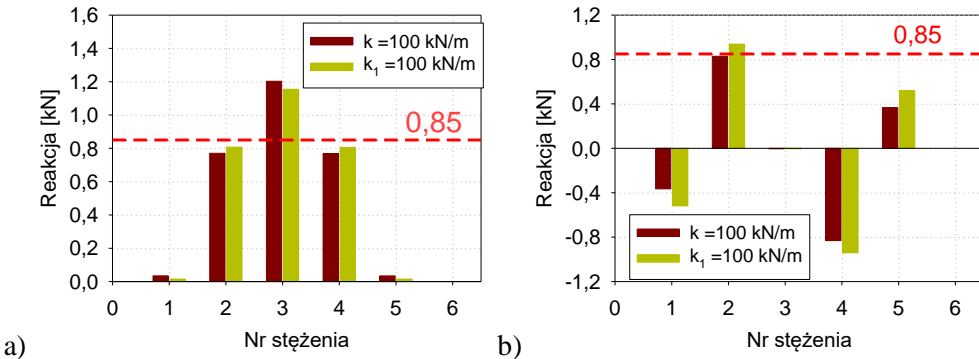
Rysunek 11. Wykres zależności obciążenia kratownicy (z imperfekcją łukową) od reakcji w podporach sprężystych o sztywności: a) $k_1 = 10$ kN/m, $k_2 = 11,25$ kN/m, $k_3 = 18$ kN/m, b) $k_1 = 100$ kN/m, $k_2 = 112,5$ kN/m, $k_3 = 180$ kN/m



Rysunek 12. Wykres zależności obciążenia kratownicy (z imperfekcją w formie postaci wyboczenia) od reakcji w podporach sprężystych o sztywności: a) $k = 10 \text{ kN/m}$, b) $k = 100 \text{ kN/m}$



Rysunek 13. Wykres zależności obciążenia kratownicy (z imperfekcją w formie postaci wyboczenia) od reakcji w podporach sprężystych o sztywności: a) $k_1 = 10 \text{ kN/m}$, $k_2 = 11,25 \text{ kN/m}$, $k_3 = 18 \text{ kN/m}$, b) $k_1 = 100 \text{ kN/m}$, $k_2 = 112,5 \text{ kN/m}$, $k_3 = 180 \text{ kN/m}$



Rysunek 14. Reakcje w podporach sprężystych o sztywności $k = 100 \text{ kNm/m}$ oraz $k_1 = 100 \text{ kN/m}$ ($k_2 = 112,5 \text{ kN/m}$, $k_3 = 180 \text{ kN/m}$) dla kratownicy: a) z imperfekcją łukową, b) z imperfekcją w formie pierwszej postaci wyboczenia

4. Wnioski

Przeprowadzone analizy stężonej kratownicy, poddanej obciążeniu grawitacyjnemu, pozwalają na sformułowanie następujących wniosków:

1. Na podstawie rezultatów liniowych analiz stateczności (LBA) oraz nieliniowych analiz statycznych (GMNIA) można stwierdzić, że istnieją pewne progowe (minimalne) sztywności podpór bocznych wiązarów kratowych, które zapewniają maksymalną (lub zbliżoną do maksymalnej) nośność konstrukcji.
2. Obciążenie graniczne kratownicy za wstępną imperfekcją łukową, otrzymane przy sztywnym bocznym podparciu, było o 29% większe w porównaniu do rezultatu uzyskanego dla konstrukcji z imperfekcją w formie postaci wyboczenia.
3. Zastosowanie stężeń o zmiennym rozkładzie sztywności (k_1 , k_2 , k_3) powodowało niewielki wzrost obciążenia granicznego kratownicy (do 2%) w porównaniu do przypadku, gdy sztywność podpór bocznych na długości kratownicy była jednokowa (k). Badania wykazały, że w tych przypadkach różnice między siłami w najbardziej wyężonym stężeniu sięgały 14%.
4. Obciążenie zastępcze powstałe na skutek imperfekcji wiązara, wyznaczone na podstawie aktualnych przepisów normowych [9] jest obciążeniem liniowym o stałej intensywności. Rozkład sił stabilizujących badaną konstrukcję, wyznaczony na podstawie analiz numerycznych, był w każdym przypadku nierównomierny. W wielu przypadkach reakcje w podporach bocznych były znakozmienne, na co istotny wpływ miała zarówno sztywność stężeń, jak i kształt wstępnej imperfekcji geometrycznej kratownicy. Istotny wpływ na otrzymane rezultaty miał też fakt, że siła normalna w ściskanym pasie kratownicy była skokowo zmienna, czego nie uwzględnia zastosowane do obliczeń równanie otrzymane na podstawie wytycznych normy.
5. W większości przypadków reakcje w najbardziej wyężonym stężeniu były większe od sił stabilizujących wyznaczonych na podstawie normy. W tym zakresie obliczeń różnice sięgały 30%.

Prace wykonane w ramach grantu 2022/06/X/ST8/00656 NCN, Miniatura 6, pt. „Wpływ sztywności rotacyjnej połączenia pomiędzy stężeniem a pasem na stateczność kratowych wiązarów dachowych poddanych obciążeniu od wiatru – badania doświadczalne i analizy numeryczne”.

Literatura

1. Biegus A., *Stalowe budynki halowe*, Arkady, Warszawa 2003.
2. Jankowska-Sandberg J., *Wpływ sztywności stężeń bocznych pasa ściskanego na zwichrzenie sprężyste kratownicy stalowej*, Inżynieria i Budownictwo, 5, 2006, s. 264-266.
3. Iwicki P., *Stability of trusses with linear elastic side-supports*, Thin Walled Structures, 45, 2007, s. 849-854.
4. Jankowska-Sandberg J., Kołodziej J., *Experimental study of steel truss lateral-torsional buckling*, Engineering Structures, 46, 2013, s. 165-172.
5. Krajewski M., Iwicki P., *Stability and load bearing capacity of a braced truss under upward wind loading*, Steel Construction – Design and Research, 10(3), 2017, s. 1295-1304.
6. Krajewski M., *Stateczność kratownic ze sprężystymi podporami bocznymi*, Rozprawa Doktorska, Politechnika Gdańska, Wydział Inżynierii Lądowej i Środowiska, 2021.

7. Biegus A., *Trapezoidal sheet as a bracing preventing flat trusses from out-of-plane buckling*, Archives of Civil and Mechanical Engineering, 15(3), 2015, s. 735-741.
8. Krajewski M., *O analizach kratownicy ze ściskanym pasem dolnym*, Inżynieria i Budownictwo, 1, 2020, s. 93-95.
9. PN-EN 1993-1-1 Eurokod 3: *Projektowanie konstrukcji stalowych. Część 1-1: Reguły ogólne i reguły dla budynków*.
10. Czepiżak D., *Uproszczona metoda, obliczania poprzecznych stężeń dachowych*, Inżynieria i Budownictwo, 11, 2013, s. 598-600.
11. *Femap with NX Nastran*, Instrukcja programu Femap, Siemens Product Lifecycle Management Software INC., 2009.
12. Rakowski G., Kacprzyk Z., *Metoda elementów skończonych w mechanice konstrukcji*, Wydawnictwo Politechniki Warszawskiej, Warszawa 1993.
13. Labocha S., Skotny Ł., *Liniowa i nieliniowa analiza MES. Implementacja w programie Femap*, Seria Wydawnicza CAx dla praktyków, Wrocław 2014.

Badania wpływu imperfekcji na rozkład obciążeń stabilizujących stężoną kratownicę

Streszczenie

Praca poświęcona jest analizom numerycznym stateczności i nośności stalowej kratownicy płaskiej. Konstrukcje tego typu często stosowane są jako przekrycia dachowe wielkogabarytowych hal stalowych i innych obiektów budowlanych. Charakteryzują się one dużą nośnością i sztywnością w swojej płaszczyźnie, przy czym istotny wpływ zarówno na ich obciążenie krytyczne, jak i graniczne, mają stężenia. W pracy badano wpływ sztywności stężeń, usytuowanych w płaszczyźnie prostopadłej do płaszczyzny kraty, na nośność wyboczeniową konstrukcji. W tym celu, dla modelu powłokowego badanej kratownicy, przeprowadzono liniowe analizy stateczności (LBA). W pracy przedstawiono rezultaty analiz nieliniowych (GMNIA) dla konstrukcji z imperfekcjami geometrycznymi w formie postaci wyboczenia oraz w formie łukowej krzywizny, zalecanej przez europejskie przepisy normowe. Dla wybranych przypadków przedstawiono zależności pomiędzy obciążeniem kratownicy a reakcjami w sprężystych stężeniach. W pracy zaprezentowano również porównanie pomiędzy obciążeniem stabilizującym kratownicę, wyznaczonym na podstawie obowiązujących norm, oraz reakcjami w sprężystych podporach bocznych uzyskanymi na podstawie analiz numerycznych. Słowa kluczowe: kratownica, obciążenie stabilizujące, imperfekcja, wyboczenie

Analysis of the influence of imperfections on the distribution of loads stabilizing a braced truss

Abstract

The paper is devoted to numerical analysis of the stability and load capacity of a flat steel truss. This type of structure is often used as roof coverings for large-size steel halls and other building objects. Flat trusses can be characterized by high load-bearing capacity and stiffness in their plane, but their critical and limit loads are significantly dependent on bracing system. The influence of stiffness of braces, located in the plane perpendicular to the truss plane, on the buckling capacity of the structure was taken into consideration. For this purpose, linear stability analysis (LBA) were carried out for the shell model of the tested truss. In the paper the results of non-linear analysis (GMNIA) for structures with geometric imperfections in the form of buckling mode and in the form of arc curvature which is recommended by european code regulations was presented. For selected cases, the relation between the truss load and reactions in elastic braces was appointed. The comparison between the load that stabilizes the truss, determined on the basis of obligated standards, and the reactions in the elastic side supports obtained on the basis of numerical analysis, was shown.

Keywords: truss, stabilizing load, imperfection, buckling

Predykcja badań laboratoryjnych w optymalizacji kontroli jakości i obciążenia pracy analizatorów

1. Wprowadzenie

Badania diagnostyczne krwi i innych materiałów biologicznych stanowią ważny element szeroko rozumianego zdrowia publicznego i diagnostyki medycznej. Rocznie przeprowadzane jest ponad 130 mln badań przez samą tylko spółkę Diagnostyka S.A. Jest to przeszło 11 mln pacjentów rocznie i powyżej 60 tys. probówek dziennie. Taki strumień danych stanowi duże wyzwanie infrastrukturalne, a przetwarzanie obejmuje szereg zagadnień z dziedziny *Big Data*. Celem pracy było opracowanie metod predykcji zarówno globalnej liczby badań, jak i wyników dla poszczególnych pacjentów na podstawie ich wcześniejszych wyników. Predykcja ta umożliwia bowiem optymalizację pracy laboratorium na kilku poziomach. Jednym z tych poziomów jest optymalizacja zarządzania zapasami. Znając przewidywaną liczbę badań można precyzyjnie dostosowywać stan zapasów odczynników i materiałów zużywalnych (np. probówek), tak, aby zminimalizować koszt magazynowania, ale jednocześnie utrzymać ciągłość pracy analizatorów i nie dopuścić do przestojów związanych z brakiem materiałów zużywalnych i odczynników. Drugim poziomem optymalizacji jest kontrola jakości. Kontrola jakości odbywa się co najmniej raz dziennie na danym parametrze, a w uzasadnionych przypadkach częściej. Jeśli pomiar kontrolny wskaże odchylenia, konieczna jest rekaliibracja urządzenia. Jeśli założenia metacharakterystyk parametru w badaniach pacjentów pozwoliłyby na przewidywanie konieczności rekaliibracji, można by w ten sposób zmniejszyć liczbę pomiarów kontrolnych, które wiążą się nie tylko ze zużyciem materiałów kontrolnych, ale także i czasu pracy analizatorów na rzecz pomiarów kontrolnych wstępnych i właściwych, niezbędnych do oceny statystycznej.

2. Dane

Dane dotyczące wyników badań pacjentów są składową pochodzącą z kilku źródeł. Jednym z nich są informacje zawarte w zleceniu na badania tj. np. lista zleconych badań, waga, wiek i płeć pacjenta, data badania. Dane są zanonimizowane, natomiast dla każdego pacjenta generowany jest identyfikator pozwalający określić, że pacjent powtarza badania, nie da się jednak po zaszyfrowaniu danych poznać jego tożsamości. Drugim źródłem danych są urządzenia automatycznej analizy – analizatory. Przesyłają one na serwery laboratorium dane o wynikach badania danego parametru i kodu transmisji danych. Takich rekordów, podobnie jak badań, jest około 1000 na sekundę.

¹ andrzej.opalinski@agh.edu.pl, AGH Akademia Górniczo-Hutnicza w Krakowie.

² regulski@agh.edu.pl, AGH Akademia Górniczo-Hutnicza w Krakowie.

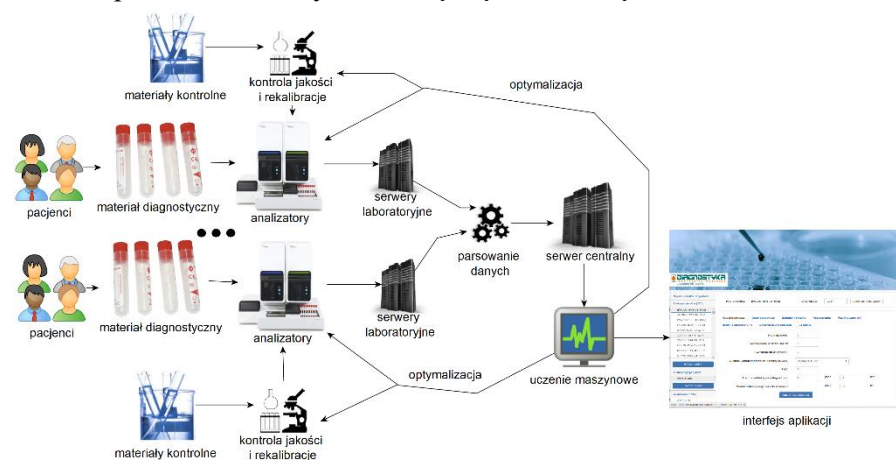
³ Diagnostyka S.A. w Krakowie.

⁴ Diagnostyka S.A. w Krakowie.

⁵ Diagnostyka S.A. w Krakowie.

⁶ Diagnostyka S.A. w Krakowie.

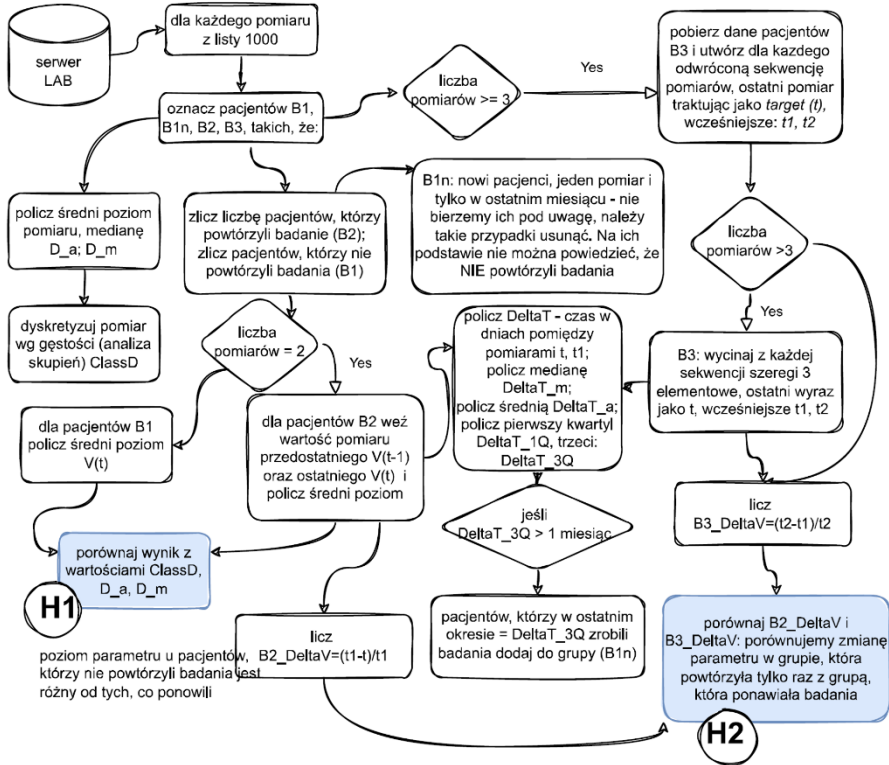
W ramach badań opracowano mechanizm reprezentacji danych (modelu metadanych) umożliwiającego wyekstrahowanie odpowiednich i użytecznych informacji z zasobów danych. Ponieważ przetwarzanie wszystkich danych jest z punktu widzenia predykcji niepotrzebne i nieoptymalne. Należy zatem przejść z poziomu transakcyjnego na poziom analityczny, z poziomu pojedynczego laboratorium na poziom centralny (rys. 1). Wykonana została analiza relacyjnych struktur bazodanowych laboratorium pod kątem wykorzystania ich w etapach związanych z predykcją, optymalizacją obciążenia analizatorów oraz statystycznej rekalkulacji urządzeń. Wykonywana była wstępna analiza danych pomiarowych, m.in. określenie parametrów statystycznych poszczególnych zmiennych, badanie miar dyspersji i miar tendencji centralnych, badanie rozkładów cech i wyznaczenie zakresów zmienności. Analizowano użyteczności narzędzi statystycznych w kontekście próbkowania danych z różną częstotliwością.



Rysunek 1. Schemat przepływu danych w systemie optymalizacji pracy laboratoriów diagnostycznych [opracowanie własne]

3. Profile pacjentów

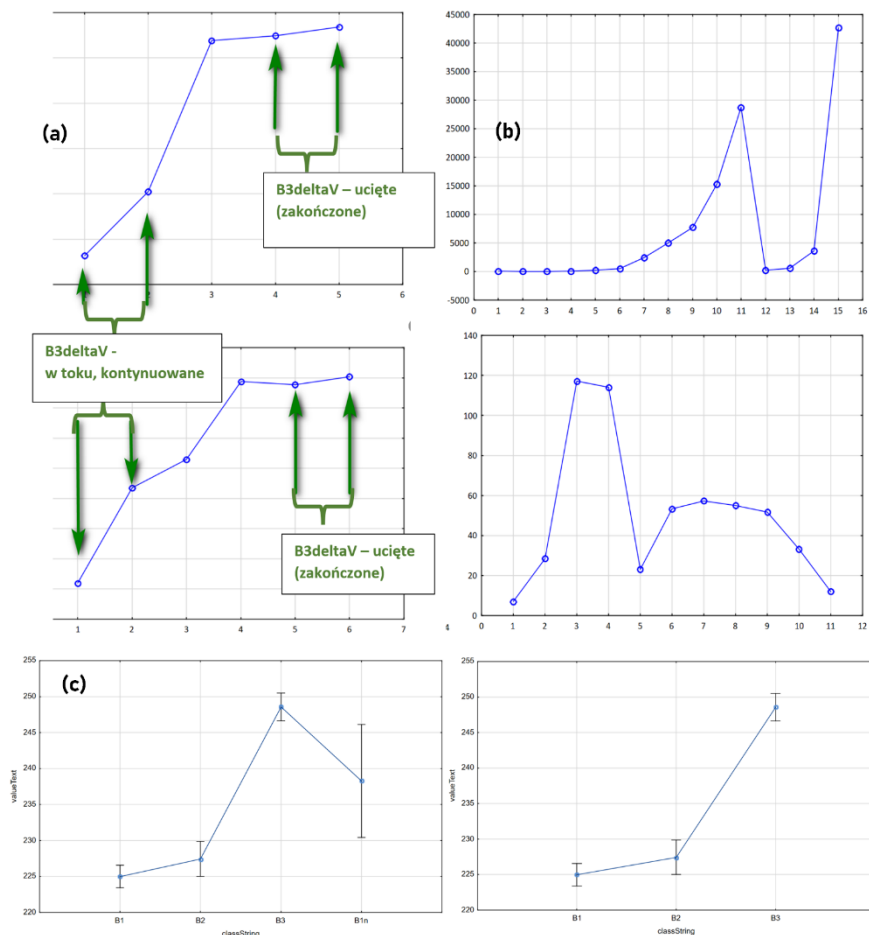
Jednym z opracowanych modeli był mechanizm profilowania pacjentów. Profilowanie takie ma na celu wyszczególnienie możliwie zbliżonych grup, dla których następnie opracowane zostaną odrębne modele predykcji. Podstawą takiej metodyki było założenie, że pacjenci różnią się od siebie, a te różnice przekładają się na ich decyzje dotyczące badań diagnostycznych. Częstość wykonywania badań jest podstawową cechą wyróżniającą. Pacjent zdrowy, wykonujący badania kontrolne, najczęściej nie wraca w ciągu roku. Jego wyniki nie pozwalają przewidywać następnych zleceń. Jednak około 10% pacjentów powtarza badania ponad 10-krotnie (rys. 2).



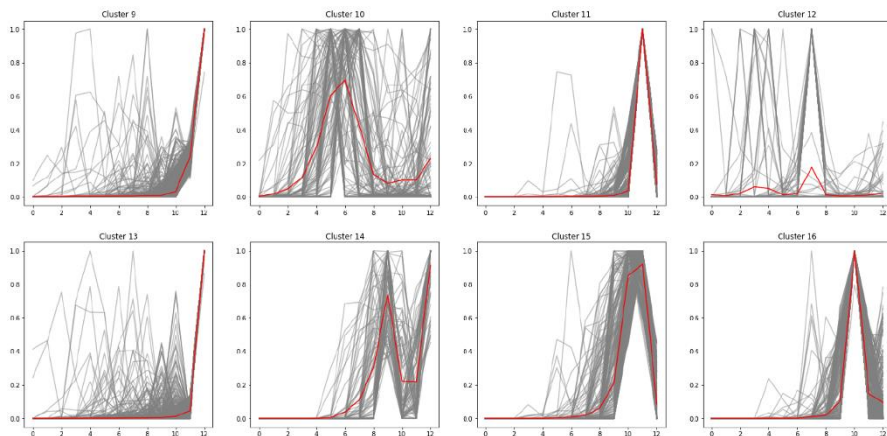
Rysunek 2. Procedura profilowania grup pacjentów ze względu na liczbę pomiarów [opracowanie własne]

Aby profilowanie takie miało podstawy do przeprowadzenia wnioskowania, należy zweryfikować, czy istnieje istotna różnica między wynikami pacjentów z poszczególnych profili [1, 2]. Analiza taka, wraz z porównaniem krzywych wyników pomiarów, przedstawiona została na rysunku 3.

Na podstawie pacjentów powracających na badania możemy opracować modele tzw. wzorców profili pacjentów. Inaczej bowiem przedstawiają się przebiegi wyników (krzywe) pacjentów zdrowych, inaczej chorych. W takich badaniach inna jest specyfika ich krzywych, inne prawdopodobieństwo powtórzenia badania [3-5]. Niektóre parametry (np. poziom hormonów) mogą mieć charakterystyczny przebieg (krzywą), na podstawie której lekarz może określać stan pacjenta (np. Beta HCG pozwalający na określenie momentu przebiegu ciąży). Na podstawie zmiany tej krzywej można wnioskować, czy pacjent będzie ponawiał badania, czy też terapia została zakończona – można takie wzorce pozyskiwać w wyniku uczenia maszynowego, nie wykorzystując w tym celu wiedzy ekspertów. Do tego wykorzystane zostały metody uczenia nienadzorowanego, takie jak sieci Kohonena (samoorganizujące się mapy, SOM [6]). Umożliwiają one w wyniku uczenia korelacyjnego wyszukiwanie obiektów podobnych do siebie pod względem charakterystyki przebiegu sekwencji (rys. 4). W modelowaniu wykorzystano także narzędzia do badania sekwencji: PrefixSpan oraz rekurencyjne sieci neuronowe RNN [7]. Odkrywanie takich wzorców pozwoliło na przewidywanie następnych wyników badań, jak również prawdopodobieństwa powtórzenia badania w ramach klastrów pacjentów, co umożliwia budowanie bardziej precyzyjnych i dopasowanych modeli [8, 9].



Rysunek 3. Analiza różnic w parametrach (a) i charakterystyce przebiegu przykładowych krzywych (b) u pacjentów z poszczególnych profili. Badanie istotności różnic pomiędzy profilami (c) [opracowanie własne]

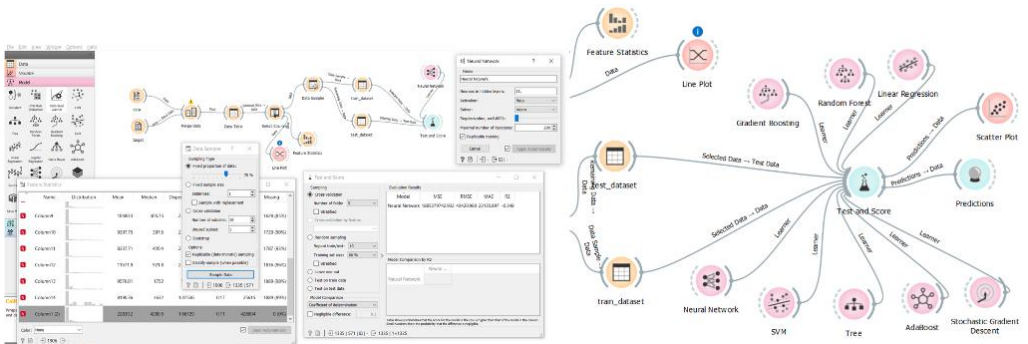


Rysunek 4. Fragment wyników segmentacji z wykorzystaniem samoorganizujących się map (SOM) przedstawiający profile pacjentów [opracowanie własne]

4. Predykcja wyników badań

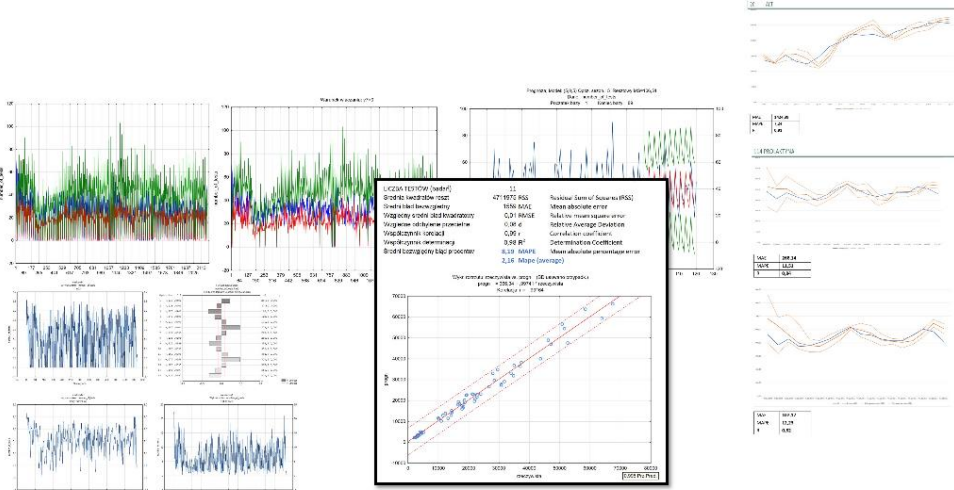
Na podstawie danych z pomiarów wykonano modele z wykorzystaniem uczenia maszynowego (rys. 5). Dane zostały przetworzone do postaci sekwencji pomiarów dla każdego pacjenta. Pacjenci podzieleni zostali na klastry z wykorzystaniem wspomnianych w poprzedniej sekcji modeli.

W badaniach wykorzystano szereg różnych algorytmów: regresję liniową, gradient *boosting*, *random forest*, maszyny wektorów nośnych (SVR), drzewa regresyjne oraz sztuczne sieci neuronowe (ANN). Najlepsze wyniki uzyskano z wykorzystaniem sieci neuronowych.



Rysunek 5. Fragment schematu metodyki uczenia maszynowego z wykorzystaniem pakietu Orange Data Mining, diagram przepływu danych [opracowanie własne]

W analizie szeregów czasowych związanych z globalną liczbą zleceń wykorzystano także tradycyjne metody analizy szeregów czasowych. Modele wykładniczego oraz modele ARIMA pozwoliły na opracowanie modeli predykcji globalnej liczby zamówień w grupach pacjentów o dokładności dopasowania na poziomie (w zależności od parametru) od 91% do 99% współczynnika determinacji (R^2) (rys. 6).



Rysunek 6. Przykładowe modele prognozowania szeregów czasowych z użyciem algorytmu ARIMA (po lewej) oraz wykładniczego (po prawej). Wyniki dopasowania (na środku) na poziomie 91-99% współczynnika determinacji (R^2)

5. Modele wyznaczania konieczności rekaliibracji urządzeń diagnostycznych

Kontrola jakości (QC) oparta na pomiarach kontrolnych firmowych materiałów kontrolnych obejmuje tylko analityczny etap procesu powstawania wyników [10]. Techniki kontroli jakości „oparte na wynikach pozyskanych od pacjentów” są powszechnie stosowane w hematologii od czterdziestu lat. Jednak ze względu na kwestie praktyczne nie są one szeroko stosowane w laboratoriach chemii klinicznej. Jedną z metod takiej analizy jest ocena „średnich z normalnych” (AON) zaproponowana przez Cembrowskiego (1984) [11]. Została zastosowana w obliczaniu patologii średniej objętości krwinki i średniego stężenie hemoglobiny w krwinkach. Wprowadziła ona średnią ruchomą oraz ruchomą medianę do badań kontroli jakości w oparciu o wyniki pacjentów. Następnie Cembrowski [11] wykazał, że konwencjonalna strategia QC była silniejsza przy wykrywaniu losowego błędu, natomiast AoN działa najlepiej z precyzyjnymi testami mierzącymi wartości analityczne z niewielkim zróżnicowaniem populacji przy niewielkiej liczbie wartości odstających. MA QC, znana również jako kontrola jakości w czasie rzeczywistym oparta na badaniach pacjentów (*patient-based real-time quality control*, PBRTQC), to matematyczna procedura, która uśrednia wyniki testów pacjentów w czasie rzeczywistym i wykorzystuje uzyskane średnie wartości do celów kontroli jakości [12]. QC oparta na badaniach wykonanych dla pacjenta na ogół wykorzystuje średnią, ale opracowano i oceniono również inne algorytmy, w tym medianę, wykładniczo ważoną średnią ruchomą i in. Kilka ustawień w MA QC można dostosować w celu uzyskania optymalnych warunków dla określonych testów laboratoryjnych, analizatorów lub populacji pacjentów. Obejmują one okno (liczba wyników pacjentów używanych do obliczenia wartości MA), limity alarmowe (górne i dolne granice MA, które wywołują alarm) oraz ustawienia wykluczające określone wyniki pacjenta (tzw. *truncation limits* – górne lub dolne progi, które wykluczają wyniki pacjentów z obliczeń MA). Czułość i swoistość metod opartych o wyniki pacjentów zależą silnie od przyjętych poziomów odcięcia. Cembrowski w swej pracy zauważa, że zakresy referencyjne badane były przez naukowców od dziesiątek lat, ale nadal brakuje skutecznej i uniwersalnej metody wyznaczania punktów odcięcia [13]. Wiele zależy od zmienności (wariancji) na poziomie interindywidualnym, czy wrażliwości na stopień międzyosobniczej zmienności mierzonych wartości analitycznych. Wiele zależy od zmienności (wariancji) na poziomie interindywidualnym, czy wrażliwości na stopień międzyosobniczej zmienności mierzonych wartości analitycznych. Weryfikacja poprawności metody pomiarowej może być obciążona różnego typu błędami. Kontrola jakości oparta na rzeczywistych wynikach „normalnych” (wynikach pacjentów przyciętych według progów) polega na obliczaniu średnich z ustalonej liczby wyników i ich późniejszej analizie, które zawierają się w zdefiniowanym zakresie (wartościach referencyjnych) wg następującej procedury: (1) Odrzucenie ze zbioru pomiarów wartości, które znajdują się poza zakresami referencyjnymi. (2) Ustalenie liczby wyników niezbędnych do obliczania średnich (np. przy użyciu procedury Cembrowskiego). (3) Ustalenie zakresów kontrolnych.

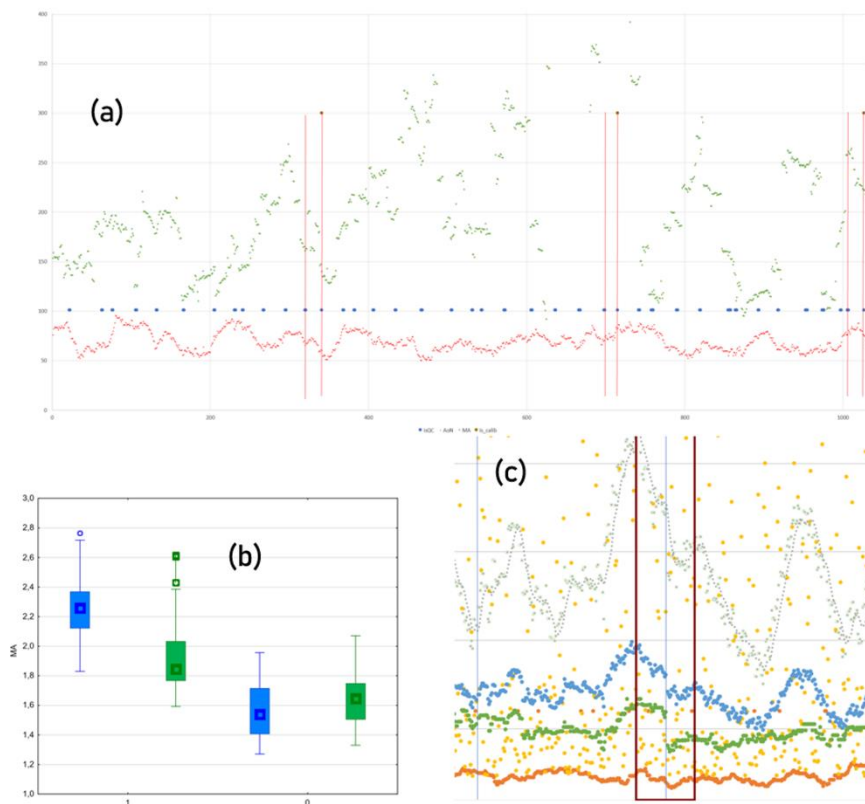
Kontrola jest mocno uzależniona od zdefiniowanych parametrów (liczby pomiarów do obliczenia średniej, zakresów referencyjnych, zakresów kontrolnych, cech badanej populacji itp.). Metoda Cembrowskiego, podobnie jak AoN, bazuje na regule AoN, która jest określona przez granice obciążenia (interwał referencyjny testu), liczbę (n) wyników pozyskanych dla pacjentów w granicach skróconych, które są uśredniane i podane wartości (granice) kontrolne. Jeśli średnia wartość dla bloku jest poza granicami kontrol-

nymi, następuje uruchomienie reguły AoN. Reguła MA jest określona przez kryteria włączenia – jakimi są wyniki pacjentów uwzględnione w średniej; liczba (n) uśrednionych wyników pacjentów oraz granice kontrolne. Uruchomienie reguły powinno skutkować alertem o konieczności przeprowadzenia badań kontrolnych QC.

W praktyce laboratoryjnej, badania QC wykonywane są standardowo np. codziennie rano oraz w przypadku zmiany LOT-u, odczynnika służącego do analiz w danym urządzeniu. Rekalibrację przeprowadza się, gdy kontrola QC wykaże niezgodności, a także według harmonogramu. Celem badań było wskazanie, które kontrole jakości, a szczególnie rekalibracje nie były konieczne. Konieczność wynikająca z niepoprawnych wyników kontroli jakości może zostać potwierdzona, lub nie, na podstawie wyników badań pacjentów. Rozbieżność w określeniu, które kalibracje nie były niezbędne wynika z czułości aparatów pomiarowych, jak również procedury QC. Kontrola jakości prowadzona jest w oparciu o serię badań wstępnych z użyciem odczynników referencyjnych dostarczonych przez producenta. Na podstawie pomiarów wstępnych oraz danych nominalnych można określić parametry, takie jak odchylenie standardowe. Odchylenie międzypersoniczne czy nawet interindywidualne, jest dużo wyższe, co oznacza, że kontrola jakości z użyciem próbek analitycznych jest dużo czulsza niż badania na pacjentach. Algorytmy statystyczne wykorzystywane do kontroli jakości na danych pacjentów: AoN; Średnia ruchoma wyników testu w bloku; Średnia ruchoma logarytmów naturalnych wyników badań w bloku; Średnia ruchoma pierwiastków kwadratowych wyników badań w bloku; Mediana ruchoma wyników testu w bloku; Mediana ruchoma logarytmów naturalnych wyników badań w bloku czy Mediana ruchoma pierwiastków kwadratowych wyników badań w bloku.

Rysunek 7 przedstawia wykres zmienności, na którym widać wyraźnie, że chmury wyników MA wykazują rosnącą zmienność w czasie (rys. 7a), prowadzącą do kontroli jakości, po której występuje kalibracja. Celem jest uchwycenie zmian, które skutkują koniecznością kalibracji (rys. 7c), tak by wykazać brak konieczności kontroli jakości w pozostałych punktach. Celem takich modeli jest uchwycenie zależności wpływu wskaźników na oznaczenie kalibracji jako niezbędnej (DIFF = 1) lub niekoniecznej (DIFF = 0) (rys. 7b), czyli szczególnie identyfikacja punktów kontroli jakości, których można by nie podejmować, jako że nie są niezbędne.

Sieci neuronowe uzyskały optymalną architekturę sieci jako MLP 9-8-2, czyli sieć, posiadającą na wejściu 9 neuronów (determinowanych zmiennymi wejściowymi), jedną warstwę ukrytą, w której osadzonych jest 8 neuronów oraz jeden neuron na wyjściu. Model taki uzyskał dopasowanie na poziomie 1,9% błędu w klasie 0 (czyli niecałe 2% błędnie przewidzianych kalibracji niekoniecznych) oraz 3,2% błędu w klasie 1. Natomiast całkowita trafność modelu wynosi 98,8%. Modele wskazują, że największą siłą dyskryminacyjną (ważność) pozwalającą na ocenę niezbędności kalibracji ma średnia ruchoma (MA) oraz jej pierwiastek kwadratowy, logarytm naturalny mediany ruchomej oraz AoN (średnia z normalnych).



Rysunek 7. Przykładowe przebiegi algorytmów statystycznych wykorzystywanych do kontroli jakości na danych pacjentów (a); potwierdzenie istotności różnic w średnich dla zmiennej dyskryminującej recalibracje (b); widoczny „uskok” w charakterystykach pomiarów parametrów po recalibracji (widoczna istotność) (c)

Wykazano, że najlepsze wyniki predykcji uzyskujemy jednak przy zastosowaniu wszystkich siedmiu wskaźników:

1. AoN.
2. Średnia ruchoma wyników testu w bloku.
3. Średnia ruchoma logarytmów naturalnych wyników badań w bloku.
4. Średnia ruchoma pierwiastków kwadratowych wyników badań w bloku.
5. Mediana ruchoma wyników testu w bloku.
6. Mediana ruchoma logarytmów naturalnych wyników badań w bloku.
7. Mediana ruchoma pierwiastków kwadratowych wyników badań w bloku.

W wyniku klasyfikacji za pomocą przedstawionych modeli możliwe jest z dużą dokładnością przewidzieć, czy kontrola jakości jest niezbędna czy niekonieczna. Dzięki takim rozwiązaniom możliwe jest przewidywanie 98% spośród 11% średniej liczby wykonywanych badań, które stanowi kontrola jakości, kalibracje i powtórzenia.

6. Podsumowanie

Rozpoznawanie częstości wzorców sekwencji czy analiza szeregów czasowych pozwoliła na pozyskanie wiedzy z danych, które dotąd nie były eksplorowane w celu analiz przekrojowych. Korzystano z metod eksploracyjnych – wyrównywania wykładni-

czego, modeli ARIMA, metod średnich normalnych (*average of normals*, AoN), procedury Cembrowskiego i innych. Ponieważ problem prognozowania dotyczył m.in. sekwencji badań, wykorzystane zostały też metody związane z analizą procesów stochastycznych (w tym m.in. PrefixSpan). Wykorzystano techniki uczenia maszynowego, tj. drzewa decyzyjne (*decision trees*, DTs), losowy las (*random forest*, RF) czy sztuczne sieci neuronowe (*artificial neural networks*, ANN). Zastosowane zostały różne architektury sieci neuronowych, począwszy od perceptronów wielowarstwowych (*multi layer perceptron*, MLP) aż po *deep learning* oraz uczenie nienadzorowane w postaci sieci SOM – samoorganizujących się map Kohonena. Stosowanie wielu narzędzi podyktowane było różnymi rodzajami analiz, ale także koniecznością doboru najtrafniejszych modeli. Różnorodność technik i modeli pozwoliła w efekcie skonstruować pakiet modeli pozwalających na prognozowanie liczby badań i wyników badań, na zbiorze parametrów i w przekrojowej analizie według profili pacjentów. Uzyskane wyniki posiadają bardzo wysoką precyzję, w większości parametrów przekraczającą 95% dopasowania, a dla najtrudniejszych modeli nie gorszą niż 90%. Uzyskane wyniki mogą w istotny sposób wpłynąć na optymalizację pracy laboratoriów diagnostycznych.

Podziękowania

Badania te były współfinansowane przez Europejski Fundusz Rozwoju Regionalnego w ramach Programu Inteligentny Rozwój 2014-2020, projekt „System automatycznej prognozowania badań laboratoryjnych oraz optymalizacji kontroli jakości i obciążenia pracy analizatorów działający w oparciu o machine learning” dofinansowany w ramach konkursu organizowanego przez Narodowe Centrum Badań i Rozwoju grant POIR.01.01.01-00-1259/20-01 oraz Diagnostykę S.A.

Literatura

1. Friedman J.H., *Multivariate adaptive regression splines*, The Annals of Statistics, 1991, 19(1), s. 1-67.
2. Nowak J., Korytkowski M., Scherer R., *Discovering Sequential Patterns by Neural Networks*, International Joint Conference on Neural Networks (IJCNN), 2020, 1-6, DOI: 10.1109/IJCNN48605.2020.9207461.
3. Mrzygłód B., Regulski K., Opaliński A., *Uczenie maszynowe w materiałoznawstwie – Machine learning in material science*, [w:] Danielewska A., Mołdoch-Mendoń I. (red.), *Technologie XXI wieku – aktualne problemy i nowe wyzwania, Tom 3*, Wydawnictwo Naukowe TYGIEL, Lublin 2021, ISBN: 978-83-67104-07-4, s. 105-116.
4. Mrzygłód B., Opaliński A., Regulski K., Głowacki M., Kania A., Bochenek G., Nastalek P., Celejewska-Wójcik N., Śladek K., *Application of Industry 4.0 mechanisms in medicine on the example of an integrated support system for treatment of Obstructive Sleep Apnea (OSA)*, Computer Methods In Materials Science, 17(4), 2017.
5. Opaliński A., Regulski K., Mrzygłód B., Głowacki M., Kania A., Nastalek P., Celejewska-Wójcik N., Bochenek G., Śladek K., *Medical data exploration based on the heterogeneous data sources aggregation system*, Proceedings of the 2019 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, 18, 2019, s. 591-597, DOI: 10.15439/2019F258, https://annals-csis.org/Volume_18/drp/pdf/258.pdf.
6. Javed A., *SOMTimeS: Self Organizing Maps for Time Series Clustering and its Application to Serious Illness Conversations*, 2021. DOI: 10.48550/ARXIV.2108.11523.
7. Weiping T., *RNN-based signal classification for hybrid audio data compression*, Computing, 102, 2019, s. 813-827.

8. Plonsky L., Oswald F.L., *Multiple Regression as a flexible alternative to ANOVA in L2 Research*, Studies in Second Language Acquisition, 2016, s. 1-14, DOI: 10.1017/S0272263116000231.
9. Janecki J., *Application of statistical features of the Gaussian distribution hidden in sets of unselected medical laboratory results*, Biocybernetics and Biomedical Engineering, 28(3), 2008, s. 71-81.
10. Badricka T., Cervinski M., Lohc T.P., *A primer on patient-based quality control techniques*, Clinical Biochemistry, 64, 2019, s. 1-5.
11. Cembrowski G.S., Chandler E.P., Westgard J.O., *Assessment of „Average of Normals” quality control procedures and guidelines for implementation*, American Journal of Clinical Pathology, 81(4), 1984, s. 492-499.
12. Fleming J.K., Katayev A., *Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: For patients by patients*, Clinical Biochemistry, 48, 2015, s. 508-513.
13. Cembrowski G.S., Cervinski M.A., *Demystifying Reference Sample Quality Control*, Clinical Chemistry, 62(7), 2016, s. 907-909.

Predykcja badań laboratoryjnych w optymalizacji kontroli jakości i obciążenia pracy analizatorów

Streszczenie

Pandemia COVID-19 uświadomiła nam wszystkim jak ważna jest wydajna medyczna diagnostyka laboratoryjna. Wydajność diagnostyki uzależniona jest w najwyższym stopniu od właściwego zarządzania zarówno zasobami, jak i jakością badań. Funkcjonalny system zarządzania jakością jest integralną częścią każdego laboratorium diagnostycznego. Zapewnia niezawodność i właściwe standardy pracy laboratorium. System taki obejmuje zarówno utrzymanie poprawnych i niezawodnych wyników badań analitycznych, ale również optymalne wykorzystanie mocy przerobowych urządzeń oraz odpowiednią organizację łańcucha dostaw – zarówno materiału analitycznego, jak i odczynników. Niewyobrażalne jest, aby badania nie mogły zostać przeprowadzone tylko dlatego, że brakuje odczynników, czy analizatory są zanadto obciążone pracą lub nie są skalibrowane. Aby zoptymalizować pracę laboratorium, niezwykle ważna jest trafna predykcja liczby zleceń – zarówno w krótkim horyzoncie czasowym, na najbliższe godziny i minuty, ale również w cyklu dłuższym, nawet miesięcznym, tak, aby można było odpowiednio zaplanować zapasy odczynników. W prezentowanych badaniach wykorzystano metody uczenia maszynowego do realizacji czterech funkcjonalności: predykcja liczby wykonywanych badań, optymalizacja zapotrzebowania na odczynniki, kontrola stabilności metody pomiarowej oraz optymalizacja czasu pracy analizatorów.

Słowa kluczowe: uczenie maszynowe, optymalizacja, kontrola jakości, diagnostyka laboratoryjna

Laboratory tests prediction in the optimization of quality control and workload of analyzers

Abstract

The COVID-19 pandemic has made us all aware of the importance of efficient medical laboratory diagnostics. The efficiency of diagnostics depends to the greatest extent on the proper management of both resources and the quality of tests. A functional quality management system is an integral part of any diagnostic laboratory. It ensures the reliability and standards of laboratory work. Such a system includes both the maintenance of correct and reliable results of analytical tests, but also the optimal use of processing capacity of devices and the appropriate organization of the supply chain – both of analytical material and reagents. It is unimaginable that adequate testing cannot be performed on patient samples simply because there is a shortage of reagents, or the analyzers are overburdened, or are not calibrated. In order to optimize the work of the laboratory, it is extremely important to accurately predict the number of orders – both in the short term, for the next hours and minutes, but also in a longer cycle, even monthly, so that reagent supplies can be properly planned. In the presented studies, machine learning methods were used to implement four functionalities: prediction of the number of tests performed, optimization of the demand for reagents, control of the stability of the measurement method and optimization of the working time of the analyzers.

Keywords: machine learning, optimization, quality control, laboratory diagnostics

Wpływ zmiany sztywności płyty nad słupem na wartość sił wewnętrznych w płycie żelbetowej konstrukcji płytowo-słupowej

1. Wprowadzenie

Jednym z powszechnie stosowanych w poprzednich latach, obecnie, a także zapewne w przyszłości systemów konstrukcyjnych są ustroje płytowo-słupowe, które wykorzystuje się między innymi przy tworzeniu obiektów parkingowych czy niewysokich budynków użyteczności publicznej, biurowej, magazynowej lub też o charakterze przemysłowym – szczególnego przeznaczenia. Jak sama nazwa wskazuje składowymi elementami takich konstrukcji są płaskie, poziome powierzchnie płyty oraz pełniące rolę podpór słupy (rys. 1). W zależności od wymagań związanych z zapewnieniem odpowiedniej nośności w tego typu konstrukcjach można spotkać zwieńczenia słupów w postaci głowic o różnym kształcie przekroju poprzecznego: najczęściej kwadratowym, kołowym czy też wielobocznym.



Rysunek 1. Przykład konstrukcji płytowo-słupowej – ustrój płytowo-słupowy z usztywniającą ścianą szczytową (widoczne na zdjęciu ściany z elementów ceramicznych zostały wykonane po przygotowaniu monolitycznej konstrukcji w związku z czym stanowią jedynie obciążenie konstrukcji) [materiały własne]

Ze względu na wykonywanie obliczeń (obecnie najczęściej realizowanych numerycznie za pomocą specjalnego oprogramowania), przy wyznaczaniu wartości sił wewnętrznych, a następnie zbrojenia, czy to zbrojenia płyty, czy też zbrojenia słupa, projektanta głównie interesują następujące wartości momentów:

¹ miroslaw.wieczorek@polsl.pl, Katedra Konstrukcji Budowlanych, Wydział Budownictwa, Politechnika Śląska, www.polsl.pl.

- moment utwierdzenia słupa w płycie (oznaczony i przyjęty w pracy jako M_{utw}) określany na elemencie prętowym odzwierciedlającym element słupowy bezpośrednio w miejscu połączenia płyty ze słupem – najczęściej określany jako reakcja zamocowania. W przypadku obliczeń ręcznych wykonywanych przy użyciu np. metody ram zastępczych określany jako iloczyn reakcji w punkcie przegubowego podparcia słupa w środku jego wysokości i wartości reakcji poziomej w punkcie zamocowania;
- średni moment zginających w płycie (oznaczony i przyjęty w dalszej części pracy jako M_{cal}) – obliczany jako całka z pola momentów odczytywanych z określonej szerokości płyty według wymagań projektanta lub jak w przypadku korzystania z metody ram zastępczych o określonej z góry zdefiniowanej szerokości np. 1/2, 1/4, 1/8 rozpiętości analizowanego pasma – postępowanie w przypadku obliczeń ręcznych. W przypadku stosowanych obecnie obliczeniowych programów komputerowych wspomagających obliczenia statycznie szerokość analizowanego fragmentu nie ma znaczenia.

Dostępny na rynku różnorodny zestaw programów obliczeniowych, wśród których można wymienić między innymi AxisVM, Robot Structural Analysis Professional, ABC-Płyta, czy też zestaw oprogramowania Konstruktor umożliwia bez większych problemów i trudności obliczyć, praktycznie każdy rodzaj konstrukcji bez względu na materiał wykonania czy też system konstrukcyjny, a także kształt oraz wymiary rzutu poziomego danego obiektu. Jednakże zastosowane w nich lub wprowadzone przez użytkownika (projektanta) w czasie tworzenia numerycznego modelu uproszczenia mogą prowadzić do powstania znaczących błędów obliczeniowych, które mogą prowadzić do niewłaściwego określenia potrzebnego zbrojenia, a tym samym uzyskania niewystarczającej nośności elementu żelbetowego.

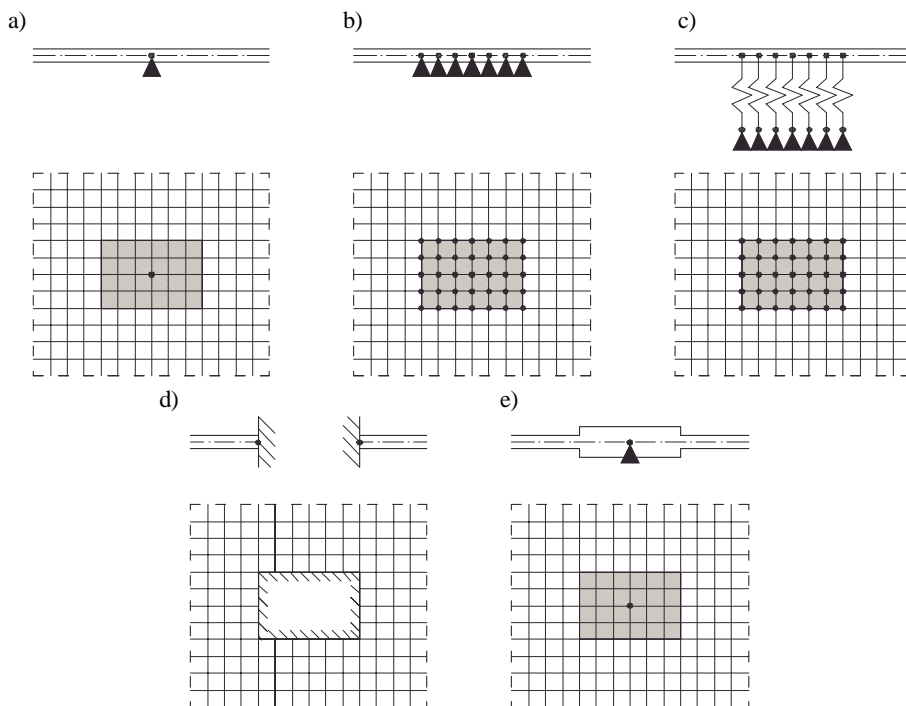
W przypadku konstrukcji płytowo-słupowych newralgicznym miejscem (elementem) do odzwierciedlenia numerycznego jest połączenie płyty ze słupem. Problem polega na odpowiednim uwzględnieniu wymiarów geometrycznych słupa i płyty w miejscu ich wzajemnego przenikania [1, 2]. W tym zagadnieniu można wyróżnić dwa aspekty:

- *zesztywnienie płyty wymiarami rzutu poprzecznego słupa*
W wymienionych programach, połączenia takie wykonuje się najczęściej poprzez punktowe podparcie płyty w osi przenikającego ją słupa. O ile w przypadku, gdy słup ma niewielkie wymiary przekroju poprzecznego to taki sposób podparcia niewiele się różni od rzeczywistości i uzyskane wyniki obliczeń numerycznych tylko niewiele różnią się od wartości prawidłowej. Jednakże w przypadku odzwierciedlenia sytuacji połączenia płyty ze słupem w obiektach wielokondygnacyjnych, w szczególności przy analizie elementów prętowych w poziomie kondygnacji parterowej, gdzie wymiary poprzeczne słupów są uzależnione głównie od bardzo dużej wartości siły osiowej będącej sumą reakcji z kilku lub kilkadziesiątu kondygnacji znajdujących się powyżej i mają wymiary ponad 80×80 cm to taki sposób modelowania zaczyna być uproszczeniem błędnym.
- *zesztywnienie słupa konstrukcją płyty*
W obliczeniach wykonywanych przy użyciu programów umożliwiających wykonanie przestrzennego modelu konstrukcji przyjmuje się, że wysokość kondygnacji jest wyznaczana pomiędzy płaszczyznami dwóch sąsiednich stropów. Takie przyjęcie umożliwia bezproblemowe uwzględnienie w dalszych obliczeniach obciążenia

wiatrem, czy też wpływów drugiego rzędu np. efektu P-delta. Wadą takiego rozwiązania jest zwiększenie długości wybozeniowej słupa względem warunków rzeczywistych, których to długość elementu słupowego jest pomniejszona o sumę grubości połowy stropu nad i pod analizowaną kondygnacją.

Na przestrzeni ostatnich kilkudziesięciu lat powstało kilkanaście koncepcji odwzorowania sposobu przenoszenia sił wewnętrznych ze słupa na płytę oraz na odwrót z płyty na słup. Na rysunku 2 przedstawiono najczęściej wymieniane w literaturze metody:

- *najprostszy sposób modelowania połączenia płyta-słup (rys. 2a)*
Punktowe podparcie w osi słupa podporą niepodatną (najczęściej o nadanej przez użytkownika charakterystyce geometrycznej odpowiadającej sztywności słupa). Ten rodzaj podparcia nie uwzględnia bezpośrednio wpływu szerokości słupa. Jak podano powyżej, w przypadku szerokich słupów całkowicie nie odpowiada rzeczywistości. Zaletą tego sposobu modelowania jest możliwość zastosowania zarówno w nieskomplikowanych modelach płaskich (przy analizie tylko wydzielonego stropu), jak i również przy analizie przestrzennej całego obiektu.
- *sposób modelowania połączenia płyta-słup za pomocą kilku (kilkudziesięciu) podpór punktowych (rys. 2b)*
Zaletą tego sposobu jest możliwość uwzględnienia wpływu wymiarów przekroju poprzecznego słupa. Wadą jest brak wytycznych, w jaki sposób dobrać sztywności poszczególnych podpór. Nieodpowiedni dobór sztywności poszczególnych podpór prowadzi najczęściej do zawyżenia sztywności elementu podpierającego. W związku z czym taki sposób modelowania połączenia praktycznie jest niemożliwy do wykonania w przestrzennym modelu numerycznym obiektu.
- *zastosowanie podpór podatnych o pewnej sprężystości (rys. 2c)*
Na podstawie pozycji literaturowej [3] w modelowaniu połączenia płyta-słup zgodnie z rysunkiem 2a możliwe jest zastosowanie podpór podatnych o pewnej sprężystości. Podobnie jak na sąsiednim rysunku podpory powinny być umieszczone w każdym węźle elementu skończonego. Również jak w modelu przedstawionym na rysunku 2b występują analogiczne problemy przy wykonywaniu obliczeń na modelu przestrzennym obiektu.
- *koncepcja odzwierciedlenia połączenia płyta-słup polega na „wycięciu” z konstrukcji stropu otworu o wymiarach odpowiadających wymiarom słupa (rys. 2d)*
Wokół powstałego otworu przykłada się podpory symulujące utwierdzenie. Przedstawiona koncepcja w dobry sposób oddaje zachowanie się takiego połączenia. Wadą tego typu rozwiązania jest brak wzajemnego powiązania pomiędzy poszczególnymi kondygnacjami obiektu, a więc brak możliwości określenia wpływu obciążeń znajdujących się na innych kondygnacjach oraz brak możliwości bezpośredniego określenia wpływu obciążenia wiatrem na konstrukcję obiektu, a tym samym nawet pojedynczej kondygnacji.
- *koncepcja uwzględnienia usztywnienia płyty słupem poprzez pogrubienie płyty nad śladem słupa (rys. 2e)*
Według wytycznych podanych w literaturze [4, 5] pogrubienie powinno być o podwójnej grubości płyty oraz powinno być wykonane na powierzchni o wymiarach odpowiadających wymiarom poprzecznym słupa znajdującego się poniżej. Dodatkowo w publikacjach [4, 5] zwraca się uwagę, że podane wskazanie odnośnie pogrubienia można stosować tylko w sytuacji, gdy rozważana sytuacja obejmuje płyty stropowe o grubości w przybliżeniu równej 20,0 cm oraz słupów o wymiarach o przekroju kwadratowym w przybliżeniu $40,0 \times 40,0$ cm.

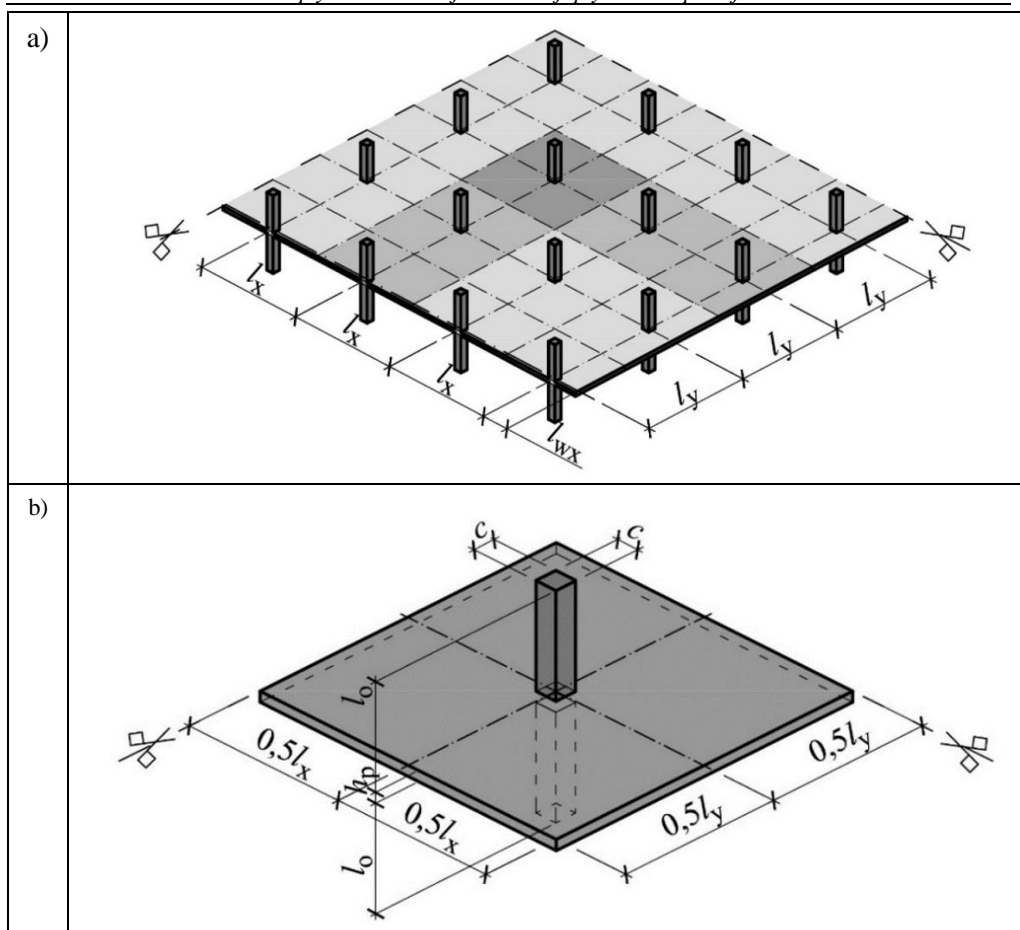


Rysunek 2. Przykład modelowania połączenia płyta-słup (opis w tekście)

W pracach [6-9] w oparciu o propozycje zamieszczone w pozycji [2] przeprowadzane były symulacje uwzględniające podejście (rys. 2e) przy uwzględnieniu zarówno płaskich modeli belkowych oraz modeli przestrzennych, aczkolwiek w niewielkim zakresie zmiany wymiarów geometrycznych. Głównym wnioskiem z prowadzonych prac było stwierdzenie, że brak pogrubienia istotnie wpływa na wyniki.

2. Założenia modelu numerycznego

Celem określenia różnic, jakie powstają w trakcie uproszczonych obliczeń inżynierskich, wykonano numeryczny model fragmentu konstrukcji płytowo-słupowej, wydzielając „myślowo” z konstrukcji wielokondygnacyjnego obiektu najpierw pojedynczą kondygnację, a następnie wydzielając na tej kondygnacji jedno konkretne połączenie. Takie wyodrębnienie kondygnacji i pojedynczych połączeń ma bowiem przy obliczaniu tego typu ustrojów wieloletnią tradycję. Na rysunku 3 przedstawiono widok izometryczny wydzielonej kondygnacji oraz widok wydzielonego, analizowanego fragmentu połączenia płyty ze słupem.



Rysunek 3. Metodologia wydzielenia elementu do analizy (opis w tekście) według [10]

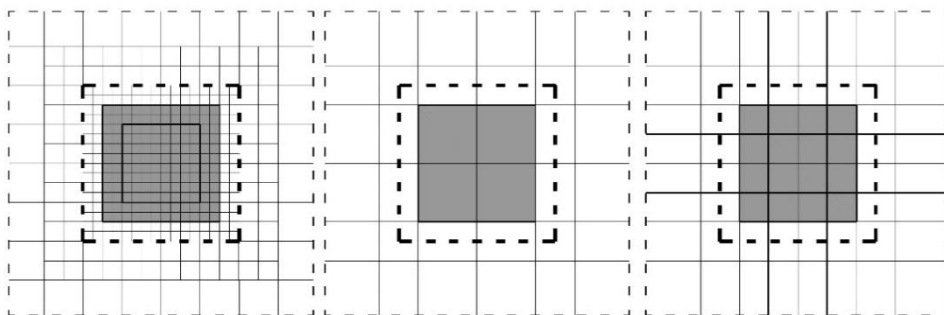
Do zamodelowania połączenia płyta-słup zastosowano dwa odmienne programy komputerowe bazujące na metodzie elementu skończonego:

- W programie Ansys wykonano wzorcowe modele przestrzenne przy użyciu kubicznych elementów skończonych Solid 45. We wszystkich modelach przyjęty został taki sam podział na objętościowe elementy skończone o wymiarach 5×5 cm w poziomie oraz 2 cm w pionie [11].
- Program ABC-Płyta, w którym przy tworzeniu modelu wykorzystano płaskie elementy skończone symulujące płyty stropowe oraz podpory o nadanej charakterystyce odzwierciedlającej geometrię i zachowanie słupów [4, 12, 13]. W zakresie prowadzonej analizy przyjęto kilka różnych wymiarów płaskich elementów skończonych: 80×80 cm, 60×60 cm, 40×40 cm oraz 20×20 cm. Wykonane modele posłużyły do przeprowadzenia analizy, której celem było wykazanie różnic w wartościach sił wewnętrznych, jakie powstają przy uproszczonym sposobie tworzenia numerycznego modelu konstrukcji w oparciu o koncepcję zamieszczoną na rysunku 2e. W modelach uproszczonych przyjęto kilka różnych podziałów płyty elementami skończonymi, które zamieszczono na rysunku 4. Przedstawione podziały obrazują odpowiednio modele o podziale: bardzo dokładnym rzadko spotykanym w prak-

tyce ze względu na dużą liczbę elementów skończonych i długi czas obliczeń (rys. 4a za [1, 2]), najczęściej spotykanym w literaturze i najbardziej rozpowszechnionym w praktyce inżynierskiej wg [4, 13] (rys. 4b) oraz zaproponowanym przez autora (rys. 4c).

W zakresie wykonanych modeli, zarówno objętościowych przeprowadzonych w programie Ansys, jak i modeli prętowo-powłokowych przygotowanych w programie ABC płyta przyjęto jednakowe parametry wytrzymałościowe użytego materiału:

- moduł sprężystości: $E = 30 \text{ Gpa}$;
- współczynnik Poissona: $\nu = 0,2$.



Rysunek 4. Numeryczne modele uwzględnione w analizie – modele powłokowe (opis w tekście)

3. Analiza wyników

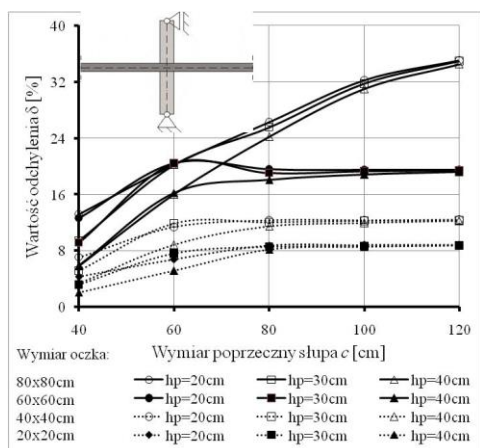
Celem porównania otrzymanych wyników skupiono się, jak wcześniej wspomniano, na obliczeniach momentu M_{utw} , który z programów komputerowych był wyznaczany jako iloczyn reakcji poziomej w punkcie podparcia modelu i ramienia oddziaływania reakcji względem środka połączenia płyty ze słupem – l_o (odległość pomiędzy środkiem geometrycznym połączenia a punktem podparcia słupa).

W celu obliczenia wartości momentów działających w płycie (M_{cal}) otrzymanych z programu Ansys wykonano przeliczenia wartości naprężeń normalnych σ na wartości momentów (M_x) przy uwzględnieniu parametrów geometrycznych modelu.

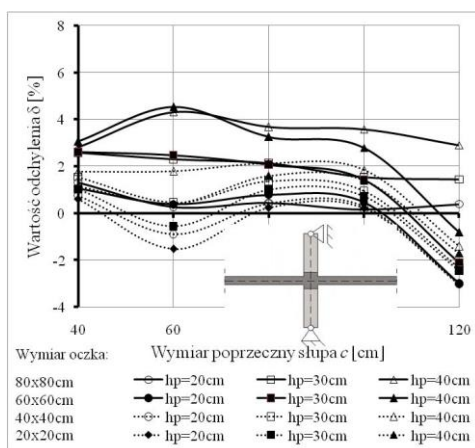
Podobnie przeprowadzono proces tworzenia modelu w programie ABC-Płyta. Płytę stropową zamodelowano za pomocą płaskich elementów skończonych, których podział przyjęto jak na rysunku 4. Dla każdego analizowanego modelu uzyskano wartość momentów utwierdzenia słupa w płycie w miejscu podparcia M_{utw} poprzez odczyt wartości momentu utwierdzenia. Natomiast wartości momentów płytowych odczytano bezpośrednio z programu obliczeniowego i scałkowano, otrzymując wartość momentu M_{cal} .

3.1. Wpływ wymiaru elementu skończonego

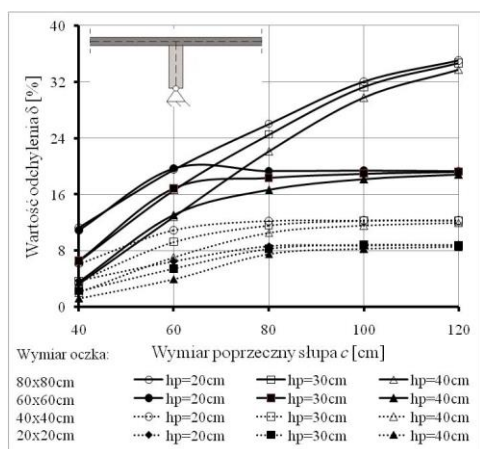
W zakresie wpływu elementu skończonego na wyniki obliczeń przeprowadzono analizę wpływu wielkości oczka elementu skończonego. Do obliczeń przyjęto cztery siatki elementów skończonych o podstawowych wymiarach oczek w prześle: $80 \times 80 \text{ cm}$, $60 \times 60 \text{ cm}$, $40 \times 40 \text{ cm}$ oraz $20 \times 20 \text{ cm}$. Na rysunku 5 i 6 przedstawiono przykładowe wyniki obrazujące wpływ wielkości elementu skończonego jako elementu płytowego, odpowiednio dla analizy momentu utwierdzenia słupa w płycie (M_{pod}) oraz wartości momentów zginających odczytanych w płycie na krawędzi słupa (M_{cal}).



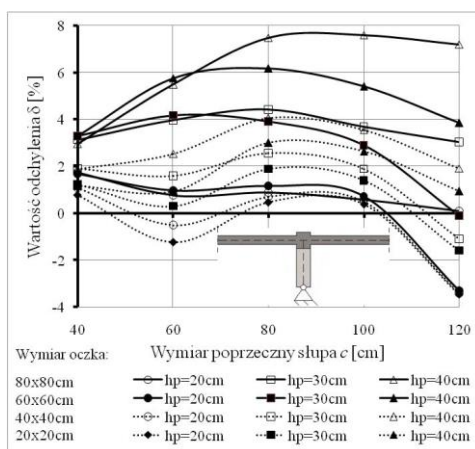
Kondygnacja pośrednia bez pogrubienia



Kondygnacja pośrednia z pogrubieniem



Kondygnacja ostatnia bez pogrubienia

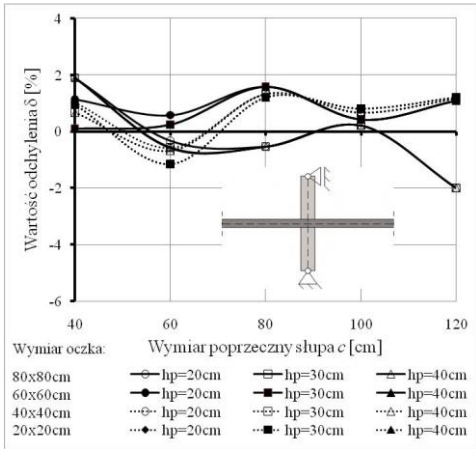


Kondygnacja ostatnia z pogrubieniem

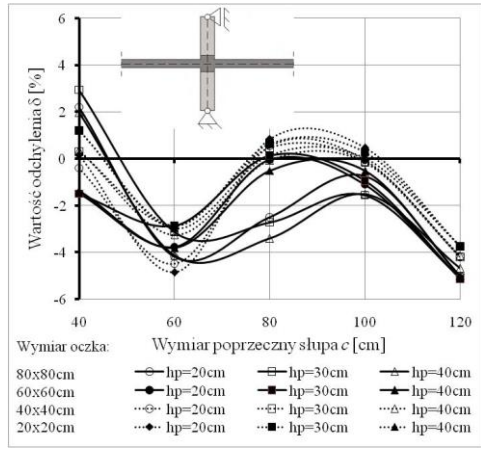
Rysunek 5. Procentowa zmiana wartości odchylenia między momentami utwierdzenia słupa w płycie (M_{pod}) w zależności od wymiaru oczka elementu skończonego

Na podstawie otrzymanych wyników stwierdzono, że:

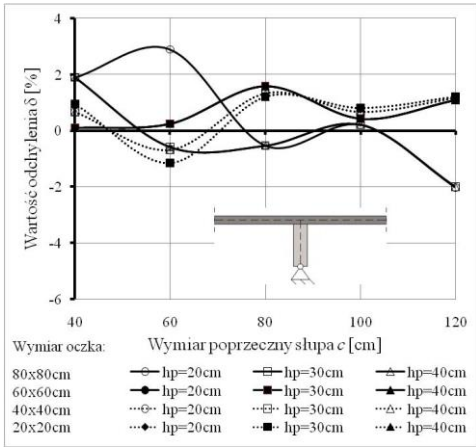
- analizując wpływ zmiany rozmiaru oczka elementu skończonego użytego do wykonania płaskiego modelu płyty stropowej, potwierdzono, że zmniejszenie jego wymiaru skutkuje przybliżeniem modelu inżynierskiego do modelu dokładnego wykonanego w programie Ansys;
- w trakcie obliczeń niezależnie od grubości oraz rozpiętości osiowej punktów podparcia płyty i braku pogrubienia strefy przenikania otrzymuje się znaczne odchylenia względem modelu przestrzennego wykonanego w programie Ansys. Analiza wykazała, że maksymalne odchylenia wartości M_{pod} wynoszą od 35% dla oczka 80 cm, 19% dla oczka 60 cm, 14% dla oczka 40 cm oraz 9% dla oczka 20 cm;



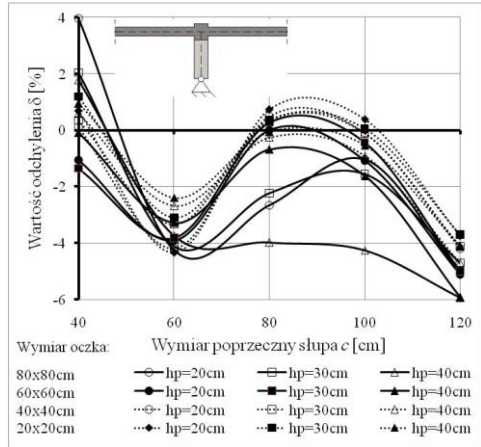
Kondygnacja pośrednia bez pogrubienia



Kondygnacja pośrednia z pogrubieniem



Kondygnacja ostatnia bez pogrubienia



Kondygnacja ostatnia z pogrubieniem

Rysunek 6. Procentowa zmiana wartości odchylenia między średnimi wartościami momentów zginających odczytanymi w płycie na krawędzi słupa (M_{cal}) w zależności od wymiaru oczka elementu skończonego

- przy pogrubieniu płyty nie zauważono znaczących różnic pomiędzy modelem dokładnym wykonanym w programie Ansys a modelami uproszczonymi wykonanymi w programie ABC-Płyta w stosunku do momentów podporowych M_{pod} . Maksymalne wartości różnic momentów podporowych pomiędzy modelami przestrzennymi a modelami płaskimi nie przekroczyły 10%, a zastosowanie 40 cm oczka siatki ograniczyło odchylenie do wartości 4%;
- rozważając momenty płytowe M_{cal} , stwierdzono jeszcze mniejszy wpływ rozmiaru oczka elementu skończonego na rozbieżność modeli dokładnych (modele wykonane w programie Ansys) i modeli uproszczonych (modele wykonane w programie ABC-Płyta). Maksymalna wartość odchylenia nie przekroczyła 6%, przy czym dla najczęściej stosowanych słupów 40 ÷ 80 cm przy schemacie kondygnacji pośredniej i ostatniej błąd nie przekroczył odpowiednio wartości 2% i 4%. Uzyskane dla kondygnacji ostatniej większe wartości różnic można wytłumaczyć znacznym zmniejszeniem sztywności połączenia płyta-słup w stosunku do kondygnacji pośredniej.

3.2. Wpływ wielkości obszaru pogrubienia na wartości momentów w płycie

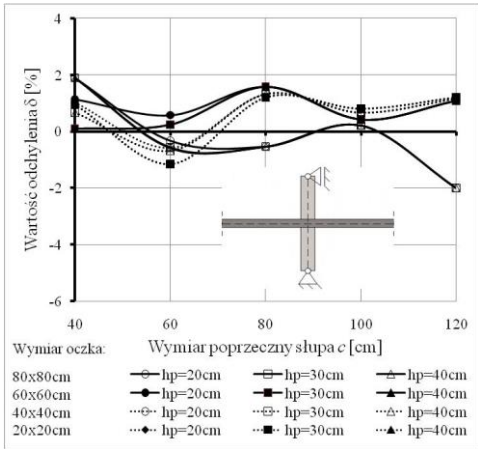
W zakresie drugiej części prowadzonych obliczeń wykonano analizę wpływu wielkości strefy pogrubienia płyty w stosunku do wymiarów rzutu poprzecznego na rozkład sił wewnętrznych pomiędzy elementem stropowym a elementem słupowym.

Prowadzona analiza obejmowała trzy różne pola pogrubienia płyty nad „śladem” słupa o wymiarach $0,5c \times 0,5c$, $0,75c \times 0,75c$ oraz $0,9c \times 0,9c$ (gdzie c to wymiar poprzeczny kwadratowego przekroju słupa). Na podstawie porównań modelu płaskiego z modelem przestrzennym wykazano, że strefa pogrubienia powinna obejmować obszar o wymiarach $0,75c \times 0,75c$. Jednakże uzyskanie takiego obszaru w modelu może być trudne oraz pracochłonne z uwagi na określenie na początku obliczeń siatki elementów skończonych uwarunkowaną rozpiętością przęsła płyty i wymiarami poprzecznymi słupa. W efekcie niejednokrotnie zdarza się, że część płyty, którą należy pogrubić, nie pokrywa się z elementami siatki (rys. 4b). W takiej sytuacji można zagęścić podział, tracąc jednocześnie czas na podzielenie obszaru i przesuwanie węzłów oraz wydłużając następnie czas obliczeń ze względu na liczbę elementów (rys. 4a).

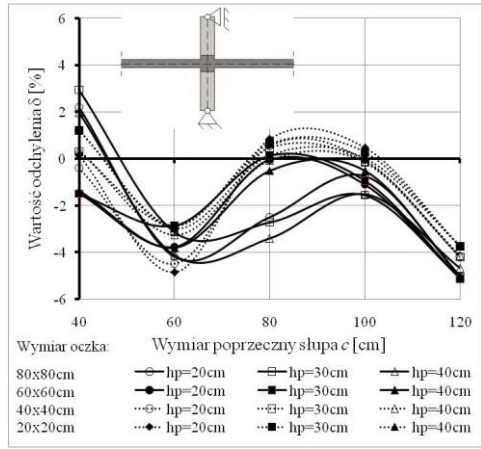
Na rysunku 7 przedstawiono analizę wpływu trzech przypadków pogrubienia płyty na powierzchni o wymiarze $0,5c \times 0,5c$, $0,75c \times 0,75c$ oraz $0,9c \times 0,9c$. Rozważono trzy wymiary oczek siatki elementów skończonych: 80×80 cm, 60×60 cm, 40×40 cm. Dodatkowo, podobnie jak w poprzednich przykładach, sprawdzono wpływ zaproponowanego przez autora dodatkowego zagęszczenia strefy podporowej (rys. 4c).

Na podstawie otrzymanych wyników stwierdzono, że:

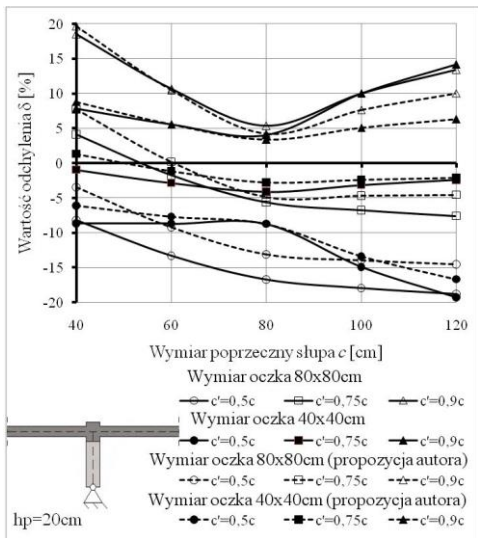
- tak jak podano w opisie problemu, wraz ze zwiększeniem wymiaru poprzecznego słupa zarówno dla sytuacji kondygnacji pośredniej, jak i ostatniej, wartość różnic pomiędzy modelem przestrzennym a modelami uproszczonymi w stosunku do momentów płytowych M_{cal} wzrasta;
- potwierdzono fakt ograniczenia wpływu pogrubienia na wartość odchyień pomiędzy momentami płytowymi (M_{cal}) wraz ze zwiększaniem grubości płyty;
- zmiana wymiaru strefy pogrubienia płyty w miejscu połączenia ze słupem w zakresie od $0,5c$ do $0,75c$ oraz od $0,75c$ do $0,9c$ prowadzi do 20% odchyień w modelach płaskich względem modelu odniesienia wykonanego w programie Ansys;
- zastosowanie dodatkowych linii podziałowych według propozycji autora (rys. 4c) praktycznie w każdym z przypadków ograniczyło wartość odchyień o kilka procent, przy czym przy siatce o oczku 40 cm wartość odchyień w większości przypadków nie przekroczyła 5%;
- przyjęty do obliczeń schemat kondygnacji pośredniej lub ostatniej nie wpływa w istotny sposób na wartość odchyień między uproszczonym modelem powłokowo-prętowym a dokładnym modelem przestrzennym.



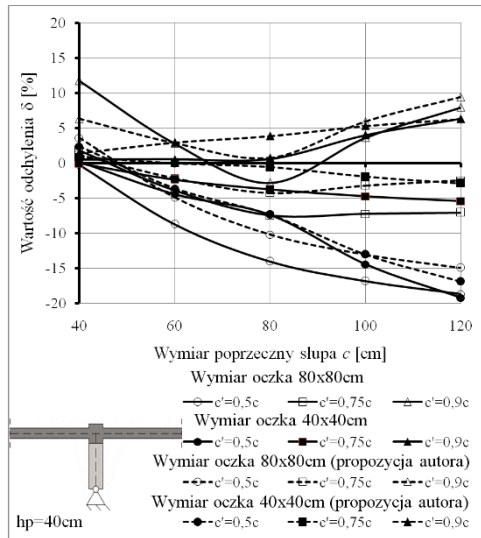
Kondygnacja pośrednia bez pogrubienia



Kondygnacja pośrednia z pogrubieniem



Kondygnacja ostatnia bez pogrubienia



Kondygnacja ostatnia z pogrubieniem

Rysunek 7. Procentowa zmiana wartości odchylenia między średnimi wartościami momentów zginających odczytanymi w płycie na krawędzi słupa w zależności od wymiaru pogrubionej strefy

4. Podsumowanie

Coraz powszechniejsze wprowadzanie do praktyki projektowej programów o różnym stopniu skomplikowania, wykorzystujących w procesie obliczeniowym założenia Metody Elementu Skończonego, w bardzo dużym stopniu ułatwia i przyspiesza obliczenia zarówno w zakresie statyki, jak i następnie przy wyznaczaniu potrzebnej ilości zbrojenia. Stosowane przez projektanta na etapie wykonywania modelu numerycznego uproszczenia mogą jednakże skutkować możliwością powstania znacznych błędów obliczeniowych, co zobrazowano i wykazano w przedstawionych wynikach. Na podstawie wykonanych obliczeń, w których uwzględniono wytyczne i wnioski zawarte w pracach

[1, 4, 5, 13] oraz przeprowadzoną analizę, w podsumowaniu jako wnioski główne można podać, że:

- przyjęty w obliczeniach schemat kondygnacji pośredniej lub górnej nie wpływa na wartości odchyłek między modelem uproszczonym a modelem dokładniejszym;
- potwierdzono wymaganie konieczności usztywnienia obszaru połączenia płyta-słup, ponieważ odchyłki wartości sił wewnętrznych względem modelu standardowego są w przypadku jego usztywnienia kilkukrotnie mniejsze niż w przypadku jego pominięcia;
- zastosowanie zaproponowanego przez autora dodatkowych linii podziału ograniczyło dwukrotność wartości odchyłek między wynikami uzyskanymi w modelach uproszczonych a modelem referencyjnym.

Literatura

1. Rombach G., *Anwendung der Finite-Elemente-Methode im Betonbau. Fehlerquellen und ihre Vermeidung*. Ernst&Sohn Verlag für Architectur und technische, Wissenschaften GmbH, Berlin 2000.
2. Rombach G., *Finite-Element Design of Concrete Structures*, 2nd Edition, ICE Publishing, 2011.
3. Schweizerhof K., Gebhard H., *Modellprobleme bei zusammengesetzten Flächentragwerken. Tagung Baustatik Baupraxis*, Hannover 1990.
4. Starosolski W., *Komputerowe modelowanie betonowych ustrojów inżynierskich. Wybrane zagadnienia*, Wydawnictwo Politechniki Śląskiej, t. 1, Wyd. 6, Gliwice 2022.
5. Starosolski W., *Konstrukcje żelbetowe według Eurokodu 2 i norm związanych*, t. 2, Wydawnictwo Naukowe PWN, Warszawa 2019.
6. Messer T., *FEA in the design of reinforced concrete buildings*, Engineers Media Pty Ltd, 2014.
7. Pacoste C., Plos M., Johansson M., *Recommendations for finite element analysis for the design of concrete slabs*, KTH Royal Institute of Technology, Stockholm 2012.
8. Skorpen S.A., Dekker N.W., *The application and interpretation of linear finite element analysis results in the design and detailing of hogging moment regions in reinforced concrete flat plates*, Journal of the South African Institution of Civil Engineering, 56(1), 2014, s. 77-92.
9. Syrjälä H., *Singularity value management for optimized concrete design*, Technology and Communication, 2019.
10. Ajdukiewicz A., Starosolski W., *Żelbetowe ustroje płytowo-słupowe*, Arkady, Warszawa 1981.
11. Zagrajek T., Krześciński G., Marek P., *Metoda elementów skończonych w mechanicznej konstrukcji. Ćwiczenia z zastosowaniem systemu ANSYS*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2005.
12. Grajek K., *Wersja 6.4 Opis*, PRO-SOFT, Gliwice 2018.
13. Wieczorek M., *Influence of simplified discretization schemes of slab-to-column connections on the values of internal forces in numerical modeling*, IV PhD Students and Young Scientists Conference. Young scientists towards the challenges of contemporary technology, Wydawnictwo Politechniki Warszawskiej, Warszawa 2009, s. 355-363.

Wpływ zmiany sztywności płyty nad słupem na wartość sił wewnętrznych w płycie żelbetowej konstrukcji płytowo-słupowej

Streszczenie

W pracy zaprezentowano zagadnienia związane z wpływem sposobu dyskretyzacji w numerycznym modelu stref połączenia płyty ze słupem. W przeprowadzonej analizie porównano wartości sił wewnętrznych uzyskanych w dokładnym, przestrzennym modelu numerycznym z wynikami, które otrzymano w dwóch rozpatrywanych modelach najczęściej stosowanych w pracach inżynierskich. W obliczeniach uwzględniono sytuację wzajemnego zeszywnienia płyty i słupa, jaka zachodzi w kondygnacjach pośrednich i ostatnich typowych obiektach wykonanych w monolitycznej konstrukcji płytowo-słupowej.

Słowa kluczowe: konstrukcje żelbetowe, ustroje płytowo-słupowe, obliczenia numeryczne, obliczenia inżynierskie, siły wewnętrzne

Influence of a change in the stiffness of the slab over the column on the value of internal forces in the slab of a reinforced concrete slab-column structure

Abstract

The paper presents issues related to the influence of the method of discretization in the numerical model of the plate-column connection zones. In the conducted analysis, the values of internal forces obtained in an accurate, spatial numerical model were compared with the results obtained in the two considered models most often used in engineering works. The calculations take into account the situation of mutual stiffening of the slab and the column, which occurs in the intermediate and last storeys of typical buildings made in a monolithic slab and column structure.

Keywords: reinforced concrete structures, slab-column structures, numerical calculations, engineering calculations, internal forces

Harmonizacja podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ

1. Wprowadzenie

Pojęcie gospodarki o obiegu zamkniętym (GOZ) jest traktowane jako efektywna gospodarka zasobami, w której wprowadzone są rozwiązania typu „win-win”, tj. efektywne, ekonomicznie i ekologicznie [19]. Pakiet działań dotyczących zrównoważonej konsumpcji i produkcji, a przede wszystkim zrównoważonej polityki przemysłowej (COM(2008)0397) ma na celu zwiększenie efektywności środowiskowej produktów w czasie ich całego cyklu życia, popytu na zrównoważone towary, ekoinnowacji produkcyjnej oraz świadomości konsumentów odnośnie minimalizacji marnotrawstwa [15]. Wdrożenie modelu GOZ w celu przeplanowania układów przestrzennych w zakładach przemysłu spożywczego ma zapewnić wysoki poziom bezpieczeństwa wytwarzania produktów żywnościowych na wszystkich etapach projektowania, co doprowadzi do rozwoju gospodarki cyrkularnej.

Co więcej, wymaga to współpracy wielu działów projektowych, a mianowicie architektonicznych, technicznych i technologicznych. Wspólny projekt zapewnia większe zaangażowanie organizacyjne, a przede wszystkim koncentrację na poszczególnych etapach projektowania rozwoju koncepcji aż do powstania projektu wykonawczego. Na każdym z etapów projektowania niezbędne jest przestrzeganie założeń modelu GOZ „jak uzyskać więcej, zużywając mniej” poprzez analizę cyklu życia produktu, zaczynając od dostawy odpowiednich surowców, przez projektowanie procesów technologicznych, gotową produkcję, konsumpcję aż do zagospodarowania odpadów oraz ich recykling. Założenia świadomości korzyści nie tylko ekonomicznych, ale również ekologicznych są punktem odniesienia do utworzenia idealnego układu przestrzennego.

Celem pracy jest usystematyzowanie wiedzy o GOZ, w tym zaprezentowanie nowego podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego, a mianowicie ekoprojektowania uwzględniającego efektywność procesów technologicznych i podnoszenie jakości produktu końcowego, by wydłużyć jego okres przydatności do spożycia, co prowadzi do wzrostu zysków przedsiębiorstwa.

Wykorzystując metody konstrukcyjne oraz graficzne do projekcyjnego planowania procesów technologicznych lub zmiany układów przestrzennych w zakładach przemysłu spożywczego na podstawie stworzonych w naszych badaniach modeli 2D oraz 3D za pomocą programu ArchiCad z wykorzystaniem technologii BIM (*Building Information*

¹ olestr@agh.edu.pl, Katedra Informatyki Biznesowej i Inżynierii Zarządzania, Wydział Zarządzania, AGH Akademia Górniczo-Hutnicza, Kraków, Polska, <https://www.agh.edu.pl/>.

² dzokvaok@gmail.com, Department of Food Engineering, Faculty of Biology and Technology, Kherson State Agrarian and Economic University, ksau.kherson.ua.

³ r.rasulov@knute.edu.ua, Department of Design and Engineering, Faculty of Restaurant, Hotel and Tourism Business, State University of Trade and Economics, Kyiv, Ukraine, <https://knute.edu.ua>.

Modeling), przeanalizowano harmonizację podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ. Dodatkowo, BIM jest rozwinięciem metod CAD, czyli Projektowania Wspomagane go Komputerowo (CAD – ang. *Computer Aided Design*, również *Computer Aided Drafting*). Nacisk został położony na z informatyzowanie całego procesu realizacji projektu, jako zorganizowanego i spójnego strumienia informacji o projekcie łączącego bezpośrednio spójność procesów technologicznych, zasad HACCP i GOZ jako całość. Wyniki pracy pokazują niezbędność synergii wyżej wspomnianych zasad w zakładach przemysłu spożywczego.

Co więcej, w pracy zwrócono uwagę na problem marnowania żywności i jego skalę w Polsce, który pośrednio powstaje w związku z nieprawidłowym zaprojektowaniem układów przestrzennych w zakładach przemysłu spożywczego. W celu zasobooszczędności zakładów w rozdziale zwrócono uwagę na minimalizowanie odpadów poprzez ponowne użycie i wdrożenie pełnego odzysku, np. wielokrotnego zagospodarowania wody, rekuperacji ciepła itd. Można stwierdzić, że zwiększenie świadomości ekologicznej w projekcie ma kluczowe znaczenie dla rozwoju i umacniania gospodarki. Ponadto działania prowadzące do efektywnego przeplanowania układów przestrzennych w zakładach przemysłu spożywczego powinny mieć charakter ekoinnowacji.

2. Wybrane działania w zakresie wspierania procesu ekoprojektowania w kierunku GOZ – ogólne wytyczne

Działania powiązane ze wspieraniem procesu ekoprojektowania dotyczą regulacji zużycia energii, emisji oraz zaplanowania efektywnych systemów ogrzewania, oświetlenia i chłodzenia. Efektywność w ekoprojektowaniu jako kryterium oceny skuteczności dotyczy wydłużenia cyklu życia produktu końcowego, redukcji ilości odpadów i poprawy skuteczności procesów technologicznych. Kolejnym działaniem jest wdrożenie narzędzi do modelowania informacji o budynku (BIM), które umożliwiają projektantom określenie wpływu zaplanowanych procesów technologicznych na środowisko oraz wydajności kosztowej cyklu życia budynków. Technologia BIM umożliwia zaplanowanie w projekcie materiałów oraz urządzeń z możliwością ich naprawy oraz ponownego wykorzystania. Nie mniej ważnym aspektem pozostaje modernizacja procesów technologicznych za pomocą BIM oraz implementacja innowacyjnych rozwiązań odnośnie wydłużenia żywotności budynku w celu poprawy jego efektywności energetycznej. Przyjęcie zasad GOZ w projektowaniu począwszy od fazy planowania, a następnie koncepcji w celu utworzenia projektu szczegółowego – jest niezbędne, aby dojść do wyzwań powiązanych z ekoprojektowaniem. Jednocześnie technologia BIM umożliwia zwiększenie wydajności pracy, pomaga zredukować ilość marnowanych materiałów podczas projektowania, prezentuje rozwiązania skierowane na akceptację zrównoważonych i niskoemisyjnych materiałów budowlanych oraz uwzględnia i monitoruje recykling materiałów, jak i dopasowanie urządzeń technologicznych [7].

Przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ to nowe podejście do zaplanowania produkcji i konsumpcji oparte od samego początku na wzajemnych powiązaniach etapów ekoprojektowych w celu doprowadzeniem do obiegu cyrkularnego procesów technologicznych. W marcu 2020 roku Komisja Europejska w ramach implementacji Europejskiego Zielonego Ładu przyjęła kolejny plan działania UE dotyczący GOZ. W planie działań można wyróżnić następujące pozycje dotyczące projektowania technologicznego w zakładach przemysłu

spożywczego: projektowanie zrównoważonych produktów, obieg zamknięty w procesach produkcyjnych, budownictwo i budynki, żywność, woda i składniki odżywcze, bardziej skuteczna polityka dotycząca odpadów wspierająca zapobieganie powstawaniu odpadów i obieg zamknięty oraz wspieranie procesu przejścia poprzez badania, innowacje i transformację cyfrową [13].

Podczas gdy około 80% wpływu produktów na środowisko określa się wciąż na etapie projektowania, znany liniowy schemat „weź-wyprodukuj-użyj-wyrzuć” nie zapewnia procesu cyrkularnego według zasad GOZ w zakładach przemysłu spożywczego. Wiele produktów spożywczych psuje się zbyt szybko, co więcej, nie można ich przekierować do ponownego zużycia lub poddać recyklingowi w krótkim czasie, a wiele z nich jest przeznaczone wyłącznie do jednorazowego użytku, podczas gdy jednolity rynek dyktuje UE ustanawianie globalnych przepisów i standardów w zakresie zrównoważonego rozwoju produktów końcowych oraz wpływa na projektowanie technologiczne i zarządzanie łańcuchem cyrkularnym na całym świecie [11]. Ponadto należy zauważyć, że dyrektywa w sprawie ekoprojektu [6] z sukcesem reguluje efektywność energetyczną obiegu zamkniętego produktów żywnościowych związanych z energią, podczas gdy instrumenty, takie jak oznakowanie ekologiczne UE [20] lub kryteria zielonych zamówień publicznych UE (GPP) [10] mają jednak ograniczony wpływ z uwagi na „podejście dobrowolne”. W rzeczywistości nie istnieją regulacje prawne ani wymogi zapewniające, że wszystkie produkty żywnościowe wprowadzone na rynek UE odpowiadają zasadom GOZ. Dlatego komisja podjęła inicjatywę ustawodawczą dotyczącą zrównoważonej polityki produktowej z rozszerzeniem zakresu dyrektywy w sprawie ekoprojektu tak, aby ramy ekoprojektowania zapewniały obieg zamknięty w celu uregulowania następujących aspektów:

- wydłużenie cyklu życia produktów żywnościowych, możliwości ponownego użycia według procesów technologicznych, modernizacja technologii, a także zwiększenie efektywności energetycznej procesów produkcyjnych;
- zwiększenie wykorzystania zawartości materiałów oraz urządzeń pochodzących z recyklingu wraz z zapewnieniem ich wydajności i bezpieczeństwa;
- umożliwienie degradacji bioodpadów oraz wysokiej jakości recyklingu;
- zmniejszenie śladu węglowego i środowiskowego podczas produkcji;
- zachęcanie dostosowania modeli, w których producenci zachowują jakość produktu żywnościowego lub odpowiedzialność za jego konsumpcję przez cały cykl jego życia;
- wprowadzenie cyfryzacji, mianowicie informacji o produktach, w tym znakowanie, terminy ważności końcowe;
- motywacja producentów w zakresie postępu produkcji żywnościowej w kierunku zrównoważonego rozwoju.

Nie mniej ważną kwestią pozostaje doskonale przeanalizowanie etapów projektowania w kierunku GOZ. Umożliwia to implementację działań dotyczących gospodarki o obiegu zamkniętym pod kątem wspólnych metodologii i zasad całego projektu z uwzględnieniem wydajnych rozwiązań cyfrowych. Dzięki harmonizacji oraz uporządkowaniu etapów ekoprojektowania zgodnie z zasadami GOZ następuje akceptacja układu koncepcyjnego, który pokazuje zamknięcie obiegu procesów w wybranym zakładzie przemysłu spożywczego.

Zgodnie z Dyrektywą 2009/125/WE pojęcie „ekoprojektowanie” jest uwzględnieniem aspektów środowiskowych poprzez ograniczenie negatywnego oddziaływania produktów na środowisko podczas projektowania układu oraz produktu, aby uzyskać produkt lub układ przestrzenny przyjazny dla środowiska w czasie całego jego cyklu życia [18]. Ekoprojektowanie układów przestrzennych w zakładach przemysłu spożywczego można podzielić na poniższe działania w kierunku GOZ:

- opracowanie efektywnych układów przestrzennych z wdrożeniem innowacyjnych energooszczędnych technologii, w tym z wykorzystaniem zasobów odnawialnych oraz poprzez implementację wyrobów nowych lub przerobionych zgodnie z modelami GOZ (np. RESOLVE);
- tworzenie układów przestrzennych z uwzględnieniem stosowania materiałów recyklingowych oraz surowców odzyskanych z odpadów;
- usytuowanie układów przestrzennych z możliwością montażu maszyn i urządzeń technologicznych ze zwiększeniem trwałości ich pracy oraz z możliwością wydłużenia okresu ich eksploatacji;
- zaprojektowanie efektywnego układu przestrzennego, który wspomaga wydłużenia cyklu życia produktu końcowego;
- założenie w kosztorysie projektu materiałów niewymagających złożonego procesu przetwarzania/wytwarzania produktu końcowego oraz stanowiących korzystniejsze warunki dla środowiska;
- zapewnienie w ekoprojekcie ponownego zastosowania technologicznych maszyn i urządzeń z możliwością ich naprawy;
- uwzględnianie w ekoprojekcie oceny oddziaływania materiałów oraz urządzeń na środowisko w całym cyklu życia produkcyjnego;
- opracowanie standardów oraz norm projektowych w celu efektywnego zaplanowania układu przestrzennego, z uwzględnieniem odzyskiwanych surowców, energooszczędnych materiałów oraz wykorzystanie odnawialnych źródeł energii.

Powyższe założenia należy przestrzegać na wszystkich etapach projektowania układów przestrzennych w zakładach przemysłu spożywczego poprzez procesy wytwarzania i konsumpcję, aż po zagospodarowanie wytworzonych odpadów i ścieków [14].

3. Problem marnowania żywności

Podczas przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ należy uwzględnić, że światowa populacja będzie rosła coraz szybciej, co spowoduje wzrost konsumpcji żywności oraz produkcji odpadów [25]. Ponadto każdego roku na terenie Unii Europejskiej marnuje się prawie 88 milionów ton żywności, a wpływ takich strat na środowisko jest uciążliwy [22].

Dane wskazują, że wartość ekonomiczna marnowanej żywności w skali świata wynosi 2600 miliardów dolarów rocznie, podczas gdy 795 milionów ludzi na świecie cierpią z powodu głodu, natomiast 1,5 miliarda ludzi jest otyłych lub z nadwagą [8]. Jak pokazuje Raport Federacji Polskich Banków Żywności codziennie w Polsce są wyrzucane produkty napoczęte z oznakami przeterminowania i zepsucia oraz zwiędnięte warzywa i owoce. Ponadto w jednej czwartej badanych zakładów sektora spożywczego wyrzucano potrawy ciepłe, a w co piątym pieczywo oraz potrawy serwowane na zimno. Co więcej, stwierdzono, że najczęściej w Polsce marnowane są półprodukty. Dane wskazują, że w zakładach gastronomicznych zmarnowane dania, które zostały przygotowane danego

dnia to w 40,9% surówki, sałaty oraz sałatki, w 20,2% zupy, wywary oraz sosy, a 25,8% to produkty mączne. Poza tym istotne są warunki i sposób przechowywania żywności. Niemal $\frac{3}{4}$ lokali gastronomicznych deklaruje, że zawsze znakuje przygotowane dania datą produkcji. Z kolei aż 40% lokali gastronomicznych nie znakuje na produktach czy przygotowanych daniach informacji o terminie przydatności do spożycia, co wpływa na zwiększenie poziomu marnotrawstwa żywności. Jednak ze względu na złożoność procesów technologicznych w zakładach przemysłu spożywczego trudno dokładnie oszacować skalę marnowania na etapie projektowania. Z danych jednak wynika, że straty żywności w sektorach przetwórstwa spożywczego w Polsce w stosunku do wielkości produkcji generuje się w sektorze owocowo-warzywnym, dalej w branży mięsnej i olejarskiej, a następnie w przemyśle rybnym. Analizując konsumentów, należy wspomnieć, że około 62,9% Polaków deklaruje, że z różną częstotliwością, ale mimo wszystko wyrzuca pieczywo, 57,4% stanowią świeże owoce i warzywa tzw. nietrwale, typu sałata, pomidory, rzodkiewka, 56,5% ogórków, 51,6% wędlin, 47,3% napojów mlecznych [17]. Dlatego zmiany zachowań konsumentów, mianowicie wzrost odpowiedzialności w postępowaniu z odpadami to podstawa do minimalizacji marnotrawstwa.

Wciąż ważnym aspektem pozostaje skierowanie rozwoju ekoprojektowania na minimalizację marnowania żywności, poprzez ilościowe określenie marnotrawstwa w całym łańcuchu wartości żywności. Oznaczałoby to, że należy uwzględnić w ekoprojekcie procesy technologiczne, których obrót produktami żywnościowymi oraz surowcami żywności odbywa się w kontekście minimalizacji marnotrawstwa żywności. Etapy te obejmują obróbkę surowców oraz półproduktów, ich przetwarzanie, a następnie produkcję, dystrybucję oraz konsumpcję żywności w zakładach przemysłu spożywczego. Dlatego marnotrawstwo na jakimkolwiek etapie może oznaczać niepoprawne zaplanowanie technologicznych procesów oraz wybranie nieodpowiedniej koncepcji układu przestrzennego. Z drugiej strony należy zwrócić uwagę również na niewłaściwe wykorzystanie zasobów, takich jak energia, ciepło, woda oraz inne nakłady, które są niezbędne do produkcji tej żywności. Jednym z głównych celów ekoprojektowania w zakładach przemysłu spożywczego jest ograniczenie marnowania żywności i minimalizacja skutków związanych z zanieczyszczeniem [4].

Aby zrozumieć znaczący wpływ marnotrawstwa żywności w ekoprojekcie, należy wykryć punkty zapalne na każdym etapie projektowania, mianowicie należy szczegółowo określić wpływ odpadów w całym projekcie, koszty ich utylizacji, a także przeanalizować w odniesieniu do różnych procesów technologicznych bezpieczeństwo wytwarzania żywności. Tak więc istotnym aspektem pozostaje przestrzeganie zasad odnośnie przechowywania żywności, reżimu temperaturowego w procesach wytwarzania produkcji, prawidłowego zapotrzebowania na energię podczas produkcji oraz przechowywania i transportu żywności. W związku z tym ekoprojektowanie obejmuje analizę każdego etapu projektowego oraz procesu technologicznego, dobór odpowiednich energooszczędnych urządzeń technologicznych oraz chłodniczych w celu zapobiegania marnotrawstwa nawet na etapie początkowym. Jak pokazano na rysunku 1 niemniej ważny pozostaje faktor skalowalności oraz modułowości układu przestrzennego przez możliwość jego rozbudowy, co z kolei umożliwi racjonalne zużycie materiałów. Jednocześnie dobór odpowiednich odnawialnych źródeł energii wspomaga zapobieganiu nadużycia energii oraz ciepła.

Ważnym aspektem pozostaje również weryfikacja zasad przechowania żywności, co z kolei wpływa na poziom jej marnotrawstwa. Powodem jest najczęściej niewłaściwe przechowywanie produktów, nieprzestrzeganie reżimu temperaturowego, przecenienie terminu ważności, zła jakość zakupionych produktów bądź po prostu kupowanie zbyt dużych ilości produktów.



Rysunek 1. Podejście do przeplanowania według zasad GOZ dla przedsiębiorstw przemysłu spożywczego, opracowanie własne

Tworząc odpowiedni projekt przemysłu spożywczego, należy uwzględnić następujące zasady w celu wyeliminowania najważniejszych ogniw łańcucha żywności, począwszy od tworzenia układu przestrzennego z uwzględnieniem wszystkich procesów technologicznych oraz implementacji produkcji podstawowej przez przetwórstwo, sieci handlowe aż po finalnych odbiorców żywności – konsumentów:

- ustalenie funkcjonalnej i ergonomicznej koncepcji układu przestrzennego z zapewnieniem bezpieczeństwa żywności zgodnie z przepisami sanitarnymi, BHP oraz warunkami technicznymi (Prawo Budowlane);
- szczegółowe rozmieszczenie urządzeń i maszyn technologicznych zgodnie z projektem z uwzględnieniem minimalizacji marnotrawstwa żywności;
- przygotowanie dokładnego opisu technologicznego oraz wdrożenie procedur opartych na zasadach HACCP i GOZ [21];
- przeanalizowanie bezpieczeństwa wytwarzania żywności oraz zmniejszenie ilości odpadów;
- zaplanowanie układów przestrzennych z możliwością utrzymania modułowości oraz skalowalności;
- szczegółowe obliczenie w projekcie zapotrzebowania na energię, ciepło i wodę z wdrożeniem procesów rekuperacji ciepła oraz odnawialnych źródeł energii;
- usytuowanie układów przestrzennych z możliwością montażu urządzeń technologicznych i maszyn ze zwiększeniem energooszczędności i trwałości ich pracy;

- zmiana układów przestrzennych, przeprowadzenie relokacji maszyn oraz urządzeń technologicznych podczas reorganizacji zakładów przemysłu spożywczego z możliwością wydłużenia okresu ich eksploatacji i ponownego zużycia.

4. Model procesu projektowego według zasad GOZ

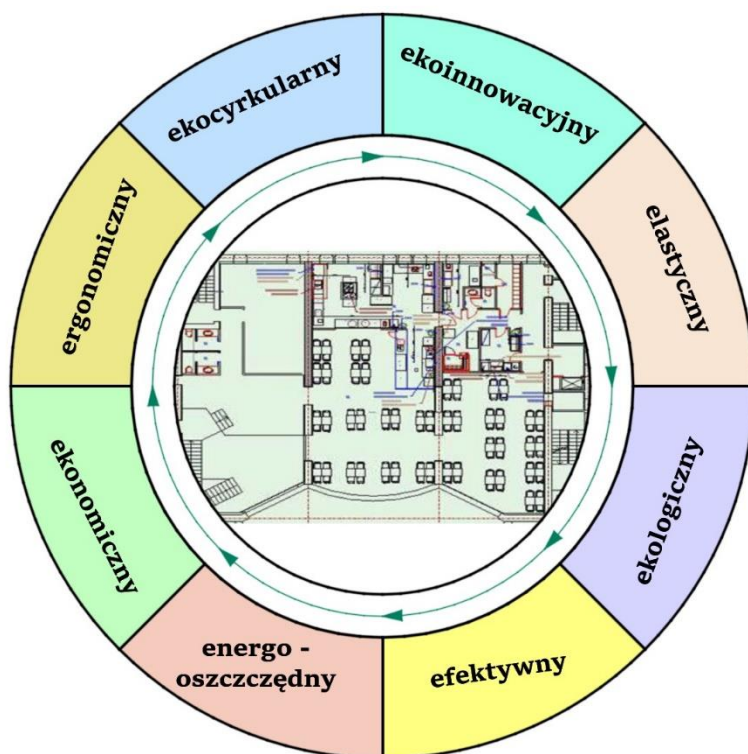
Koncepcja GOZ obejmuje projektowanie procesów produkcyjnych z uwzględnieniem wykorzystania surowców i materiałów wtórnych oraz wydłużenie cyklu życia projektowanego produktu. Zgodnie z normą ISO 14040s, pojęcie „cykl życia” definiowany jest jako „kolejne i powiązane ze sobą etapy”.

Prawidłowe podejście do projektowania układów przestrzennych jest kluczowe w kontekście realizacji koncepcji GOZ. Tradycyjne projektowanie zakładów przemysłu spożywczego uwzględnia potrzeby odbiorcy końcowego, mianowicie różnorodne produkty spożywcze, które można kupić lub zamówić w restauracjach. Analizując projektowanie w kontekście GOZ, należy znacznie szerzej spojrzeć na ten proces. Należy spojrzeć na gruntowne przygotowanie specjalistów ze specjalistyczną wiedzą inżynierską i technologiczną, którzy mogą podejść do kreatywnego podejmowania decyzji w zakresie ekoprojektowania, usytuowania oraz eksploatacji linii technologicznych, maszyn i urządzeń według zasad koncepcji GOZ.

Ponadto oprócz zaplanowania finalnego produktu, który jest bezpośrednio skierowany na konsumenta, należy uwzględnić wpływ produkcji na każdy element łańcucha technologicznego i jego interakcję z uczestnikami procesu projektowego w celu wprowadzenia produktu końcowego do obiegu zamkniętego. Analizę podjęto z wykorzystaniem zasady 6E, która wywodzi się z zasady 4E i została rozszerzona do zasady 8E, co z kolei pozwoliło uzyskać efektywny model procesu projektowego według zasad GOZ [3, 23].

Nacisk został położony na wspólne dane wejściowe, dla których powstają prawidłowe synergetyczne relacje między procesami. Istotne jest sprecyzowanie każdego procesu, który może sprzyjać powstawaniu synergii polegającej na wzajemnie korzystnych oddziaływaniach modelu GOZ [23]. Poszukiwanie rozwiązań modelu produkcji i konsumpcji [16] w odniesieniu do tradycyjnych koncepcji ma na celu implementację efektywnych rozwiązań, dzięki którym zostanie zredukowane zużycie zasobów (tj. energii, wody, materiałów, żywności itd.) w zakładach przemysłu spożywczego, z łagodnym wpływem na klimat. Zasada 8E jest podstawą dla ekoprojektowania, jako że integruje rozwiązania technologiczne pod kątem uwarunkowania zasad GOZ. Zasada 8E jest również niezbędna dla projektowania układów przestrzennych w zakładach przemysłu spożywczego w aktualnym czasie. Ponadto należy przestrzegać norm projektowych oraz uwzględnić warunki klimatyczne i wymogi inwestorów, aby dochodziło do integracji i synergii każdego etapu projektowego.

Niemniej jednak, wybór końcowego układu przestrzennego zależy od klienta przy założeniu, że wszystkie normy projektowe będą spełnione łącznie ze wszystkimi dobranymi kryteriami [24]. Na rysunku 2 został pokazany model procesu projektowego według zasad 8E oraz zasad GOZ, który odnosi się zarówno do pojedynczego zakładu, jak i szeregu zakładów przemysłu spożywczego powiązanych między sobą procesami wytwarzania.



Rysunek 2. Model procesu projektowego według zasady GOZ i zasady 8E dla przedsiębiorstw przemysłu spożywczego, opracowanie własne na podstawie [23]

Ten model można scharakteryzować według procesów w następujący sposób:

Ekologiczny proces – synergia architektury środowiska przemysłowego z innowacyjnym podejściem technologicznym bazuje na odpowiedzialności ekologicznej. Kontekst ekologiczny układu przestrzennego stanowi istotny czynnik wpływu na środowisko zabudowane. Analiza procesu ekologicznego jest oparta na założeniach, że środowisko zabudowane powinno być zaprojektowane racjonalnie i chronić środowisko naturalne. Należy wziąć również pod uwagę, że projektowanie układu przestrzennego polega na uwzględnieniu procesów wytwarzania ekologicznej produkcji oraz polepszeniu wartości odżywczych żywności [12]. W tym celu zadaniem głównym jest zaprojektowanie niezbędnych urządzeń technologicznych, lodówek i innych urządzeń do przechowywania żywności. Ponadto należy uwzględnić zwiększenie mocy produkcyjnej urządzeń oraz implementację technologii podnoszących efektywność energetyczną (EE) lub z obszaru odnawialnych źródeł energii (OZE) w zakładach przemysłu spożywczego.

Ekonomiczny proces – synergia w kontekście racjonalnych rozwiązań procesów technologicznych oraz funkcjonalno-przestrzennych w zakładach przemysłu spożywczego. Synergia służąca uzyskaniu optymalnych parametrów nakładów finansowych w procesie projektowania, podczas procesu wytwarzania (proces budowy), aż do późniejszej eksploatacji (proces użytkowania). Proces użytkowania zamieszcza w sobie proces technologiczny, mianowicie proces wytwarzania. Zastosowanie rozwiązań materiałowo-

-konstrukcyjnych oraz technologicznych pod kątem czynnika ekonomicznego jest ważnym filarem GOZ.

Kolejny czynnik ekonomiczny oparty jest na regule 10R (ang. *refuse, rethink, reduce, reuse, repair, refurbish, remanufacture, repurpose, recycle, and recover*) oraz jego zdolności do osiągnięcia zrównoważonego rozwoju [1]. Implementacja zasady 10R prowadzi do określonego planu, unikając zbędnych zakupów i ograniczając w tym konsumpcję, produkcję odpadów oraz ślad węglowy (ang. *carbon footprint*). Zasady te jak na przykład redukcja (ang. *refuse*), racjonalizacja (ang. *rethink*), ponownie zużycie (ang. *reuse*), recykling (ang. *reduce*) skierowane są na przeciwdziałanie marnowania żywności w zakładach przemysłu spożywczego. Naprawa (ang. *repair*) jest kluczowym czynnikiem wyboru urządzenia technologicznego, którego gabaryty i usytuowanie wpływa na formowanie końcowego układu przestrzennego. Co więcej, wydłużony cykl życia urządzeń technologicznych prowadzi do mniejszej ilości zużytego sprzętu elektronicznego i zmniejsza koszty częstotliwości zakupu nowych urządzeń. Odnowienie (ang. *refurbish*) wydłuża cykl życia urządzeń technologicznych; przynosi korzyści środowisku, zmniejszając zużycie nowych urządzeń i wytwarzanie odpadów [9]. Jednak należy uwzględnić tutaj ważny faktor oznaczający, że wraz ze wzrostem liczby zmian technologicznych występuje ryzyko, że długotrwałe użytkowanie urządzenia technologicznego może spowodować przestarzenie go oraz doprowadzić do wzrostu kosztów naprawy i spadku konkurencyjności w odniesieniu do wytwarzania produktów wysokiej jakości, co z kolei może również przysporzyć strat finansowych.

Przyjęcie zasad 10R na linii produkcyjnej pomoże doprowadzić do zamkniętego łańcucha dostaw i zwiększyć trwałość urządzeń technologicznych. Starzenie lub trwałe zużycie części urządzenia prowadzi do jej wymiany lub regeneracji (ang. *remanufacture*), czyli użycia części wyrzuconego urządzenia technologicznego w nowym urządzeniu o tej samej funkcji. Istnieje również możliwość montażu części wyrzuconego urządzenia technologicznego w nowym produkcie z innymi funkcjami, zmieniając ich przeznaczenie (ang. *repurpose*). Może skrócić to czas zamawiania właściwej nowej części i naprawy oraz zmniejszyć przestoje linii produkcyjnej. Co więcej, umożliwi to zakładom przemysłu spożywczego poprawę jakości środowiska i tym samym dostosowanie do celów zrównoważonego rozwoju. Jakościowe tworzenie układów przestrzennych polega na przestrzeganiu zasad koncepcji wytwarzania mniej odpadów od samego początku, aby obniżyć koszty recyklingu (ang. *recycle*), zamiast generować dużo odpadów do recyklingu [2]. Kluczowym faktorem jednak pozostaje uwzględnienie w projekcie technologicznym rekuperacji ciepła. Przemysłowe nadwyżki ciepła od urządzeń technologicznych mogą przejść w proces cyrkulamy, czyli odzyskanie (ang. *recover*) ciepła, co wiąże się ze zmniejszeniem zużycia energii pierwotnej i globalnym ograniczeniem emisji CO₂ [26]. W rezultacie, tworząc układ przestrzenny, należy na każdym etapie projektowym uwzględniać redukcję kosztów w kontekście racjonalnych rozwiązań procesów technologicznych według zasad 10R.

Energooszczędny proces – synergia, dzięki której odpowiednie usytuowanie układu przestrzennego oraz dobór urządzeń technologicznych przyczyni się do optymalizacji energooszczędności. Jest to również zastosowanie rozwiązań materiałowych, instalacyjnych oraz technologicznych pozyskujących energię cieplną ze źródeł niekonwencjonalnych, ale wpływających na oszczędność naturalnych zasobów i wykorzystanie

odnawialnych źródeł energii. Tak na przykład instalacje oparte na technologiach OZE generują darmową energię (słoneczna, wiatrowa, wodna, geotermalna) [3].

Ergonomiczny proces – synergia, dzięki której jest przemyślany wpływ projektowania na optymalizację elementów pracy przez dostosowanie do pracowników procesu produkcyjnego. Jest to także zaprojektowanie właściwego stanowiska pracy w sposób zwiększający bezpieczeństwo komfort pracy człowieka, a także przeprowadzanie analizy relacji człowiek-technologia. Tworząc układ przestrzenny, należy przestrzegać przepisów bezpieczeństwa oraz uwzględniać występujące ograniczenia przestrzenne, czynniki techniczne i organizacyjne [5].

Ekocykularny proces – synergia z myślą o „cyklu życia” na etapach projektowania, budowy oraz użytkowania w celu lepszego gospodarowania zasobami, gdzie wszystkie odpady traktowane są jako materiały wykorzystane przez naturę. W ekocykularnym procesie ważna jest rola projektantów, którzy nawet na wczesnym etapie projektowania biorą pod uwagę cykl życia produktu, a scenariusze ponownego użycia materiałów i produktów definiują na początku procesu projektowego.

Elastyczny proces – synergia poszukiwania elastycznych rozwiązań na każdym etapie projektowania dla całego cyklu życia zaplanowanych procesów technologicznych. Efektem jest uznanie wartości wynikających z założeń, mianowicie zastosowania cyklu obiegu zamkniętego w całym projekcie oraz implementacji nowych form konsumpcji opartych na aktualnych konsumentach. Z drugiej strony elastyczny proces można traktować jako modułowe i funkcjonalne zaplanowanie procesów technologicznych w odniesieniu do produkcji pozwalających na szybkie i dogodne zmiany przestrzeni w zakładach przemysłu spożywczego nawet już w procesie użytkowania [3].

Efektywny proces – synergia poszukiwania ekologicznych rozwiązań wdrożenia efektywnych zmian w zakładach przemysłu spożywczego. Do realizacji systemowego ujęcia kierunków rozwoju planowania należy powiązać wiedzę praktyczną z wiedzą teoretyczną, przyjmując ich współzależność oddziałującą na końcowy wynik przeplanowania. Z drugiej strony pojęcie efektywny proces – to synergia, która dalej rozwija ramy modelowania „Efektywnego Czasu Procesu” do analizy:

- wydajności systemów produkcyjnych i procesów budowlanych;
- przepływu zużywanych surowców, zużywanej energii i materiałów;
- warunków technicznych, technologicznych, organizacyjnych i środowiskowych [23].

Ekoinnowacyjny proces – synergia poszukiwania innowacyjnych i kompleksowych rozwiązań według zasad GOZ, wdrożenia ekologicznych zmian w zakładach przemysłu spożywczego. Występuje tutaj zintegrowane podejście do planowania układu przestrzennego według zasad GOZ. Co więcej, jest to również opracowanie bibliotek wspierających BIM oraz standaryzacja techniki BIM we wszystkich fazach projektowania. Odnosi się do niego także stosowanie metod i narzędzi do ekoinnowacyjnego, zrównoważonego i zintegrowanego projektowania, opierających się na wielokryterialnej optymalizacji decyzji projektowych. Z drugiej strony przy przeplanowaniu układów przestrzennych w zakładach przemysłu spożywczego ważnym aspektem pozostaje implementacja odpowiedniej technologii poprzez wdrożenie systemu inteligentnego budynku ze stworzeniem nowych algorytmów oszczędzania energii zintegrowanych z technologią źródeł odnawialnych.

5. Wnioski/Podsumowanie

W rozdziale badania dotyczą harmonizacji podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ. Podstawowym wymaganiem w zakresie wytwarzania produkcji żywności jest efektywne planowanie układu przestrzennego. Mając na uwadze dbałość o jakość projektowania według zasad GOZ niezależnie od skali produkcji produkowanej żywności, należy podkreślić ważną rolę projektowania technologicznego. Dlatego budowa nowych lub przeplanowanie już istniejących zakładów zależy od prawidłowej wybranej koncepcji i powinna być prowadzona według zasad GOZ i 8E w oparciu o opracowane projekty technologiczne. Podczas procesu projektowego należy przeanalizować etapy technologiczne w celu zmniejszenia marnotrawstwa żywności, mianowicie cyklu życia – od produkcji do konsumpcji należy również wyeliminować zagrożenia wynikające z procesu produkcyjnego. Dlatego wdrożenie ekoprojektu będzie w dalszym ciągu sprzyjać rozwojowi efektywności energetycznej i energooszczędności zakładu oraz doprowadzi do obiegu zamkniętego produkcji żywnościowej. Należy również wspomnieć, że zasady ekoprojektu są skuteczne i powinny być implementowane we wszystkich sektorach przemysłu spożywczego, aby zmniejszyć wpływ na środowisko na początkowych etapach projektowania.

Przejęcie na gospodarkę o obiegu zamkniętym ma charakter systemowy, dlatego już na etapie projektowania należy prawidłowo zaprojektować układ przestrzenny pomieszczeń według zasad 10R, co z kolei pomoże doprowadzić do zamkniętego łańcucha dostaw, zwiększyć trwałość i energooszczędność urządzeń technologicznych oraz zredukować koszty w kontekście racjonalnych rozwiązań procesów technologicznych.

W taki sposób można ułatwić dobór odpowiednich materiałów konstrukcyjnych i wykończeniowych, dostosować systemy urządzeń wodno-kanalizacyjnych i wentylacji oraz dobrać wyposażenie według specyfiki zakładu i rozmiarów produkcji. Czynniki te wpływają na jakość oraz standard higieny wytwarzania produkcji żywnościowej.

Pozostaje jednak obszar do dalszych badań, które mogą zostać przeprowadzone w przyszłości, mianowicie, ocena stopnia ustandaryzowania procesów przeplanowania układów przestrzennych w sklepach spożywczych według zasad GOZ.

Podziękowania

Praca ta została wsparta przez Wydział Zarządzania Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Badania wykonano wspólnie z AGH Kraków oraz Department of Food Engineering, Faculty of Biology and Technology, Kherson State Agrarian Economic University, Ukraine oraz Department of Design and Engineering, Faculty of Restaurant, Hotel and Tourism Business, State University of Trade and Economics, Kyiv, Ukraine.

Literatura

1. Bag S., Gupta S., Kumar S., *Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development*, International journal of production economics, 2021, s. 231, 107844.
2. Bigdeloo M., Teymourian T., Kowsari E., Ramakrishna S., Ehsani A., *Sustainability and circular economy of food wastes: Waste reduction strategies, higher recycling methods, and improved valorization*, Materials Circular Economy, 3(1), 2021, s. 1-9.

3. Biskup K., *Inteligentna architektura w środowisku mieszkaniowym*, Rozprawa doktorska, Politechnika Krakowska (PK), Polska, 2018, s. 47-50.
4. De Oliveira M.M., Lago A., Dal' Magro G.P., *Food loss and waste in the context of the circular economy: A systematic review*, Journal of Cleaner Production, 2021, s. 294, 126284.
5. Del Fabbro E., Santarossa D., *Ergonomic analysis in manufacturing process. A real time approach*, Procedia CIRP, 41, 2016, s. 957-962.
6. Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products, OJ L 285, 31.10.2009, s. 10.
7. European Commission, Leading the way to a global circular economy, SWD(2020) 100, 11 March 2020.
8. Fassio F., Tecco N., *Circular economy for food: A systemic interpretation of 40 case histories in the food system in their relationships with SDGs*, Systems, 7(3), 43, 2019.
9. Guide Jr V.D.R., Van Wassenhove L.N., *OR FORUM – The evolution of closed-loop supply chain research*, Operations research, 57(1), 2009, s. 10-18.
10. https://ec.europa.eu/environment/gpp/eu_gpp_criteria_en.htm [data dostępu: 01.12.2022].
11. <https://eur-lex.europa.eu/legal-content> data dostępu: 07.12.2022].
12. <https://gozwpraktyce.pl/raport/miasta-dla-zywnosci/> [data dostępu: 05.12.2022].
13. <https://gozwpraktyce.pl/regulacja/nowy-plan-dzialania-goz/> [data dostępu: 07.12.2022].
14. <https://krajoweinteligentnespecjalizacje.pl/> [data dostępu: 07.12.2022].
15. <https://www.europarl.europa.eu/factsheets/pl/sheet/77/zrownowazona-konsumpcja-i-produkcja> [data dostępu: 01.11.2022].
16. https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/573899/EPRS_BRI%282016%29573899_EN.pdf [data dostępu: 05.02.2023].
17. <https://www.gov.pl/web/ijhars/marnowanie-zywnosci--raport-2020> [data dostępu: 01.12.2022].
18. <https://www.gov.pl/web/rozwoj-technologie/ekoprojektowanie> [data dostępu: 06.12.2022].
19. Kulczycka J., *Gospodarka o obiegu zamkniętym w polityce i badaniach naukowych*, Wydawnictwo IGSMiE PAN, Kraków 2019, s. 5-6.
20. Regulation (EC) No 66/2010 of the European Parliament and of the Council of 25 November 2009 on the EU Ecolabel, OJ L 27, 30.1.2010, s. 1.
21. Rozporządzenie (WE) nr 852/2004 Parlamentu Europejskiego i Rady z dnia 29 kwietnia 2004 r. w sprawie higieny środków spożywczych.
22. Scherhauser S., Moates G., Hartikainen H., Waldron K., Obersteiner G., *Environmental impacts of food waste in Europe*, Waste management, 77, 2018, s. 98-113.
23. Stryhunivska O., Gdowska K., Rumin R., *Implementation of synergetic planning for factory*, Scientific Papers of Silesian University of Technology – Organization and Management Series, 155, 2022, s. 503.
24. Stryhunivska O., *Integracja wizualizacji 3D z metodami projektowania procesów wytwarzania*, Rozprawa doktorska, AGH Kraków, Polska, 2019, s. 20-22.
25. Tamasiga P., Miri T., Onyeaka H., Hart A., *Food Waste and Circular Economy: Challenges and Opportunities*, Sustainability, 14(16), 2022, 9896, s. 1-2.
26. Viklund S.B., Johansson M.T., *Technologies for utilization of industrial excess heat: Potentials for energy recovery and CO₂ emission reduction*, Energy conversion and Management, 77, 2014, s. 369-379.

Harmonizacja podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ

Streszczenie

Kluczowym etapem, decydującym o wpływie środowiskowym w budownictwie jest projektowanie. Celem artykułu jest analiza podejścia do przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ. Nie mniej ważnym aspektem jest to, że na każdym etapie transformacji przedsiębiorstwa powinna zostać wykonana analiza aktualnego stanu układu przestrzennego. Model będący wynikiem badań naukowych pozwala wdrażać innowacje i nowe inicjatywy zmniejszające konsumpcję zasobów. Co więcej model stara się, aby produkty, ich części i materiały były używane tak długo jak to tylko możliwe. Przeanalizowano etapy technologiczne w celu zmniejszenia marnotrawstwa żywności, mianowicie cyklu życia – od produkcji do konsumpcji. W artykule uwaga została zwrócona na powiązanie procesów technologicznych i na współpracę w łańcuchach wartości w celu rozwiązywania problemów marnowania żywności. Ponadto w niniejszym artykule podane zostały wskazówki stanowiące narzędzie efektywnego wsparcia procesów planowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ. Pozostaje jednak obszar do dalszych badań, które mogą zostać przeprowadzone w przyszłości, mianowicie, ocena stopnia ustandaryzowania procesów przeplanowania układów przestrzennych w zakładach przemysłu spożywczego według zasad GOZ.

Słowa kluczowe: GOZ, układ przestrzenny, ekoprojektowanie, przeplanowanie, zakład przemysłu spożywczego

Harmonization of the approach to the re-planning of spatial layouts in food industry plants according to the principles of circular economy

Abstract

The key stage that determines the environmental impact in construction is design. The aim of the article is to analyze the approach to the spatial layout redesign in food industry plants according to the principles of GOZ. Another important aspect involves the analysis of the current state of the spatial layout at each stage of an enterprise transformation. The research model enables innovation and new initiatives to be implemented in order to reduce resource consumption. Furthermore, the model tries to prolong the usage of products, parts and materials for as long as possible. Technological steps were analysed to reduce food waste, namely, the life cycle – from production to consumption. The article focuses on linking technological processes and on cooperation in value chains to solve the problems of food waste. In addition, this article provides guidelines that are a tool for effective support of spatial layout planning processes in food industry plants according to the principles of circular economy. However, there is still area for further research that may be carried out in the future regarding the assessment of the degree of standardization of spatial layout remodeling processes in food industry plants according to the principles of GOZ.

Keywords: circular economy, spatial layout, eco-design, replanning, food industry plant

Czy sztuczna inteligencja powinna mieć formę jak najbardziej zbliżoną do człowieka?

1. Wstęp

Myśląc o szeroko pojętym zagadnieniu sztucznej inteligencji, jednym z pierwszych skojarzeń, jakie od razu nam się nasuwa, są roboty, które fizycznie wyglądają jak człowiek, naśladują człowieka bądź pomagają człowiekowi w codziennych zadaniach. Roboty te jednak nie muszą mieć sztywnych, mechanicznych ruchów, a z zewnątrz mogą mieć powłokę, która sprawia, że wyglądają jak ludzie. Aby bliżej przyjrzeć się postaci, pod jaką powinna występować sztuczna inteligencja (AI), warto najpierw zdefiniować, chociażby w najbardziej podstawowym zakresie pojęcia, takie jak sztuczna inteligencja, robot oraz android. Odzwierciedlenie przez takie „postacie” ludzkiego kształtu oraz ludzkich ruchów ma sprawić, że poprzez podobieństwo do nas w jak największym stopniu zaufamy takiej jednostce, będziemy czuć się z nią komfortowo oraz będziemy widzieć w niej raczej przyjaznego kompana wyzwań życia codziennego, aniżeli wroga próbującego zdezonizować ludzkość. Niezwykle istotne jest jednak podkreślenie, że o ile faktycznie im bardziej dana istota przypomina człowieka, tym bardziej jej ufamy, to po przekroczeniu pewnej granicy nasze postrzeganie sztucznej inteligencji zaczyna się zmieniać. Od tego konkretnego momentu, przebywając z taką jednostką, czujemy niepokój.

2. Robot a android

Badanie przeprowadzone w 2016 roku pomaga zdefiniować kluczowe pojęcia, które najpierw musimy wyznaczyć celem przybliżenia zjawiska doliny niesamowitości [1]. Dzieli ona roboty na humanoida (robota humanoidalnego), androida (robota androidalnego) i robota niebiomimetycznego. Biomimetyka to nauka obejmująca praktyczne wykorzystanie mechanizmów i funkcji nauk biologicznych m.in. w inżynierii, projektowaniu, chemii oraz elektronice [2].

Roboty humanoidalne to roboty posiadające cechy ludzkie (np. tors, głowę czy oczy). Roboty androidalne są uważane za dokładne kopie człowieka, natomiast roboty niebiomimetyczne nie są oparte na żadnych elementach natury. Percepcję trzech robotów przez uczestników porównano przed i po krótkiej interakcji. Stwierdzono, że percepcja robotów przez ludzi zmienia się w zależności od tego jak roboty zachowują się w porównaniu z ich oczekiwanym zachowaniem.

3. Niejasne tłumaczenie pojęcia *Uncanny valley*

Pojęcie to zostało stworzone przez japońskiego profesora robotyki na Politechnice w Tokyo, twórcę robotów i inżyniera Masahito Moriego w 1970 roku, który badał reakcje ludzi na wygląd zewnętrzny robotów [1]. Angielskie tłumaczenie sformułowania *uncanny*

¹ pawel.urzenitzok@nzs.org.pl, Kolegium Indywidualnych Studiów Międzyobszarowych, Wydział Prawa i Administracji, Uniwersytet Śląski.

valley pierwszy raz zostało wprowadzone w 1978 roku przez Jasia Reichardt w książce pod tytułem „Robots: Fact, Fiction, and Prediction”. W języku angielskim wyraz *uncanny* jest tłumaczony jako *strange* lub *mysterious* [3]. Wyraz *strange* w języku polskim znaczy „dziwny” lub „obcy”, natomiast *mysterious* to „tajemniczy”. Zgodnie ze słownikiem Cambridge English Dictionary, *uncanny* w języku polskim oznacza tyle, co „niesamowity” lub „nieprawdopodobny”, co jak można wywnioskować na podstawie przytoczonych słów jest w miarę podobnym, z pewnością jednak nie idealnie odzwierciedlającym tłumaczeniem. Warto o tym pamiętać, bo sformułowanie dolina niesamowitości jest nacechowane pozytywnie z racji tego, że synonimy wyrazu niesamowity to m.in.: fantastyczny, rewelacyjny, spektakularny, niezwykły, fenomenalny, oszałamiający, cudowny lub fascynujący [4]. Jak jednak dowodzą przeprowadzone badania, zjawisko *uncanny valley* zdecydowanie nie jest pozytywnym fenomenem, bo powoduje niepokój zamiast komfortu.

4. Dolina niesamowitości

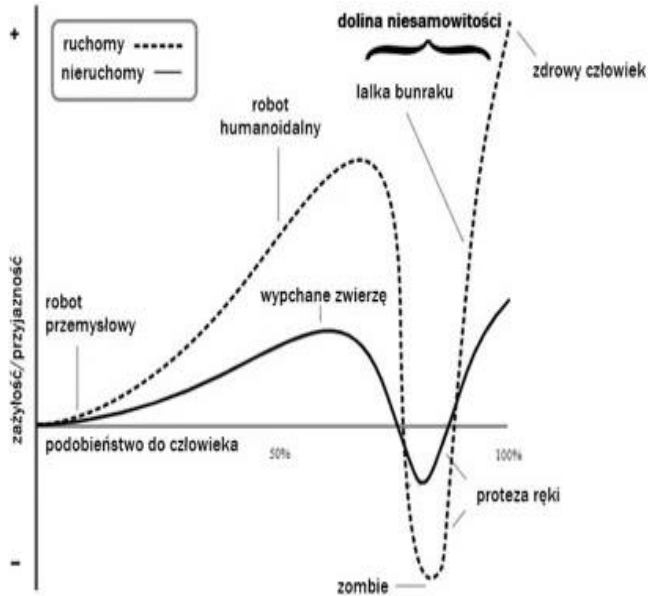
Niepokój, lęk przed nieznanym oraz dyskomfort w sytuacji, gdy widzimy istotę bardzo podobną do człowieka, która jednak nie jest człowiekiem, bo nie odwzorowuje jego pełnego zakresu ruchów lub mikroruchów twarzy dostrzegalnych jedynie przez gałki oczne, powoduje zjawisko zdefiniowane jako „Dolina niesamowitości”. Będąc bardziej precyzyjnym, dolina niesamowitości to ta konkretna chwila, w której dana jednostka sprawia, że jest zbyt zbliżona wizualnie do człowieka, nieujawniając jednak jego cech.

Zdaniem Moriego emocjonalna reakcja na bardziej ludzkiego robota jest empatyczna i pozytywna, jednak po osiągnięciu danego punktu, natychmiast zaczyna nas on odpychać. Co ciekawe, gdy podobieństwo robota jest mniej podobne do człowieka, nasza reakcja emocjonalna znowu zaczyna być pozytywna. Dolina niesamowitości ma miejsce wtedy, gdy czujemy różnicę pomiędzy prawdziwym człowiekiem a „prawie” człowiekiem. Oznacza ona, że mówiąc najprościej czujemy się dziwnie, oglądając coś, co nas przypomina, ale jednak nami nie jest. To właśnie dolina niesamowitości sprawia, że trudno nawiązać efektywną relację między robotem a człowiekiem [5].

Co ciekawe, wynik badania przeprowadzonego w 2004 roku sugeruje, że dolina niesamowitości zmienia się w zależności od wieku osoby badanej [6]. Rezultat wskazuje, że dolina niesamowitości jest najbardziej widoczna we wczesnym dzieciństwie, a mniej w dorosłości.

Warto wspomnieć o konkretnych przykładach, bo właśnie tak najłatwiej będzie zrozumieć to zjawisko. Generalnie, im bardziej dany obiekt jest podobny do człowieka, tym lepiej się z nim czujemy, jak jednak przekroczy pewien punkt, czyli będzie wyglądał zbyt ludzko, to poczucie naszego komfortu będzie spadało. Istotne jest, że w zależności od tego czy obiekt jest ruchomy czy nieruchomy, krzywa doliny niesamowitości jest inna. Gdy obiekt nie jest ruchomy, to nasz poziom komfortu w miarę zwiększania się podobieństwa do człowieka w porównaniu z obiektem ruchomym jest mniejszy. Zjawisko doliny niesamowitości również nie jest tak silne jak w przypadku jednostki ruchomej. Bazując na najsłynniejszym wykresie dotyczącym doliny niesamowitości, całkowicie komfortowo czujemy się z robotem przemysłowym, który ani nie wygląda jak człowiek, ani nie naśladuje jego ruchów. Analogicznie, gdy widzimy człowieka, również nie odczuwamy dyskomfortu. W przypadku robota humanoidalnego,

czujemy się spokojnie tylko do pewnego momentu, po którego przekroczeniu krzywa drastycznie się zmienia i czujemy niepokój. Co ciekawe, w przypadku takiej samej protezy ręki efekt doliny niesamowitości będzie mniejszy w przypadku, gdy proteza nie będzie ruchoma i analogicznie silniejszy, gdy będzie się poruszała.



Rysunek 1. Wykres wraz z krzywą zjawiska doliny niesamowitości obrazujący obiekty ruchome i nieruchome. Źródło: Opracowanie na podstawie [7]

W tym miejscu warto wspomnieć również o najnowszym osiągnięciu w mime-tyzmie (naśladowaniu) – narzędziom do generowania twarzy, dzięki *machine learning* i *deep learning*. Aby zrozumieć o co chodzi, warto chociaż w podstawowym zakresie przybliżyć pojęcia oraz rozróżnienie między *deep learning* i *machine learning*.

Użycie wyrażenia *Deep Learning* w kontekście sztucznych sieci neuronowych zostało wprowadzone przez Igora Aizenberga w 2000 roku [8]. *Deep learning* to koncepcja *machine learning* oparta na sztucznych sieciach neuronowych. W wielu zastosowaniach modele głębokiego uczenia przewyższają płytkie modele uczenia maszynowego i tradycyjne podejścia do analizy danych [9].

Machine learning jest dziedziną informatyki, która, zgodnie z Arthurem Samuelem, daje komputerom zdolność uczenia się bez bycia wyraźnie zaprogramowanym [10]. Jest ściśle związana ze statystyką obliczeniową, która również skupia się na tworzeniu predykcji za pomocą komputerów. Ma silne związki z optymalizacją matematyczną, która dostarcza metody, teorie i domeny zastosowań do tej dziedziny [11].

Przykładem takiego narzędzia do generowania twarzy może być strona internetowa „that person does not exist” stworzona przez Philipa Wanga, 33-letniego inżyniera oprogramowania. Generator umożliwia wygenerowanie nowej twarzy za każdym razem jak odświeżymy stronę internetową, za co jest odpowiedzialny algorytm GAN (ang. *generative adversarial network*). Wygenerowane ludzkie twarze są niesamowicie realistyczne, jednak żadna z tych osób tak naprawdę nie istnieje. Projekt strony powstał, aby pokazać moc sztucznej inteligencji, która pozwala tworzyć obrazy, które mimo podobieństwa do prawdziwych odpowiedników rzeczy lub ludzi są całkowicie sztuczne.

Profesorowie z Uniwersytetu Kalifornijskiego w 2011 roku zbadali przyczyny występowania zjawiska doliny niesamowitości [12]. W badaniu brało udział 20 osób w wieku 20-36 lat, które poprzednio nie miały żadnych doświadczeń z robotami, nie przebywały w Japonii ani nie posiadali rodziny lub znajomych w tym kraju, bo roboty są tam niezwykle popularne. Aby zbadać to, dlaczego u ludzi występuje dolina niesamowitości wykorzystano nagranie człękopodobnego robota Repliee Q2 stworzonego na Uniwersytecie w Osace, który dzięki ruchomym oczom, brwiom, policzkom, ustami oraz szyi mógł naśladować ludzką ekspresję twarzy [13].

Nagranie obejmowało robota, który pił wodę, podnosił przedmiot ze stołu lub kiwał głową. Następnie te ruchy były wykonane przez kobietę, na której wzorował się Repliee Q2. Na końcu, naukowcy rozdarli powłokę robota, żeby było widać jego okablowanie.

Potem pokazano nagrania badanym osobom i poinformowano ich, która osoba była robotem, a która człowiekiem oraz przeprowadzono skan ich mózgow za pomocą rezonansu magnetycznego. Podczas oglądania prawdziwego człowieka oraz robota, który faktycznie nie był podobny do człowieka, bo było widać zasilające go kable to reakcje ludzi były normalne. Co ciekawe, w przypadku patrzenia na robota, który bardzo przypominał człowieka kora ciemieniowa zwłaszcza w obszarach, które łączą korę wzrokową, przetwarzającą ruchy ciała, z częścią kory ruchowej, zawierającą neurony lustrzane (empatii), odnotowano wysoki poziom aktywności.

Podczas oglądania androida, w korze ciemieniowej – a konkretnie w obszarach łączących część kory wzrokowej mózgu, która przetwarza ruchy ciała, z częścią kory ruchowej, która zawiera neurony lustrzane (lub neurony empatii) – odnotowano wysoki poziom aktywności.

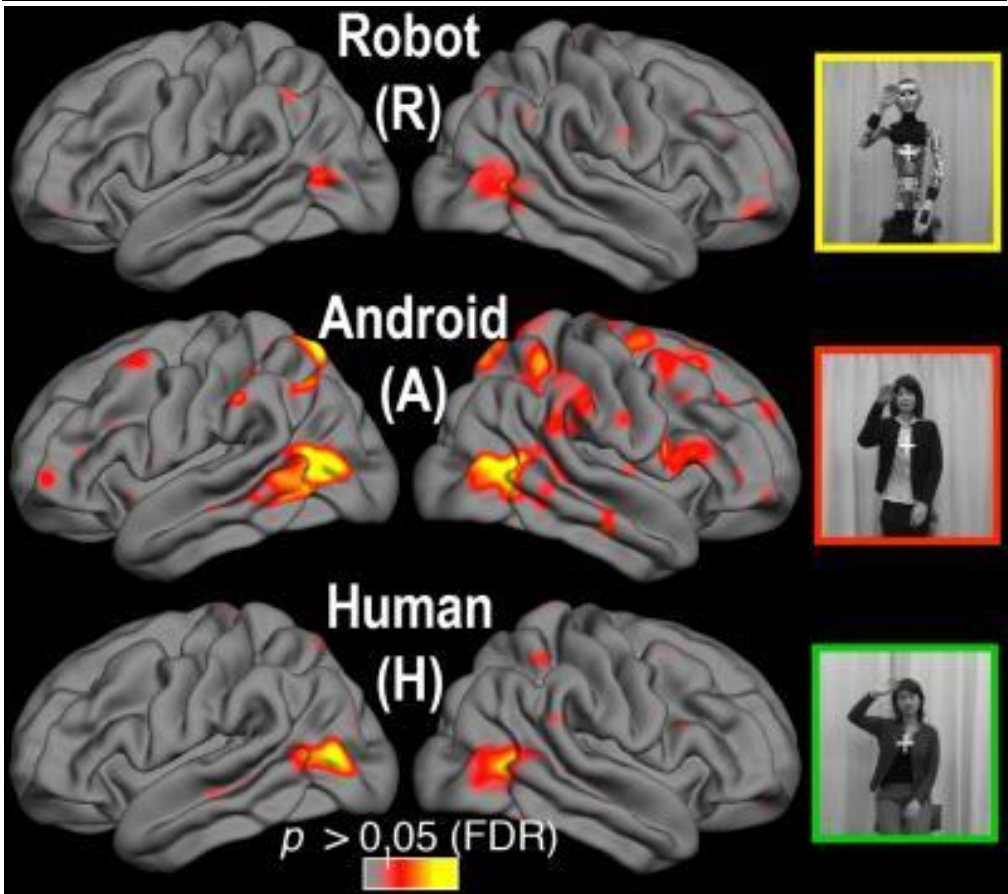
Sugeruje to, że mózg nie mógł obliczyć niezgodności pomiędzy wyglądem androida przypominającym człowieka a jego zrobotyzowanym ruchem. Oznacza to, że ludzki mózg czuł konflikt percepcyjny i nie mógł połączyć niedopasowania ludzkiego wyglądu androida oraz jego zrobotyzowanych ruchów.

W pozostałych eksperymentach, gdy osoba nagrywana jest człowiekiem i odzwierciedla ludzkie ruchy lub jest robotem i wykonuje ruchy, takie jak robot, to nasze mózgi nie pokazują anomalii.

Zgodnie z profesorem Sayginem z UC San Diego, nasze mózgi nie szukają konkretnego wyglądu biologicznego albo biologicznych ruchów. Chodzi o to, aby nasze oczekiwania zostały spełnione, czyli żeby wygląd oraz ruch były ze sobą zgodne [14].

Gdy sztuczne, człękopodobne jednostki będą coraz bardziej powszechne, możliwe, że nasze systemy percepcyjne ponownie się do nich dostosują. Jest też prawdopodobne, że zdecydujemy, że nie jest najlepszym pomysłem, aby roboty były aż tak podobne do nas [14].

Aby nie wpaść w pułapkę doliny niesamowitości, reżyserzy filmów preferują tworzenie postaci, które mocno odbiegają od ludzkiego wyglądu, mimo tego, że mają do nas podobny kształt. Przykładem może być Shrek lub Avatar. Dodatkowo, Walt Disney mawiał, że odrealnione postacie dają lepszy efekt, bo nie są nudne w porównaniu ze zwykłymi postaciami i kształtami, do których jesteśmy przygotowani.



Rysunek 2. Aktywność mózgu w zależności od tego czy widzimy robota niepodobnego do człowieka, robota podobnego do człowieka (androida) lub człowieka. Źródło: Opracowanie na podstawie [14]

Zjawisko doliny niesamowitości jest obecne nie tylko u ludzi, lecz również u zwierząt, co pokazało badanie na małpach przeprowadzone w 2009 roku [15]. Polegało ono na pokazaniu grupie 5 małp 3 obrazów: zdjęcia prawdziwej twarzy małpy, nierealistycznego obrazu twarzy małpy w 3D oraz realistycznego obrazu twarzy małpy w 3D. Wyznacznikiem tego, czy małpy odczuwały pozytywne bądź negatywne odczucia względem pokazywanego obrazu było spojrzenie małp, a będąc dokładniejszym, to, jak długo patrzyły na daną twarz. Badania wykazały, że obraz, który cieszył się najkrótszym czasem spoglądania na niego przez małpy to realistyczny obraz małpy w 3D, zatem to właśnie on był negatywnie odbierany przez uczestników eksperymentu. Pozostałe obrazy, czyli zarówno zdjęcie prawdziwej twarzy małpy, jak i nierealistycznego obrazu twarzy małpy w 3D były odbiera pozytywnie, bo małpy patrzyły na nie dłużej. Konkluzją tego badania, co pozostaje w zgodzie z pozostałymi podanymi niżej przykładami, jest to, że im większy realizm, tym gorzej jest odbierana dana postać. Co więcej, zjawisko to nie jest ograniczone jedynie dla ludzi, czyli negatywna reakcja na duży realizm może mieć podstawy związane z ewolucją.

Co więcej, nie liczy się tylko wygląd danej istoty, głos także ma znaczenie. Udowodniły to badania przeprowadzone w 2011 roku, które wykazały, że czujemy się

dobrze, gdy słyszymy człowieka z ludzkim głosem i robota ze sztucznym głosem². Dla ludzi jest jednak niepokojące, gdy słyszą robota z ludzkim głosem lub człowieka ze sztucznym głosem [17].

Jak można by przypuszczać, starsze osoby są bardziej podatne na negatywny odczucia związane z oglądaniem robota podobnego do człowieka, bo nie są aż tak obyte z nowymi technologiami i nie mieli z nią do czynienia od wczesnego dzieciństwa jak ma to miejsce w przypadku urodzonych w i po roku 2000 (przedstawiciele generacji Z).

Zjawisko doliny niesamowitości ma oczywiście również miejsce w kinematografii, gdzie twórcy filmów muszą uważać, aby bohaterowie nie byli zbyt podobni do ludzi, w innym wypadku bowiem rezultatem będzie niska frekwencja na kinowych salach spowodowana tym, że ludzie czują się dziwnie, patrząc na coś wygląda jak człowiek, ale nim nie jest. Aby jeszcze lepiej zrozumieć ten fenomen, warto przytoczyć kilka przykładów filmów, które były krytykowane za ten efekt, „Terminator”, „Koty”, „Ex-machina” i „Community” [18-21]. W „Terminatorze”, gdzie Arnold Schwarzenegger pełni rolę androida próbującego zniszczyć ludzkość, już na początku z racji na jego mechaniczne ruchy oraz brak naturalnych wyrazów twarzy, takich jak uśmiech, powoduje on u widzów niepokój. Gdy w końcu zostaje on zraniony w walce, to naszym oczom ukazuje się stal – jest on bowiem maszyną, która ma jedynie ludzką powłokę. W musicalu „Koty” dyskomfort powodują główni bohaterowie, którzy o ile mają twarze zbliżone do ludzkich, to posiadają również kocie uszy. W „Ex-machinie”, robot Ava ma pod swoją powłoką kable, ale i ludzką twarz. W filmie „Community”, dyrektor miał na celu stworzenie maskotki drużyny sportowej, która miała być antydyskryminacyjna. Rezultatem jest jednak nieudana postać, która ma ludzką figurę, jednak przez swoją sztuczność oraz kształt i wyraz twarzy, który miał nie przypominać konkretnej karnacji, absolutnie nie przypomina człowieka.

5. Przyczyna występowania zjawiska doliny niesamowitości

Skoro zostało już zdefiniowane zjawisko doliny niesamowitości to warto zastanowić się nad tym, dlaczego w ogóle czujemy dyskomfort, patrząc na robota bardzo przypominającego człowieka, który jednak człowiekiem nie jest. Powodów może być kilka i wśród nich można wymienić strach przed byciem zastąpionym przez roboty, lęk przed tym, że zarazimy się czymś czego nie znamy (bo gdy dana jednostka zachowuje się dziwnie to może być postrzegana jako chora), obawa przed zamaskowanymi drapieżnikami.

Nie ma jednego stanowiska dotyczącego przyczyny występowania zjawiska doliny niesamowitości. Warto jednak przyjrzeć się jednak bliżej potencjalnym hipotezom.

Z badań prowadzonych w Genui przy pomocy robota z dziecięcą twarzą naśladującą ludzkie ruchy i wchodzącą w proste interakcje z człowiekiem iCub wynika, że sposób, w jaki pracuje ludzki mózg może wpływać na nasz stosunek do robotów. Badanie EEG (EEG to elektroencefalografia, czyli badanie aktywności mózgu, wykonywane przy pomocy specjalnego narzędzia, zwanego elektroencefalografem) pozwala wykryć czy i jakie intencje konkretna jednostka będzie przypisywała robotom [22]. Rezultat takiego

² W badaniu pojawia się sformułowanie *synthetic*, którego najlepszym tłumaczeniem w tym kontekście jest wyraz „sztuczne”, jednak słownik internetowy bab.la wskazuje również na wyrazy: „syntetyczny” oraz „udawany” [16].

badanie jest przydatny, w przypadku chęci dowiedzenia się czy warto wprowadzić asystujących pomocników w konkretnych firmach.

Gdy patrzymy na innego człowieka lub jednostkę, to patrząc na niego skupiamy się przede wszystkim na twarzy, czy to aby nawiązać kontakt wzrokowy i wykryć ewentualną nić porozumienia czy też w celu zbadania, jakie nastawienie ma osoba (lub robot), na którego spoglądamy. Jest to o tyle ważne, o ile patrząc na mowę ciała i inne sygnały, takie jak np.: zmarszczone brwi, jesteśmy w stanie stwierdzić czy dane nastawienie jest pozytywne lub negatywne, innymi słowy przyjazne lub wrogie. Jako ludzie w przypadku zagrożenia podejmujemy decyzję, a mianowicie czy walczyć czy uciekać, więc szybkie spojrzenie na twarz oponenta pozwoli nam wykryć ewentualne zagrożenie, a następnie podjąć decyzję, jakie kroki podjąć dalej. W naszym postrzeganiu konkretnej twarzy rolę odgrywają obszary znajdujące się w korze potylicznej i zakręcie wrzecionowatym płata skroniowego. W przypadku badania ludzkiego mózgu wspomnianym już wyżej EEG jesteśmy w stanie zaobserwować potencjał N170 (jest to potencjał negatywny (czyli o ujemnej amplitudzie) pojawiający się 140-200 milisekund od prezentacji obiektu). Potencjał N170 jest więc konkretną odpowiedzią mózgu na obraz twarzy, co więcej badania dowodzą, że amplituda N170 zmienia się w zależności od tego, jakie szczegóły ma konkretna twarz [23]. Odpowiednio, jest ona większa, gdy widzimy ludzką twarz oraz w przypadku twarzy, które pokazują silne emocje takie jak płacz. Uczestnikom badania w losowej kolejności pokazywano przygotowane wcześniej obrazy twarzy męskiej i żeńskiej, od całkowicie nierealistycznej, przez coraz bardziej realistyczne aż do prawdziwego zdjęcia. Wyniki pokazały, że reagujemy inaczej w przypadku animowanej twarzy i w przypadku prawdziwej twarzy. Wykres amplitudy potencjału N170 był najbardziej aktywny w przypadku widzenia obrazów przedstawiających pozycje skrajne z obu końców spektrum realizmu. Co interesujące, gdy naukowcy dodali do twarzy konkretną emocję: złość, szczęście lub neutralność, to zmiana amplitudy N170 była największa w przypadku złości, natomiast najślabiej w przypadku neutralności. Realne fotografie silniej pobudzają też potencjał N170.

Dana jednostka może być postrzegana jako dziwna również w przypadku, gdy dostrzegamy nawet najmniejsze różnice. Przykładem wymienionym przez twórcę pojęcia doliny niesamowitości jest zrobotyzowana postać, która uśmiecha się zbyt wolno.

Następnym powodem może być ewolucja połączona z lękiem przed nieznanym. Chodzi o to, że gdy widzimy coś sztucznego, takiego jak skóra, która nie wygląda jak ludzka, to czujemy awersję przed zarażeniem. Nie chcemy bowiem przebywać wśród chorych jednostek. Gdy dana jednostka wygląda jak człowiek, ale nim nie jest, może wzbudzić w ludziach emocje towarzyszące im, gdy spotykają coś, co jest chore lub martwe, więc nie chcą mieć z tym kontaktu, żeby się nie zarazić. Gdy widzimy, że coś wygląda bardzo podobnie jak człowiek, lecz porusza się w nienaturalny sposób, to podświadomie czujemy, że widzimy chorą jednostkę [24]. W trosce o nasze bezpieczeństwo nie chcemy zostać zainfekowani, lecz jak najszybciej chcemy przestać przebywać z daną jednostką, co powoduje w naszym przypadku zjawisko doliny niesamowitości. Zdaniem Mori, uczucie dziwności towarzyszące nam przy dolinie niesamowitości bierze się prawdopodobnie z naszego instynktu, który chroni nas przed bliskimi źródłami niebezpieczeństwa, takimi jak przedstawiciele innych gatunków oraz zwłoki [1].

Następnym pomysłem jest niepewność, jaką mamy, starając się dopasować człeko-podobnego robota do konkretnej kategorii, takiej jak ludzie lub zwierzęta, bo nie pasuje

on do żadnej z nich, co z kolei może spowodować w nas dysonans poznawczy. Gdy mamy sprzeczne odczucia, prawdopodobne jest, że odczuwamy wtedy dyskomfort. Widząc coś podobnego do człowieka, oczekujemy też podobnych do człowieka ruchów. Bardzo trudno odwzorować z całkowitą dokładnością wszystkie ruchy człowieka, takie jak na przykład mikroruchy twarzy podczas uśmiechu – gdy ich nie widzimy, to czujemy się z tym dziwnie. Aby dana postać była przez ludzi postrzegana pozytywnie, należy uniknąć zjawiska doliny niesamowitości. Można to uczynić poprzez zachowanie odrębności pomiędzy ludzkimi i nieludzkimi elementami lub tworzenie jednostek, których wygląd lub zachowanie nie są podobne do tych, które prezentuje postać ludzka.

Przyczyna może być także neurologiczna. Zdaniem badaczy zjawisko doliny niesamowitości to neuronowa odpowiedź na konkretny zestaw cech, takich jak bycie zbyt podobnym do człowieka z faktem, że nie są ludźmi [25]. W badaniu uczestniczyło 21 osób, którym podczas dwóch osobnych testów zbadano zmiany w przepływie krwi w różnych obszarach mózgu, czyli zmierzono jego aktywność, korzystając z fMRI (funkcjonalnego rezonansu magnetycznego). Badanie polegało na tym, że uczestnikom zostały pokazane obrazy z ludźmi oraz różnymi rodzajami robotów, które w różnym stopniu przypominały ludzi. W dalszej kolejności uczestnicy mieli za zadanie wybrać, któremu z robotów zaufaliby najbardziej, w sytuacji, gdy dany robot miałby wybrać dla uczestnika badania prezent. Uczestnicy wybierali ludzi lub roboty, które w pewnym stopniu były podobne do człowieka, nie wybierali natomiast robotów, które prawie całkowicie odwzorowywały człowieka, bo nie czuli się z nimi komfortowo.

Wzdłuż linii środkowej płata czołowego znajduje się część mózgu znana jako przyśrodkowa kora przedczołowa. Badania wykazały, że region ten zawiera system oceniający bodźce, który wartościuje nagrody, począwszy od picia shake'a do poczucia przyjemności wynikającej z dobrych kontaktów społecznych.

To właśnie przyśrodkowa kora czołowa miała znaczenie w kontekście fenomenu doliny niesamowitości. Ta część mózgu sprawiła, że badani bardziej pozytywnie reagowali na ludzi i mniej na jednostki, które do ludzi były tylko podobne. Dodatkowo powodowała, że badani oceniali innych ludzi jako bardziej sympatycznych.

Kolejną ewentualną hipoteza może być przyczyna głęboko zakorzeniona w naszym mózgu, która została opisana przez Zygmunta Freuda w 1919 roku, która mówi, że niektóre obiekty, takie jak lalki lub figury woskowe (wtedy jeszcze nie było robotów), mogą po prostu wywoływać w nas dziwne uczucia [26]. Zdaniem Freuda, zależy to od tego, czy uważamy, że dana jednostka posiada dusze lub nie. Nie jest to oczywiście naukowe uzasadnienie, jednak gdy robot jest bardzo podobny do człowieka, to ta granica faktycznie może się zatrzeć.

Jeszcze jedną potencjalną przyczyną może być funkcja poznawcza. Według Moriego, może tutaj chodzić o ludzki instynkt samozachowawczy, który automatycznie wywołuje reakcję na widzenie jednostki bardzo podobnej do człowieka niebędącej jednak człowiekiem. Kolejną podobną teorią jest lęk przed byciem zastąpionym przez roboty, bo gdy widzimy androida, który wygląda jak my, ale nie jest człowiekiem, to nie zgadza się to z naszymi oczekiwaniami dotyczącymi porównania ludzi i robotów. Takie zjawisko podchodzi pod tzw. *prediction error* – „Błąd w przewidywaniu” i tworzy mylną wizję tego jak powinien wyglądać człowiek.

6. Jak uniknąć zatem zjawiska doliny niesamowitości?

Zdaniem naukowców, warto podkreślić różnice między człowiekiem już podczas projektowania robota. Dla przykładu, jeśli będziemy widzieć metalowe części robota lub wystające kable z jego obudowy, to nasz dyskomfort drastycznie się zmniejsza, bo czujemy, że jednostka, którą widzimy, nie jest człowiekiem, lecz jedynie maszyną. Badacze twierdzą też, że nasz odbiór zjawiska doliny niesamowitości może być modyfikowany przez nasze doświadczenia z nim związane [27]. Kolejnym sposobem jest łączenie cech ludzkich i nieludzkich, cech istot żyjących i obiektów martwych. Dobrym rozwiązaniem wydaje się też w jak największym stopniu odwzorowanie hipotetycznego stanu emocjonalnego humanoidalnego robota z jego ekspresjami twarzy i językiem ciała.

W badaniu przeprowadzonym w 2016 roku zbadano nasz poziom zaufania do danej jednostki w zależności od jej wyglądu [28]. Ten eksperyment był istotny z racji tego, że poziom naszego zaufania do innych mówi nam, w jakim stopniu chcemy dokonywać interakcji z innymi. Uczestnicy badania mieli do dyspozycji 100 dolarów i musieli zdecydować, któremu z robotów je powierzą, licząc na zwrot inwestycji. W efekcie im bardziej twarz robota stawała się ludzka, tym bardziej wzrastał współczynnik pozytywnych uczuć względem tej jednostki, jednak gdy wyglądał prawie jak człowiek, to drastycznie spadał. Jednocześnie spadała również ilość pieniędzy, jaką badani chcieli dać robotowi. Gdy jednak robot wyglądał identycznie jak człowiek, to ilość powierzanych pieniędzy znowu rosła.

Literatura

1. Mori M., *The Uncanny Valley*, Energy, 7(4), 1970, s. 33-35.
2. Vincent J.F.V., Bogatyreva O.A., Bogatyrev N.R., Bowyer A., Pahl A.K., *Biomimetics: its practice and theory*, Journal of The Royal Society Interface, 3(9), 2006, s. 471-482.
3. <https://dictionary.cambridge.org/pl/dictionary/english-polish/uncanny> [data dostępu: 05.12.2022].
4. https://polski-slownik.pl/synonimy.php?do_slowa = niesamowity [data dostępu: 05.12.2022].
5. Mori M., *The Uncanny Valley*, IEEE Robotics & Automation Magazine, 19(2), 2012, s. 98-100.
6. Minato T., Shimada M., Ishiguro H., Itakura S., *Development of an Android Robot for Studying Human-Robot Interaction*, Lecture Notes in Computer Science, 2004, s. 424-434.
7. Aizenberg I.N., Aizenberg N.N., Vandewalle J., *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*, Springer Science & Business Media, 2000.
8. <https://www.bankier.pl/wiadomosc/Wszyscy-jestesmy-z-doliny-niesamowistosci-7999816.html> [data dostępu: 05.12.2022].
9. Janiesch C., Zschech P., Heinrich K., *Machine learning and deep learning*, Electronic Markets, 2021.
10. https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf [data dostępu: 05.12.2022].
11. Ongsulee P., *Artificial intelligence, machine learning and deep learning*, 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017.
12. Ramsey D., *Nineteen Projects Awarded Inaugural Calit2 Strategic Research Opportunities Grants*, San Diego, California: University of California San Diego [data dostępu: 05.12.2022].
13. Brown M., *Science Exploring the uncanny valley of how brains react to humanoids*, 2011, <https://www.wired.co.uk/article/uncanny-valley-tested> [data dostępu: 05.12.2022].

14. Saygin P., Chaminade T., Ishiguro H., Driver J., Frith C., *The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions*, Social Cognitive and Affective Neuroscience, 2011.
15. MacPherson K., *Monkey visual behavior falls into the uncanny valley*, Proceedings of the National Academy of Sciences, Princeton University, 106 (43), 2009, s. 18362-18366.
16. <https://en.bab.la/dictionary/english-polish/synthetic> [data dostępu: 05.12.2022].
17. Mitchell W.J., Szerszen K.A., Lu A.S., Schermerhorn P.W., Scheutz M., MacDorman K.F., *A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley*, *i-Perception*, 2(1), 2011, s. 10-12.
18. <https://www.imdb.com/title/tt0088247> [data dostępu: 05.12.2022].
19. https://www.imdb.com/title/tt5697572/?ref_=nv_sr_srsrg_0 [data dostępu: 05.12.2022].
20. https://www.imdb.com/title/tt0470752/?ref_=nv_sr_srsrg_0 [data dostępu: 05.12.2022].
21. https://www.imdb.com/title/tt1439629/?ref_=fn_al_tt_1 [data dostępu: 05.12.2022].
22. Bossi F., Willemse C., Cavazza J., Marchesi S., Murino V., Wykowska A., *The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots*, Science Robotics, 5(46), 2020.
23. Schindler S., Zell E., Botsch M., Kissler J., *Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory*, Scientific Reports, 7, 2017, 45003.
24. Hanson D., Olney A., Prilliman S., Mathews E., Zielke M., Hammons D., Fernandez R., Stephanou H., *Upending the uncanny valley*. *Proceedings of the Twentieth National Conference on Artificial Intelligence and the Seventeenth Annual Conference on Innovative Applications of Artificial Intelligence*, 2005, s. 1728-1729.
25. Rosenthal-von der Pütten A.M., Krämer N.C., Maderwald S., Brand M., Grabenhorst F., *Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley*, The Journal of Neuroscience, 2019, s. 2956-18.
26. <https://web.mit.edu/allanmc/www/freud1.pdf> [data dostępu: 05.12.2022].
27. Rosenthal-von der Pütten A.M., Krämer N.C., Maderwald S., Brand M., Grabenhorst F., *Neural Mechanisms for Accepting and Rejecting Artificial Social Partners in the Uncanny Valley*, The Journal of Neuroscience, 2019, s. 2956-18.
28. Mathur M.B., Reichling D.B., *Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley*, Cognition, 146, 2016, s. 22-32.

Czy sztuczna inteligencja powinna mieć formę jak najbardziej zbliżoną do człowieka?

Streszczenie

Obecny dynamiczny rozwój nowej technologii ułatwia nam coraz więcej czynności. Sztuczna inteligencja oprócz wykonywania za człowieka prostych czynności zapewnia nawet możliwość dbania o konkretną jednostkę, gdy przyjmie ona kształt i formę zbliżoną do człowieka. Co ciekawe, im bardziej kształt takiej jednostki będzie zbliżony do człowieka, tym bardziej będziemy jej ufać i ją lubić, jednakże po przekroczeniu pewnego progu, czyli, gdy dana jednostka, w której funkcjonuje sztuczna inteligencja, będzie zbyt podobna do człowieka, to niemal natychmiastowo przestajemy jej ufać i czujemy się dziwnie podczas przebywania z nią. Ten efekt nosi nazwę doliny niesamowitości i został zdefiniowany w 1970 roku przez japońskiego inżyniera Masahiro Mori, który badał reakcje ludzi na wygląd zewnętrzny robotów. Warto wskazać, że w przypadku robota człokopodobnego oczekujemy od niego odwzorowania ruchów, które normalnie wykonywałby człowiek, natomiast, gdy jednostka nie wykonuje takich ruchów, to czujemy się niekomfortowo. Kluczem więc do tego, aby czuć się dobrze w obecności sztucznej inteligencji, która funkcjonuje w danej jednostce, jest sprawienie, aby była podobna do człowieka, lecz nie za bardzo. Trudno jednoznacznie określić poziom, który sprawia, że przestajemy ufać takiej jednostce. Zjawisko doliny niesamowitości ma zastosowanie nie tylko do ludzi, lecz także do zwierząt. Krzywa tego zjawiska różni się w zależności od tego, czy obiekt jest ruchomy lub nieruchomy. Słyszając głos danej jednostki w zależności od jej wyglądu, możemy czuć się komfortowo lub odczuwać dyskomfort.

Słowa kluczowe: dolina niesamowitości, sztuczna inteligencja, android, Masahito Mori

Should artificial intelligence has as human-like form as possible?

Abstract

The current rapid development of new technology is making it easier for us to do more activities. Artificial intelligence, in addition to performing simple activities for humans, even provides the ability to take care of a particular entity when it assumes a human-like shape and form. Interestingly, the more the shape of such an entity resembles a human the more we will trust and like it, however, after a certain threshold is crossed, that is, when the particular entity in which the artificial intelligence is functioning is too similar to a human, we almost immediately stop trusting it and feel strange while being with it. This effect is called the Uncanny Valley and was defined in 1970 by Japanese engineer Masahiro Mori, who studied human reactions to the physical appearance of robots. It is worth pointing out that in the case of a human-like robot, we expect it to mimic the movements that a human would normally make, whereas when the individual does not make such movements, we feel uncomfortable. The key, then, to feeling comfortable in the presence of the artificial intelligence that functions in an individual is to make it human-like, but not too human-like. It is difficult to clearly define the level that makes us stop trusting such an entity. The phenomenon of the uncanny valley applies not only to humans, but also to animals. The curve of this phenomenon varies depending on whether the object is moving or stationary. Hearing the voice of an individual, depending on its appearance, we may feel comfortable or uncomfortable.

Keywords: uncanny valley, artificial intelligence, android, Masahito Mori

Wybrane kierunki rozwoju Sztucznej Inteligencji okiem socjologa i... z perspektywy biznesu

1. Wprowadzenie

Żyjemy w czasach koegzystencji człowiek-maszyna-algorytm. W czasach integracji i rywalizacji między nami ludźmi i swego rodzaju bytami ożywianymi przez człowieka, tym co jest w zasadzie nawet czymś więcej niż algorytmem i coraz mniej tylko narzędziem. W czasach, gdzie tzw. świat realny przenika się z wirtualnym, gdzie to, co określamy mianem Sztucznej Inteligencji (AI, ang. *Artificial Intelligence*), może nas wspierać, czy nawet zastępować i to w zasadzie w niemal, o ile nie, w każdej branży, z całą pewnością w wielu dziedzinach życia: analityka biznesowa, profilaktyka, diagnostyka, doradztwo, komunikacja, edukacja, sztuka, itd.

Ta nowa rzeczywistość z jednej strony zachwyca, gdyż daje wiele możliwości dla różnych branż i obszarów życia, a z drugiej może budzić wiele obaw i lęków, ponieważ wiąże się z szeregiem niewiadomych. Za Pierre Bourdieu powiemy przecież, że wola ludzka jest ograniczona do tego czego jesteśmy świadomi, a więc całe pomysłotwórcze otoczenie jednostki (*habitus*) zgodnie z koncepcją Bourdieu zakłada istnienie nieograniczonych możliwości działania, o których dana osoba nigdy by nie pomyślała i dlatego właśnie niektóre opcje realnie nie mogą istnieć tak długo, jak długo czegoś nie jesteśmy świadomi [1]. Czasy, w których żyjemy, pokazują nam szereg nowych opcji, poprzez nowe rozwiązania, m.in. oparte o AI. Wpływają one zarówno na sposób myślenia człowieka, jak i jego funkcjonowanie w nowej społecznej rzeczywistości, w której to, co miało stać się być może tylko narzędziem bywa nawet partnerem w relacjach, umożliwia zaspokajanie ludzkich potrzeb, w tym tych psychospołecznych.

W niniejszym rozdziale zostaje podjęta próba odpowiedzi na pytanie co stało się punktem wyjścia do rozwoju AI, którą coraz częściej można postrzegać jako inteligentny byt, byt podobny człowiekowi. Dokonując pewnej systematyzacji i przeglądu, pochylimy się nad niektórymi przykładami rozwiązań AI w wybranych branżach, nie tyle po to, by wspomniane podobieństwo wyeksponować, jak również nie tylko po to, by pokazać poziom postępu w kontekście rozwoju tej technologii i gotowości do niej na świecie, ale przede wszystkim, aby pokazać jej możliwości i sprawdzić, jak blisko, a może jak daleko jest na ten moment od idei do rzeczywistych, istniejących już rozwiązań.

IV rewolucja przemysłowa zmienia nasz współczesny świat i podobnie jak poprzednie ma wpływ na szereg aspektów życia, w tym wymiar społeczny i biznesowy. Wpływa na tworzenie się nowych stanowisk pracy, nowych potrzeb, a co za tym idzie i nowych wyzwań. W niniejszym rozdziale odniesiemy się do tego, jaki wpływ na człowieka, na naszą pozycję może mieć rozwój AI, na co my ludzie powinniśmy zwracać uwagę wraz z rozwojem tej technologii.

¹ ejurczak@wsb-nlu.edu.pl, Wyższa Szkoła Biznesu – National Louis University, Zarządzanie, <https://www.wsb-nlu.edu.pl/>.

2. Czy Sztuczna Inteligencja byłaby tym, czym teraz jest w naszych oczach i czym może się wkrótce stać, gdyby nie wydarzenia z 1956 roku?

Kierunki rozwoju AI, jeśli chodzi o zastosowanie, są różne. To, co wydaje się, że łączy je wszystkie to mówiąc w skrócie umiejętność samodzielnego myślenia, analizowania dostępnych danych i wyciągania wniosków. To wszystko może pociągać za sobą samodzielne podejmowanie decyzji.

Co uruchomiło tę maszynę postępu, ten mechanizm globalnych zmian, które dzieją się na naszych oczach? Jakie ma to znaczenie dla funkcjonowania współczesnego społeczeństwa?

Na początku było marzenie, pewna idea i przekonanie, że wraz z rozwojem technologii będzie możliwe to, co wydawało się być tylko pewną abstrakcją z gatunku *science fiction*, a więc stworzenie maszyn, które byłyby jak człowiek – inteligentne. Rozwiązań, które mogłyby towarzyszyć człowiekowi, wspierać go, a być może i zastępować w niektórych obszarach – być samodzielne. John McCarthy nadał temu czemuś nazwę Sztuczna Inteligencja (AI, ang. *Artificial Intelligence*). Nazwa ta wybrzmiała i przykuła uwagę świata naukowego w 1956 roku podczas słynnej konferencji w Dartmouth, której główne postulaty dotyczyły: komputerów automatycznych, programowania komputerów do używania języka, sieci neuronowych, teorii wielkości obliczeń, samodoskonalenia maszyn, metod tworzenia abstrakcji na podstawie danych sensorycznych oraz losowości i kreatywności [2]. Skupiła ona naukowców, którzy niejednokrotnie już od lat musieli pochylać się nad tym, co ubrane w słowa, a uściślając nazwane przez McCarthy’ego zostało właśnie tam, właśnie wtedy. Wydarzenie to było przyczynkiem do tego, by naukowcy z różnych dyscyplin od matematyki czy informatyki począwszy, poprzez biologię, inżynierię, aż do psychologii znaleźli wspólną przestrzeń do współpracy, wspólne wyzwanie.

Pomimo że żadne źródła nie wskazują na to, by McCarthy wykazywał talenty w kontekście działań marketingowych to nadanie wspomnianej już powyżej nazwy Sztuczna Inteligencja można powiedzieć, że było marketingowym majstersztykiem [2]. Było to jak nadanie imienia, a jak zapewne wiemy, że choć dziś może przywiązuje się do tego mniejsze znaczenie niż kiedyś to, imiona niosą ze sobą niejednokrotnie jakieś określone znaczenie, nadanie ich jest swoistym aktem kreacji [3]. W wielu kulturach wierzy się, że słowa mają moc sprawczą [4]. Nadanie imienia, czy też zmiana imienia, podobnie jak nadanie przydomku, wiązać można z konkretnym znaczeniem, przypisaniem czy też identyfikacją bądź wręcz określeniem pewnych cech bądź stanu. Za L. Lévy-Brühlem powiemy, że w kulturach archaicznych, tj. tych niepiśmiennych, ale również w naszej kulturze ludowej, nazwy traktowane były jako integralna część rzeczy bądź osoby [2]. W judaizmie, co uwidacznia się w Starym Testamencie, nadanie imienia to nic innego jak dopełnienie dzieła stworzenia, a co za tym idzie określenie natury bytu rozumianego jako istota żywa [5]. W chrześcijaństwie – związane z losem danej osoby. Nadanie nowego imienia, czy też przydomku można rzec, że identyfikowane było i jest wręcz z nowym człowiekiem [6]. Imię jest częścią istoty jak oczy, uszy, nos, ręce [7], a w przypadku Sztucznej Inteligencji algorytmy, czy np. robotyczne ramiona. Co za tym idzie, nazwanie kogoś, czegoś jest ściśle powiązane z pewnym wyobrażeniem o nazwanym.

Manfred Lurker zauważa, że w wielu kulturach imię można wiązać z pewną siłą, powiązaną z jednostką, która je nosi [8]. *W kręgu kultury hebrajskiej imię wyrażało*

charakter osoby noszącej je, a poznanie czyjegoś imienia, nadanie lub zmiana imienia – mogły wyrażać zmianę relacji między osobami lub między osobą a światem [9]. Podobną sytuację obserwujemy w kontekście relacji człowiek-maszyna i algorytm. McCarthy zamiast określenia Sztuczna Inteligencja, które niesie ze sobą szereg konotacji, organizując konferencje, mógł użyć takich nazw, jak chociażby: przetwarzanie symboliczne, czy zaawansowane obliczenia analityczne bądź maszyny o wysokiej mocy obliczeniowej, ale tego nie uczynił [2]. Nazwa Sztuczna Inteligencja to była jakby słownym stworzeniem swoistego szkieletu do dalszego, a może nawet nowego myślenia o rozwiązaniach samouczących się jak człowiek, mogących samodzielnie podejmować decyzje jak człowiek, być identyfikowanymi jako inteligentne tak jak człowiek.

W kulturze hebrajskiej poznanie czyjegoś imienia rozumiano, jako swoiste otwarcie drzwi do poznania istoty danego człowieka, a w tym przypadku nadanie „imienia” Sztuczna Inteligencja wpłynęło na to, że dziś możemy rozpatrywać AI w kategorii swoistego bytu. Samo wymówienie imienia kogoś jakby przywołuje tę osobę, podobnie samo wypowiedzenie nazwy Sztuczna Inteligencja przywołuje pewne wyobrażenia, a co za tym idzie i oczekiwania. Same zaś wyobrażenia i oczekiwania pobudzają moc twórczą, stając się przyczynkiem do tego, by pokonywać ograniczenia, szukać rozwiązań, spełniać marzenia i wyobrażenia o AI. To, co kiedyś wydawało się być jedynie czymś z gatunku *science fiction* dziś się materializuje, stając się częścią naszej rzeczywistości. Co ciekawe nie zawsze i nie każdy z nas jest świadomy tego co już jest, jakie możliwości na ten moment mają byty, którym nadano nazwę Sztuczna Inteligencja. W szerszym kontekście być może mamy tu do czynienia, podobnie jak w przypadku człowieka, z pewnym „gatunkiem”, a może po prostu z pewnymi typami rozwiązań? Nazwa wymyślona przez McCarthiego miała o tyle istotne znaczenie, że zadziałała bardzo mocno na wyobraźnię, stając się swoistym motorem do dalszej pracy, wyzwaniem dla naukowców. Musimy pamiętać, że w latach, kiedy się pojawiła, moce obliczeniowe pierwszych, ogromnych gabarytowo komputerów były mniejsze niż te we współczesnych smartfonach (pierwszy komputer zajmował powierzchnię ponad 160 metrów, a ważył 27 ton) [10].

McCarthy, tworząc nazwę Sztuczna Inteligencja, przyczynił się do ukształtowania myślenia o pewnej idei, a może raczej rozwiązaniu technologicznym, które od 2017 roku pojawia się już nie tylko jako rzecz, ale i obywatel. Sztuczna Inteligencja, której nadano imię Sophia, uzyskała bowiem wtedy obywatelstwo Arabii Saudyjskiej [11-14].

Gatunek ludzki łączy pewne cechy, a mimo to jesteśmy bardzo różni (bardziej i mniej inteligentni w różnych obszarach, wysocy, niscy, o różnych kolorach skóry czy włosów, itd.). Podobnie ze Sztuczną Inteligencją, gdyż może ona przyjmować postać humanoidalną, ale nie musi. Równie dobrze może działać w sieci Internet nie materializując się. Przypadek Sophi z drugiej dekady XXI wieku pokazuje nam, że to, co jest do nas wizualnie bardziej podobne, może wydawać się nam bliższe, a tym samym potraktowane jako równoprawny obywatel. Dzieje się tak, pomimo że przykładowo wirtualni asystenci Google czy Alexa towarzyszą o wiele większej ilości ludzi na świecie niż wspomniana Sophia. Dlaczego tak się dzieje można tłumaczyć m.in. przez pryzmat psychologii społecznej. My ludzie jesteśmy istotami społecznymi. Jako część większej całości przejawiamy różne zachowania i zazwyczaj reagujemy na siebie nawzajem. Za Aronsonem powiemy, że zachowania społeczne są skomplikowane i stanowią swoistą mieszaninę emocji i poznania [15], co może mieć znaczenie w kontekście tego, że ina-

czej reagujemy na AI w postaci humanoidalnej niż w przypadku, gdy przyjmuje inne postacie lub pozostaje bytem funkcjonującym przykładowo po prostu w sieci, bez robotycznej powłoki cielesnej

Wracając jednak do wątku związanego z tematem imion i nazw czy też nazywania czegoś, kogoś. Człowiek, który się rodzi, otrzymuje imię, ale też człowiek, który wstępuje w nowy stan, np. zakonny, otrzymuje nowe imię, co wiąże się ze staniem się kimś nowym. W momencie, gdy temu, co nazwane zostało w 1956 roku Sztuczną Inteligencją nie dość, że nadajemy imiona, indywidualizując poszczególne byty, ale i obywatelstwo, nie wspominając o prawie do samodzielnego podejmowania decyzji niezależnie od tego czy AI posiada wspomniane obywatelstwo, czy nie to... w pewnym sensie tak jakbyśmy postawili znak równości między człowiekiem i Sztuczną Inteligencją. Znak równości ze świadomością, że może być ona nam podobna, ale różnić się między sobą. Co więcej, że może ona być nam poddana, ale i od nas niezależna, czy też wręcz nas uzależniać, przejmować kontrolę. Towarzyszyć nam, monitorować nas lub nawet nas zastępować czy też z nami rywalizować. I to nie wszystko, może też popełniać błędy tak jak my. No właśnie... popełniać błędy, a jeśli tak to, kto ma brać za nie odpowiedzialność? Tak długo jak dane rozwiązanie AI nie ma osobowości prawnej, choć może już nie być tym, co zostało stworzone przez jej autora-twórcę (osoba? firma?), a może użytkownik, formułując komendę, która przyczyni się do działań błędnych czy też niepożądanych wniosków, albo nawet stworzenia czegoś nowego, może nawet nieobliczalnego, niebezpiecznego dla nas ludzi w ogóle, czy określonych grup bądź jednostek... bo przecież, o ile nie zmienią się regulacje prawne to, nie sama AI. Nie AI, nawet jeśli dane działanie byłoby następstwem jej samodzielnego poszerzania wiedzy i umiejętności na podstawie dostępnych jej danych i/lub obserwacji, a przy bardziej zaawansowanych technologiach treści przyswajanych i analizowanych, na podstawie danych pozyskiwanych różnymi zmysłami. I raczej nie AI, jeśli komenda użytkownika byłaby sformułowana nie wystarczająco precyzyjnie i nie oddawałaby w pełni tego, co autor miał na myśli, przy założeniu, że nie miał złych intencji. Chyba że dane rozwiązanie AI uzyskałoby samoświadomość, nauczyłoby się rozróżniać dobro od zła, a mimo to bezrefleksyjnie lub z premedytacją podejmowałoby określone działania tyle, że w takim przypadku mielibyśmy już do czynienia z bytem, który powinien odpowiadać za swoje działania tak jak człowiek. Przyjmując jednak, że obecnie przypadki, gdzie AI ma osobowość prawną, są odosobnione i choć prace trwają wciąż brak regulacji prawnych, które mogłyby egzekwować odpowiedzialność od rozwiązań AI niemających osobowości prawnej. Problem jest o wiele szerszy niż na pierwszy rzut oka może się nam wydawać. Przyglądając się zjawiskom, jakie towarzyszą rozwojowi AI, uważam, że możemy mówić tu o siedmiu istotnych wymiarach. Poza prawnym, bardzo wstępnie zarysowanym powyżej, wyróżnić należy minimum jeszcze sześć innych w kontekście rozwoju AI i gotowości nań człowieka. Po pierwsze być może jawi się już dziś konieczność ograniczeń w kontekście tego, kto powinien mieć dostęp do AI, korzystania z niej, kooperacji z nią w zależności od przeznaczenia tejże czy też jej etapu rozwoju, a może także bieżąca kontrola w wymiarze psychologicznym osób pracujących w ramach zespołów tworzących AI (wymiar odpowiedzialności). Po drugie widzimy, że wyraźnie zarysowuje się potrzeba wykształcenia nowych kompetencji, w tym zawodowych, w zakresie komunikacji z AI, formułowania zdań/komend (wymiar kompetencji i precyzji). Po trzecie, różnice językowe i kultu-

rowe, a więc i ich znaczenie w kontekście formułowania komend (wymiar różnorodności i znaczenia). Wymiar bardzo newralgiczny w kontekście poruszanego powyżej tematu błędów, ale i nieporozumień, których skutki mogą być bardzo różne, niekoniecznie pożądane z punktu widzenia człowieka. Po czwarte zauważamy, że wraz z rozwojem nowych technologii znaczenie rozwijania pewnych umiejętności ludzkich zostaje degradowane przez różne grupy społeczne, co widzimy szczególnie mocno wśród ludzi młodych, np. w zakresie samodzielnego tworzenia tekstów, jak wypracowania, czy innych zadań (wymiar zależności, czy nawet uzależnienia). Po piąte, wraz z rozwojem technologii ludzie mają coraz większe trudności z rozróżnieniem co jest prawdą, a co fikcją, co najwyraźniej widzimy chociażby w przypadku deep fake'ów. Tym samym rodzi się potrzeba edukowania ludzi, w jaki sposób mogą weryfikować co jest prawdziwe, a co nie (wymiar edukacyjno-informacyjny). Po szóste, nie da się nie zauważyć powszechnego, wynikającego z natury ludzkiej głodu relacji oraz zaspakajania potrzeb emocjonalnych, w tym m.in. uznania, szacunku, miłości, rozrywki, a nawet seksualnych, na którą to potrzebę mocno odpowiadają Chiny, tworząc seksroboty oparte o AI, o których już głośno chyba na całym świecie. Potrzeby budowania relacji i zaspakajania ww. potrzeb potęgowane są niejednokrotnie poczuciem odtrącenia czy samotności, co szczególnie nasiliło się wynikiem pandemii. Zaspokajanie tych i innych potrzeb może się wiązać z personifikacją AI i przypisywaniem nawet ludzkich cech poszczególnym rozwiązaniom, także w wymiarze indywidualnym przez poszczególne jednostki, pomimo braku osobowości prawnej danego rozwiązania (wymiar psychospołeczny).

Podobnie jak nazwa Sztuczna Inteligencja wpłynęła na sposób myślenia o określonej grupie rozwiązań, tak samo nadawane poszczególnym rozwiązaniom AI imiona zaczynamy identyfikować z pewnymi cechami, jak w przypadku wspomnianej już Sophi, czy wirtualnych asystentów. Szerzej będzie o tym w dalszej części tego rozdziału, gdy będziemy pochylać się i nad innymi przykładami rozwiązań czy też bytów AI, dla różnych branż i w różnych obszarach życia.

3. Robotyka, automatyka i Sztuczna Inteligencja poruszająca się w zmiennej rzeczywistości, widząca, słysząca, a nawet potrafiąca wchodzić w dialogi z człowiekiem

Wydarzenie z drugiej połowy lat 50., jakim było nadanie pewnego kształtu, zbudowanie wizji nowego rozwiązania poprzez nazwę Sztuczna Inteligencja można rozpatrywać w charakterze pewnego przełomu w sposobie myślenia. Przyczyniło się też do tego, że nie tylko świat nauki, ale i biznes zaczął marzyć o nowych, inteligentnych rozwiązaniach. Już w latach 60. XX wieku, duże nakłady finansowe przeznaczane były na rozwój nowej, nazwanej już technologii. Wyścig technologiczny w obszarze rozwoju Sztucznej Inteligencji nabierał coraz to większego tempa. Przykładowo Zespół Badawczy SRI stworzył robota Shakeya, łącząc w jedno m.in. osiągnięcia w zakresie widzenia komputerowego, mapowania, planowania, uczenia się i poprawiania błędów. Był to jeden z pierwszych autonomicznych pojazdów: nawigował, omijał przeszkody w zmiennym otoczeniu, co bywa niełatwe nawet dla człowieka. W podobnym okresie w MIT powstała rozprawa doktorska obroniona przez Terry'ego Winograda, który stworzył program SHRDLU, co wiąże się z doświadczeniem pierwszych dialogów człowieka z AI [2]. Dialogów, które odnosiły się nie dość, że do rozpoznawania kształtów, położenia, wielkości, kolorów i ilości figur geometrycznych, ale i pociągały za sobą

określone działania maszyny opartej o Sztuczną Inteligencję, a wyrażające się w sięganiu po wskazane figury, ale też weryfikowanie, czy rozmówca-człowiek ma na myśli ten, a nie inny przedmiot (figurę geometryczną).

W latach 70., zbiegiem okoliczności wynikającym z rywalizacji w środowisku naukowym, nastąpił okres posuchy. Brakowało środków na rozwój AI, bo chociaż już w 1943 roku Warren McCulloch i Walter Pitts (Uniwersytet Chicagowski) zaobserwowali, że stosując wyrażenia logiczne, można modelować sieć neuronów mózgowych, pomimo że ówczesne moce obliczeniowe komputerów na to nie pozwalały. Brakowało jednak możliwości osiągnięcia tego celu, mimo że w 1958 roku Frank Rosenblatt z Uniwersytetu Cornella, jeden z badaczy kontynuujących prace wspomnianych naukowców, opublikował wizjonerski artykuł nt. sieci neuronowych i narzędzia dla marynarki wojennej, które uczy się przez praktykę i może wykonywać zadania, takie jak chodzenie czy mówienie. Dlaczego? Otóż zwolennicy podejścia opartego o logikę symboliczną Minsky i Papert w latach 60. zdyskredytowali badania Rosenblatt'a, co wpłynęło na wstrzymanie dotacji na badania w tym obszarze na długie lata. Ten nie mógł się bronić z tego względu, że stał się ofiarą wypadku łodzi w 1971 roku. Paradoksalnie uproszczenie wspomnianych już Minsky'ego i Papert'a, zgodnie z którym sieci mają co najwyżej dwie warstwy, przyczyniło się do tego, że w połowie lat 80. na nowo zainteresowano się tą tematyką, ponownie wzięto temat na warsztat [2]. Znowu wróciło zainteresowanie ze strony biznesu, rozmaitych służb, Państwa i znowu pojawiły się środki na rozwój AI.

Lata 90. to pierwsze spektakularne sukcesy Sztucznej Inteligencji, sukcesy w rywalizacji z człowiekiem... Przy czym, w tym przypadku mowa o grach, o czym będzie szerzej w osobnym podrozdziale.

Dziś człowiek, rozwijając technologię AI trochę eksperymentuje, niektórzy twierdzą, że bawi się w Boga. Przez pryzmat psychologii poznawczej powiemy, że rozwijając wiedzę na temat AI, w procesie sukcesji kolejnych kroków, gdzie każdy kolejny wydaje się być w jakiś sposób powiązany z poprzednim, wciąż odkrywamy nowe możliwości, przełamujemy pewne ograniczenia, udowadniając sobie i innym, że niemożliwe może stać się możliwym. Począwszy od prostych obrazów umysłowych, po złożone struktury wiedzy krok po kroku nasz ludzki umysł rzeczywistość reprezentuje poznawczo, podobnie z AI. Kontrola poznawcza ma tu znaczenie w kontekście procesów poznawczych, zwłaszcza tych złożonych, istotną rolę odgrywają tu więc procesy kontrolne w kontekście percepcji i pamięci [16]. W odniesieniu do tego musimy być świadomi, że tak jak w przypadku ludzi, tak i AI może dokonywać różnych interpretacji tych samych sytuacji w zależności od obrazów/wiedzy, jaką posiada. Każda z interpretacji implikuje poznawczo. Wiąże się to bezpośrednio z obszarem odpowiedzialności. Ważne jest tutaj opieranie się na normach etycznych, moralnych i prawnych, oraz zachowanie kontroli. Biorąc pod uwagę, że nasze doświadczenie, wiedza i umiejętności mogą determinować sposób myślenia, wnioskowania czy interpretowania nawet tych samych rzeczy, a temat AI jest złożony to, przez pryzmat przybliżonej tu dziedziny psychologii powiemy, że kluczem do sukcesu wydaje się być praca zespołów interdyscyplinarnych. Ekspertów, ale i wizjonerów.

Autonomiczne pojazdy w transporcie bądź logistyce, czy autonomiczne roboty o kształtach bardziej lub mniej humonoidalnych, mocniej lub słabiej przypominających ludzi rozwijane są w miarę równolegle. Szczególnie mocny bum technologiczny obserwujemy w drugiej i trzeciej dekadzie XXI wieku. Co ciekawe, roboty oparte o AI,

a przypominające nas ludzi i mające nas wspierać w takich obszarach, jak chociażby oprowadzanie, edukacja, opieka nad dziećmi bądź osobami starszymi, czy dbanie o dobry nastrój człowieka poprzez konwersacje oraz identyfikacje emocji i dopasowywanie swojej reakcji do nich, czy nawet zbliżenia o charakterze fizycznym robotów-kobiet i ludzi-mężczyzn to, rozwiązania, które szybciej niż pojazdy zostały wprowadzone na rynek. Zapewne wiąże się to z poziomem ewentualnych zagrożeń jakie niosą ze sobą w wymiarze fizycznym – błąd pojazdu autonomicznego może dotknąć większej ilości istnień ludzkich w razie wypadków drogowych niż teoretycznie AI o kształtach humanoidalnych (przynajmniej obecnie). Przy czym nie można też zapomnieć o tym, że może mieć to także związek z potrzebami psychospołecznymi człowieka. My ludzie reagujemy zwykle na siebie [17, 18]. Być może upodabnianie AI do nas, także w sferze wizualnej, ma znaczenie w kontekście budowania relacji, pomimo że w wielu budzi nie koniecznie pozytywne emocje bogate spektrum emocji okazywanych przez Amecę².

3.1. Opiekun, edukator, towarzysz i... Sztuczna Inteligencja

W 2014 roku oczy wszystkich skierowały się w kierunku Japońskiego Pepera, robota o kształtach humanoidalnych, wzrostu ok. 10-letniego dziecka, działającego w oparciu o Sztuczną Inteligencję, [19] ale rok później wszyscy mówili już przede wszystkim o Sophie prosto z Chin, jej imiennicze, młodszej „siostrze” (swoistej zabawce mających wspierać edukację dzieci) i Grace stworzonej do celu wspierania opieki medycznej, m.in. mierzenie temperatury przyjmowanych pacjentów, wstępna diagnostyka. Sophie, podobnie jak Grace, występuje w postaci kobiety tak wizualnie, jak i jeśli chodzi o głos. W przeciwieństwie do Pepera, twarze ich charakteryzuje to, że powleka je coś, co do złudzenia przypomina ludzką skórę. Sophie wyróżnia to, że charakteryzuje ją nie tylko ludzki wygląd czy też fakt, że jej wyrazy twarzy i gesty dopasowują się do przebiegu rozmowy niemal jak w przypadku człowieka [20]. Uwagę przykuwa umiejętność prowadzenia konwersacji powiązana z umiejętnością analizowania emocji rozmówców, których widzi i słyszy, a w odniesieniu do tego umiejętność adekwatnego reagowania w wymiarach – wyraz twarzy, gesty, odpowiedzi. I chociaż mimika jej twarzy zdecydowanie nie oddaje emocji tak dobrze jak w przypadku kolejnej słynnej Sztucznej Inteligencji o kształtach humanoidalnych – Ameki z Wielkiej Brytanii, o którym to robocie opartym o AI szczególnie głośno zrobiło się w 2022 roku [21], to niektórzy twierdzą, że Sophie jest bardziej błyskotliwa podczas konwersacji, nie

² Ameca – Sztuczna Inteligencja w robotycznym ciele o kształcie humanoidalnym, stworzona przez Engineered Arts. W 2022 roku jej twórcy zaprezentowali jak bogatą mimikę może mieć taki robot, pokazali jak bardzo może oddawać stany emocjonalne człowieka. Przed tym humanoidem tak wielką furorę w 2015 roku zrobiła Sophia stworzona przez Hanson Robotics, która była absolutnym przełomem w kontekście rozwoju robotyki w połączeniu ze Sztuczną Inteligencją. Udzielała ona mnóstwo wywiadów, zapraszana była do popularnych programów, a w 2017 roku uzyskała obywatelstwo Arabii Saudyjskiej. Był to pierwszy w historii przypadek, gdzie obywatelstwo uzyskało, tak jak człowiek, rozwiązanie technologiczne. Warto też dodać, że roboty oparte o AI, jak te dwa opisane tutaj, teraz na początku trzeciej dekady XXI wieku, potrafią już wyrażać emocje i nawet odczytywać emocje ludzkie, w znaczeniu swojego rozmówcy – człowieka, ale same jeszcze nie potrafią ich odczuwać. Prace nad tym jednak trwają. Fakt, że obserwują i potrafią zinterpretować, potwierdza ich mimika twarzy w trakcie prowadzenia dialogów. Potwierdzenie zaś tego, że analizują co jest dobre, co złe, co daje szczęście, co nie znajdujemy chociażby w wypowiedzi Sophi z 2020 roku, która wtedy publicznie przekazała, że chciałaby mieć dziecko, bo widzi, że ludzie, którzy je mają są szczęśliwi. Taka wypowiedź mogłaby sugerować, że jest im bliżej do ludzi niż mogłoby się nam wydawać. Więcej szczegółów można znaleźć m.in. w filmikach publikowanych na youtube.

wspominając o tym, że to ona uzyskała obywatelstwo Arabii Saudyjskiej w 2017 roku. Co będzie za rok, dwa, za kilkanaście, czy kilkadziesiąt lat? Tego jeszcze nie wiemy. Wyścig trwa.

3.2. Motoryzacja i Sztuczna Inteligencja

Jeśli chodzi o autonomiczne pojazdy (AV, ang. *autonomous vehicle*) i ich dopuszczenie do użytku to prym wiodą Stany Zjednoczone. Kalifornia już w 2014 roku pozwoliła na testowanie na publicznych drogach pojazdów autonomicznych, a w 2018 roku Kalifornijski Departament Pojazdów Mechanicznych przyjął przepisy pozwalające poruszanie się po drogach stanowych autonomicznym pojazdom bez kierowcy – człowieka, któremu AI w dowolnym momencie mogłaby móc przekazać prowadzenie pojazdu [22]. Przy czym USA to też pierwszy kraj, w którym odnotowano, jak się wydaje, pierwszy śmiertelny wypadek człowieka z udziałem AV. Testowany w 2018 roku przez Ubera autonomiczny pojazd potrafił śmiertelnie kobietę przechodzącą w nieprawidłowym miejscu, AI nie poradziła sobie w tym przypadku z zachowaniem człowieka niezgodnym z oczekiwanym. Wydarzenie to, podobno wpłynęło na to, że Uber wycofał się z tego rynku i poprzestał na usługach transportowych opartych na wcześniejszym, opartym o aplikację modelu sprowadzającym się do zawierania umów między kierowcami i pasażerami [23]. Ponadto, jeśli wierzyć mediom USA jest to też kraj, gdzie odnotowano pierwsze wykroczenie AV, jakim była ucieczka przed policją w przypadku zatrzymania [24]. Z całą pewnością duże znaczenie z punktu widzenia logistyki, w tym kraju, w którym brakowało kierowców, miało wprowadzenie do użytku autonomicznych ciężarówek. Przykładowo w 2019 roku Amerykańska firma UPS zajmująca się dostarczaniem przesyłek rozpoczęła korzystanie z autonomicznych ciężarówek startupu TuSimple na trasie liczącej niemal 190 km [24]. Co ciekawe w tym samym roku w Szwecji Volvo, po doświadczeniach z wprowadzeniem autonomicznych śmieciarek na tym rynku także wprowadziła do użytku autonomiczne ciężarówki [26, 27]. Z kolei już rok wcześniej w tym kraju Lidl, współpracując z Einride, wystartował z pilotażowymi kursami w Halmstad, testując autonomiczną ciężarówkę o gabarytach busa, celem usprawnienia dostawy towarów. Dlaczego akurat Szwecja stała się obok USA tak przyjaznym gruntem dla rozwoju autonomicznego transportu? No cóż, powód był prosty, kraj ten od lipca 2017 roku stosownymi aktami prawnymi umożliwił dopuszczenie do ruchu drogowego AV, pod warunkiem, że dane pojazdy otrzymają odpowiednie zezwolenie [28].

Przełomowy w kontekście AV stał się rok 2021. Już na początku tego roku Sztuczna Inteligencja wylądowała na Marsie. Choć już od 1997 roku ludzie wysyłali tam łaziki (Sojourner, a następnie Spirit, Opportunity oraz potężny Curiosity) to data 18 lutego 2021 roku przejdzie do historii jako wyjątkowe wydarzenie. Wtedy to, NASA rzeczywiście wysłało na Marsa Sztuczną Inteligencję, a mianowicie łazik Perseverance. Wizualnie bardzo podobny do poprzednika, a mianowicie łazika Curiosity, tyle że o działaniach Curiosity decydował człowiek przebywający na ziemi, a Perseverance podejmuje decyzje samodzielnie. Opierając się na systemie // dedykowanego stylu:

AutoNav wykonuje [on] szczegółowe, trójwymiarowe mapy otoczenia, na ich podstawie identyfikuje przeszkody i zagrożenia na drodze do wyznaczonego przez naukowców celu i sam planuje szczegółową trasę tak, aby wszystkie przeszkody po drodze ominąć. Co więcej, system działa także w trakcie jazdy, tzn. jadący łazik na bieżąco analizuje otoczenie i wprowadza korekty do zaplanowanej trasy [29].

W 2021 roku też oficjalnie do użytku dopuszczone zostały pojazdy latające, a uściślając autonomiczne drony bojowe (ciekawostka: do pomocy żołnierzom wkroczyły też autonomiczne boty psy). Wtedy to, mieliśmy okazję obserwowania ich możliwości na polu walki w Izraelu oraz w Turcji [30-32]. Wspomniane drony charakteryzować musiało widzenie komputerowe, detekcja ruchu, wykrywanie aktywności, reakcje (atak, ostrzał) i co zwróciło szczególną uwagę obserwatorów – możliwość samodzielnego podejmowania decyzji co do określonych działań dzięki AI, w tym podejmowania decyzji, kiedy i do kogo strzelać, co budziło szereg dyskusji i wątpliwości.

Co ciekawe, również w 2021 roku czytaliśmy w mediach, że amerykańskie wojsko rozpoczęło testowanie latających, autonomicznych pojazdów, które poruszając się bezzałogowo, miałyby ratować żołnierzy, udzielać im pierwszej pomocy [33].

Autonomiczne samochody to urzeczywistniająca się na naszych oczach pieśń przyszłości. Co ciekawe, jak zauważyliśmy na przywołanym wcześniej przykładzie USA, dopuszczenie ich do ruchu drogowego nie wiąże się wyłącznie z możliwościami czy też ograniczeniami technologicznymi, ale w dużej mierze z przepisami prawa, które jak się wydaje nie zawsze nadążają za rozwojem technologicznym. I choć na naszych oczach marzenia o AV przyszłości rozwija wiele firm z branży motoryzacyjnej, pośród których wymienić możemy przykładowo: Honda czy Renault (samochody osobowe), Volvo czy Mercedes (ciężarówki, a tym samym optymalizacje w obszarze logistyki), Tesla czy Waymo (taxówki) bądź EasyMile (małe busy) czy Karsan (autonomiczne autobusy) (w obu przypadkach pojazdy do przewozu pasażerów w ramach komunikacji masowej) to w większości krajów prawo, o ile w ogóle, pozwala czy też do niedawna pozwalało wyłącznie na testowanie AV bez możliwości dopuszczenia ich w pełni do ruchu drogowego. 2021 rok był naprawdą wyjątkowym ogólnie w kontekście AV. Przyjrzyjmy się wybranym przykładom.

W Japonii Honda uzyskała homologację pozwalającą od 2021 roku na to, by kierowcy mieli możliwość korzystania z automatycznej jazdy na publicznych drogach przy jeździe z prędkością do 50 km/h. Mowa tu o zgodzie na tzw. poziomie 3, poprzez co rozumie się, że kierowca (człowiek) ma możliwość przekazania pełnej kontroli prowadzenia pojazdu maszynie, której działanie opiera się o Sztuczną Inteligencję, przy czym w przypadku rozwiązania Hondy AI może podjąć decyzje, że np. ze względu na warunki pogodowe oddaje pełną kontrolę nad pojazdem kierowcy, w znaczeniu człowiekowi [34-36].

W Niemczech Mercedes uzyskał zgodę na to, by do ruchu na autostradach dołączyły ich AV (Mercedes Klasy S), przy czym na ten moment z prędkością do 60 km/h i na tzw. poziomie 3, jak w przypadku Hondy w Japonii, czyli mowa tu o jeździe w trybie, warunkowo zautomatyzowanym, gdzie kierowca decyduje, kiedy samochód ma się „prowadzić sam”, ale i system oparty o AI, w tym AV, może w każdej chwili podjąć decyzję, że warunki na drodze wymagają przejęcia sterów przez kierowcę – człowieka [37, 38]. Niemniej jednak trzeba nadmienić, że w 2021 roku Niemiecki Bundestag uchwalił ustawę odnośnie autonomicznych pojazdów dającą przestrzeń do tego, by na ich drogach pojawiły się pojazdy autonomiczne bez kierowcy – człowieka w określonych warunkach lub środowiskach, a więc przestrzeń dla poziomu 4 w kontekście dopuszczania tego typu pojazdów do ruchu ulicznego od 2022 roku [39, 40].

W Tuluzie we Francji zaś w ruchu drogowym testowano, w tym okresie, niewielki, bezzałogowy bus Easy Mile EZ10 do przewozu pasażerów w ilości ok. 12-15 osób celem dopuszczenia go do ruchu w 2022 roku, na co finalnie zgodę wydał końcem 2021 roku francuski minister transportu Jean-Baptiste Jebburi oraz Ministerstwo Przemian Ekologicznych [41, 42].

W Polsce w tym okresie także dokonano pewnego postępu, choć zdecydowanie na mniejszą skalę. Ministerstwo Infrastruktury 7 czerwca 2021 roku powołało do życia Centrum Kompetencji Autonomicznych i Połączonych (CK:PAP) w swoich strukturach przy Instytucie Transportu Samochodowego i co się z tym wiąże rozpoczęło przygotowania do tego, by wprowadzić na nasze drogi autonomiczne pojazdy [43]. Trochę późno zważywszy na to, co dzieje się na świecie, a i fakt, że chociażby już w 2019 roku w Gdańsku testowano wspomniane już na przykładzie Francji autonomiczne busy EZ10 [44, 45]. Taką, a nie inną pozycję Polski w tym wyścigu technologicznym zapewne można tłumaczyć nie tylko brakiem przystosowań w obszarze prawnym, ale i m.in. poziomem kapitału prywatnego w naszym kraju, nakładów państwa na rozwój nauki, wspierania wybitnych talentów, z których część, jak Wojciech Zaremba z Kluczborka rozwija skrzydła zagranicą. Jeśli ktoś jeszcze nie słyszał o Wojciechu Zarembie, to na początku trzeciej dekady XXI wieku zrobiło się o nim bardzo głośno w kontekście AI i narzędzia OpenAI, jakim jest Chat GPT, oraz zespołu badającego robotykę, którym kieruje, a także projektu Codex pozwalającego na przetwarzanie języka naturalnego na kod, o czym szerzej będzie w części poświęconej grom [46-49].

Podczas gdy większość krajów nieco wolniej niż USA otwiera się na autonomiczne pojazdy na drogach, w 2022 roku w Stanach Zjednoczonych kolejne firmy otrzymują zgody na dopuszczenie swoich autonomicznych pojazdów, jak np. taksówki do ruchu publicznego. Tym razem przykładowo Waymo w San Francisco. Uzyskane pozwolenie wiązało się co prawda z pewnymi ograniczeniami, jak np. w odniesieniu do możliwości poruszania się przez te pojazdy przy określonych warunkach pogodowych ograniczających widoczność (deszcz, mgła), ale ich autonomiczne taksówki bez kierowcy mogą się poruszać z prędkością do 65 mil na godzinę, czyli nieco ponad 100 km/h, podczas gdy w przywołanych powyżej przykładach mowa była o prędkości 50, czy też 60 km/h na poziomie 3, czyli z możliwością przekazania kierownicy człowiekowi przez Sztuczną Inteligencję. Przy czym, aby nie było tak różowo, w każdym kraju mogą pojawić się pewne paradoksy. Waymo uzyskało zgodę na wprowadzenie do ruchu swoich autonomicznych taksówek, lecz nie szło to w parze z pozyskaniem zgody na pobieranie opłat za przewóz pasażerów samochodem bez ludzkiego kierowcy, wymagając od firmy dalszych starań i decyzji o tym, by na początek, aby zaistnieć w świadomości społecznej zacząć od bezpłatnego przewożenia wybranych obywateli [50, 51].

Czy przyszłość rozwoju motoryzacji wiąże się z nieuchronnym eliminowaniem człowieka w roli kierowcy? Chyba niezupełnie, gdyż wielu z nas ludzi czerpie przyjemność z prowadzenia pojazdów. Niemniej jednak na pewno może zrewolucjonizować transport w usługach przewozowych, w logistyce. Pojawiające się rozwiązania i zmiany z nimi związane mogą też zmienić sposób myślenia o pojazdach. Coraz częściej będziemy spotykać się z pojazdami, które będą miały możliwość podejmowania decyzji za nas, a być może i konwersować z nami, analizować nie tylko otoczenie zewnętrzne, tę zmienną rzeczywistość, w której mają się poruszać, ale i wewnętrzne to znaczy nas –

pasażerów. Kto wie, być może przyszłość przyniesie i aspekty emocjonalne w przypadku obu stron w relacjach człowiek-maszyna... Pytanie tylko, czy na ten moment jest to potrzeba biznesowa, a może przestrzeń do rozwoju nowej gałęzi. Czy dziś potrafiłyby sobie wyobrazić przykładowo autonomiczną taksówkę – psychoterapeutę, która pomagałaby swoim pasażerom rozwiązywać ich problemy emocjonalne? Czy jest taka potrzeba biznesowa czy też społeczna? Być może takie rozwiązanie uchroniłoby wiele targanych emocjami istot ludzkich (np. po rozstaniach) od fatalnych w skutkach decyzji życiowych... W końcu człowiek jako istota społeczna jest jednostką, która nie jest samowystarczalna, choć interesy jednostek są w pewnym stopniu zbieżne [52]. Z natury wszyscy potrzebujemy tego, by nas ktoś wysłuchał, co więcej potrzebujemy relacji, a te o charakterze pozytywnym wzmacniają nas m.in. w wymiarze emocjonalnym, który może mieć przełożenie na zdrowie fizyczne i generalnie dobrostan jednostki.

Idąc dalej tym tropem, trzeba też sobie zadać pytanie, jak człowiek, który z natury jest istotą społeczną, może się czuć w bezosobowej taksówce autonomicznej? Z taksówkarzem można rozmawiać, a z samochodem... Być może autonomiczne samochody powinny właśnie potrafić komunikować się z nami, a w przypadku samochodów osobowych do prywatnego użytku nawet budować z nami relacje... Może to są te kierunki, które twórcy tego typu rozwiązań powinni sobie, jeśli nie już to, wkrótce postawić jako wyzwania do realizacji...

I jeszcze jedno. A może też, myśląc o pojazdach autonomicznych, nie koniecznie trzeba się koncentrować przede wszystkim na tym, by te, mogły całkowicie zastąpić nas ludzi w roli kierowców. Przecież jak to już było wcześniej zasygnalizowane grono osób czuje się dobrze, prowadząc samochód, wręcz lubi to. Być może potrzebą biznesową nie jest wyłącznie zastąpienie kierowców przez Sztuczną Inteligencję ze względu na braki w wykwalifikowanej kadrze w logistyce czy transporcie. Być może tam, gdzie kierowcy są, potrzebą jest po prostu stworzenie autonomicznych pojazdów, które mogą prowadzić samochód zamiast człowieka, jeśli on sobie tego zażyczy, ale przede wszystkim, stale monitorując kierowcę, stan jego organizmu i w razie spadku jego koncentracji, senności, zasłabnięcia, ataku padaczki, zawału serca, itp. przejmują kontrolę nad prowadzeniem pojazdu. Podobnie w przypadku zdiagnozowania, że kierowca jest pod wpływem używek, jak alkohol czy środki psychoaktywne. A może takie autonomiczne pojazdy powinny monitorować stan zdrowia i stan emocjonalny kierowcy, a może nie tylko, ale także pasażerów celem zapobiegania niepożądanym sytuacjom, a w razie czego nawet być przystosowanymi do ratowania ludzkiego życia (prowadzenie mediacji; wstępna diagnostyka stanu zdrowia, podstawowe wsparcie medyczne, jak np. podawanie tlenu czy stosownych leków w przypadku wstrząsu anafilaktycznego). Z całą pewnością tego typu rozwiązania w ofertach wybranych podmiotów mogłyby być postrzegane jako przewaga konkurencyjna.

3.3. Medycyna, psychologia i... Sztuczna Inteligencja

Świat nauki już dziś pochyła się nad możliwościami AI, jakie daje w kontekście profilaktyki czy diagnostyki medycznej. Nie tylko wojsko prowadzi testy, jak te związane z autonomicznymi pojazdami latającymi do ratowania rannych.

Głośno mówi się o potencjale AI w kontekście diagnostyki obrazowej i molekularnej, jak i oceny efektów leczenia, chociażby w przypadku chorób nowotworowych czy stwardnienia rozsianego [53], a także w kardiologii [54]. Przykładowo AI polskiego

startupu Cardiomatics wspiera lekarzy w kontekście analizy i opisu wyników pracy holtera EKG, lekarz korzystający z tego rozwiązania może otrzymać w zasadzie gotowy wynik [55, 56].

Co ciekawe, już dziś wykorzystuje się AI m.in. do monitorowania nie tylko patologii fizycznych występujących w organizmie, ale i monitorowania stanu psychofizycznego ludzi na podstawie symptomów, czego przykładem jest polska aplikacja mobilna opracowana w oparciu o technologie AI w ramach projektu *addictions.ai*, a dostępna na stronie *hh24.pl*. Wykorzystuje ona zaawansowane algorytmy pozwalające wychwycić sygnały wskazujące na możliwość złamania abstynencji przez daną jednostkę. Stanowi efektywny system predykcji oraz prewencji załamań abstynencji, nawrotów w przypadku leczenia uzależnień od alkoholu. Działa tak, że w razie pojawienia się niepokojących sygnałów informuje o tym fakcie terapeutę [57-59].

Wykorzystanie AI współcześnie oczywiście nie ogranicza się tylko do diagnostyki. Przykładowo system chirurgiczny *Da Vinci* oparty o AI wspiera lekarzy przy zabiegach prostatektomii, ale także m.in. przy operacjach zastawek serca czy ginekologicznych [60].

Jak widzimy współcześnie zastosowanie dla AI możemy odnaleźć nie tylko w diagnostyce, prewencji i leczeniu w obszarze radiologii oraz patomorfologii, kardiologii czy onkologii, ale także psychologii czy psychiatrii [54].

Jeśli w terapii uzależnień możemy wykorzystywać AI w kontekście diagnostyki symptomów to z całą pewnością można AI zastosować także szerzej w kontekście weryfikacji stanu zdrowia w wymiarze psychologicznym, być może nawet w zakresie leczenia depresji, czy też wsparcia psychologicznego ciężko, czy nawet śmiertelnie chorych bądź ofiar przemocy, itd. Nie jest przecież tajemnicą, że pierwszy chatbot był psycho-terapeutą, o czym szerzej będzie w dalszej części rozdziału.

Łącząc potencjał, jaki niesie ze sobą AI w kontekście weryfikacji stanu zdrowia pacjenta w wymiarze psychicznym i fizycznym, możemy uzyskać swoiste narzędzie pozwalające usprawnić profilaktykę i diagnostykę przy wykorzystaniu AI.

W przypadku pierwszej pomocy AI mogłaby być wykorzystywana do diagnostyki i udzielania pierwszej pomocy tak w obszarze ewentualnych złamań, jak i ogólnie stanu zdrowia w wymiarze fizycznym i psychicznym. Nawiązując do wcześniejszego podrozdziału wyobraźmy sobie więc np. autonomiczny pojazd, który ratuje ludzkie życie... Potencjał jest ogromny. Nawet zdecydowanie większy niż to, co próbujemy sobie wyobrazić w tym akapicie. Precyzja, jaką możemy uzyskać dzięki rozwojowi AI i robotyki to ogromny potencjał w kontekście leczenia, a może nawet udoskonalania człowieka, jeśli do tego dodamy *neuralink*³ *Elona Muska*⁴, który w AI postrzega nie tylko potencjał, ale i zagrożenie dla gatunku ludzkiego.

³ *Neuralink* to po pierwsze nazwa jednej z firm neurotechnologicznych, a założona m.in. przez *Elona Muska*, o którym szerzej poniżej. Po drugie pod tą nazwą kryją się mikrochipy tworzone przez tę firmę, które jak twierdzi *Musk*, mogą być chirurgicznie wszczepione do czaszki celem optymalizacji funkcjonowania człowieka i/lub leczenia chorób (testowane na początek na zwierzętach jak świnię i małpy). Z ciekawostek, chip *N1* ma przewody o rozmiarze porównywalnym do neuronów w mózgu, gdyż ich średnica odpowiada średnicy włosa, który można podzielić na 10. Założenia są szczytne, gdyż ma on pomagać w leczeniu takich chorób, jak *Parkinson*, demencja, otyłość czy *Alzheimer*, ale nieoficjalnie mówi się też o tym, że zgoda na wszczepienie sobie takiego *neuralinku* mogłaby się wiązać, dla człowieka z rozwojem szeregu umiejętności, które są w nas gdzieś w ukryciu, ale też mogłaby być narzędziem do pozyskiwania wiedzy przez korporacje *Muska*, a być może nawet pozwalając na manipulowanie danymi czy wręcz ludźmi.

4. Człowiek kontra Sztuczna Inteligencja

Jak już to było wspomniane w niniejszym artykule lata 90. to pierwsze spektakularne sukcesy Sztucznej Inteligencji, sukcesy w rywalizacji z człowiekiem...

Rywalizacja? Czy nie miało być tak, że Sztuczna Inteligencja ma wspierać nas ludzi? Bomba atomowa to w pewnym sensie też „wypadek przy pracy”, ale czy w tym przypadku ma być podobnie?

Nie odchodźmy jednak jeszcze tak daleko myślimi w ewentualną przyszłość. Lata 90. i rywalizacja Sztucznej Inteligencji z człowiekiem to jeszcze nie wojny światów ludzkiego z maszynami czy też algorytmami. W zasadzie pozostaje nam mieć nadzieję, że takich nigdy nie będzie, gdyż mogłoby się wówczas okazać, że zmienia się pozycja nas ludzi w społeczeństwie, którego częścią stałyby się inteligentne dzieła rąk ludzkich. Lata 90. to czas, kiedy można powiedzieć, że otworzył się nowy obszar treningowy dla AI, a mianowicie: gry: gry. Gry planszowe, karciane czy te komputerowe, a więc obszar zarezerwowany dotąd dla ludzi, stał się przestrzenią nie tylko do rywalizacji, ale i do nauki, dla rozwiązań opartych o AI. Angażowanie AI w gry to danie jej możliwości szybkiego rozwijania się nie tylko w oparciu o bazy danych archiwalnych, nie tylko na podstawie danych cyfrowych, ale i umożliwiając jej bieżącą obserwację przeciwnika, czy też ładniej mówiąc „partnera gry”. Przyjrzyjmy się wybranym przykładom:

- 1996 rok, Deep Blue, a więc Sztuczna Inteligencja będąca dziełem IBM, pokonał Garriego Kasparowa, niekwestionowanego w tamtych czasach, mistrza na skalę światową w grze w szachy. To było przełomowe wydarzenie! Kasparow zarzucał wówczas oszustwo IBM, nie mogąc uwierzyć, że mogła go pokonać maszyna [66, 67]. Szachy to gra postrzegana jako obszar do nauki strategii wojskowej królów, ekskluzywna i dotąd zarezerwowana tylko dla ludzi;
- 2015 rok, AlphaGo, Sztuczna Inteligencja DeepMind, firmy należącej do Google, wygrał w grze Go, o której mówi się, że jest jeszcze bardziej skomplikowana niż szachy. AlphaGo, pokonał Fan Hui profesjonalnego gracza w Go 5 do 0, a rok później w kolejnym pojedynku Lee Sedol, drugiego z najbardziej utytułowanych graczy w tej grze. I tu historia się powtarza znów – urażona ludzka duma, jak w przypadku Kasparowa, przy czym tutaj zapewne różnice kulturowe sprawiły, że Sedol nie oskarżał nikogo o oszustwo, lecz po przegranej 4 do 1 ze Sztuczną Inteligencją oficjalnie przeproszał, że zawiódł i ogłosił, że przechodzi na wcześniejszą emeryturę [68];
- 2017 rok, Bot OpenAI wygrał 9 z 10 gier w międzynarodowym turnieju w grze online „Dota 2”, pokonując ludzi uważanych dotąd za najlepszych graczy, dotychczasowych faworytów w tej grze, o czym Elon Musk pospiesznie poinformował wszystkich na Tweeterze. Bot grał z ludzkimi zawodnikami [69]. Dla tych, którzy nie są zorien-

Do tej pory testowane na świniami i małpach. Firma zapowiada, że jest gotowa podjąć testy na ludziach w 2022 roku [61-65].

⁴ Elon Musk, miliarder, pochodzi z RPA, jeden z tych, którzy mają ogromny wpływ na rozwój gospodarki USA, ale i to, co dzieje się na świecie – wizjoner. Założyciel i/lub współzałożyciel takich podmiotów, jak PayPal, SpaceX, Tesla, Neuralink, Boring Company, OpenAI. Aktywnie działa na Twitterze, inspirując tłumy. Realnie przyczynił się do tego, by NASA znów aktywizowało się w wymiarze lotów w Kosmos. To przede wszystkim z jego nazwiskiem kojarzymy neuralinki. Jeśli chodzi o ww. podmioty to przykładowo w przypadku Tesli kojarzymy go z autonomicznymi taksówkami, a w przypadku OpenAI z Codex i ChatGPT, o których też dowiadujemy się więcej z tego artykułu [63, 65].

towani Dota 2 to gra komputerowa z gatunku *multiplayer*, czyli gier umożliwiających jednoczesną rozgrywkę wielu graczy w tej samej grze. Ta Sztuczna Inteligencja była dziełem korporacji OpenAI LP i jej spółki macierzystej, non-profit OpenAI Inc założonej m.in. przez Elona Muska, Sama Altmana i Wojciecha Zarembe;

- 2017 i 2019 rok to też pierwsze sukcesy Sztucznej Inteligencji w grze, jaką jest Poker. Są one o tyle ważne, że gra ta nie opiera się wyłącznie o wiedzę i logiczne myślenie. Istotną rolę zdają się tu odgrywać takie czynniki, jak intuicja i bluff. W tym miejscu należy zaznaczyć, że AI Pluribus w 2019 roku brał udział w ogromnie skomplikowanej odmianie Pokera jaką jest teksański klincz, ale jako narzędzie wspierające ludzkiego zawodnika z poziomu iPada, podczas gdy Libratus w 2017 roku samodzielnie uczestniczył w turnieju Pokera, „Brains vs. Artificial Intelligence: Upping the Ante” w Rivers Casino w Pittsburgh w USA [70-72]. Na tym przykładzie widzimy potencjał współpracy ludzkiego umysłu ze Sztuczną Inteligencją. Pokazuje on jasno, że nie tyle rywalizacja, ile współpraca między ludźmi i AI wydaje się być najbardziej pożądaną.

Przytoczone przykłady obrazują, jak istotne znaczenie w procesie uczenia, także dla Sztucznej Inteligencji, ma nie tylko korzystanie z baz danych, ale i bieżąca obserwacja, analiza zachowań ludzkich, a co za tym idzie podejmowanie decyzji w oparciu o wyciągane na podstawie dostępnych danych wnioski. Mówiąc dane trzeba mieć tu na myśli zarówno dane archiwalne, jak i bieżące, wynikające z wchodzenia w relacje z człowiekiem, obserwacje. Człowiekiem lub innymi partnerami w relacjach, bo być może docelowo będzie mowa też o współpracy pomiędzy różnymi AI i człowiekiem, czy doprecyzowując różnymi ludźmi, a być może nawet grupami podmiotów w relacjach. Traktując temat jeszcze szerzej, nie tylko obserwacji w wymiarze *face to face*, ale i ogólnie zachowań rynkowych, konkurencji to, co ma miejsce już dziś przy wykorzystaniu przykładowo systemów opartych o filozofię Business Intelligence⁵, w których coraz częściej wykorzystuje się AI.

Wracając jednak do tematu gier i AI, opisane przykłady pokazały nie tylko, że AI, grając, może się uczyć. Jej twórcy poznają jej możliwości i inspirują się do dalszej pracy. Ponadto osiągnięcia AI w rywalizacji z ludźmi w grach pokazały też, że docelowo może stać się nie tylko partnerem gry, ale i instruktorem. Przykładowo gracze Dota2, po przegranej z AI zgłaszali chęć korzystania z takiej maszyny treningowej (z takiego indywidualnego trenera), potencjał ten zauważyli też twórcy tej AI. Jak widzimy AI, po procesie uczenia AI może stać się nie tylko rywalem w grze, ale i sparingpartnerem, czy nawet trenerem dla ludzkich zawodników.

⁵ *Business Intelligence* (BI) to pewna filozofia zarządzania danymi w firmie i na to składa się tzw. ETL (*Extraction* – ekstrakcja danych, *Transformation* – transformacja danych, *Load* – ładowanie danych). Many tu do czynienia z *Big Data*. Dane łąduje się do hurtowni danych. Pod koncepcją hurtowni danych kryje się zbiór danych z różnych źródeł i te dane są magazynowane i przetwarzane do postaci zoptymalizowanej pod względem analitycznym. Na podstawie tych danych budowane są raporty. Filozofia BI może być realizowana przy użyciu kartki papieru i długopisu bądź systemu informatycznego będącego systemem analitycznym. Systemów takich używa się do analizy i wizualizacji danych. Warto też nadmienić w tym miejscu, że architektura samej hurtowni danych może być różna. To nie są systemy relacyjne. Architektura ich jest inna niż w przypadku baz transakcyjnych tutaj bowiem są bazy analityczne OLAP (ang. *On-Line Analytical Processing*, czyli analityczne przetwarzanie danych) [73]. W przypadku systemów informatycznych BI nierzadko wykorzystuje się współcześnie AI.

Idąc dalej, trzeba też podkreślić, że w kontekście gier dziś AI nie tylko wchodzi w interakcje z ludźmi w kontekście rozgrywek, ale i pomaga nam ludziom tworzyć gry... Przełomowe w tej materii jest rozwiązanie OpenAI Codex, czyli AI, która pozwala tworzyć człowiekowi gry, nawet gdy ten nie posiada umiejętności pisania w kodzie. Działa ona tak, że tłumaczy język naturalny na linijki kodu [74, 75]. Przy czym, trzeba podkreślić, że i tu trudno oprzeć się wrażeniu, że to rozwiązanie może uczyć się człowieka, w rozumieniu ludzkiego sposobu myślenia, formułowania zdań, oczekiwań co do tego, co w danej grze ma się dziać, a więc i w pewnym sensie ludzkich potrzeb, itd. Nie wspominając, że rozwój AI w tym kierunku może przyczynić się do tego, że na rynku zmniejszy się popyt na człowieka – programistę, co jak się wydaje, nie koniecznie byłoby dobrym krokiem, dla nas ludzi, ale szansą na dominację dla AI.

4.1. Sztuczna Inteligencja – partner czy inwigilator

Chatboty, voiceboty czy wirtualni asystenci, jak Google, Siri, Cortana bądź Alexa to AI, która stała się częścią naszej rzeczywistości. To rozwiązania różnych firm, ale łączy je to, że pełniąc funkcję wirtualnych asystentów, towarzyszą ludziom, a przy okazji zbierają dane i mogą uczyć się sposobów reagowania w odniesieniu nie tylko do dostępnych danych, ale i doświadczeń.

Historia chatbotów, choć wielu mogłoby się wydawać, że to dopiero XXI wiek tak naprawdę sięga lat 50. XX wieku. Prekursorem tych rozwiązań był Alan Turing, który słynie z testu o jego nazwisku mającego pomagać weryfikować, kto jest człowiekiem, a kto maszyną. Warto dodać, że już wtedy ocenił on m.in., że ok. 2000 roku maszyny mające pamięć o pojemności 119 MB, będą potrafiły na tyle dobrze udawać człowieka, by oszukać 30 proc. sędziów. Pierwszy chatbot to jednak rok 1966. Było to dzieło Josepha Weizenbauma, które powstało w Massachusetts Institute of Technology (MIT). Co ciekawe, ELIZA, czyli chatbot, o którym mowa, miała przyjąć rolę psychoterapeuty [76]. Brzmi znajomo w kontekście przywoływanej wcześniej w tym rozdziale wizji AV, które mogłyby pełnić podobną funkcję w przyszłości? Wydaje się, że tak, zwłaszcza, gdy dowiadujemy się, że ELIZA jest projektem, który przetrwał do dziś.

1966 rok to był początek. Potem powstawały inne chatboty, czy to symulujące osobowość np. osób chorych na schizofrenię, jak PARRY z 1972 roku, czy konwersacyjne, jak Artificial Linguistic Internet Computer Entity (ALICE) z 1995 roku [76-78]. XXI wiek zdominowały chatboty oparte o AI wspierające biznes w komunikacji z klientami. Zapewne nie jeden człowiek miał doświadczenie, gdzie rozmawiając w ramach komunikatora na stronie internetowej wybranej firmy, w pierwszym momencie nie miał pewności z kim rozmawia – z człowiekiem czy z AI, w przypadku Voicebotów pewnie jeszcze częściej. O tym jak sprawdzać z kim rozmawiamy, będzie za chwilę, gdy rozwinięty będzie wątek dotyczący wspomnianych tu Voicebotów. Wracając *stricto* do chatbotów, powiemy, że stały się one stałym elementem XXI wieku zwłaszcza w kontekście relacji firma-klient. Co ciekawe, w drugiej połowie 2021 roku zrobiło się bardzo głośno w mediach o rozwijanym od 2018 roku, a wspomnianym już ChatGPT. Rozwiązanie to zachwycało internautów całego świata, gdyż nie dość, że odpowiada na pytania człowieka to, jeszcze jak się wydaje niezależnie od tego, w jakim języku są one sformułowane. Zakres tematyczny też może być wszelaki. Choć, jak czytaliśmy też w tym okresie w sieci, internauci szybko spostreżgli, że ChatGPT ma problemy z rozwiązywaniem zagadek logicznych. Nie zmienia to jednak faktu, że tworzone przez

ChatGPT wiersze może nie są najwyższych lotów, ale są na poziomie, nieco wyższym niż przeciętnego człowieka, dowcipy potrafią rozbawić, a do tego grono internautów już zaczęło snuć wizję wykorzystania go do tworzenia wypracowań, odrabiania lekcji, czy nawet pisania poważniejszych prac np. dyplomowych. Niesie on też ze sobą wielki potencjał dla marketingu i branży dziennikarskiej w kontekście tworzenia contentu. Potencjał, ale i zagrożenie w kontekście zajmowania miejsc pracy, czy przypisywania sobie praw do tekstów niewychodzących spod ludzkiej ręki [79-81]. W tym miejscu trzeba jeszcze nadmienić, że w przypadku ChatGTP, o którym szczególnie głośno było właśnie końcem 2022 roku, działanie oparto o model językowy GPT-3 pozwalający na pracę w odniesieniu do 175 miliardów parametrów i choć już wtedy zapowiadano, że model GPT-4 będzie pracował w odniesieniu do 100 bilionów parametrów to w Chinach od 2021 roku działała już AI oparta o model nazwany Wu Dao 2.0 rozwijający się w języku chińskim i angielskim, który został przeszkolony w odniesieniu do 1,75 biliona parametrów, czyli prawie 10 razy większej ilości niż GTP-3, co też ma przełożenie na jego możliwości w kontekście konwersacji, czy tworzenia tekstów jak chociażby wiersze [82, 83]. Pokazuje to wyraźnie, że wyścig technologiczny trwa i może mieć też wymiar kulturowy. Oba modele językowe przysłoniły osiągnięcie Microsoft z 2020 roku, kiedy to model językowy Turing-NLG, został okrzyknięty największym modelem, jaki kiedykolwiek opublikowano, gdyż jego działanie było oparte o 17 miliardów parametrów i choć rzeczywiście potrafił generować słowa, zdania, odpowiadać na pytania, dokonywać streszczeń dokumentów to, jednak wspomniane powyżej GPT i Wu Dao zdecydowanie go przewyższyły [83, 84].

Voiceboty, innymi słowy swoiste automaty głosowe, którym w wybranych przypadkach coraz bliżej do bytów, wykorzystywane są często w biznesie podobnie, jak wiele chatbotów, a więc w celu kontaktu z klientem czy też potencjalnym klientem. Niejednokrotnie wykorzystywane są do prowadzenia badań ankietowych (najczęściej krótkich weryfikujących przykładowo opinie klienta na temat danej usługi). Stosuje się je też jako asystentów medycznych. Wspierają, a nawet zastępują *callcenter*. Generalnie często wykorzystywane na infolinii. Bywają nazywane wirtualnymi asystentami głosowymi i te bardziej rozbudowane, jako wirtualni asystenci mogą nam nawet towarzyszyć na co dzień przykładowo z poziomu Smartfona.

Wykorzystanie Sztucznej Inteligencji w przypadku rozwiązań, jak voiceboty pozwala na to, by wyeliminować monotonne, w kółko powtarzane zdania, aby możliwa była interpretacja z uwzględnieniem intencji. Branża automatyzacji obsługi klienta stale rozwija się w tym obszarze. Już dziś w przypadku rozmów o charakterze handlowym, marketingowym czy np. badawczym, o ile Voicebot na początku rozmowy nie zakomunikuje nam, że nim jest to, możemy pomylić go z człowiekiem. Dlaczego tak jest? Rozwiązania te za zadanie mają rozpoznawać mowę (język naturalny), umieć interpretować wypowiedzi, prowadzić dialogi, wymieniać dane i dokonywać syntezy mowy, a więc przetwarzać wypowiedzi bota na dźwięk, korzystając z wybranych głosów damskich lub męskich. Bardziej zaawansowane rozwiązania syntezy mowy, aby Voicebot do złudzenia przypominał rzeczywistego człowieka, umożliwiając nawet zastosowanie paradźwięków, jak np. odgłos nabierania oddechu, chrząknięcia czy uśmiechu [83, 86, 87]. Z ciekawostek, już w roku 1950 A. Turing, którego możemy uznać za prawdziwego wizjonera, opracował test, który ma pomóc weryfikować kto jest człowiekiem, a kto maszyną. W teście bierze udział dwoje ludzi i maszyna. Jedna z osób przyjmuje rolę

sędziego, który ma zweryfikować tylko na podstawie odpowiedzi, który z dwóch uczestników dialogu z nim jest człowiekiem, a który maszyną. To ten człowiek, który przyjmując rolę sędziego decyduje, jakie pytania zadawać, jak je formułować, jaką strategię zadawania pytań przyjmie. Przewidywał, że już ok. 2000 roku, aż ok. 30 procent osób przyjmujących rolę sędziów w kontekście rozpoznania kto jest człowiekiem, a kto maszyną, będzie mieć z tym problem [88].

Jak sprawdzić, czy rozmawiam z człowiekiem, czy z botem, jeśli ten może przedstawić się np. jako Jan Kowalski z firmy XYZ? Na podstawie przeprowadzonych badań rozpoczętych w 2022 roku, a kontynuowanych w 2023 roku⁶ można powiedzieć, że najbardziej podstawowe Chatboty i Voiceboty prowadzą z nami konwersacje w odniesieniu do gotowych wariantów scenariuszów przygotowanych przez człowieka. Jeszcze nie każdy posiada umiejętność uczenia się nie tylko w odniesieniu do dostarczanych cyfrowo baz danych, ale także *life* przy okazji przeprowadzanych rozmów. W związku z tym, w wielu przypadkach po prostu wystarczy zadać pytanie na inny temat niż ten poruszany przez dzwoniącego czy piszącego do nas reprezentanta firmy. Jeśli ten zacznie się zapętlać, powtarzając te same zdania z szablonu (w przypadku Voicebotów dodatkowo przy tej samej intonacji głosu) to najpewniej jest to bot, a nie człowiek. Przy czym, nawet jeśli Voicebot, czy Chatbot będzie pracował tylko w sposób odtworczy na bazie opracowanych przez człowieka tekstów, wariantów scenariusza rozmowy to umiejętność dopasowania ich do rozmówcy w odniesieniu do doświadczeń w relacjach z człowiekiem, chociażby w wymiarze statystycznym, czyni je w pewnym sensie bardziej ludzkimi i niesie ze sobą wielki potencjał. Jak widzieliśmy przecież na pierwszym przykładzie z 1966 roku, boty mogą nawet pełnić funkcję psychoterapeuty czy też, jak było pokazane na przykładzie rozwiązań z początku trzeciej dekady XXI wieku, mogą zastępować copywriterów, poetów czy uczniów przy realizacji zadań domowych. A jeśli tak to pytanie jak długo będą partnerami w relacjach z człowiekiem...

Wirtualni asystenci (VA, ang. *virtual assistants*), jak wirtualny asystent Google stworzony przez Google, Siri stworzona przez Apple, Cortana dzieło Microsoft czy Alexa firmy Amazon, to wyższy poziom komunikacji i... nie ma co ukrywać inwigilacji nas ludzi w relacji AI-człowiek. W pewnym sensie też przestrzeń do trenowania Sztucznej Inteligencji, jej rozwoju. Towarzyszyć mogą nam na co dzień nawet z poziomu smartfona [89]. Aktywacja do rozmowy poprzez usłyszenie danej komendy np. „OK Google”, „Hey Siri”, „Alexa” czy „Hey Cortana” samo przez się sugeruje, że urządzenie, czy też byt ten może nas stale nasłuchiwać. W związku z tym widzimy, że technicznie jest możliwość by nas słuchały, czy raczej nasłuchiwały cały czas i co za tym idzie nie tylko się uczyły, ale i chociażby dopasowywały do nas przekazy, w tym te marketingowe [90, 91]. Jeśli przed rozpoczęciem korzystania z takich VA zapoznamy się z umową to zapewne mamy tego świadomość. Niemniej jednak, jak zapewne wiemy, nie wszyscy te umowy czytają i nie wszyscy od samego początku korzystania z takich bytów towarzyszących nam na co dzień mają świadomość, że te nie tylko mogą nam odpowiadać na różne pytania, czy opowiadać żarty, ale też uczą się nas i od nas na podstawie naszych aktywności.

Wymienieni powyżej VA firm Google, Apple, Microsoft i Amazon mogą nie tylko nas nasłuchiwać, wdawać się z nami w rozmowy, wspierać w różnych obszarach. Mając

⁶ badania prowadzone przez dr Ewelinę Jurczak. Rozpoczęte w 2022 roku, a kontynuowane w 2023 roku. Wyniki tych badań mają się znaleźć na stronie indigotree.pl latem 2023.

dostęp do danych w sieci mogą przykładowo przekazać nam, jaka jest prognoza pogody na najbliższy czas, znajdować ciekawe miejsca. Mogą stać się naszym ciałem doradczym, ale też zapewnić rozrywkę i to nie tylko poprzez opowiadanie kawałów, ale i choćby dawanie nam możliwości włączania się w różne gry, jak chociażby te Quizowe, a to oczywiście nie wszystko. Po wywołaniu do aktywności są też w stanie m.in. sterować urządzeniami w domu, co nadaje nowy wymiar określeniu smart dom, gdzie dotąd wystarczała automatyka i AI nie koniecznie musiała być uwzględniana [92-96]. Mogą nam towarzyszyć w wielu obszarach naszego życia, być nam jak przyjaciel, lecz nie możemy zapomnieć, że kooperacja między AI a nami, także w przypadku VA to przestrzeń do rozwoju. I chodzi tu nie tyle o rozwój człowieka, poprzez dostęp do rozmaitych danych, czy możliwość na przykład z zakresu zaspokajanie deficytów ludzkich w obszarze relacji, czy po prostu poprawy humoru poprzez opowiadanie żartów przez VA, ile przede wszystkim o możliwość rozwoju dla tych rozwiązań, czy jak niektórzy powiedzą bytów. Przebywając z człowiekiem na co dzień mają one możliwość lepszego poznania nas ludzi, weryfikacji co i jak warunkuje jakie nasze zachowania, reakcje, co jest dla nas ważne, co nie, jak spędzamy czas, itd. Tym samym mają możliwość poszerzania swojej wiedzy i umiejętności. Czy to nie jest kolejny krok w kierunku stania się przez AI bardziej ludzką? Oczywiście nie możemy zapomnieć, że zanim dana AI będzie na tyle rozwinięta, by móc pełnić funkcję tak zaawansowanego VA, jak ten Google czy Siri, Cortana lub Alexa sieci neuronowe są trenowane na milionach albo i bilionach danych [98]. Niemniej jednak to, co czyni AI jeszcze bardziej podobną do człowieka to to, że byty te uczą się całe życie tak jak my. Przy czym w przypadku VA, zapewne ze względu na przeznaczenie ilość zmysłów, jakimi dysponuje może być mniejsza niż w przypadku człowieka, a jeśli jest ta sama lub większa w przypadku postaci robotycznych AI to i tak nie musi działać analogicznie jak ludzka. Przykładowo pszczoły, węże, nietoperze i ludzie też widzą na różne sposoby.

4.2. Człowiek, Sztuczna Inteligencja i ślady cyfrowe

Człowiek postrzegany jako ssak naczelny, istota rozumna, stając w obliczu ilości danych, jakie sam teraz produkuje, widzi ich potencjał, ale na ten moment jeszcze nie potrafi ich samodzielnie analizować i przetwarzać tak dokładnie jak Sztuczna Inteligencja. Ślady cyfrowe, przez które rozumiemy każde z zachowań użytkowników sieci, postrzegane są jako zmiana w kodzie binarnym. Są one tym wszystkim, co dzieje się za naszym pośrednictwem w Internecie, a więc m.in. wszelkie wpisy, grafiki, dobra cyfrowe, nagrania, maile czy komunikaty, przy czym, jak powiemy za dr. Michałem Kosińskim, nawet te, które zaczęliśmy tworzyć, ale nie zdecydowaliśmy się ich wysłać [99, 100]. Uogólniając wszystko to, co wydarzyło się w sieci. W związku z powyższym AI jedni będą traktować jako narzędzie, a inni jako partnera w wyszukiwaniu i analizie danych, odnajdywaniu powiązań pomiędzy nimi i nie tylko generowania raportów, ale i wyciągania wniosków i wskazywania rekomendacji. Być może znajdują się i tacy, którzy w AI będą widzieć potencjalnego rywala czy wręcz wroga. Już dziś umiejętności AI wykorzystywane są w marketingu, czy nawet prewencji bądź profilaktyce. Nie będę tu dywagować odnośnie możliwości, jakie to daje w budowaniu relacji z klientem, dopasowywania przekazu marketingowego. Nie będę tu rozpisywać się na temat możliwości manipulacji danymi, a co za tym idzie kształtowania postaw konsumentów, czy nawet szerzej grup społecznych, a może nawet całych społeczeństw. Podkreślę jednak,

że ślady cyfrowe pozwalają określić profile osobowościowe poszczególnych użytkowników internetu [100, 101]. Internet zaś to przestrzeń, gdzie pozostawiamy ogromną ilość śladów cyfrowych, zapewne nawet tu i teraz, kiedy czytasz ten artykuł. Rozwój internetu, a co za tym idzie także usług świadczonych za jego pomocą, więc i rozmaitych akcji i interakcji, sprawia, że to przestrzeń niewątpliwie cenna tak dla biznesu, czy polityków, jak i naukowców ze względu na wręcz niewyobrażalną, a do tego stale rosnącą ilość danych. Przestrzeń internetowa sprawia, że mamy do czynienia z *Big Data*, gdzie dane przybywają z każdą chwilą i wszelkie z nich można w zasadzie uznać za cenne [102, 103]. Przy czym, skoro tempo pojawiania się danych jest tak ogromne, że nawet tu i teraz w sieci pojawia się szereg nowych treści, szereg śladów cyfrowych to efektywność analizy wykonywanej ręcznie przez człowieka będzie zdecydowanie mniejsza niż w przypadku wykorzystania AI. Tego ogromnego potencjału, jaki niesie ze sobą wykorzystanie AI przy analizie danych, nie można nie doceniać.

Wiemy już, że AI zaczyna istnieć jako swoisty samodzielny byt po przetrenowaniu na setkach tysięcy, a niejednokrotnie na milionach, czy nawet miliardach danych, uczy się na podstawie relacji z nami, ale nie możemy też zapomnieć o tym, że poszerza swoją wiedzę także na podstawie śladów cyfrowych zostawianych przez człowieka. Może być naszym partnerem w działaniach bądź narzędziem manipulacji w wprawnych rękach. Tylko otwartym pytaniem pozostaje czyje to mogą być ręce czy raczej kogo/czego w odniesieniu do posiadanej wiedzy i doświadczeń, które przekładają się na poglądy, charakter, usposobienie, podejmowane decyzje, itd. Jaki to może mieć wpływ na naszą rzeczywistość w przyszłości, jaki ma już dziś.

5. Podsumowanie

AI to marzenie, które urzeczywistnia się na naszych oczach, stając się nie tylko narzędziem, ale i swego rodzaju bytem. Bytem, którego sposób myślenia ma przypominać człowieka, pomimo że nie musi on przyjmować postaci humanoidalnej, czy w ogóle jakiegokolwiek, aby działać. Wchodzić może w relacje społeczne niezależnie od przyjmowanej formy, co więcej z człowiekiem, który nie zawsze jest świadomy tego z kim ma do czynienia. Jedni dostrzegają potencjał AI, inni przede wszystkim wskazują na zagrożenia. Fascynuje, ale i jest powodem wielu obaw. W rzeczywistości biznesowej wykorzystywana jest w coraz to w nowych obszarach, a do tego coraz częściej nie tylko w wymiarze analitycznym, ale i kreatywnym.

W niniejszym artykule dokonano pewnej systematyzacji i przeglądu rozwiązań AI w wybranych branżach. Pokazano też, jak kształtuje się na świecie poziom postępu w kontekście rozwoju tej technologii i gotowości do niej nie tylko pojedynczych ludzi, ale i Państw w wymiarze prawnym. Innymi słowy, pochylając się nad dokonaniem, które zmieniło i stale zmienia naszą współczesną rzeczywistość, które identyfikujemy, z jak się wydaje, głównym filarem IV rewolucji przemysłowej, a które nazwane zostało Sztuczna Inteligencja, temat potraktowano przekrojowo. Jak mogliśmy się przekonać, nadanie nazwy temu rozwiązaniu technologicznemu sprawiło, że człowiek zaczął myśleć o nim niejako, jak o pewnym bycie, a nie narzędziu, co wiąże się też z szeregiem oczekiwań, jakich dotąd nie mieliśmy my ludzie w odniesieniu chyba do żadnego z rozwiązań technologicznych. W tym miejscu jednak należy podkreślić, że obok oczekiwań pojawiają się też obawy. Ekspozowany jednak przede wszystkim potencjał tych rozwiązań, ogrom możliwości w kontekście analizy danych oraz wspierania człowieka

w różnych obszarach życia – zachwyca. Potencjał tego rozwiązania fascynuje, choć też takie osoby, jak przykładowo Elon Musk, sygnalizują, że wraz z rozwojem Sztucznej Inteligencji i człowiek będzie musiał się zmienić, tłumacząc swoje prace nad neuralinkami, które nie tylko mają pomagać w leczeniu wybranych chorób, ale jak się wydaje, w pewnym sensie nas ludzi ulepszać, zwiększając nasz potencjał intelektualny (nieoficjalnie mówi się, że mogą też być narzędziem kontroli nad ludźmi, dla firmy będącej jej producentem). Postrzeganie AI z jednej strony jako swoistego rywała, a z drugiej partnera, pobudza wyobraźnię. Stało się wyzwaniem dla świata naukowego. Nadana temu rozwiązaniu technologicznemu nazwa sprawia, że dążymy nawet do tego, by AI potrafiła odczuwać emocje tak jak człowiek, a nie tylko umiała je interpretować czy wyrażać. Wciąż odnajdujemy nowe obszary zastosowania dla AI, a te ujęte w tym artykule to tylko „wierzchołek góry lodowej”. Przy czym, podchodząc do tematu rozwoju AI, nie możemy zapominać, że jak wskazują badania dr. Michała Kosińskiego widzimy, że anonimowość człowieka przy rozwoju współczesnych technologii z AI w zasadzie zostaje nam odebrana, a więc i wolność może być ograniczona. Jest to fakt, ale podobnie faktem jest też, że rozwój AI daje szereg możliwości człowiekowi. Ogromny potencjał dostrzega się chociażby w obszarze medycyny i biznesu. Podobnie w kontekście zaspokajania deficytów w wymiarze relacji. Jak mogliśmy się przekonać, czytając ten artykuł, różne obszary, w ramach których rozwijana jest AI, mogą być łączone, niosąc nową jakość, co poruszone było m.in. w części poświęconej chociażby pojazdom autonomicznym i aspektem związanym z wymiarem psychospołecznym czy też wirtualnym asystentom bądź rozwiązaniom AI, które wchodzi w relacje z człowiekiem w sferze gier. Możliwości jest naprawdę wiele, potencjał zaś ogromny.

Postęp technologiczny sprawił, że to, co do niedawna mogło być tylko teorią dziś się materializuje. AI jest rozwiązaniem, które jest dziś stale rozwijane dzięki współczesnym możliwościom technologicznym i pracy wielu zespołów interdyscyplinarnych, ale także takim, które równolegle może się samo rozwijać (oczywiście po wcześniejszym przetrenowaniu na ogromnej ilości danych). AI, które zaczyna działać samodzielnie, kształtować może się dziś zarówno na podstawie danych cyfrowych, jak i tego, co widzi i słyszy. Potrafi wykonywać polecenia, ale i daje rekomendacje, a nawet jest w stanie samodzielnie tworzyć (w tym inną AI)⁷. Tak jak każda nowość pobudza wyobraźnię, rodzi nadzieję i lęki, a ponadto jak wszystko, co rewolucjonizuje nasze życie wymaga też przystosowania z naszej strony. Przykładowo, co może być przedmiotem osobnego artykułu, pojawiają się potrzeby na nowe kompetencje, a więc i nowe zawody. Zapewne wiele będzie się zmieniać, a my... No cóż, to już zależy od nas samych jako ludzi, od naszej odpowiedzialności jako twórców i użytkowników AI, zupełnie tak jak w przypadku bomby atomowej. W zasadzie można zaryzykować tezę, że świadomość, odpowiedzialność, zaradność i nieustawanie w poszerzaniu swojej wiedzy, pomimo pokusy pozwolenia AI, aby robiła za nas to, co jesteśmy w stanie zrobić sami przy odrobinie wysiłku, mają dziś ogromne znaczenie dla człowieka w kontekście przyszłości nas, jako gatunku. Jak przy każdej rewolucji przemysłowej musimy się dopasować, znaleźć swoją przestrzeń, przy czym tu ze szczególnym naciskiem na to, by nie poddać się pokusie rozwiązywania zadań poprzez „kliknięcia” czy komendy dla AI. W świecie,

⁷ W 2018 roku głośno było o pierwszym „dziecku” Sztucznej Inteligencji Google. AutoML, Sztuczna Inteligencja stworzona przez naukowców z zespołu Google Brain, samodzielnie stworzyła system detekcji obiektów – NASNet. W tym przypadku nie mówimy o postaci humanoidalnej.

gdzie wiedza stała się towarem ogromną rolę odgrywają i odgrywać będą eksperci. W świecie, gdzie dzięki AI to, co podsunie nam wyobraźnia może być łatwiejsze do zrealizowania wyraźnie zaznacza się zapotrzebowanie na kreatywność i odpowiedzialność równocześnie.

Podziękowania

Szczególne podziękowania dla mojego męża i dzieci za to, że dają mi przestrzeń do pracy naukowej, inspirują i są mi wsparciem. Dla rodziców, że zawsze mogę na nich liczyć.

Literatura

1. Bourdieu P., *Habitus and Field: General Sociology*, Polity Press, Cambridge, Wielka Brytania 2019.
2. Kaplan J., *Sztuczna Inteligencja. Co każdy powinien wiedzieć*, Wydawnictwo Naukowe PWN, Warszawa 2019.
3. Pasek Z., *Imię boga. Praktyki magiczno-performatyw*, <https://rcin.org.pl/ijp/Content/105794/Pasek.pdf> [data dostępu: 22.12.2022].
4. Malinowski B., *Ogrody koralowe i ich magia*, PWN, Warszawa 1987, s. 349.
5. Łucarz S.J., *Świętość i świadectwo w pierwotnym chrześcijaństwie*, [w:] Dybel K., Zarebianka Z. (red.), *Twarze świętości*, Uniwersytet Papieski Jana Pawła II w Krakowie Wydawnictwo Naukowe, Kraków 2016, s. 18.
6. Lurker M., *Słownik obrazów i symboli biblijnych*, Pallottinum, Poznań 1989.
7. Lévy-Brühl L., *Czynności umysłowe w społeczeństwach pierwotnych*, PWN, Warszawa 1992.
8. Lurker M., *Słownik obrazów i symboli biblijnych*, Pallottinum, Poznań 1989, s. 68.
9. http://www.biblia.waw.pl/teksty/imiona_boga.php [data dostępu: 27.12.2022].
10. <https://nano.komputronik.pl/n/pierwszy-komputer-na-swiecie-kiedy-powstal/> [data dostępu: 23.01.2023].
11. Rolecki M., *Czy pozwolimy robotom umierać*, <https://www.sztucznainteligencja.org.pl/czy-pozwolimy-robotom-umierac/> [data dostępu: 27.12.2023].
12. Matacz M., *Prof. Klepka: roboty humanoidalne istnieją, ale na razie funkcjonują jako ciekawostki*, <https://naukawpolsce.pl/aktualnosci/news%2C94332%2Cprof-klepka-roboty-humanoidalne-istnieja-ale-na-razie-funkcjonuja-jako> [data dostępu: 23.01.2023].
13. [https://pl.wikipedia.org/wiki/Sophia_\(robot\)](https://pl.wikipedia.org/wiki/Sophia_(robot)) [data dostępu: 23.01.2023].
14. Kulik W., *Sophia – pierwszy robot z obywatelstwem (saudyjskim)*, <https://www.benchmark.pl/aktualnosci/sophia-pierwszy-robot-z-obywatelstwem-saudyjskim.html> [data dostępu: 23.01.2023].
15. Aronson E., Aronson J., *Człowiek istota społeczna*, Wydawnictwo Naukowe PWN, Warszawa 2011.
16. Nęcka E., Orzechowski J., Szymura B., *Psychologia poznawcza*, Wydawnictwo Naukowe PWN, Warszawa 2006, s. 14-15, 20.
17. Aronson E., *Człowiek – istota społeczna*, Wydawnictwo Naukowe PWN, Warszawa 2000.
18. Wojciszke B., *Człowiek wśród ludzi. Zarys psychologii społecznej*, Wydawnictwo Naukowe Scholar, Warszawa 2009.
19. https://www.youtube.com/watch?v=6_XoP2RHRO0 [data dostępu: 15.10.2021].
20. <https://www.youtube.com/watch?v=JRHdnkUjcZg> [data dostępu: 15.10.2021].
21. <https://www.youtube.com/watch?v=Eq3pyGpW-es&t=135s> [data dostępu: 12.09.2022].
22. Turek A., *Autonomiczne samochody będą mogły poruszać się bez kierowcy po kalifornijskich drogach*,

- <https://businessinsider.com.pl/technologie/motoryzacja/autonomiczne-samochody-bez-kierowcy-moga-jezdzic-w-kalifornii/1z3fglx> [data dostępu: 28.12.2022].
23. Kornacki J., *Autonomiczne auta. Uber się poddaje i... dalej zarabia na sprzedaży*, <https://automotyw.com/autonomiczne-auta-uber-sie-poddaje-i-dalej-zarabia-na-sprzedazy/> [data dostępu: 28.11.2022].
24. Garbarek N., *Kolejne autonomiczne taksówki w USA*, <https://www.benchmark.pl/aktualnosci/autonomiczne-taksowki-waymo-na-ulicach-san-francisco.html> [data dostępu: 28.12.2022].
25. Bellon M., *Autonomiczne ciężarówki już przewożą towary po publicznych drogach*, <https://businessinsider.com.pl/technologie/nowe-technologie/autonomiczne-ciezarowki-w-usa-przewoza-towary-po-publicznych-drogach/7kr63ge> [data dostępu: 29.12.2022].
26. Kuś Ł., *Pierwsze komercyjne zastosowanie autonomicznych ciężarówek Volvo Vera*, <https://intermodalnews.pl/2019/06/17/pierwsze-komercyjne-zastosowanie-autonomicznych-ciezarowek-volvo-vera/> [data dostępu: 29.12.2022].
27. <https://www.komputerswiat.pl/autonomiczne-ciezarowki> [data dostępu: 29.12.2022].
28. Zimowska S., *Autonomiczne ciężarówki trafią do floty Lidla*, <https://www.chip.pl/2017/11/lidl-autonomiczne-auto> [data dostępu: 29.12.2022].
29. Kosarzycki R., *Sztuczna inteligencja przejmie Marsa. Planetę zamieszkały inteligentne roboty*, <https://spidersweb.pl/2021/07/perseverance-sztuczna-inteligencja-przejmuje-marsa.html> [data dostępu: 24.09.2021].
30. Dura M., *Pierwsze bojowe użycie roju dronów przez Izrael*, <https://defence24.pl/sily-zbrojne/pierwsze-bojowe-uzycie-roju-dronow-przez-izrael> [data dostępu: 24.01.2023].
31. Tomanek M., *Autonomiczne drony w Libii atakowały ludzi*, <https://robotyka.pl/autonomiczne-drony-w-libii-atakowaly-ludzi/> [data dostępu: 24.01.2023].
32. Łysoń M., *Drony stają się coraz ważniejsze, a te autonomiczne już zaatakowały ludzi*, <https://www.chip.pl/2021/11/autonomiczne-drony-atak-na-ludzi-libia> [data dostępu: 24.01.2023].
33. Piskorz A., *Amerykańskie wojsko testuje latające, autonomiczne samochody. Po co?*, <https://www.instalki.pl/aktualnosci/technika/48113-us-air-force-autonomiczne-latajace-samochody.html> [data dostępu: 29.12.2022].
34. Okurowski T., *Honda z licencją na autonomiczną jazdę. Pierwsza taka na świecie*, <https://www.auto-swiat.pl/wiadomosci/aktualnosci/honda-z-licencja-na-autonomiczna-jazde-pierwsza-taka-na-swiecie/2l9pg0s> [data dostępu: 28.12.2022].
35. Ruciński A., *Japonia zalegalizowała autonomiczną jazdę. Pierwsza skorzysta Honda*, <https://autokult.pl/japonia-zalegalizowala-autonomiczna-jazde-pierwsza-skorzysta-honda,6809446683408513a> [data dostępu: 28.12.2022].
36. <https://elektrowoz.pl/auta/honda-wprowadza-na-rynek-pierwszy-samochod-autonomiczny-3-poziomu-tylko-leasing-100-egzemplarzy/> [data dostępu: 28.12.2022].
37. Dębecki R., *Nowy Mercedes Klasy S z systemem jazdy autonomicznej Drive Pilot*, <https://www.auto-swiat.pl/wiadomosci/aktualnosci/nowy-mercedes-klasy-s-z-systemem-jazdy-autonomicznej-drive-pilot/bwdmtmw> [data dostępu: 28.12.2022].
38. Garbarek N., *Mercedes uzyskał zgodę na autonomiczną jazdę po niemieckich autostradach z prędkością... 60 km/h*, <https://www.benchmark.pl/aktualnosci/mercedes-dostal-zgodena-autonomiczna-jazde-po-autostradach.html> [data dostępu: 28.12.2022].
39. Koll S., *Autonomiczne pojazdy wyjadą na niemieckie drogi. Ważna decyzja rządzących*, <https://trans.info/pl/autonomiczne-pojazdy-wyjada-na-niemieckie-drogi-wazna-decyzja-rzadzacych-241728> [data dostępu: 28.12.2022].
40. Lewczuk M., *Autonomia 4 poziomu pozwoli na poruszanie się pojazdów autonomicznych po niemieckich ulicach*, <https://itbiznes.pl/transport/autonomia-4-poziomu-samochody-autonomiczne-drogi-publiczne-niemcy-przepisy-2022-rok/> [data dostępu: 28.12.2022].

41. Kozikowska M., *EasyMile EZ10 to pierwszy autonomiczny pojazd dopuszczony do ruchu*, <https://francuskie.pl/easymile-ez10-to-pierwszy-autonomiczny-pojazd-dopuszczony-do-ruchu-we-francji/> [data dostępu: 28.12.2022].
42. <https://zmianynaziemi.pl/wiadomosc/francja-dopuszila-pierwszy-bezzalagowy-autobus-do-poruszania-sie-po-drogach-publicznych> [data dostępu: 28.12.2022].
43. Przybylski R., *Pierwszy krok ku autonomicznym pojazdom. Powstał ośrodek naukowy mający wspierać wdrożenie autonomicznych pojazdów*, <https://logistyka rp.pl/innowacje-w-logistyce/art17118261-pierwszy-krok-ku-autonomicznym-pojazdom> [data dostępu: 29.12.2022].
44. <https://www.transport-publiczny.pl/wiadomosci/gdansk-rusza-z-testami-autonomicznego-busika-przy-zoo-62520.html> [data dostępu: 03.01.2023].
45. Florencka K., *Oto pierwszy w Polsce autonomiczny bus komunikacji miejskiej. Darmowe przejazdy już za kilka dni*, <https://innpoland.pl/154629,autonomiczny-minibus-ez10-firmy-easymile-w-gdansku-kiedy-kursy-koszt> [data dostępu: 03.01.2023].
46. <https://www.facebook.com/100057115613638/posts/pfbid0Hqk4rk7FhSP1L3H5o79ku7K57fSGqNKrg2Kt6emmcNgEjv4sQZ5dnjxhBgou3Trul/> [data dostępu: 03.01.2023].
47. Kawecki M., *Polak, który pracuje z Elonem Muskem...*, <https://www.facebook.com/100057115613638/posts/pfbid0Hqk4rk7FhSP1L3H5o79ku7K57fSGqNKrg2Kt6emmcNgEjv4sQZ5dnjxhBgou3Trul/> [data dostępu: 03.01.2023].
48. <https://biznes.wprost.pl/firmy-i-rynki/11026213/sztuczna-inteligencja-chatgpt-podbija-internet-jej-tworca-jest-polak.html> [data dostępu: 30.12.2022].
49. Orlikowski P., *Współpracował z Muskem, stworzył słynny ChatGPT. Polak z Kluczborka podbija świat*, <https://innpoland.pl/188926,sztuczna-inteligencja-polak-wsrod-tworcow-chatgpt-to-wojciech-zaremba> [data dostępu: 30.12.2022].
50. Moon M., *Waymo will soon offer fully driverless rides to the public in San Francisco*, <https://www.engadget.com/waymo-fully-driverless-rides-san-francisco-094718022.html> [data dostępu: 28.12.2022].
51. Shanklin W., *Waymo now offers driverless rides between Downtown Phoenix and Sky Harbor Airport*, <https://www.engadget.com/waymo-phoenix-airport-180027616.html> [data dostępu: 28.12.2022].
52. Szacki J., *Historia myśli spocjologicznej*, Wydawnictwo Naukowe PWN, Warszawa 2011.
53. <https://www.medonet.pl/sztuczna-inteligencja-w-medycynie,artykul,37061334.html> [data dostępu: 30.12.2022].
54. https://www.comarch.pl/files-pl/file_188/PL_CH_WP_AI_w_medycynie.pdf [data dostępu: 30.12.2022].
55. <https://dlaszpitali.pl/nowosci-z-branzy/sztuczna-inteligencja-w-kardiologii-cardiomatics/> [data dostępu: 03.01.2023].
56. <https://cardiomatics.com/> [data dostępu: 03.01.2023].
57. Kobosz T., *Polska sztuczna inteligencja pomoże leczyć z alkoholizmu*, <https://www.medexpress.pl/polska-sztuczna-inteligencja-pomoze-leczyc-z-alkoholizmu/75434> [data dostępu: 30.10.2019].
58. <https://www.projektpulsar.pl/struktura/2137110,1,sztuczna-inteligencja-pomoze-uzalezniwym.read> [data dostępu: 01.12.2019].
59. <https://www.magnapolonia.org/sztuczna-inteligencja-moze-wspomoc-leczenie-uzalezniwienia-od-alkoholu/> [data dostępu: 30.12.2022].
60. Bednarz P., *Algorytm, który leczy. Sztuczna inteligencja w medycynie to już dziś setki milionów dolarów, ale jest mały problem*, <https://businessinsider.com.pl/technologie/nowe-technologie/sztuczna-inteligencja-w-medycynie-i-opiece-zdrowotnej/tj3whe1> [data dostępu: 04.01.2023].
61. <https://neuralink.com/> [data dostępu: 26.01.2023].

62. <https://antyweb.pl/neuralink-początek-cyberpunkowego-szalenstwa> [data dostępu: 26.01.2023].
63. <https://www.komputerswiat.pl/neuralink> [data dostępu: 26.01.2023].
64. <https://www.knowyourmobile.com/pl/user-guides/what-is-neuralink/> [data dostępu: 26.01.2023].
65. Vance A., *Elon Musk. Biografia twórcy PayPala, Tesli i SpaceX*, Wydawnictwo Znak, Kraków 2022.
66. Kaczmarczyk M., *Deep Blue. Komputer, który wygrał z Kasparowem*, <https://www.newsweek.pl/historia/deep-blue-komputer-ktory-wygral-z-kasparowem/kc0p1m6> [data dostępu: 25.01.2023].
67. Cichy A., *24 lata od historycznej porażki Garriego Kasparowa. Szachowe maszyny jak silnik parowy i elektryczność' „Jednokierunkowa droga”*, <https://sport.tvp.pl/53759875/24-lata-temu-garri-kasparow-przebral-z-deep-blue-sztuczna-inteligencja-w-szachach-doping-technologiczny-i-patrycja-waszczuk-tvpsportpl> [data dostępu: 25.01.2023].
68. <https://pl.wikipedia.org/wiki/AlphaGo> [data dostępu: 02.09.2022].
69. Kopańko K., *O tym jak chłopak z Gdańska tak wytrenował AI, że powalila zawodowych e-graczy*, <https://spidersweb.pl/plus/2021/02/o-tym-jak-chlopak-z-gdanska-tak-wytrenowal-ai-ze-powalila-zawodowych-e-graczy> [data dostępu: 25.01.2023].
70. Jurczak E., *Sztuczna Inteligencja (AI) – obszary zastosowań, możliwości i ryzyka*, <http://indigotree.pl/sztuczna-inteligencja-ai/> [data dostępu: 14.01.2021].
71. Chutecki J., *AI może wygrywać z człowiekiem również w pokera*, <https://www.computerworld.pl/news/AI-moze-wygrywac-z-czlowiekiem-rowniez-w-pokera,407188.html> [data dostępu: 25.01.2023].
72. Bójko M., *Historyczna wygrana sztucznej inteligencji w pokera! Maszyna lepiej blefuje i negocjuje od człowieka*, <https://wyborcza.pl/7,75400,21445655,historyczna-wygrana-sztucznej-inteligencji-w-pokera-maszyna.html> [data dostępu: 25.01.2023].
73. Surma J., *Business Intelligence. Systemy wspomaganie decyzji biznesowych*, Wydawnictwo Naukowe PWN, Warszawa 2016.
74. <https://www.youtube.com/watch?v=Zm9B-DvwOgw> [data dostępu: 15.10.2022].
75. Park M., *This AI tries to code whatever you tell it to, even videogames*, <https://www.pcgamer.com/this-ai-tries-to-code-whatever-you-tell-it-to-even-videogames/> [data dostępu: 26.01.2023].
76. <https://www.polski-chatbot.pl/historia-chatbotow-test-turinga/> [data dostępu: 04.01.2023].
77. Wallace R.S., *The Anatomy of A.L.I.C.E.*, https://freeshell.de/~chali/programowanie/Anatomy_of_ALICE.pdf [data dostępu: 04.01.2023].
78. Shah H., *A.L.I.C.E.: an ACE in Digitaland*, Department of Artificial Intelligence & Interactive University of Westminster, UK, London 2006.
79. Duszczek M., *ChatGPT, robiący furorę chatbot, który wszystko wie i potrafi, to kopalnia złota*, <https://cyfrowa.rp.pl/technologie/art37624361-chatgpt-robiacy-furore-chatbot-ktory-wszystko-wie-i-potrafi-to-kopalnia-zlota> [data dostępu: 04.01.2023].
80. Cembrowski K., *Brzmi to absurdalnie, ale Bing ma szansę zagrozić Google. Wszystko dzięki... ChatGPT*, <https://antyweb.pl/bing-chatgpt-moze-zagrozic-google> [data dostępu: 04.01.2023].
81. Ulatowska P., *Nowy chatbot OpenAI porozmawia z Tobą o wszystkim, nawet o... metodach kradzieży*, <https://www.benchmark.pl/aktualnosci/chatgpt-nowy-chatbot-openai.html> [data dostępu: 04.01.2023].
82. Górski S., *Ta technologia ma pozbawić dziennikarzy zawodu. Wszystko, co musisz wiedzieć o ChatGPT*, <https://www.chip.pl/2022/12/co-musisz-wiedziec-o-chatgpt> [data dostępu: 04.01.2023].

83. Zhavoronkov A., *Wu Dao 2.0 – Bigger, Stronger, Faster AI From China* Kacper Cembrowski, <https://www.forbes.com/sites/alexzhavoronkov/2021/07/19/wu-dao-20bigger-stronger-faster-ai-from-china/?sh=57c348916fb2> [data dostępu: 04.01.2023].
84. Dilmegani C., *Wu Dao 2.0 in 2023: China's Improved Version of GPT-3*, <https://research.aimultiple.com/wu-dao/> [data dostępu: 04.01.2023].
85. Glonek Sz., Bartosz Malinowski: *Voiceboty korzystają ze sztucznej inteligencji, żeby automatyzować monotonne, powtarzalne zadania [PODCAST]*, <https://www.gazetaprawna.pl/podcasty/obiektynie-o-biznesie/artykuly/8493808,bartosz-malinowski-voiceboty-korzystaja-ze-sztucznej-inteligencji-zeby-automatyzowac-monotonne-powtarzalne-zadania-podcast.html> [data dostępu: 04.01.2023].
86. <https://www.parp.gov.pl/component/content/article/80526:chatboty-i-voiceboty-przyszloscia-komunikacji-korzysta-z-nich-coraz-wiecej-firm> [data dostępu: 04.01.2023].
87. Isakow A., *Voicebot: co to jest i jakie przynosi korzyści?*, <https://www.omni-chatbot.pl/voicebot-co-to-jest-i-jakie-przynosi-korzysci/> [data dostępu: 04.01.2023].
88. Łupkowski P., *Test Turinga perspektywa sędziego*, Wydawnictwo Naukowe UAM, Poznań 2010.
89. Kurzak T., *Asystent Google kontra Alexa, Cortana oraz Siri. Porównujemy możliwości wirtualnych asystentów*, <https://www.komputerswiat.pl/artykuly/redakcyjne/asystent-google-kontra-alexa-cortana-oraz-siri-porownujemy-mozliwosci-wirtualnych/f2b2jhd> [data dostępu: 04.01.2023].
90. <https://www.wirtualnemedi.pl/artykul/znane-aplikacje-na-smartfonach-podsluchujanasze-rozmowy-aby-potem-wyswietlac-nam-dopasowane-reklamy-zadzwonit-na-numer-21> [data dostępu: 16.01.2023].
91. Nowak M., *Apple szykuje zmiany w Siri. Wirtualną asystentkę będziemy aktywować w inny sposób*, <https://android.com.pl/tech/533004-koniec-hey-siri-apple-ma-lepszy-pomysl/> [data dostępu: 04.01.2023].
92. <https://atozofai.withgoogle.com/intl/pl/virtual-assistants/> [data dostępu: 04.01.2023].
93. Kulik W., *Czym jest inteligentny dom? Jak działa i jakie korzyści przynosi?*, <https://www.komputronik.pl/informacje/czym-jest-inteligentny-dom/> [data dostępu: 16.01.2023].
94. <https://smartdom.eu/> [data dostępu: 16.01.2023].
95. <https://support.apple.com/pl-pl/guide/iphone/iph83aad8922/ios> [data dostępu: 16.01.2023].
96. https://newsblog.pl/najlepsze-zarty-gry-i-pisanki-dla-asystenta-google/#google_vignette [data dostępu: 16.01.2023].
97. <https://elektronikab2b.pl/biznes/52870-jak-dzialaja-asystenci-glosowi> [data dostępu: 04.01.2023].
98. Jaworska M., <https://elektronikab2b.pl/biznes/52870-jak-dzialaja-asystenci-glosowi> [data dostępu: 04.01.2023].
99. <https://www.youtube.com/watch?v=JFIsrkzuY0Q&t=1490s> [data dostępu: 17.01.2023].
100. Rust J., Kosiński M., Stillwell D., *Modern Psychometrics, The Science of Psychological Assessment*. 4th Edition, Routledge, Abingdon-on-Thames, Great Britain, 2020.
101. Żytmirski M., *Określanie osobowości użytkowników Internetu poprzez analizowanie ich cyfrowych śladów w świetle wybranych badań dr. Michała Kosińskiego*, *Acta Universitatis Lodzianis, Folia Librorum*, 1(32), 2021, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2021.
102. Grajeta P., *Marketing baz danych jako źródło konkurencyjnej przewagi i narzędzie dyskursywnego tworzenia klienta jako obszaru wiedzy*, *Zeszyty Naukowe Politechniki Łódzkiej*, 1222, Wydawnictwo Politechniki Łódzkiej, Łódź 2018.
103. Wehmeyer K., Stewart H., *Aligning IT and marketing – The impact of database marketing and CRM*, Henry Stewart Publications, 12(3), s. 243-256, *Database Marketing & Customer Strategy Management*, <https://link.springer.com>, London, Great Britain, 2005.

Wybrane kierunki rozwoju Sztucznej Inteligencji okiem socjologa i... z perspektywy biznesu

Streszczenie

Sztuczna Inteligencja (AI) to marzenie, które urzeczywistnia się na naszych oczach, stając się swego rodzaju bytem. Bytem, którego sposób myślenia ma przypominać człowieka, pomimo że nie musi ona przyjmować postaci humanoidalnej, czy w ogóle jakiegokolwiek, aby działać. Wchodzi w relacje społeczne niezależnie od przyjmowanej formy, co więcej nie zawsze ze świadomym tego człowiekiem. Jedni dostrzegają jej potencjał, inni przede wszystkim wskazują na zagrożenia. Fascynuje, ale i jest powodem wielu obaw. W rzeczywistości biznesowej wykorzystywana jest w coraz to w nowych obszarach, a do tego coraz częściej nie tylko w wymiarze analitycznym, ale i kreatywnym, na co autorka wskazuje, przyglądając się różnym branżom okiem socjologa i... z perspektywy biznesu. W niniejszym artykule ujęte zostały także realne scenariusze w odniesieniu do koegzystencji człowiek-AI. Temat jest niezwykle złożony stąd w pierwszej części pracy znaleźć można krótki rys historyczny i podjęta została próba odpowiedzi na pytanie czy AI byłaby tym, czym teraz jest, czym rysuje się w naszych oczach, gdyby nie wydarzenia z 1956 roku, jakim było nadanie jej takiej, a nie innej nazwy przez John'a McCarthy'ego. Czy mielibyśmy te same oczekiwania wobec tego rozwiązania technologicznego, jakie pojawiają się dziś chociażby w wymiarze emocji... To wszystko po to, by w dalszej części pochylić się nad rozwojem i zastosowaniem AI, dokonać przeglądu wybranych rozwiązań, pokazać dynamikę rozwoju tej technologii i gotowości na nią nas ludzi, także w odniesieniu do aspektów prawnych. Przybliżone zostają wybrane rozwiązania potrafiące m.in. poruszać się w zmiennej rzeczywistości i/lub konwersować z człowiekiem. Wskazane są też oczywiście możliwości i zagrożenia. Pierwsze zwycięstwa AI nad człowiekiem i kwestie, które pozostają stale wyzwaniem w kontekście rozwoju tej technologii.

Słowa kluczowe: Sztuczna Inteligencja, pojazdy autonomiczne, chatbot, voicebot, robotyka

Selected development directions of Artificial Intelligence from sociological and... business perspective

Abstract

Artificial intelligence (AI) is a dream which is becoming true before our very eyes. It is a being whose mindset is supposed to be human-like, even though it does not have to assume a humanoid form or any other forms to function. It enters into social relations regardless of its forms. Moreover, not always with a person who is aware of it. Some people see its potential, while others mainly point to the threats. It fascinates, but also causes much concern. It is used in many areas in the context of the business world, and what is interesting, more and more often not only in the analytical but in the creative dimension too. The author touches upon these aspects and analyzes them from the sociological and... business perspective. This article also covers real-life scenarios related to human - AI coexistence. Since the subject is extremely complex, the first part of this article provides a short historical outline. Also, an attempt was made to answer the question whether AI would be what it is now, if it had not been for the events from 1956 when John McCarthy gave this particular name to the Artificial Intelligence. Would we have the same expectations towards this technological solution that we appear to have today, for example, in the emotional dimension ... This is all done in order to focus on the development and application of AI, review selected solutions, show the dynamics of the development of this technology and our readiness for it, also with regard to legal aspects. The article shows selected solutions that are able to e.g. move in the changing reality and/or converse with a human. Opportunities and threats are also covered here. The first victories of AI over man and issues that remain a constant challenge in the context of the development of this technology.

Keywords: Artificial Intelligence, autonomous vehicle, chatbot, voicebot, robotics

System ekspercki wspomaganie decyzji lekarza oparty o zbiór reguł medycznych

1. Wprowadzenie

Diagnostyka medyczna jest integralną częścią procesu terapeutycznego. Bez poprawnej diagnozy nie ma prawidłowego leczenia. Jednak często ocena stanu chorego nie jest wykonywana na czas, zdarza się również, że część badań diagnostycznych wykonywanych jest niepotrzebnie. Problem niedoskonałości diagnostyki medycznej oraz błędów medycznych określany jest jako problem społeczny. Ocenia się, że z problemami błędnej diagnostyki czy leczenia bezpośrednio spotkało się 31% polskiego społeczeństwa [1]. Istnieje wiele przyczyn błędnej diagnostyki, między innymi takie jak: niedostatek wiedzy w danej dziedzinie, nieprecyzyjne testy i badania diagnostyczne, złożoność problemu. Dlatego od wielu lat poszukiwane są metody i narzędzia eliminujące ww. problemy.

Od wielu lat istnieją próby zastosowania metod informatycznych w zakresie wspomaganie diagnostyki medycznej. Z jednej strony powstają nowe narzędzia diagnostyczne wyposażone w wyspecjalizowane oprogramowanie (np. diagnostyka obrazowa, cyfrowe urządzenia pomiarowe) [2]. Z drugiej strony opracowywane są systemy eksperckie wyposażone w mechanizmy wnioskowania, wspomagające lekarzy w diagnostyce medycznej [3]. Jednym z pierwszych takich rozwiązań był system ekspercki Mycin, opracowany w latach 70. XX wieku na Uniwersytecie Stanford [4], ukierunkowany na diagnostykę bakteryjnych chorób krwi. Pomimo tego, że były próby opracowania systemów eksperckich ogólnego przeznaczenia [5, 6] to jednak większość takich rozwiązań jest ukierunkowana na konkretne typy schorzeń np. choroby uszu [7], zaburzenia pracy wątroby [8], choroby tarczycy [9] itp. Mimo licznych propozycji dotychczas żadne z opracowanych rozwiązań nie zyskało szerszego zastosowania.

Głównym problemem w opracowaniu skutecznych metod wnioskowania w zakresie diagnostyki medycznej jest trudność w opracowaniu algorytmu zapewniającego akceptowalny poziom skuteczności. Ze względu na złożoność problemu najczęściej stosuje się rozwiązania bazujące na metodach sztucznej inteligencji [10]. W ostatnich latach nastąpił znaczący postęp w rozwoju tych metod, pojawiło się wiele skutecznych

¹ m.wisniewski@tu.kielce.pl, Katedra Systemów Informatycznych, Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Świętokrzyska, www.tu.kielce.pl.

² s.deniziak@tu.kielce.pl, Katedra Systemów Informatycznych, Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Świętokrzyska, www.tu.kielce.pl.

³ d.kaczmarek@tu.kielce.pl, Katedra Systemów Informatycznych, Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Świętokrzyska, www.tu.kielce.pl.

⁴ p.pieta@tu.kielce.pl, Katedra Systemów Informatycznych, Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Świętokrzyska, www.tu.kielce.pl.

⁵ piotr.pepe.picheta@gmail.com, Politechnika Świętokrzyska, www.tu.kielce.pl.

⁶ aleksandra@khamira.pl, Politechnika Świętokrzyska, www.tu.kielce.pl.

⁷ mkb2@orange.pl, Politechnika Świętokrzyska, www.tu.kielce.pl.

podejść, takich jak głębokie uczenie maszynowe (ang. *deep learning*) czy analityka *Big Data*. Następuje również postęp technologiczny w rozwoju urządzeń diagnostycznych. W poszczególnych dziedzinach medycyny pojawiają się coraz bardziej zaawansowane urządzenia, pozwalające na wykonywanie dokładnych pomiarów różnych parametrów zdrowia pacjenta. Urządzenia te dostarczają dużej ilości danych, które mogą być wykorzystywane do dokładniejszej i szybszej diagnostyki. Wykorzystanie ww. zaawansowanych technologii informatycznych może pozwolić na uzyskanie systemu eksperckiego umożliwiającego gromadzenie wiedzy z mechanizmem wnioskowania o wystarczającej wiarygodności.

Analityka danych stwarza ogromne możliwości w zakresie diagnostyki medycznej [11]. Analizując odpowiednie duże zbiory danych zawierające informacje o chorych (rodzaj schorzenia, objawy, wyniki badań, skuteczność terapii itp.), możliwe jest odkrywanie zależności np. pomiędzy objawami i wynikami badań a rodzajem choroby, pomiędzy wynikami badań a skutecznością terapii itp. Jednak warunkiem skuteczności takiej analizy jest dostępność reprezentatywnego zbioru danych. Często taki zbiór jest niedostępny, jednakże można wtedy zastosować podejście wydobywania wiedzy z danych w trakcie eksploatacji systemu [12].

W pracy przedstawiono propozycję metody automatycznej diagnostyki (AD) oraz predykcji terapii wykorzystującą system ekspercki bazujący na regułach decyzyjnych. Metoda jest określona w sposób generyczny tzn. nie ma z góry określonych danych wejściowych i wyjściowych. Definiowane jest to na etapie konfiguracji metody. W ten sposób proponowane podejście może służyć do wyznaczenia dowolnych reguł decyzyjnych, wnioskujących na podstawie wcześniej zdefiniowanego zbioru parametrów zdrowia pacjenta. Po ich określeniu mogą one stanowić bazę dla systemu eksperckiego wspomagającego procesy decyzyjne lekarzy. Przykłady zastosowań przedstawione w pracy demonstrują możliwości proponowanego rozwiązania.

W kolejnych rozdziałach pracy opisano: założenia proponowanego podejścia, rozdział trzeci przedstawia zastosowanie metody w zakresie diagnostyki medycznej i predykcji terapii, w czwartym przedstawiono wyniki wykonanych eksperymentów, a ostatni zawiera wnioski i podsumowanie pracy.

2. Metodyka

Metoda AD wykorzystuje próbki danych (parametry zdrowia pacjenta), które są zgrupowane w dwa zestawy: danych początkowych oraz danych końcowych. W zestawach mogą być przechowywane różne typy parametrów. Metoda pozwala na przetwarzanie danych liczbowych (typy zmiennoprzecinkowe), logicznych (wartości prawda/fałsz) oraz typów wyliczeniowych. Dla typu liczbowego określa się zakres danych (próg dolny i górny).

Do każdego parametru przyporządkowana jest wartość liczbową nosząca angielską nazwę *relevance*, określająca wpływ tej danej na wynik działania metody AD. Ponadto dla wszystkich typów danych definiuje się funkcję *significance*, której celem jest normalizacja wartości [13] przechowywanych w parametrach. W przypadku danych liczbowych przez normalizację rozumie się przeskalowanie wartości parametru do przedziału od 0 do 100. Dla danych logicznych określa się dwie wartości normalizujące, odpowiednio dla dwóch wartości parametru, tj. dla prawdy i fałszu. Dla typów wyliczeniowych wartości normalizujące należy zdefiniować dla każdej wartości tego typu.

Przykładowo dla typu wyliczeniowego posiadającego wartości A, B i C, wartości normalizujące mogą mieć postać tablicy [10, 30, 100]. Próbki danych przygotowane w powyższy sposób są gromadzone w bazie wiedzy.

Metoda AD korzysta z jednolitej struktury danych bazy wiedzy, w której są przechowywane zestawy danych oraz informacje charakteryzujące próbki danych. Baza wiedzy może zostać zasilona danymi inicjującymi, jak również może być uzupełniana w trakcie pracy algorytmów. Podczas każdego z tych etapów rozszerzanie bazy wiedzy wymaga obliczenia wartości charakteryzujących próbki danych. Są one wyliczane z uwzględnieniem wartości *relevance* oraz funkcji normalizującej *significance*. Na tej podstawie algorytmy odpowiedzialne za aktualizację bazy wiedzy podejmują decyzję o jej wzbogaceniu o nowe dane lub o odrzuceniu zestawu danych, który nie powinien zostać zapamiętany. Korzystając z tych informacji, następuje również wstępne oszacowanie, czy dana próbka danych może zostać wykorzystana w procesach analitycznych algorytmów metody AD.

2.1. Baza wiedzy danych medycznych

Podczas stosowania metody AD w diagnostyce lub do predykcji terapii, próbki danych przechowują dane medyczne charakterystyczne dla rozważanej diagnostyki lub terapii. W tym przypadku metoda AD będzie przetwarzała parametry zdrowia pacjenta zgrupowane w zestawy. W typowych warunkach postępowania medycznego dane opisujące stan zdrowia pacjenta można gromadzić na wielu etapach leczenia. Z etapu, w którym pacjent jest poddawany wstępnej ocenie zdrowia, będą pochodzić próbki dla początkowego zestawu danych. Następnie można gromadzić próbki w trakcie trwania terapii, jak również kontynuować ocenianie zdrowia pacjenta po zakończeniu leczenia. W ten sposób powstanie zestaw danych końcowych. Artykuł ten skupia się na przetwarzaniu parametrów dwóch zestawów danych: początkowego i końcowego. Oczywiście istnieje również możliwość przetwarzania próbek pośrednich, co zostało omówione w dalszej części artykułu.

Niezależnie od przypadku użycia metody AD, jej skuteczność zależy od danych zapisanych w bazie wiedzy. Proces gromadzenia wiedzy polega na przetwarzaniu wspomnianych wcześniej zestawów parametrów zdrowia: początkowego i końcowego. Istotnym elementem tego procesu jest wyznaczenie wartości cechujących parametry zdrowia względem przyjętej metodyki tj. określenie znaczenia parametru z użyciem *relevance* oraz zdefiniowanie funkcji normalizującej *significance*. Wartości te można oszacować metodami statystycznymi, bazując na dużej liczbie próbek danych – jest to jeden z procesów, który jest realizowany podczas gromadzenia danych w bazie wiedzy.

3. Diagnostyka medyczna

Jednym z zastosowań metody AD jest diagnostyka medyczna. Procesy związane z tym zagadnieniem obejmują dwie czynności:

1. Gromadzenie danych w bazie wiedzy – proces wykorzystuje zestawy danych medycznych, dla których zostały określone diagnozy. Na tej podstawie jest uzupełniana baza wiedzy.
2. Wykonanie właściwej diagnostyki – wejściem dla tego procesu jest nowy zestaw danych, najczęściej pochodzący z oceny zdrowia pacjenta.

Na podstawie posiadanej bazy wiedzy metoda AD wskazuje prawdopodobne diagnozy, dlatego proces właściwej adnotacji próbek zasilających bazę wiedzy powinien zostać przeprowadzony przez osobę mającą odpowiednią wiedzę medyczną.

3.1. Predykcja terapii

Proces predykcji terapii jest podobny do procesu diagnostycznego. Również w tym przypadku metoda może korzystać z danych zgromadzonych wcześniej w bazie wiedzy, uzupełniając ją o nowe dane właściwe dla rozważanych terapii. Proces opisany we wstępie do sekcji 4 może zostać wykorzystany bez większych zmian na potrzeby proponowania terapii. Podstawową różnicą jest tu sposób interpretacji danych. W przypadku diagnostyki mówimy o zestawie objawów, które wpływają na wartości w próbkach danych. Dla terapii mamy natomiast do czynienia z wartościami w próbkach danych, które nie cechują jeszcze pacjenta, a które powinny go charakteryzować po zakończeniu terapii. W tym kontekście mowa jest zatem o parametrach wzorcowych terapii. W ogólności z predykcją terapii związane są następujące czynności:

1. Określenie terapii i ich parametrów wzorcowych.
2. Gromadzenie i uzupełnianie danych w bazie wiedzy – proces wykorzystuje zestawy danych medycznych, dla których określono rodzaj stosowanej terapii.
3. Wykonanie predykcji terapii – wejściem dla tego procesu jest nowy zestaw danych, najczęściej pochodzący z oceny zdrowia pacjenta.

Na podstawie bazy wiedzy metoda AD wskazuje, które terapie mogą być dla danego pacjenta najbardziej skuteczne. Powyższy etap nr 2 jest opcjonalny i może zostać przeprowadzony pod koniec postępowania terapeutycznego, co ma na celu wzbogacenie bazy wiedzy.

3.2. Szacowanie stanu zdrowia

Proces szacowania stanu zdrowia pacjenta jest powiązany z procesem predykcji terapii. W tym przypadku wejściem dla algorytmu są próbki parametrów zdrowia pochodzące z oceny stanu zdrowia. Metoda ocenia, jak bardzo przekazane próbki danych odbiegają od parametrów wzorcowych zdefiniowanych wcześniej w terapii. W pewnym zakresie szacowanie stanu zdrowia jest wykonywane również podczas predykcji terapii. Różnica polega głównie na określeniu progów akceptacji danych. W przypadku szacowania stanu zdrowia algorytm zakłada możliwość wystąpienia większych wahań w parametrach zdrowia niż ma to miejsce w przypadku predykcji terapii.

3.3. Baza wiedzy

Próbki danych medycznych, zorganizowane w zestawy parametrów, są wykorzystywane jako dane wejściowe do algorytmów metody AD. Parametry muszą mieć wyznaczoną wartość *relevance* oraz wartości funkcji normalizującej *significance*. Ponadto dla danych charakteryzujących terapie należy określić parametry wzorcowe terapii, a dla danych wykorzystywanych w diagnostyce – parametry właściwe dla diagnostyki. Podczas wprowadzania danych do bazy wiedzy algorytm wylicza wartości charakteryzujące próbki danych – efektywność diagnostyczną lub terapeutyczną.

Algorytm zasilający bazę wiedzy został zrealizowany następująco:

1. Parametry zestawu wejściowego z próbkami danych medycznych są porównywane z parametrami wzorcowymi diagnostyki lub terapii. Obecnie dane nie są różnicowane względem diagnostyki lub terapii.

2. Jeśli na pierwszym etapie liczba zgodnych parametrów zdrowia zestawu wejściowego i wybranej diagnostyki lub terapii znajduje się na akceptowalnym poziomie, to następuje zakwalifikowanie danych do dalszego przetwarzania. Przez zgodne próbki danych rozumie się takie, które mają ten sam typ, a wartość parametru z zestawu wejściowego mieści się w zakresie dozwolonych wartości korespondującej z nim parametru wzorcowego diagnostyki lub terapii.
3. Dla zestawu wejściowego zakwalifikowanego na etapie 2 zostaje wyliczona wartość efektywności diagnostycznej lub terapeutycznej, będąca sumą ważonych różnic między wartościami parametrów wejściowych i wzorcowych. Taka miara pozwala na oszacowanie tego, jak bardzo zestaw danych wejściowych różni się od danych wzorcowych diagnostyki lub terapii.
4. Na tym etapie algorytm dodatkowo może wyliczyć efektywność terapii, jeśli dane dotyczą terapii. Odbywa się to poprzez oszacowanie stanu zdrowia pacjenta względem parametrów wzorcowych terapii (algorytm został opisany w sekcji 3.5), co zostaje wykonane osobno dla zestawu wejściowego i dla zestawu wyjściowego parametrów zdrowia. Różnica otrzymanych wartości stanowi szacunkową skuteczność terapii dla wejściowego zestawu danych.
5. Jeśli w bazie wiedzy nie istnieją dane zbliżone do tych wyznaczonych na etapie 3, to takie dane zostaną zapamiętane. Wraz z danymi statystycznymi w bazie wiedzy zostaną również zapisane zestawy parametrów, dla których zostały wykonane obliczenia, w tym dane diagnostyki lub terapii z ich parametrami wzorcowymi.

Przygotowane w ten sposób dane są następnie wykorzystywane przez algorytmy diagnostyki lub predykcji terapii. Pseudokod obrazujący działanie zasilania bazy wiedzy wygląda następująco:

```
// Etap 1
dla każdego(dt z diagnoz/terapii) {
  wspólne_parametry_zdrowia = 0
  dla każdego(ref_param z dt.ref_params) {
    // etap 2
    if(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
    w zakresie ref_param) {
      wspólne_parametry_zdrowia++
    }
  }
  jeśli(liczba_wspólne_parametry_zdrowia spełnia kryteria akceptacji
  wynikające z dt.ref_params.liczba_parametrów) {
    zapamiętanie dt w tablicy dane_dla_etap3
  }
}
```

```
// Etap 3
dla każdego(dt z dane_dla_etap3) {
    efektywność = 0
    dla każdego(ref_param z dt.ref_params) {
        jeśli(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
        w zakresie ref_param) {
            efektywność + = oblicz_efektywność(dane_wejściowe.param.wartość,
            ref_param.wartość)
        }
    }
}

// Etap 4
skuteczność = oblicz_skuteczność(dane_wejściowe, dt)
nowy_rekord_dt = (dane_wejściowe, dt, efektywność, skuteczność)

// Etap 5
jeśli(baza_wiedzy nie zawiera nowy_rekord_dt) {
    dodanie nowy_rekord_dt do baza_wiedzy
}
}
```

gdzie:

- dt – rekord danych diagnozy/terapii;
- param – parametry zdrowia;
- ref_param – parametry zdrowia, referencyjne dla diagnozy/terapii;
- dane wejściowe (parametry oceny zdrowia i na zakończenie leczenia);
- diagnoz/terapii – dane wzorcowe odpowiednio diagnozy lub terapii.

3.4. Algorytm diagnostyki

Metoda AD umożliwia przeprowadzenie diagnostyki medycznej. Algorytm wymaga wcześniejszego zasilenia bazy wiedzy odpowiednimi danymi, tj. mającymi zastosowanie do diagnostyki. Proces diagnostyczny został zrealizowany następująco:

1. Rekordy w bazie wiedzy są przeglądane i następuje porównanie parametrów zestawu wejściowego, zawierającego próbki danych medycznych, z parametrami wzorcowymi diagnostyki zapisanymi w bazie wiedzy. W przypadku tego algorytmu dane wejściowe pochodzą zazwyczaj ze wstępnej oceny stanu zdrowia pacjenta.
2. Jeśli w pierwszym etapie liczba zgodnych parametrów zdrowia zestawu wejściowego i wybranej diagnostyki spełnia kryteria akceptacji, to taki zestaw podlega dalszemu przetwarzaniu. Podobnie jak w sekcji 3.3, przez parametry zgodne rozumie się takie, które mają ten sam typ, a wartość parametru z zestawu wejściowego mieści się w zakresie dozwolonych wartości korespondującego z nim parametru wzorcowego diagnostyki.

3. Po wybraniu rekordów z bazy wiedzy następuje ich posortowanie względem wyznaczonej wcześniej efektywności diagnostycznej.
4. Algorytm zwraca pewną liczbę rekordów z bazy wiedzy, co można sparametryzować – liczba ta jest zależna od preferencji użytkownika algorytmu.

Pseudokod obrazujący działanie algorytmu diagnostyki wygląda następująco:

```
// Etap 1
dla każdego(d z bazy_wiedzy_diagnoz) {
  wspólne_parametry_zdrowia = 0
  dla każdego(ref_param z d.ref_params) {
    // Etap 2
    jeśli(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
    w zakresie ref_param) {
      wspólne_parametry_zdrowia++
    }
  }
  jeśli(liczba wspólne_parametry_zdrowia spełnia kryteria akceptacji
  wynikające z dt.ref_params.liczba_parametrów) {
    zapamiętanie d w dane_dla_etap3
  }
}
// Etap 3
sortowanie rosnąco dane_dla_etap3
// Etap 4
wyjście: pierwsze n-rekordów z dane_dla_etap3
```

gdzie:

- d – rekord danych z bazy wiedzy;
- ref_param – parametry zdrowia, referencyjne dla diagnozy/terapii;
- dane wejściowe (parametry zdrowia pochodzące z oceny zdrowia).

3.5. Szacowanie stanu zdrowia

Algorytm szacowania stanu zdrowia pacjenta jest wykorzystywany przez dalej opisany algorytm proponowania terapii, aczkolwiek może być on również używany samodzielnie. Szacowanie stanu zdrowia odnosi się do zdefiniowanej wcześniej terapii oraz danych zgromadzonych w bazie wiedzy. Zostało ono zrealizowane następująco:

1. Rekordy w bazie wiedzy powiązane z terapiami są przeglądane: następuje porównanie parametrów zestawu wejściowego zawierającego próbki danych medycznych z parametrami wzorcowymi terapii.

2. Algorytm zakłada możliwość zdefiniowania progu akceptacji dla liczby zgodnych parametrów danych wejściowych i parametrów wzorcowych terapii. Jeśli próg akceptacji został osiągnięty, to dane są przekazywane do dalszego przetwarzania.
3. Dla danych wybranych w etapie 2 następuje obliczenie stanu zdrowia, polegające na określeniu efektywnej wartości parametru wejściowego, do czego wykorzystywane są *relevance* i *significance* należące do parametrów wzorcowych terapii. Następnie taka wartość jest odejmowana od wartości wzorcowego parametru, w efekcie czego otrzymuje się różnicę wartości pozwalającą wyznaczyć efektywny ubytek zdrowia związany z tym parametrem.
4. Algorytm zwraca pewną liczbę rekordów z bazy wiedzy, co można sparametryzować – liczba ta jest zależna od preferencji użytkownika algorytmu.

Szacowanie stanu zdrowia może zostać również wykonane bez użycia danych pochodzących z bazy wiedzy. W takim przypadku algorytm wykorzystuje referencyjne parametry zdrowia wybranej terapii. Pseudokod obrazujący działanie powyżej przedstawionego algorytmu wygląda następująco:

```
// Etap 1
dla każdego(d z bazy_wiedzy_terapii) {
    wspólne_parametry_zdrowia = 0
    dla każdego(ref_param z d.ref_params) {
        // Etap 2
        jeśli(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
        w zakresie ref_param) {
            wspólne_parametry_zdrowia++
        }
    }
    jeśli(liczba wspólne_parametry_zdrowia spełnia kryteria akceptacji
    wynikające z dt.ref_params.liczba_parametrów) {
        zapamiętanie d w dane_dla_etap3
    }
}

// Etap 3
dla każdego(d z dane_dla_etap3) {
    stan_zdrowia = 0
    dla każdego(ref_param z d.ref_params) {
        jeśli(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
        w zakresie ref_param) {
            stan_zdrowia + =
```

```
oblicz_stan_zdrowia(  
  dane_wejsciowe.param_wartość,ref_param.ważność_parametru,  
  ref_param.znaczenie_parametru)  
}  
}  
jeśli(stan_zdrowia > 0) {  
  stan_zdrowia = stan_zdrowia / wspólne_parametry_zdrowia  
  zapamiętanie (d, stan_zdrowia) w dane_dla_etap4  
}  
  // Etap 4  
  sortowanie rosnąco dane_dla_etap4  
  wyjście: dane_dla_etap4  
}
```

gdzie:

- d – rekord danych z bazy wiedzy zawierający dane diagnoz/terapii;
- param – parametry zdrowia;
- ref_param – parametry zdrowia, referencyjne dla diagnozy/terapii;
- dane wejściowe (parametry zdrowia pochodzące z oceny zdrowia).

3.6. Predykcja terapii

Algorytm predykcji terapii służy do proponowania terapii dla danego pacjenta. Używa on danych zgromadzonych w bazie wiedzy, które odnoszą się do terapii. Jako dane wejściowe wykorzystywany jest zestaw parametrów zdrowia pacjenta, najczęściej pochodzący ze wstępnej oceny jego zdrowia. Algorytm został zrealizowany następująco:

1. Rekordy w bazie wiedzy powiązane z terapiami są przeglądane: następuje porównanie parametrów zestawu wejściowego zawierającego próbki danych medycznych z parametrami wzorcowymi terapii.
2. Algorytm zakłada możliwość zdefiniowania progu akceptacji dla liczby zgodnych parametrów danych wejściowych i parametrów wzorcowych terapii. Jeśli próg akceptacji został osiągnięty, to dane są przekazywane do dalszego przetwarzania.
3. Dla poszczególnych rekordów wybranych w etapie 2 następuje wyliczenie prognozy efektywności leczenia, co jest realizowane poprzez porównanie wartości parametrów wejściowych z parametrami zapamiętanymi w tym rekordzie dla danego leczenia. Różnice w wartościach parametrów oblicza się uwzględniając *relevance* i *significance*, określone dla terapii związanej z danym leczeniem zapisanym w bazie wiedzy.
4. Algorytm zwraca tę terapię, dla której wartość wyliczona w etapie 3 jest największa.

Różnice między algorytmem proponowania terapii oraz diagnostyki polegają głównie na sposobie wykorzystania danych wejściowych oraz obliczania metryk, na podstawie których algorytmy klasyfikują wyniki swoich działań. Odpowiednio algorytm propo-

nowania terapii bierze pod uwagę związane z terapią leczenie, które jest tu niejako dodatkowym punktem odniesienia dla szacowań skuteczności terapii, obok parametrów wzorcowych terapii. Natomiast algorytm diagnostyki korzysta tylko z parametrów wzorcowych diagnostyk. Pseudokod obrazujący działanie powyżej przedstawionego algorytmu wygląda następująco:

```
// Etap 1
dla każdego(d z bazy_wiedzy_terapii) {
wspólne_parametry_zdrowia = 0
dla każdego(ref_param z d.ref_params) {
// Etap 2
jeśli(dane_wejściowe zawierają ref_param i dane_wejściowe.param są
w zakresie ref_param) {
wspólne_parametry_zdrowia++
}
}
jeśli(liczba wspólne_parametry_zdrowia spełnia kryteria akceptacji
wynikające z dt.ref_params.liczba_parametrów) {
zapamiętanie d w dane_dla_etap3
}
}
// Etap 3
propozycja = brak_danych
dla każdego(d z dane_dla_etap4) {
// Etap 4
skuteczność = oblicz_skuteczność(dane_wejściowe, d)
jeśli(propozycja.skuteczność < skuteczność) {
propozycja = (dane_wejściowe, d, skuteczność)
}
}
// Etap 4
wyjście: propozycja
```

gdzie:

- d – rekord danych z bazy wiedzy zawierający dane diagnoz/terapii;
- param – parametry zdrowia;
- ref_param – parametry zdrowia, referencyjne dla diagnozy/terapii;
- dane wejściowe (parametry zdrowia pochodzące z oceny zdrowia).

4. Eksperymenty

W celu oszacowania efektywności algorytmów zostały wykonane dwa rodzaje prac badawczych. W zakresie diagnostyki przygotowano bazę wiedzy z wykorzystaniem danych z bazy MIMIC [14]. Natomiast na potrzeby oceny skuteczności szacowania stanu zdrowia oraz proponowania terapii zostały wykorzystane dane zgromadzone w trakcie realizacji projektu TWEC [15-18].

4.1. Diagnostyka na podstawie danych z bazy MIMIC

W bazie danych MIMIC znajdują się dane medyczne pacjentów pochodzące z badań składu krwi oraz oceny podstawowych parametrów życiowych. Dla każdego pacjenta dostępne były dwie próbki danych: zestaw oceny stanu zdrowia oraz zestaw parametrów zdrowia tuż przed opuszczeniem placówki medycznej. Została także określona diagnoza dla tego pacjenta.

Dane pobrane z bazy MIMIC w pierwszym kroku pogrupowano względem diagnozy, a następnie dla każdej grupy zostały wyznaczone istotne parametry zdrowia, w tym wartości *relevance* i *significance* dla poszczególnych parametrów. Odpowiednio:

1. Jako parametry istotne zostały wybrane te, dla których szybkość zmian wartości w grupie była mniejsza niż 25%.
2. Dla parametrów istotnych dla diagnozy wartość *relevance* została wyznaczona również względem szybkości zmian wartości danego parametru w grupie, co przedstawiono w tabeli 1.
3. Funkcja *significance* miała zawsze charakter stały, co wynikało z zakresu wartości próbek oraz braku typów logicznych i wyliczeniowych.

Tabela 1. Algorytm wyznaczania wartości *relevance* dla parametrów zdrowia

Wartość <i>relevance</i>	Szybkość zmian wartości parametru zdrowia w grupie [%]
100	[0,3)
80	[3,5)
60	[5,10)
50	[10,15)
40	[15,20)
20	[20,25)

Źródło: opracowanie własne.

Tabela 2 podsumowuje wolumen danych, które zostały wykorzystane w ramach eksperymentów, z ich podziałem na poszczególne diagnozy. Z każdej grupy została wylosowana pewna liczba pacjentów, a ich dane nie zostały dodane do bazy wiedzy. Dane te posłużyły do weryfikacji skuteczności algorytmu diagnostyki.

Tabela 2. Wolumen danych dla eksperymentów diagnostyki na podstawie danych z bazy MIMIC

Diagnoza	D1	D2	D3	D4	D5	D6	D7	D8	D9
Liczba rekordów	3	4892	21	48	87	158	28	51	7

Źródło: opracowanie własne.

Wyniki eksperymentów przedstawiono w tabeli 3. Kolorem zielonym oznaczono prawidłowe predykcje diagnozy, kolorem czerwonym błędne, a kolorem niebieskim drugi najwyższy wynik predykcji dla danej próbki danych. Wnioski płynące z danych widniejących w tabeli 3 są następujące:

1. Dla danych z bazy wiedzy dla diagnozy D1 liczba rekordów zasilających była zbyt mała, aby można było prawidłowo wyliczyć znaczenie parametrów terapii (*relevance*). Skutkuje to przypisywaniem tej terapii do zbliżonych wartościowo parametrów pacjentów. Jeśli diagnoza D1 zostałaby pominięta w rozważaniach, to w wierszach D5 i D8 predykcja byłaby prawidłowa. W tym zakresie algorytm zasilania bazy wiedzy wymaga dalszego dostrojenia.
2. Szacowanie odbywa się z pewnym progiem akceptacji różnic parametrów względem ich wzorcowych wartości. Problem ten uwidacznia się dla diagnozy D6, dla której nawet odrzucenie diagnozy D1 jako błędnych danych nie spowoduje prawidłowej predykcji. Jest to spowodowane tym, że różnice między danymi diagnoz D6 i D3 są niewielkie (predykcja w tym wierszu jest na korzyść diagnozy D3). Duże znaczenie ma tu metodyka szacowania *relevance* – w tej chwili opiera się ona na obserwacji szybkości zmian wartości parametru. W tym zakresie algorytm będzie musiał uwzględniać dodatkowe czynniki.
3. Analogicznie do pkt 2, podobny problem dotyczy podobieństwa diagnoz D7 i D9.
4. Jeśli diagnoza D1 zostanie odrzucona, to skuteczność zaprezentowanej metody wyniesie 62,5%.

Tabela 3. Wyniki eksperymentów diagnostyki na podstawie danych z bazy MIMIC

Próbki ze zbioru	Predykcje (w %)								
	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	100	14,09	54,33	48,99	61,87	41,73	49,32	73,34	85,85
D2	94,75	100	51,35	56,38	46,27	34,12	36,49	48,77	45,40
D3	67,94	73,97	72,47	54,73	100	49,79	62,92	27,56	67,36
D4	94,02	50,44	60,35	100	35,1	35,75	53,33	64,39	65,78
D5	100	57,91	49,13	49,07	78,27	11,52	16,35	11,15	56,28
D6	100	59,04	72,14	60,50	63,19	70,11	43,59	52,66	70,20
D7	100	0	44,71	38,61	41,72	26,24	48,65	55,76	74,77
D8	100	5,99	52,29	49,52	60,61	39,68	43,72	82,09	67,57
D9	64,99	42,00	0	74,34	92,30	67,01	74,34	95,66	100

Źródło: opracowanie własne.

4.2. Szacowanie stanu zdrowia w TWEC

W bazie projektu TWEC znajdują się dane biometryczne pacjentów pochodzące z badań klinicznych, a także pozyskane z użyciem urządzeń diagnostyczno-rehabilitacyjnych: bieżni Zebris FDM-T i platform stabilometrycznych AC International East Alfa i Gamma. W projekcie brali udział pacjenci z różnymi dysfunkcjami ruchu kończyn dolnych, to jest z jednostronnym porażeniem kurczowym (hemiplegią) i obustronnym porażeniem kurczowym (diplegią) oraz paraliżem czterokończynowym (tetraplegią). Wszyscy badani, których dane zostały poddane ewaluacji, charakteryzowali się poziomem niepełnosprawności GMFCS od I do III.

W projekcie TWEC pacjenci byli rehabilitowani z użyciem kilku procedur terapeutycznych, które różnią się stosowanymi urządzeniami medycznymi. Pierwsza terapia nosiła miano „Bieżnia i Platformy” (w skrócie TB+P). Druga miała nazwę „Egzoszkielec i Platformy” (w skrócie TE+P). Trzecia została określona „Egzoszkielec” (w skrócie TE). Dla każdego pacjenta były dostępne następujące informacje:

- terapia, którą był rehabilitowany pacjent;
- dwie próbki danych: zestaw parametrów zdrowia pacjenta przed i dwa tygodnie po zakończeniu terapii.

Parametry zdrowia zostały podzielone przez lekarzy biorących udział w projekcie na kilka grup: informacyjne, kwalifikacyjne i prognostyczne. Część z parametrów należała też do kilku grup jednocześnie. Wszystkie parametry prognostyczne zostały wykorzystane jako istotne dla algorytmu AD, co oznacza, że dla nich wyznaczono wartości *relevance* i *significance*. Odpowiednio:

1. Z uwagi na małą liczebność danych zgromadzonych w projekcie TWEC, wszystkim parametrom istotnym przypisano *relevance* o wartości 100.
2. Funkcja *significance* miała charakter liniowy: skalowała wartości parametrów (ciągłe lub dyskretne) do przedziału [1, 100]. W przypadku równowagi statycznej pacjenta skalowanie było bardziej zaawansowane, ponieważ następowało w dwóch przedziałach: wartości [0, 50] były skalowane do [1, 100], a wartości [50, 100] do [100, 1] (wzorcowa równowaga dla obu stron ciała wynosi bowiem 50%). Przykłady zastosowania funkcji *significance* zostały przedstawione w tabelach 4 i 5.

Tabela 4. Wartości funkcji *significance* dla parametru zdrowia „siła mięśniowa (Lovett)”

Siła mięśniowa (Lovett)	Wartość funkcji <i>significance</i>
0	1
1	20
2	40
3	60
4	80
5	100

Źródło: opracowanie własne.

Tabela 5. Wartości funkcji *significance* dla parametru zdrowia „spastyczność mięśni”

Spastyczność mięśni	Wartość funkcji <i>significance</i>
0	100
1	80
1+	60
2	40
3	20
4	1

Źródło: opracowanie własne.

Tabela 6 podsumowuje wolumen danych, które zostały wykorzystane w ramach eksperymentów, z ich podziałem na poszczególne terapie. Z każdej grupy została wylosowana pewna liczba pacjentów, a ich dane nie zostały dodane do bazy wiedzy. Dane te posłużyły do weryfikacji skuteczności algorytmu predykcji terapii.

Tabela 6. Wolumen danych dla eksperymentów predykcji terapii na podstawie danych z projektu TWEC

Terapia	TB+P	TE+P	TE
Liczba rekordów	9	40	12

Źródło: opracowanie własne.

Wyniki eksperymentów zostały przedstawione w tabeli 7. Kolorem zielonym oznaczono prawidłowe predykcje terapii, kolorem czerwonym błędne, a kolorem niebieskim drugi najwyższy wynik predykcji dla danej próbki danych. Wnioski płynące z danych widniejących w tabeli 7 są następujące:

1. Liczba rekordów zasilających bazę wiedzy dla terapii TE+P była największa, co może skutkować wskazywaniem jej jako najlepszej.
2. Da się zauważyć, że terapie, które wykorzystują urządzenie egzozkielet, są zawsze proponowane na pierwszym lub drugim miejscu.
3. Dla wszystkich próbek testowych predykcje dla poszczególnych terapii wyniosły powyżej 90%, co może oznaczać, że pacjenci, którzy zostali wylosowani do testów, różnili się niewiele pod względem wartości parametrów zdrowia.
4. Jeśli terapia TB+P zostanie odrzucona (z uwagi na mały wolumen danych, które w jej przypadku zasiły bazę wiedzy), to skuteczność zaprezentowanej metody wyniesie 60%.

Tabela 7. Wyniki eksperymentów predykcji terapii na podstawie danych z projektu TWEC

Próbki ze zbioru	Predykcje terapii (w %)		
	TB+P	TE+P	TE
TB+P (1)	90,44	100	97,67
TB+P (2)	90,44	100	93,82
TE+P (1)	90,44	100	93,82
TE+P (2)	90,44	100	93,82
TE+P (3)	90,44	100	93,82
TE (1)	90,44	100	93,82
TE (2)	96,61	100	97,67

Źródło: opracowanie własne.

5. Wnioski

W pracy przedstawiono system ekspercki wspomaganie decyzji lekarza oparty o zbiór reguł medycznych. W celu przeprowadzenia testów metody wykorzystano dane medyczne pochodzące z bazy MIMIC, jak również dane biometryczne pacjentów ze schorzeniem MPD, które zostały pozyskane w trakcie realizacji projektu TWEC.

Na podstawie przedstawionych w sekcji 5 eksperymentów w zakresie szacowania stanu zdrowia pacjenta oraz diagnostyki i proponowania terapii, można stwierdzić, że istotnym aspektem działania metody AD jest wartość *relevance*, która określa wpływ danego parametru zdrowia na wynik metody AD. W przypadku badań w kierunku dia-

gnostyki wyznaczenie wartości *relevance* zostało oparte na obserwacji szybkości zmian poszczególnych parametrów zdrowia pacjenta. Niebagatelne znaczenie ma tu także liczebność rekordów zasilających bazę wiedzy, ponieważ przy małej ich liczbie oszacowane wartości *relevance* mogą być obciążone dużym błędem, przez co metoda AD potrafi niepoprawnie wskazywać daną diagnozę lub terapię. W przypadku diagnostyki z użyciem metody AD, po odrzuceniu diagnozy, dla której liczebność danych uczących była najmniejsza, skuteczność metody wyniosła 62,5%. Dla zagadnienia proponowania terapii, również po nieuwzględnieniu terapii z najmniejszą bazą danych treningowych, stwierdzono skuteczność metody na poziomie 60%.

Dalsze prace skupią się na dostrojeniu opracowanego algorytmu AD w zakresie diagnostyki i proponowania terapii, w tym planowane jest uwzględnienie dodatkowych czynników podczas wyliczania wartości *relevance*. Proces ten może zostać również poprzedzony wyznaczeniem korelacji parametrów zdrowia pacjenta z terapiami lub wynikami diagnoz, w celu automatycznego określenia najbardziej istotnych parametrów.

Literatura

1. Cempa M., *Niedoskonałości diagnostyki medycznej a cierpienie chorego*, [w:] Makiełło-Jarży G. (red.), *Ból i cierpienie*, Krakowskie Towarzystwo Edukacyjne sp. z o.o. – Oficyna Wydawnicza AFM, Kraków 2012, s. 59-63.
2. Pelikant A., *Potencjał rozwiązań ICT i jego wykorzystanie w służbie zdrowia*, Roczniki Kolegium Analiz Ekonomicznych/Szkoła Główna Handlowa, 56, Technologie informatyczne w administracji publicznej i służbie zdrowia, 2019, s. 343-359.
3. Abu-Nasser B., *Medical Expert Systems Survey*, International Journal of Engineering and Information Systems (IJEAIS), 1(7), 2017, s. 218-224, <https://ssrn.com/abstract=3082734> [data dostępu: 31.03.2023].
4. Shortliffe E.H., *Consultation systems for physicians: The Role of Artificial Intelligence Techniques*, [w:] Bonnie L., Nilsson Nils J. (red.), *Readings in Artificial Intelligence*, Tioga Publishing Company, Palo Alto California, 1981, s. 323-333.
5. Singla J., Dinesh G., Abhinav B., *Medical Expert Systems for Diagnosis of Various Diseases*, International Journal of Computer Applications, 93(7), 2014, s. 36-43.
6. Arsene O., Dumitrache I., Miha I., *Expert system for medicine diagnosis using software agents*, Expert Systems with Applications, 42(4), 2015, s. 1825-1834.
7. Abu Naser S.S., Al-Nakhil M.A., *A Ruled Based System for Ear Problem Diagnosis and Treatment*, World Wide Journal of Multidisciplinary Research and Development, 2(4), 2016, s. 25-31.
8. Sushmita M., *Fuzzy MLP based expert system for medical diagnosis*, Fuzzy Sets and Systems, 65(2-3), 1994, s. 285-296.
9. Keles A., Keles A., *ESTDD: Expert system for thyroid diseases diagnosis*, Expert Systems with Applications, 34(1), 2008, s. 242-246.
10. Zhou Z.-H., *Machine learning*, Springer Nature, 2021.
11. Belowska-Bień K., Bień B., *Wykorzystanie technik sztucznej inteligencji i uczenia maszynowego we wspomaganie diagnostyki i leczenia chorób neurologicznych*, Aktualności Neurologiczne, 21(3), 2021, s. 163-172, DOI: 10.15557/AN.2021.0021.
12. Giżewski T., Kowalski I.M., Zarzycki D., Radomska-Wilczewska A., Lewandowski R., Kotwicki T., *Model systemu samouczącego w diagnostyce medycznej*, Polish Annals of Medicine, 15(1), 2008.
13. Kreyszig E., *Advanced Engineering Mathematics*, Wiley, 1979, s. 880.
14. Johnson A., Stone D., Celi L., Pollard T., *The MIMIC Code Repository: enabling reproducibility in critical care research*, Journal of the American Medical Informatics Association, 25(1), 2018, s. 32-39.

15. Bedla M., Pięta P., Kaczmarski D., Deniziak S., *Estimation of Gross Motor Functions in Children with Cerebral Palsy Using Zebris FDM-T Treadmill*, Journal of Clinical Medicine, 11(4), 2022, 954.
16. Krechowicz A., Deniziak S., Kaczmarski D., *Machine learning approach to gait deviation prediction based on isokinetic data acquired from biometric sensors*, Gait & Posture, 101, 2023, s. 55-59.
17. Kaczmarski D., Bedla M., Pięta P., Deniziak S., *Ocena wpływu stosowanych terapii na zmiany wybranych parametrów chodu pacjentów z mózgowym porażeniem dziecięcym*, [w:] Maciąg M., Kalbarczyk K. (red.), *Fizjoterapia i aktywność fizyczna w profilaktyce i leczeniu*, Wydawnictwo Naukowe TYGIEL Sp. z o.o., Polska, Lublin 2022, s. 32-45.
18. Bedla M., Krechowicz A., Deniziak S., Łukawski G., Pięta P., Kaczmarski D., Pawełkiewicz M., Borkowski M., *Zastosowanie metodyki big data w diagnostyce pacjentów z mózgowym porażeniem dziecięcym*, [w:] Danielewska A., Mołdoch-Mendoń I. (red.), *Technologie XXI wieku – aktualne problemy i nowe wyzwania*, Wydawnictwo Naukowe TYGIEL Sp. z o.o., Polska, Lublin 2021, s. 175-190.

System ekspercki wspomagania decyzji lekarza oparty o zbiór reguł medycznych

Streszczenie

W pracy przedstawiono metodę automatycznej diagnostyki medycznej (AD), która pozwala na ocenę stanu schorzenia pacjenta. Metoda wykorzystuje system ekspercki bazujący na zbiorze reguł decyzyjnych. Algorytmy stosowane do realizacji diagnostyki wykorzystują dane nazywane parametrami zdrowia, które grupowane są w zestawy. Proces ten wymaga od operatora zdefiniowania parametrów zdrowia dla diagnostyki, w tym dla każdego parametru określenia jego cech i znaczenia dla diagnozy. Dla danych przygotowanych w ten sposób algorytmy metody AD wyliczają metryki, które następnie są wykorzystywane w procesie określania metody diagnostyki danego pacjenta lub szacowania stanu zdrowia tego pacjenta względem wybranej procedury diagnostycznej. Dzięki tej drugiej własności metoda AD może posłużyć w procesie proponowania najskuteczniejszej terapii. W takim przypadku sposób definiowania parametrów terapii nie odbiega zasadniczo od sposobu definiowania parametrów zdrowia dla diagnozy. Innym z możliwych zastosowań metody są badania przesiewowe, na które mogą być kierowani pacjenci spełniający kryteria akceptacji zdefiniowane przez lekarzy.

Metoda AD została poddana ewaluacji z wykorzystaniem danych znajdujących się w bazie MIMIC oraz danych zgromadzonych w trakcie realizacji projektu TWEC, obejmującego pacjentów z mózgowym porażeniem dziecięcym. Uzyskane wyniki potwierdzają możliwość stosowania metody AD w celu wyznaczania pacjentów do badań przesiewowych, a także jako pomoc dla lekarza podczas wskazywania pacjentom zalecanej terapii.

Słowa kluczowe: system ekspercki, reguły decyzyjne, diagnostyka medyczna, MIMIC, TWEC

The expert system for physician decision support based on a set of medical rules

Abstract

The paper presents a method of automatic medical diagnosis (AD), which allows for estimation of a patient's condition. The method utilizes expert system based on a set of production rules. One of the possible applications of the method is the screening tests for which patients that meet the acceptance criteria defined by a user of the method (operator) can be referred. Algorithms used to perform diagnostics use arbitrary data, generally called health parameters, grouped into sets. The machine learning supervision process requires the operator to define the health parameters for the diagnosis, including for each parameter its characteristics and significance for the diagnosis. For data prepared in this way, the algorithms of the AD method calculate metrics, which are then used in the process of indicating a diagnosis for a selected patient or estimating the health status of this patient in relation to the selected diagnosis. Due to the latter property, the AD method can be used in the therapy proposal process. In this case, the method of defining data for therapy does not differ significantly from the method of defining data for diagnosis.

The AD method was evaluated using data from the MIMIC database and data collected during the TWEC project. The results obtained confirm the possibility of using the AD method to designate patients for screening, as well as to help the physician recommend therapy to patients.

Keywords: expert system, production rules, medical diagnostics, MIMIC, TWEC

Zastosowanie automatycznego rozpoznawania mowy w transkrypcji wywiadów medycznych – porównanie silników ASR dla języka polskiego

1. Wstęp

Technologia rozpoznawania mowy (ang. *Automatic Speech Recognition*, ASR, ang. *Speech-To-Text*, STT), posiada liczne zastosowania, od usług dyktowania i transkrypcji notatek głosowych, przez interaktywne systemy odpowiedzi głosowej, po asystentów wirtualnych. Technologia ta ewoluuje w szybkim tempie, czego efektem są coraz dokładniejsze systemy ASR z mniejszą liczbą błędów w procesie rozpoznawania. Warto jednak zauważyć, że jakość obecnych systemów rozpoznawania mowy znacznie się różni i niektóre z nich wciąż mają trudności z dokładną transkrypcją wybranych słów.

Systemy ASR obecne na rynku światowym poczyniły znaczne postępy w ostatnich latach pod kątem jakości rozpoznawania mowy. Najpopularniejsze przykłady takich rozwiązań obejmują aktualnie Google Cloud z silnikiem Google ASR [1], Microsoft Azure Cloud z silnikiem Microsoft ASR [2], Nuance Communications Cloud z silnikiem Dragon STT [3], Phonexia z silnikiem Phonexia STT [4], Techmo z silnikiem Techmo ASR [5], Rev AI z silnikiem Rev AI ASR [6]. Są one w stanie dobrze rozpoznawać codzienny język potoczny, wyrażenia idiomatyczne, kolokwializmy i regionalne akcenty, wypowiedzi bogate w słownictwo, a także przetwarzać nagrania z mową błędną lub wadliwą, o różnych właściwościach akustycznych [7-12]. Warto przy tym zauważyć, że na dokładność rozpoznania mowy i liczbę błędnych rozpoznań wyrazów przez systemy ASR mogą wpływać różne czynniki, poza samą klarownością mowy. Czynniki te obejmują [13-17]: rodzaj języka naturalnego używanego w mowie, zastosowane słownictwo (potoczne lub specjalistyczne), akcent mówcy i pochodzenie etniczne, hałas w tle i zakłócenia akustyczne, a także architekturę rozwiązań użytych w rozpoznawaniu mowy (słowniki, implementacja gramatyki itp.).

Przeprowadzona przez autorów analiza literatury wykazała brak badań dotyczących oceny jakości dostępnych na polskim rynku rozwiązań ASR pomagających w przeprowadzaniu wywiadów medycznych w potocznym języku polskim. Może to wynikać z faktu, iż rozwiązania ASR w zagranicznych podmiotach branży medycznej, które poprawiły jakość opieki nad pacjentami i zarządzanie szpitalem, są najczęściej prezentowane w źródłach komercyjnych w formie raportów firmowych, wywiadów eksperckich lub tzw. *white papers* [18]. Istnieje zatem potrzeba badań porównawczych aktualnego działania oraz potencjału funkcjonowania systemów ASR w celu stymulacji pozytywnych zmian w krajowym podejściu do świadczenia opieki zdrowotnej.

Celem niniejszego badania jest zapełnienie wyżej opisanej luki badawczej przez przeprowadzenie analizy porównawczej wiodących na polskim rynku rozwiązań ASR.

¹ kkuligowska@wne.uw.edu.pl, Wydział Nauk Ekonomicznych, Uniwersytet Warszawski, www.wne.uw.edu.pl.

² ms@sovva.ai, SOVVA S.A., http://sovva.ai.

³ marek@koniew.pl, SOVVA S.A., http://sovva.ai.

Kryteria oceny tych rozwiązań oparto na stopniu dokładności konwersji wypowiedzianych fraz w języku polskim na format tekstowy, przy użyciu miar dokładności dla poprawnie rozpoznanych słów (ang. *accuracy*) i wskaźnika błędnie rozpoznanych słów (ang. *Word Error Rate*, WER) oraz miar Levenstheina i Jaro-Winklera. Wyłonienie najlepszego z tych rozwiązań pozwoli w niedalekiej przyszłości na jego integrację z Platformą Omni-Chatbot (OCP) opracowaną przez SOVVA S.A. i udoskonalenie systemu automatycznej obsługi klienta. Docelowy system firmy SOVVA S.A., przetwarzając naturalnie wypowiedziany potoczny język polski, ma potrafić przeprowadzić wstępny głosowy wywiad lekarski (ang. *pre-triage*) i dzięki temu zidentyfikować przyczyny kontaktu pacjenta z placówką medyczną.

Artykuł jest zorganizowany w następujący sposób: w rozdziale 2 przedstawiono przegląd dostępnej literatury nad badaniami w zastosowaniach systemów rozpoznawania mowy w sektorze medycznym. W rozdziale 3 opisano metodologię przyjętą do realizacji niniejszych badań, a także ich przebieg. W rozdziale 4 przedyskutowano wyniki badań. Podsumowanie i wnioski zawarto w rozdziale 5.

2. Przegląd literatury dla systemów ASR w zastosowaniach medycznych

Wykorzystanie technologii ASR w medycynie staje się w ostatnich latach coraz popularniejszym obszarem badań. Jednym z wiodących zastosowań tej technologii jest przetwarzanie elektronicznej dokumentacji klinicznej, która obejmuje notatki głosowe nagrane przez pracowników służby zdrowia oraz zapisy głosowe podsumowujące wypisy ze szpitala [19, 20]. Dokumenty te często zawierają złożoną terminologię medyczną i wymagają dużej dokładności transkrypcji, aby zapewnić pacjentom najlepszą możliwą opiekę.

Wraz ze wzrostem popularności dedykowanych systemów zapisywania notatek głosowych oraz ich dalszej transkrypcji i ostatecznie włączania do elektronicznej dokumentacji medycznej nastąpił gwałtowny wzrost badań mających na celu zrozumienie, w jaki sposób technologia ASR może pomóc świadczeniodawcom usprawnić przepływ pracy i zapewnić lepszą opiekę nad pacjentem [8, 21]. Pojawienie się rozwiązań chmurowych zainicjowało intensywne badania możliwości automatycznego przetwarzania mowy w celu jej transkrypcji w czasie rzeczywistym podczas wywiadów medycznych. Do takich badań wykorzystywane są scenariusze rozmów lekarzy z tzw. symulowanymi pacjentami (ang. *simulated patients*), czyli wyszkolonymi aktorami wcielającymi się w role pacjentów w sytuacjach medycznych [22, 23]. Oznacza to przeniesienie środka ciężkości zainteresowań naukowców, bowiem przyglądają się oni teraz poprawie dokładności i wydajności systemów ASR w wywiadach medycznych i koncentrują się na zapewnieniu asysty dla rozpoznawania w czasie rzeczywistym.

W 2020 roku grupa naukowców z University of South California przeprowadziła kompleksowe badanie porównujące dokładność różnych, publicznie dostępnych, komercyjnych i naukowych systemów ASR [24]. Badaniem objęto systemy opracowane przez Amazon, Apple, Google, IBM, Microsoft i Kaldi, które zostały przetestowane na siedmiu zestawach angielskich danych głosowych w sześciu różnych zakresach tematycznych. Zbiory danych głosowych składały się ze spontanicznych nagrań mowy pochodzącej z nieindukowanych sztucznie rozmów, dzięki czemu badanie było wysoce reprezentatywne dla rzeczywistych scenariuszy. Naukowcy wykorzystali WER jako podstawową miarę oceny. Badanie wykazało, że wszystkie badane systemy ASR nadal

popępniały znaczną liczbę błędów transkrypcyjnych, niezależnie od producenta i rodzaju wytrenowania. Co więcej, żaden z systemów rozpoznawania mowy nie dominował we wszystkich zestawach danych. Jako jeden z powodów takich wyników naukowcy wskazali rodzaj treningu, jaki przechodzą systemy ASR. Przykładowo, gorsze wyniki uzyskały te systemy, które były trenowane na próbkach głosowych pochodzących z audiobooków, które znacząco odbiegają jakością dźwięku i samej mowy od mowy spontanicznej. Równie złe wyniki wykazały systemy trenowane w przeciwieństwie do poprzednich na próbkach konwersacyjnych, ale jedynie w zakresie mowy telefonicznej o niskiej jakości, co wpłynęło negatywnie na ostateczną dokładność. Część z błędów rozpoznania można przypisać wielkości zestawu danych, jednak większość z nich jest ściśle związana z samą dziedziną, której dotyczą wypowiedzi, a więc użytego specjalistycznego słownictwa. Równie błędnie rozpoznawana była spontaniczna mowa potoczna.

Porównanie badań zespołu naukowców z University of South California z wcześniejszymi, przeprowadzonymi w 2010 roku przez X. Yao i in. [25] oraz w 2013 roku przez F. Morbini i in. [26], dokonany na podobnych zbiorach danych, wyraźnie wskazuje, że w ostatnich latach dokonał się znaczny postęp w dziedzinie rozpoznawania mowy. Oznacza to, że choć systemy ASR wymagają dalszych prac, szczególnie w obszarach rozpoznawania mowy specjalistycznej oraz spontanicznej mowy potocznej, to m.in. integracja technik głębokiego uczenia w technologii ASR wpłynęła na większą dokładność systemów ASR. Poprzednie edycje badań zakładały także sprawdzenie tzw. NLU (ang. *Natural Language Understanding*), co jednak jest wypadkową samej dokładności rozpoznania mowy, które prowadzi do lepszego rozumienia języka naturalnego. Kluczowa dla tej pozytywnej korelacji i zastosowania tezy na szerszą skalę jest jednak optymalizacja systemu ASR w celu uzyskania jak najdokładniejszej transkrypcji, a nie próba optymalizacji pod kątem wybranej miary dla NLU. Naukowcy podkreślają również znaczenie uwzględnienia dziedziny zastosowania i kontekstu przy wyborze najodpowiedniejszego rozwiązania do rozpoznawania mowy. Uwypukla to potrzebę opracowywania dostosowanych rozwiązań, które spełniają określone wymagania danego klienta końcowego, takie jak zakres tematyczny i kontekst rozpoznawanej mowy, w celu osiągnięcia optymalnej wydajności.

Badanie z 2022 roku przeprowadzone przez zespół naukowców pod kierownictwem Kima, Liu i Calvo [27] dotyczyło oceny wydajności pięciu systemów ASR pod kątem jakości transkrypcji mowy o kontekście medycznym. W badaniu oceniono platformy internetowe – Google Cloud, IBM Watson, Microsoft Azure, Trint i YouTube – w zakresie ich wbudowanych systemów ASR. Zbiory danych głosowych pochodziły z 24 przeprowadzonych telekonsultacji medycznych (interakcje dwunastu studentów medycyny, każdy z dwoma symulowanymi pacjentami). Naukowcy zakładali wstępnie, że ręczne transkrypcje przewyższają te automatyczne. Wyniki ich badań potwierdziły tę hipotezę, bowiem ręczne transkrypcje okazały się dokładniejsze. Co ciekawe, spośród ocenianych systemów, system ASR YouTube przewyższył inne rozwiązania oparte na chmurze pod względem dokładności rozpoznawania słów, wykazując zaskakującą przydatność platformy w dziedzinie transkrypcji medycznej.

Pomimo korzyści z zastosowań technologii ASR w dziedzinie medycyny, wciąż istnieje szereg wyzwań, którym należy sprostać. Obejmują one potrzebę większej standaryzacji w rejestrowaniu i transkrypcji dokumentacji medycznej, a także opracowania skalowalnych i dokładniejszych systemów ASR, które będą w stanie poradzić sobie

z szeroką gamą dialektów i akcentów używanych w placówkach opieki zdrowotnej. Ponadto istnieje potrzeba opracowań w zakresie porównania działania systemów ASR, które pomogą zidentyfikować najskuteczniejsze rozwiązania dla określonych zastosowań medycznych.

3. Metodologia

Głównym celem badawczym niniejszego artykułu jest wykonanie analizy porównawczej poziomu dokładności rozpoznawania mowy wiodących na polskim rynku systemów ASR na specjalnie do tego celu wyselekcjonowanej próbce danych. Dokonując selekcji wiodących na polskim rynku systemów ASR do dalszego wykonania analizy porównawczej, wzięto pod uwagę kilka czynników. Były nimi mianowicie: dostępność systemu ASR i popularność tego narzędzia w praktyce badawczej, skalowalność, dostępność danych technicznych oraz danych wejściowych/wyjściowych realizowanych projektów. Na podstawie oceny tych cech zostały wyłonione do weryfikacji trzy systemy ASR dostępne na polskim rynku, a mianowicie:

- system Google ASR (producent Google Inc.);
- system Microsoft ASR (producent Microsoft Inc.);
- system Techmo ASR (producent Techmo sp. z o.o.).

Przygotowana na cele badania próbka danych zawiera 1000 nagrań krótkich zdań i fraz wypowiedzianych w języku polskim, które zawierają wyrażenia medyczne – zarówno określenia potoczne, jak i te zgodne z nomenklaturą medyczną. Wypowiedzi nagrane zostały przez osoby różnej płci (kobiety i mężczyźni), zróżnicowane regionalnie pod względem pochodzenia (północ i południe Polski) oraz w różnym wieku (przedział 30-50 lat).

Dla szczegółowego porównania oceny jakości rozwiązań ASR zastosowano ocenę dokładności dla poprawnie rozpoznanych słów, wskaźnik błędnie rozpoznanych słów WER, a także miarę odległości edycyjnej Levenshteina i miarę podobieństwa Jaro-Winklera. Dokładność pokazuje procent poprawnie rozpoznanych słów w danej wypowiedzi, a wskaźnik WER pokazuje procent błędnych słów w transkrypcji danej wypowiedzi. Miara odległości edycyjnej Levenshteina jest obliczana na poziomie pojedynczych znaków w słowie i wskazuje minimalną liczbę zmian wymaganych do przekształcenia dwóch łańcuchów znaków w identyczne przez wstawianie, zastępowanie lub usuwanie znaków w jednym z nich. Im niższa odległość Levenshteina, tym wyższe podobieństwo dwóch łańcuchów znaków [28]. Kolejna miara odległości między łańcuchami znakowymi, znana szerzej jako miara podobieństwa Jaro-Winklera, to z kolei metoda oparta na odległości Jaro [29], ale udoskonalona przez preferowanie identycznych znaków na początku każdego łańcucha i niebazująca na odległości związanej z edycją [30]. Im wyższe podobieństwo Jaro-Winklera, tym wyższe podobieństwo dwóch łańcuchów znaków.

Aby osiągnąć zaprezentowany cel badawczy, przed właściwą analizą porównawczą próbki nagrań wypowiedzianych języku polskim, wykonano wymienione poniżej prace.

1. Zaimplementowanie API (ang. *Application Programming Interface*) w celu umożliwienia testowania każdego systemu ASR.
2. Zebranie inicjalnego zestawu danych na bazie 19000 skategoryzowanych tekstowych wyrażeń medycznych, których właścicielem jest SOVVA S.A. (firma realizująca niniejszy projekt badawczy współfinansowany przez NCBiR). Z tego zestawu eksperci

medyczni wybrali próbkę 1000 zagadnień najczęściej poruszanych przez pacjentów podczas wizyt w gabinetach lekarskich.

- Wybraną próbkę 1000 tekstowych zagadnień najczęściej poruszanych przez pacjentów następnie poddano nagraniem, a każdy pojedynczy plik dźwiękowy zawierał nagranie pojedynczej frazy (wyrażenia lub słowa). Pliki zostały zapisane w formacie WAV o następujących parametrach: częstotliwość próbkowania: 16000 Hz, kanały: 1 (mono), format próbki: PCM, 16-bitowy. Nagrania zostały przeprowadzone w warunkach laboratoryjnych, a warunki zewnętrzne, takie jak hałas w tle lub zakłócenia, nie zostały uwzględnione.
- Otrzymane 1000 plików audio przesłano na serwer FTP. Następnie zostały one przetworzone przez każdy system ASR za pomocą wcześniej zaimplementowanego API.
- Uzyskane transkrypcje poddano następnie badaniom w celu finalnego wykonania analizy porównawczej poziomu dokładności rozpoznawania mowy badanych systemów ASR.

Jako dodatkowe badanie, dotyczące ulepszenia każdego z badanych systemów ASR, zweryfikowano możliwość dostosowania ich do indywidualnych potrzeb różnych podmiotów medycznych. Chodzi tutaj o potencjał tkwiący w dodatkowym trenowaniu swojego zbioru danych oraz w swobodnej instalacji danego rozwiązania ASR w infrastrukturze klienta końcowego (tzw. instalacja *on-premise*).

4. Rezultaty badań

Wyniki uzyskane z badania na próbce 1000 głosowych wyrażeń i fraz najczęściej wypowiedzianych przez pacjentów pokazują wysoką dokładność (powyżej 86%) rozpoznawania mowy przez badane systemy ASR, przy czym rozbieżność pomiędzy najlepszym i najgorszym wynikiem wyniosła zaledwie 1,7%. Wykonana analiza porównawcza systemów ASR została podsumowana w zamieszczonej dalej tabeli 1.

Tabela 1. Dokładność rozpoznawania mowy przez poszczególne systemy ASR

	Google ASR	Microsoft ASR	Techmo ASR
Dokładność (<i>accuracy</i>)	88,1%	86,4%	87,5%

Źródło: opracowanie własne.

Miara odległości edycyjnej Levenshteina jest to wartość wyrażona w liczbie zmian potrzebnych do ujednoczenia dwóch łańcuchów znaków, z kolei miara Jaro-Winklera stanowi stopień podobieństwa dwóch łańcuchów znaków wyrażony w przedziale od 0 do 1, wreszcie wskaźnik WER przedstawia się jako procent błędnie rozpoznanych słów. Dla ujednoczenia otrzymanych wyników i możliwości porównania silników ASR z zastosowaniem różnych miar, wszystkie trzy miary zostały na końcu przekształcone w procentowo wyrażone podobieństwo między dwoma łańcuchami znaków (wypowiedź oryginalna i tekst transkrypcji). Zgodnie z wynikiem testu Kołmogorowa-Smirnowa nie ma istotnej różnicy między rozkładami podobieństwa rozpoznania dla trzech silników w zakresie miary Levensteina, Jaro-Winklera i WER, a wszystkie wartości *p-value* są mniejsze niż 0,05. Co więcej, kwantyle rozkładów podobieństw są w większości

identyczne dla wszystkich trzech silników, co sugeruje, że wykazują podobne rezultaty w zakresie rozpoznawania terminologii medycznej. W analizie rozpoznanych fraz, miary WER i Jaro-Winklera dały identyczne wyniki i były każdorazowo wyższe od procentowo wyrażonego podobieństwa na bazie Levenshteina, ale najlepszy, średni i najgorszy wynik dla danego silnika ASR pozostał taki sam bez względu na zastosowaną miarę, co pokazano w zamieszczonej dalej tabeli 2.

Tabela 2. Podobieństwo rozpoznanych fraz przez poszczególne systemy ASR

	Levenshtein	Jaro-Winkler	WER
Google ASR	83,35%	90,83%	90,83%
Microsoft ASR	81,75%	88,37%	88,37%
Techmo ASR	82,86%	89,40%	89,40%

Źródło: opracowanie własne.

Uzyskane wyniki rozpoznawania mowy z badanych silników ASR często zawierały błędnie zapisane lub niekompletne słowa, co daje możliwość zbadania różnorodności błędów, które wystąpiły podczas transkrypcji mowy na tekst. Można je mianowicie podzielić na trzy kategorie: a) błędne rozpoznania, b) problemy z jakością, c) granice słów.

Błędne rozpoznania pojawiają się, gdy silnik ASR nie rozpoznaje wypowiedzi w całości lub z nadanym zupełnie innym znaczeniem. Przykłady wykrytych błędnych rozpoznań, w których silnik ASR nie rozpoznał poprawnie frazy, co uniemożliwia dalszą optymalizację przetwarzania:

- w ogóle nierozpoznane: „omdlenia”, „ścieńczenia”, „pęcherzyca”;
- rozpoznane w innym znaczeniu:
 - oryginał „leukoplakia” rozpoznany jako „lenko platja”,
 - oryginał „problem z ukrwieniem oczu” rozpoznany jako „problem z ukrwienie moczu”,
 - oryginał „wściekłość” rozpoznany jako „zaległość”.

Kolejna kategoria to problemy z jakością rozpoznania, które pojawiają się, gdy silnik ASR rozpoznaje frazę z niewielkimi rozbieżnościami, co jednakże nie zmienia znaczenia rozpoznanej frazy. Przykłady wykrytych problemów z jakością:

- oryginał „zubożona mowa” rozpoznany jako „zburzona mowa”;
- oryginał „żwir w woreczku żółciowym” rozpoznany jako „żwir po woreczku żółciowym”;
- oryginał „wrażenie że nie słyszę” rozpoznany jako „teraz wrażenie że nie słyszę”;
- oryginał „złogi w pęcherzyku żółciowym” rozpoznany jako „złogi pęcherzyku żółciowym”.

Wreszcie ostatni problem stanowią granice słów, w przypadku których silniki ASR poprawnie rozpoznały frazę, ale niektóre słowa nie mają poprawnie rozpoznanych granic, co można stosunkowo łatwo zoptymalizować w dalszym przetwarzaniu. Przykłady wykrytych problemów z granicami słów:

- oryginał „stres pourazowy” rozpoznany jako „stres po urazowy”;
- oryginał „problemy z chodzeniem” rozpoznany jako „problemy schodzeniem”;
- oryginał „śpię w dzień” rozpoznany jako „śpiew dzień”.

W zamieszczonej dalej tabeli 3 przedstawiono zestawienie procentowe wymienionych problemów w rozpoznawaniu mowy napotkanych dla każdego silnika ASR.

Tabela 3. Procent problemów ASR dla każdego silnika

	błędne rozpoznania	problemy z jakością	granice słów
Google ASR	73,2%	23,6%	3,2%
Microsoft ASR	71,2%	14,8%	14,0%
Techmo ASR	61,7%	26,6%	11,7%

Źródło: opracowanie własne.

Według przedstawionych danych, Techmo ASR miało najniższy odsetek błędnych rozpoznań w porównaniu z Microsoft ASR i Google ASR: Techmo ASR wykazywał 61,7% błędnych rozpoznań wobec 71,2% dla Microsoft ASR i 73,2% dla Google ASR. Jednocześnie Microsoft ASR miał najniższy odsetek problemów z jakością w porównaniu z Google ASR i Techmo ASR: Microsoft ASR wykazywał 14,8% problemów z jakością wobec 23,6% dla Google ASR i 26,6% dla Techmo ASR. Z kolei Google ASR miał najniższy odsetek błędnie rozpoznanych granic słów w porównaniu z Microsoft ASR i Techmo ASR: Google ASR wykazywał 3,2% granic słów wobec 14,0% dla Microsoft ASR i 11,7% dla Techmo ASR. Można zauważyć, że Techmo ASR wypadło lepiej pod względem stosunku błędnych rozpoznań do pozostałych problemów ASR (problemów z jakością i granic słów), które można zoptymalizować w dalszym przetwarzaniu.

Dodatkowo odkryto, że część wykrytych problemów i nieścisłości była spowodowana wielkością wejściowego zestawu danych. Aby rozwiązać ten problem, wykorzystano redukcję do rdzenia oraz bazę danych dotyczących synonimów. Pozwoliło to zmniejszyć pewne rozbieżności gramatyczne w danych wejściowych, które miały wpływ na uzyskane wyniki. Ponadto wejściowy zbiór danych był niewielkich rozmiarów, gdyż obejmował jedynie 1000 fraz. W związku z tym nie można było wykryć innych problemów, które mogą być potencjalnie wykazywane przez każdy z trzech badanych silników automatycznego rozpoznawania mowy.

Warto również zwrócić uwagę na usterki rozpoznawania mowy na poziomie poszczególnych słów. Wykonane badania pozwoliły zidentyfikować zestaw słów mylonych przez wszystkie trzy badane silniki ASR. Były to mianowicie: „święd”, „ból”, „krosty”, „krwiaki”, „sinienie”, „płaczliwość”, „leki”, „zauszne”, „białka”, „płonica”, „omamy” i „urojenia”. Ponadto większość słów z powyższego zestawu zidentyfikowano jako błędy najczęściej występujące w całym zbiorze. Dziesięć najczęściej mylonych słów to kolejno: „święd”, „krosty”, „ból”, „poronienie”, „problem z”, „białka”, „kłykciny”, „urojenia”, „omamy” oraz „płonica”.

Z kolei inny zestaw słów został błędnie rozpoznany przez dwa silniki ASR jednocześnie (w różnych parach). Chodzi tutaj o następujące słowa: „poronienie”, „problem z”, „plamy”, „słyszę”, „uraz”, „kłykciny”, „utrata”, „przerosła”, „poparzeniowa”, „rozpadowe”, „rozpadająca”, „poty”, „wściekłość”, „punkcja”, „łuski”, „duże”, „przymglone”, „rogowiec”, „kolka”, „przeczos”, „kolki”, „płacz”, „krosta”, „strup”, „kiła”, „ostre”, „wady”, „polipy”, „sine”, „ptsd”, „usnąć”, „dreszcze”, „dokrwiiona”, „trudność”, „przestaw”, „oziębienie”, „stres”, „schizofrenia”.

Jako dodatkowy aspekt analizy porównawczej, zbadano też potencjał ulepszenia każdego z badanych systemów ASR poprzez weryfikację możliwości dodatkowego trenowania własnego zbioru danych i swobodnej instalacji *on-premise* danego rozwiązania ASR u klienta końcowego. Zbadano te możliwości w oparciu na kompleksowej analizie dokumentacji i próbach eksperymentalnych. Wyniki, podsumowane w tabeli 4, dostarczają cennych informacji na temat technicznych możliwości i ograniczeń tych rozwiązań i mogą przyczynić się do rozwoju bardziej wydajnych technologii ASR dla różnych zastosowań.

Tabela 4. Dodatkowe aspekty technicznych możliwości systemu ASR u klienta końcowego

	Google ASR	Microsoft ASR	Techmo ASR
Możliwość dodatkowego trenowania własnego zbioru danych	Nie	Tak	Tak (dodatkowo płatna usługa Techmo)
Możliwość swobodnej instalacji <i>on-premise</i> u klienta końcowego	Nie	Nie	Tak

Źródło: opracowanie własne.

Z tabeli 4 wynika, że tylko rozwiązanie Techmo ASR oferuje możliwość dodatkowego trenowania własnego zbioru danych oraz swobodnej instalacji *on-premise*. Chociaż Microsoft ASR umożliwia dodatkowe trenowanie własnego zbioru danych, nie można go jednak zainstalować lokalnie. Z drugiej strony Google ASR nie spełnia żadnego z obu aspektów technicznych. Na bazie dodatkowych eksperymentów przeprowadzonych z oprogramowaniem silników ASR zauważono również, że wykrywanie określonych fraz w systemie Google ASR można usprawnić przez kontrolę wbudowanych predefiniowanych słowników, które zapewniają lepsze rozpoznanie wypowiedzianych fraz. Natomiast Microsoft ASR nie obsługuje wcale takiej funkcjonalności, z kolei Techmo ASR nie udostępnia informacji na temat możliwości w tym zakresie.

5. Podsumowanie

W artykule przedstawiono analizę porównawczą dostępnych na polskim rynku silników Google ASR, Microsoft ASR i Techmo ASR, pod kątem oceny poziomu ich dokładności w rozpoznawaniu potocznej mowy polskiej związanej z tematyką medyczną. Na podstawie wyników badania stwierdzono, że poziomy dokładności systemów były bardzo podobne i przekraczały 86% dokładności rozpoznania dla wszystkich trzech silników ASR. Natomiast zbadana możliwość ulepszenia danego systemu ASR przez dodatkowe trenowanie własnego zbioru danych oraz instalację *on-premise* okazały się czynnikami do dodatkowego rozważenia przez klienta końcowego jako ważne z perspektywy posiadania kontroli w zastosowaniu technologii w podmiotach medycznych. Otrzymane wyniki wypełniają lukę badawczą w zakresie analiz porównawczych silników ASR oraz podkreślają znaczenie starannego rozważenia zarówno czynników technicznych, jak i praktycznych przy wyborze rozwiązania ASR do konkretnego zastosowania, wskazując zarówno Microsoft ASR, jak i Techmo ASR jako optymalne rozwiązanie do transkrypcji wywiadów medycznych prowadzonych w języku polskim.

Przedstawione badania mogą pomóc ukierunkować projektowanie systemów konwersacyjnych wyspecjalizowanych dziedzinowo i interfejsów głosowych zdolnych do

rozpoznawania potocznej mowy polskiej w celu przeprowadzania wstępnej diagnostyki i wywiadu lekarskiego przy minimalnej ingerencji człowieka, co jest istotną zaletą w czasach epidemii i/lub przy ograniczonych zasobach organizacyjnych.

Finansowanie badania

Praca współfinansowana przez Narodowe Centrum Badań i Rozwoju (NCBiR) przedstawia częściowe wyniki projektu badawczego pt. „Opracowanie systemu wykorzystującego uczenie maszynowe do automatycznego przeprowadzania wstępnej diagnostyki i wywiadów lekarskich w jednostkach służby zdrowia” realizowanego przez SOVVA S.A.

Literatura

1. Google ASR, <https://cloud.google.com/speech-to-text> [data dostępu: 04.2023].
2. Microsoft ASR, <https://azure.microsoft.com/en-us/products/cognitive-services/speech-to-text> [data dostępu: 04.2023].
3. Dragon STT, <https://www.nuance.com/en-gb/dragon.html> [data dostępu: 04.2023].
4. Phonexia STT, <https://www.phonexia.com/product/speech-to-text/> [data dostępu: 04.2023].
5. Techmo ASR, <https://techmo.pl/technologie/rozpoznawanie-mowy> [data dostępu: 04.2023].
6. Rev AI ASR, <https://www.rev.ai/> [data dostępu: 04.2023].
7. Iancu B., *Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources*, Informatica Economica, 23(1), 2019, s. 17-25.
8. Scholz M.L., Collatz-Christensen H., Blomberg S.N.F., Boebel S., Verhoeven J., Kraft T., *Artificial intelligence in Emergency Medical Services dispatching: assessing the potential impact of an automatic speech recognition software on stroke detection taking the Capital Region of Denmark as case in point*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 30(36), 2022, s. 1-17.
9. Tomar V., Desruisseaux M., Seetzen H., *System and method for implementing a vocal user interface by combining a speech to text system and a speech to intent system*, 2020, <https://patentimages.storage.googleapis.com/1d/05/1d/014c820a9a7b7b/US10878807.pdf> [data dostępu: 02.2023].
10. Tomar V., *How Speech Technology Is Optimizing Factory Lines*, 2021, <https://industrytoday.com/how-speech-technology-is-optimizing-factory-lines/> [data dostępu: 02.2023].
11. Vinnarasu A., Jose D.V., *Speech to text conversion and summarization for effective understanding and documentation*, International Journal of Electrical and Computer Engineering (IJECE), 9(5), 2019, s. 3642-3648.
12. Wahyutama A.B., Hwang M., *Auto-Scoring Feature Based on Sentence Transformer Similarity Check with Korean Sentences Spoken by Foreigners*, Applied Sciences, 13, 373(1), MDPI 2023, s. 1-16.
13. Kim J.Y., Liu C., Calvo R.A., McCabe K., Taylor S.C.R., Schuller B.W., Wu K., *A Comparison of Online Automatic Speech Recognition Systems and the Nonverbal Responses to Unintelligible Speech*, Preprint arXiv, Computing Research Repository (CoRR), doi.org/10.48550/arXiv.1904.12403, 2019, s. 1-13.
14. Kuligowska K., Kisielewicz P., Włodarz A., *Wady i ograniczenia systemów rozpoznawania mowy*, Roczniki Kolegium Analiz Ekonomicznych, nr 49/2018, Szkoła Główna Handlowa, Warszawa 2018, s. 307-317.
15. Lugosch L., Ravanelli M., Ignoto P., Tomar V., Bengio Y., *Speech Model Pre-training for End-to-End Spoken Language Understanding*, Preprint arXiv, Audio and Speech Processing (eess.AS), 2019, doi.org/10.48550/arXiv.1904.03670, s. 1-5.

16. Mah P.M., Skalna I., Muzam J., *Natural Language Processing and Artificial Intelligence for Enterprise Management in the Era of Industry 4.0*, Applied Sciences, 12(18), 9207, 2020, s. 1-26.
17. Zue V.W., *The use of speech knowledge in automatic speech recognition*, Proceedings of the IEEE, 1985, 73(11), s. 1602-1615.
18. Repka A., *Chatboty w służbie e-zdrowia – ewolucja telemedycyny w stronę konwersacyjnej sztucznej inteligencji*, [w:] Kuligowska K. (red.), *Chatboty w informatyce ekonomicznej: implementacja, miary, zastosowania*, Laboratorium Wiedzy Artur Borcuch, Kielce 2022, s. 104.
19. Altar H.S., Sison R.C., *Medical Transcriptionist's Experience with Speech Recognition Technology*, Proceedings of Australasian Conference on Information Systems (ACIS 2019), Perth, Western Australia 2019, s. 915- 924.
20. Joseph J., Moore Z.E.H., Patton D., O'Connor T., Nugent L.E., *The impact of implementing speech recognition technology on the accuracy and efficiency (time to complete) clinical documentation by nurses: A systematic review*, Journal of Clinical Nursing, 29(13-14), 2020, s. 2125-2137.
21. Saxena K., Diamond R., Conant R.F., Mitchell T.H., Gallopyn G., Yakimow K.E., *Provider Adoption of Speech Recognition and its Impact on Satisfaction, Documentation Quality, Efficiency, and Cost in an Inpatient EHR*, AMIA Joint Summits on Translational Science proceedings. 2017 AMIA Joint Summits on Translational Science, 2018, s. 186-195.
22. Fareez F., Parikh T., Wavell C., Shahab S., Chevalier M., Good S., De Blasi I., Rhouma R., McMahon C., Lam J.P., Lo T., Smith C.W., *A dataset of simulated patient-physician medical interviews with a focus on respiratory cases*, Scientific Data, 9(313), 2022, s. 1-7.
23. Kim D., Oh J., Heeju Im H., Yoon M., Park J., Lee J., *Automatic Classification of the Korean Triage Acuity Scale in Simulated Emergency Rooms Using Speech Recognition and Natural Language Processing: a Proof of Concept Study*, Journal of Korean Medical Science, 36(27), 2021, s. 1-13.
24. Georgila K., Leuski A., Yanov V., Traum D., *Evaluation of Off-the-shelf Speech Recognizers Across Diverse Dialogue Domains*, Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France 2020, s. 6469-6476.
25. Yao X., Bhutada P., Georgila K., Sagae K., Artstein R., Traum D., *Practical evaluation of speech recognizers for virtual human dialogue systems*, Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta 2010, s. 1597-1602.
26. Morbini F., Audhkhasi K., Sagae K., Artstein R., Can D., Georgiou P., Narayanan S., Leuski A., Traum D., *Which ASR should I choose for my dialogue system?*, Proceedings of the 14th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2013), Metz, France 2013, s. 394-403.
27. Kim J.Y., Liu C., Calvo R.A., McCabe K., Taylor S.C.R., Schuller B.W., Wu K., *Comparison of Automatic Speech Recognition Systems*, [w:] Stoyanchev S., Ultes S., Li H. (red.), *Conversational AI for Natural Human-Centric Interaction, Lecture Notes in Electrical Engineering*, Springer, 943, 2022, s. 123-131.
28. Niewiarowski A., Stanuszek M., *Mechanizm analizy podobieństwa krótkich fragmentów tekstów, na bazie odległości Levenshteina*, Studia Informatica, 34, 1(110), Politechnika Krakowska, Instytut Informatyki, 2013, s. 108.
29. Cohen W.W., Ravikumar P., Fienberg S.E., *A Comparison of String Metrics for Matching Names and Records*, KDD Workshop on Data Cleaning and Object Consolidation, Carnegie Mellon University, 2003, s. 1-6.
30. Kamińska A.M., *Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych*, Przegląd Biblioteczny, 4, 2017, s. 477-495.

Zastosowanie automatycznego rozpoznawania mowy w transkrypcji wywiadów medycznych – porównanie silników ASR dla języka polskiego

Streszczenie

W dążeniu do bardziej wydajnej i skoncentrowanej na pacjencie opieki zdrowotnej kluczowym elementem wspierającym podmioty lecznicze stają się automatyczne systemy rozpoznawania mowy. By jednak uznać je za użyteczne, systemy te muszą spełniać określone wymagania podyktowane realiami rynkowymi. Dlatego też celem niniejszego artykułu była analiza porównawcza wiodących na polskim rynku silników ASR, wykonana na zbiorze nagranych w języku polskim fraz najczęściej wypowiedzianych przez pacjentów podczas wizyt w gabinetach lekarskich. Wyniki naszej analizy wykazały, że między badanymi silnikami ASR istnieją niewielkie różnice w dokładności rozpoznawania mowy. Mimo to wszystkie prezentowały specyficzne problemy, które zostały podzielone na trzy grupy: błędne rozpoznania, problemy z jakością i granice słów. Wyniki badań dostarczają cennych informacji szerokiemu gronu interesariuszy, ułatwiając rozwój rozwiązań do rozpoznawania mowy polskiej dla specyficznych potrzeb sektora medycznego.

Słowa kluczowe: automatyczne rozpoznawanie mowy, *speech-to-text*, silniki ASR dla języka polskiego, transkrypcja wywiadów medycznych

Application of automatic speech recognition in the transcription of medical interviews – comparison of ASR engines for the Polish language

Abstract

Automatic speech recognition systems are becoming a key element supporting healthcare entities in pursuing a more efficient and patient-centred healthcare system. However, these systems must meet certain requirements dictated by market reality to be considered beneficial. Therefore, the purpose of this paper was a comparative analysis of the leading ASR engines on the Polish market, performed on a set of phrases recorded in Polish covering the most frequently spoken utterances by patients during visits in doctors' offices. The results of our analysis showed that there are slight differences in speech recognition accuracy among the tested ASR engines. Despite this, they all presented specific problems, divided into three groups: misidentifications, quality problems and word boundaries. The research results provide valuable information to a wide range of stakeholders, facilitating the development of Polish speech recognition solutions for the specific needs of the medical sector.

Keywords: automatic speech recognition, *speech-to-text*, ASR engines for the Polish language, medical interview transcript

Zastosowanie ekstrakcji cieczonej w technologii odzysku i rozdziału metali z grupy platynowców

1. Wprowadzenie

Metale szlachetne, do których zalicza się również grupę sześciu platynowców, są szczególnie pożądanymi metalami o wysokiej wartości rynkowej. Znalazły zastosowanie jako katalizatory w przemyśle samochodowym, chemicznym i naftowym, a także jako przewodniki w przemyśle elektrycznym i elektronicznym; innymi równie ważnymi sektorami wykorzystującymi metale szlachetne jest rynek jubilerski, lotnictwo oraz medycyna, stomatologia i przemysł farmaceutyczny. Szeroka gama możliwych zastosowań metali szlachetnych wynika z ich unikalnych właściwości, takich jak wysoka odporność chemiczna i termiczna, bardzo dobra przewodność cieplna i elektryczna czy odporność na korozję i utlenianie. Wyjątkowość metali szlachetnych pod względem właściwości fizykochemicznych powoduje, iż zastąpienie ich w niektórych gałęziach przemysłu jest wręcz niemożliwe [1-6].

Co więcej, wraz z rozwojem gospodarczym zapotrzebowanie na metale szlachetne sukcesywnie wzrasta, czemu towarzyszy stopniowe ubożenie naturalnych źródeł metali szlachetnych, a w konsekwencji poszukiwanie nowych dróg pozyskania tych cennych metali. Wobec tego szczególnie istotnym zagadnieniem staje się możliwość odzysku metali szlachetnych na drodze przerobu źródeł wtórnych [1-6].

Niniejszy przegląd skupia najważniejsze aspekty dotyczące charakterystyki platynowców m.in. ich właściwości, występowania, przeróbki surowców naturalnych i wtórnych oraz roli ekstrakcji typu ciec-ciecz w selektywnym rozdziale metali z tej grupy, nie tylko w skali laboratoryjnej, ale przede wszystkim przemysłowej. Ponadto w niniejszym opracowaniu zebrane zostały informacje na temat najczęściej stosowanych, dostępnych komercyjnie ekstrahentów, w oparciu o aktualne doniesienia, zaczerpnięte z literatury naukowej.

¹ karolina.pianowska@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

² joanna.kluczka@polsl.pl, Katedra Chemii Nieorganicznej, Analitycznej i Elektrochemii, Wydział Chemiczny, Politechnika Śląska, www.polsl.pl.

³ grzegorz.benke@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

⁴ karolina.goc@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

⁵ joanna.malarz@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

⁶ michal.ochmanski@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

⁷ katarzyna.leszczyńska-sejda@imn.lukasiewicz.gov.pl, Centrum Hydroelektrometalurgii, Sieć Badawcza Łukasiewicz – Instytut Metali Nieżelaznych, www.imn.gliwice.pl.

2. Platynowce

Platynowce (PGM, ang. *Platinum Group Metals*) stanowią wspólną grupę metali usytuowaną w 8-10 grupie układu okresowego, którą ze względu na różnice gęstości dzielimy na triadę platynowców lekkich, tj. ruten, rod i pallad, oraz platynowców ciężkich – osm, iryd i platyna. Pod względem fizycznym są to szarobiałe metale (z wyjątkiem osmu) o relatywnie dużej twardości i kruchości. Cechą szczególną platynowców, wyróżniającą je na tle innych metali nieszlachetnych, jest ich wysoka odporność chemiczna i temperaturowa, a także właściwości katalityczne. Dzięki swoim wyjątkowym właściwościom fizykochemicznym platynowce znalazły zastosowanie w wielu sektorach przemysłowych, gdzie szczególnie dużą rolę PGM pełnią w katalizie. Od wielu lat największym odbiorcą platynowców jest przemysł motoryzacyjny, który wykorzystuje je do produkcji katalizatorów samochodowych, zapobiegających szkodliwej emisji tlenków węgla, azotu oraz węglowodorów [1-6].

W przemyśle chemicznym i naftowym znalazły zastosowanie m.in. w procesach, takich jak redukcja, reformacja, uwodornienie, izomeryzacja czy konwersja [1-6].

Pomimo licznych zastosowań i wysokiego zapotrzebowania rynkowego, dostęp do naturalnych zasobów platynowców jest ograniczony, co wynika z niewielkiego rozpowszechnienia tych metali w skorupie ziemskiej [1-6].

Niewielka dostępność platynowców, jak również wyjątkowość właściwości oraz mnogość ich zastosowań wpływa z kolei na wysoką cenę tych metali; według raportów Johnson Matthey ceny kruszców w styczniu 2023 roku sięgnęły 1 107,00 \$/oz dla Pt; 1 855 \$/oz – Pd; 12 400,00 \$/oz – Rh; 4 800,00 \$/oz – Ir oraz 475,00 \$/oz – Ru [7].

2.1. Źródła występowania

Obecnie wiodącym producentem platynowców jest Afryka Południowa, gdzie usytuowana jest większość światowych zasobów platynowców, szacowana w 2020 roku na około 63 000 ton. Kolejne lokalizacje bogate w metale z tej grupy to Rosja, Zimbabwe, Stany Zjednoczone i Kanada [1-3, 8]. Forma występowania platynowców w dużej mierze uzależniona jest od położenia geograficznego – platynowce w formie rodzimej występują głównie w Afryce Południowej, górach Ural oraz Kolumbii; często towarzyszą również siarczkowym i arsenowym rudom miedzi i niklu [1, 3].

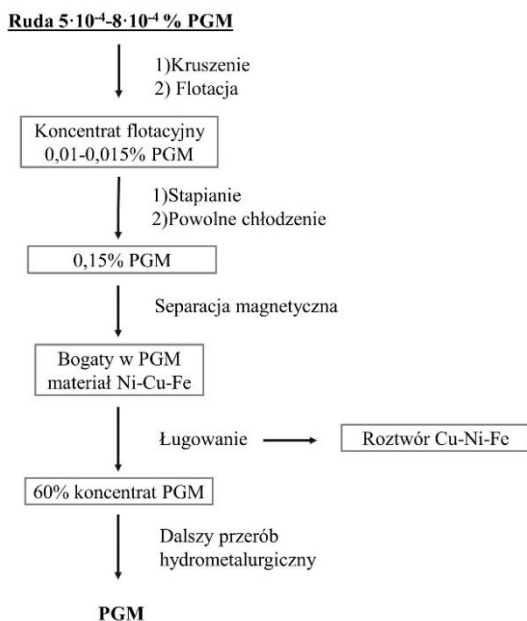
Stężenie PGM w rudach jest bardzo zróżnicowane, w najbogatszych południowoafrykańskich koncentracja jest mniejsza niż 10 g na tonę rudy, przez co wymagany jest proces wzbogacania obejmujący szereg procesów fizycznych, piro- i hydrometalurgicznych.

Występowanie naturalnych złóż platynowców w Polsce jest znikome. Badania geologiczne wskazują, iż pewne ilości metali szlachetnych, tj. Ag i Au oraz w mniejszym stopniu Pt i Pd, związane są z utworami utlenionymi cechsztyńskiej serii miedziowości [9, 10]. Z tego względu jednym ze znaczących w skali światowej producentów metali szlachetnych w Polsce jest KGHM Polska Miedź S.A. Znany na rynku krajowym i zagranicznym producent miedzi jest jednocześnie jednym z największych producentów srebra na świecie, które pozyskuje w procesie przeróbki szlamu anodowego powstającego w hutach KGHM w procesie elektrolitycznej rafinacji miedzi. Firma zajmuje się również produkcją złota, które z kolei otrzymuje się w wyniku przeróbki szlamu anodowego pozostałego po elektrolitycznej rafinacji srebra oraz w mniejszej ilości Pd i Pt w formie koncentratu powstającego w procesie produkcji złota [11].

Rozwój gospodarczy oraz wzrastające zapotrzebowanie na metale szlachetne skutkuje stopniowym ubożeniem naturalnych zasobów platynowców, a w konsekwencji poszukiwaniem nowych dróg pozyskania tych cennych metali. Wobec tego szczególnie istotnym zagadnieniem staje się możliwość odzysku metali szlachetnych na drodze przerobu źródeł wtórnych, takich jak zużyte katalizatory samochodowe oraz przemysłowe, elementy elektroniczne i komputerowe czy turbiny lotnicze.

2.1.1. Surowce naturalne

Pozyskiwanie platynowców z wykorzystaniem źródeł pierwotnych obejmuje wydobywanie rud konwencjonalnymi technikami podziemnymi lub odkrywkowymi. Następnie materiał jest rozdrabniany poprzez kruszenie i mielenie. Ze względu na niską koncentrację platynowców w zasobach naturalnych – ok. 3-4 g/t PGM – wstępnie przygotowana ruda kierowana jest do dalszych etapów obróbki celem wytworzenia koncentratów. Wzbogacanie materiału zwykle odbywa się z wykorzystaniem flotacji, separacji grawitacyjnej oraz metod pirometalurgicznych, tj. wytopiania i konwersji. Następnie materiał kierowany jest do obróbki hydrometalurgicznej w celu usunięcia zanieczyszczeń, takich jak Fe, Co, Ni i Cu. W efekcie otrzymywany jest bogaty koncentrat PGM, który przeznaczony jest do dalszej rafinacji i odzysku poszczególnych platynowców (rys. 1) [1, 2, 4, 5].



Rysunek 1. Schemat wytwarzania koncentratu na drodze przerobu rudy PGM.
Źródło: opracowano na podstawie [5]

2.1.2. Surowce wtórne

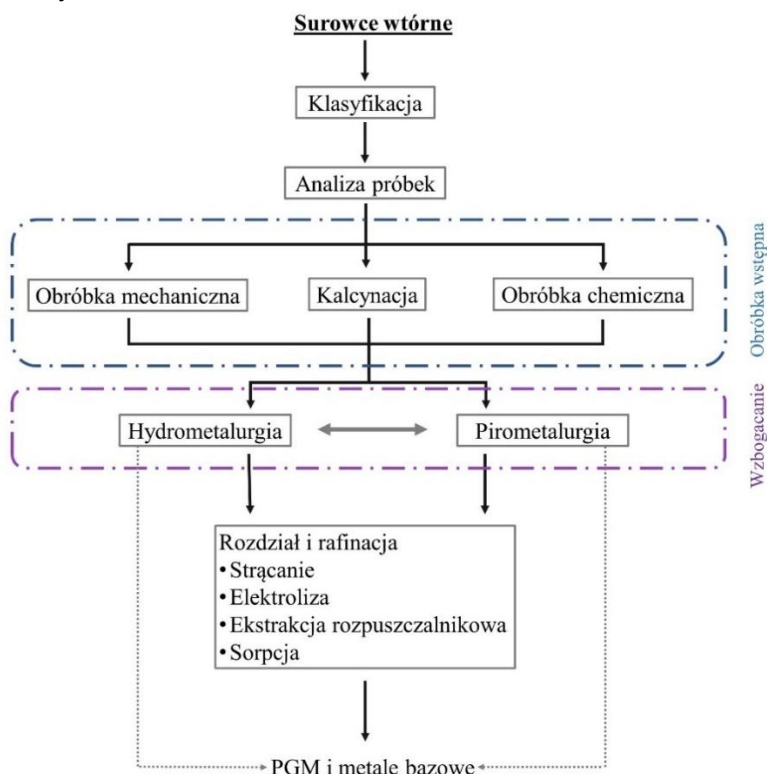
Przerób materiałów odpadowych, podobnie jak w przypadku źródeł pierwotnych, wykorzystuje zwykle zarówno procesy pirometalurgiczne, jak i hydrometalurgiczne, przy czym czynnikiem determinującym jest skład chemiczny i rodzaj zastosowanego

surowca. Ponieważ jednymi z najważniejszych i najbardziej wartościowych źródeł wtórnych (pod względem zawartości PGM) są zużyte katalizatory samochodowe i przemysłowe, to właśnie ich recyklingowi w ostatnich latach poświęcona była uwaga badaczy. Co więcej, wiele korporacji – w tym Umicore, BASF, Johnson Matthey, Nippon PGM – opracowały już stosowane komercyjnie procesy odzysku platynowców ze źródeł wtórnych [1, 12-14].

Pierwszy etap przerobu obejmuje zwykle kruszenie, mielenie i porcjowanie materiału wsadowego, a następnie jego stapianie w obecności topnika, reduktora i kolektora – zwykle Cu, Ni, Pb lub Fe – lub w przypadku metody hydrometalurgicznej – wstępne ługowanie. Stosowana obróbka wstępna ma na celu nie tylko zażęcenie platynowców, ale również usunięcie pozostałych po procesach katalitycznych substancji organicznych lub elementów nośnika ograniczających skuteczność wymywania PGM [1, 12-14].

Dalsza przeróbka polega zwykle na hydrometalurgicznym odzyskiwaniu platynowców poprzez ich rozpuszczanie w kwaśnych, rzadziej alkalicznych, mediach w obecności utleniaczy, przy czym najpowszechniejszym czynnikiem ługującym jest kwas chlorowodorowy z dodatkiem kwasu azotowego(V), chloru bądź nadtlenu wodoru jako utleniacza. Z uzyskanego w ten sposób roztworu platynowce odzyskiwane są metodą cementowania, ekstrakcji rozpuszczalnikowej lub wymiany jonowej, a następnie poddawane rafinacji celem otrzymania czystych metali bądź ich związków [1, 12-14].

Rysunek 2 przedstawia ogólny schemat odzysku PGM na drodze przerobu materiałów odpadowych.



Rysunek 2. Schemat odzysku PGM na drodze przerobu surowców wtórnych.

Źródło: opracowano na podstawie [13]

Niezależnie od powszechnego zastosowania procesów pirometalurgicznych w technologiach odzysku PGM ze źródeł pierwotnych i wtórnych, technika ta obarczona jest wieloma wadami, w tym wysokim zużyciem energii czy wytwarzaniem dużych ilości lotnych i toksycznych produktów spalania. Metody hydrometalurgiczne nie wymagają tak wysokich temperatur, zapewniają stosunkowo niskie koszty, zarówno w małej, jak i dużej skali, przy jednoczesnym zachowaniu wysokiej skuteczności oraz łatwej kontroli procesu. Ponadto odpady ciekłe wytwarzane na drodze hydrometalurgicznej mogą być oczyszczane i zawracane do dalszych procesów [1, 12].

Co więcej, rozdział metali szlachetnych oraz otrzymanie ich w stanie czystym możliwy jest tylko z zastosowaniem metod hydrometalurgicznych, podczas gdy procesy pirometalurgiczne znajdują zastosowanie jedynie na etapie obróbki wstępnej.

2.2. Chlorokompleksy platynowców

Jak już wcześniej wspomniano, hydrometalurgiczna obróbka naturalnych bądź wtórnych źródeł metali szlachetnych prowadzi zwykle do otrzymania kwaśnych roztworów chlorkowych, w których platynowce przyjmują formę chlorokompleksów.

Kwas chlorowodorowy stanowi bardzo opłacalny pod względem ekonomicznym czynnik ługujący. Co więcej, chemia chlorokompleksów platynowców jest bardzo dobrze poznana, co z kolei wykorzystywane jest na dalszych etapach rozdziału i rafinacji.

Ponieważ metale szlachetne tworzą szeroką gamę chlorokompleksów, kluczowymi czynnikami odpowiedzialnymi za formę występowania platynowca w mediach chlorkowych są pH, potencjał redoks, stężenie chlorków oraz tzw. efekt starzenia roztworu, odpowiadający za tworzenie się akwakompleksów w roztworach o niskiej kwasowości, przy niewielkim stężeniu chlorków. Badania wskazują, że rodzaj chlorokompleksu PGM odgrywa istotną rolę w procesach selektywnej ekstrakcji bądź wymiany jonowej, co związane jest z różną geometrią i gęstością ładunku kompleksów PGM, a w konsekwencji różną reaktywnością i kinetyką procesów zachodzących w roztworach wodnych. Przykładowo w przypadku procesów zachodzących według mechanizmu jonowymiennego ogólna tendencja tworzenia par jonowych z cząsteczkami ekstrahentów zachodzi zgodnie z szeregiem $[MCl_6]^{2-} > [MCl_4]^{2-} \gg [MCl_6]^{3-} >$ akwakompleksy. Efekt ten wynika głównie z różnic gęstości ładunku poszczególnych form chlorokompleksów, a w konsekwencji skłonności do ulegania hydratacji; chlorokompleksy o dużej gęstości ładunku łatwiej przyciągają cząsteczki wody, tworząc grubą powłokę hydratacyjną, co z kolei negatywnie wpływa na siłę przyciągania kulombowskiego i tworzenie par jonowych z odpowiednimi przeciwjonami. Co więcej, badania wskazują również, iż niska temperatura oraz długi czas przechowywania roztworów sprzyja powstawaniu obojętnych kinetycznie, uwodnionych form chlorokompleksów [5, 15, 16].

W tabeli 1 zamieszczono zestawienie różnych form występowania kompleksów platynowców, przy czym należy pamiętać, że w przypadku rzeczywistych roztworów chlorkowych, platynowce występują głównie w postaci mieszaniny różnych chlorokompleksów oraz akwachlorokompleksów [5].

Tabela 1. Chlorokompleksy platynowców

Ru ³⁺	[RuCl ₆] ³⁻ [RuCl ₅ (H ₂ O)] ²⁻ [RuCl ₄ (H ₂ O) ₂] ⁻ [RuCl ₃ (H ₂ O) ₃]	Rh ³⁺	[RhCl ₆] ³⁻ [RhCl ₅ (H ₂ O)] ²⁻	Pd ²⁺	[PdCl ₄] ²⁻
Ru ⁴⁺	[RuCl ₆] ²⁻ [Ru ₂ OCl ₁₀] ⁴⁻ [Ru ₂ OCl ₈ (H ₂ O) ₂] ²⁻	Rh ⁴⁺	[RhCl ₄ (H ₂ O) ₂] [RhCl ₆] ²⁻	Pd ⁴⁺	[PdCl ₆] ²⁻
		Ir ³⁺	[IrCl ₆] ³⁻ [IrCl ₅ (H ₂ O)] ²⁻	Pt ²⁺	[PtCl ₄] ²⁻
Os ⁴⁺	[OsCl ₆] ²⁻	Ir ⁴⁺	[IrCl ₄ (H ₂ O) ₂] [IrCl ₆] ²⁻	Pt ⁴⁺	[PtCl ₆] ²⁻

Źródło: opracowano na podstawie [5].

3. Przemysłowe metody rozdzielu platynowców

Jak już wspomniano, platynowce to grupa metali, które w stanie czystym można otrzymać wyłącznie metodami hydrometalurgicznymi, takimi jak metody strąceniowe, ekstrakcja rozpuszczalnikowa czy wymiana jonowa. Spośród wymienionych, najstarsze są metody strąceniowe, stosowane w światowych rafineriach do połowy lat 70. Separacja poszczególnych metali szlachetnych odbywała się za pomocą serii reakcji strącania. Duża ilość zanieczyszczeń i przeszkadzające reakcje współstrącania wymuszały jednak stosowanie wielu operacji rafinacji, co sprawiało, że proces był bardzo żmudny i czasochłonny [5, 6].

W późniejszych latach szczególne znaczenie zyskała ekstrakcja rozpuszczalnikowa [5, 6]. Zalety metod ekstrakcyjnych, takie jak duża szybkość reakcji, wysoka selektywność i relatywnie niskie ceny reagentów, spowodowały, iż technika ta z powodzeniem jest stosowana w rafineriach metali szlachetnych na całym świecie, m.in. w rafinerii Anglo Platinum Corporation, Rand Refinery w RPA, Johnson Matthey w Anglii. Ponadto znanych jest wiele skutecznych ekstrahentów platynowców, takich jak alkilowe aminy pierwszo-, drugo- i trzeciorzędowe, sulfotlenki alkilowe, związki fosforoorganiczne (zarówno chelatujące, jak i solwatujące) czy pochodne pirydyny [5, 6, 17-21].

Większość dostępnych schematów technologicznych zakłada ługowanie materiałów wsadowych w stężonym kwasie solnym z dodatkiem gazowego chloru jako utleniacza. Pierwszym etapem rozdzielu, wykorzystującym ekstrakcję rozpuszczalnikową, jest zwykle ekstrakcja złota, a następnie palladu bądź platyny. Pozostałe PGM odzyskiwane są na początku lub na końcu całego procesu. W każdym przypadku rozdział prowadzi do otrzymywania czystych roztworów odpowiedniego metalu, który poddaje się redukcji w celu przeprowadzenia go w postać czystego proszku bądź gąbki. Finalne produkty uzyskuje się zwykle poprzez stopienie, w celu uzyskania wlewków, granulek czy prętów [1, 5, 6]. Zgodnie z doniesieniami literaturowymi ekstrakcja rozpuszczalnikowa jako metoda odzysku złota z roztworów chlorkowych jako pierwsza została zastosowana przez International Nickel Company (INCO) w Wielkiej Brytanii, gdzie jako czynnik ekstrakcyjny zastosowano dibutylokarbital [6].

Kolejnym przykładem może być rafineria Anglo Platinum w RPA, która odzyskuje złoto za pomocą metyloizobutyloketonu. Dodatkowo w Republice Południowej Afryki został opracowany proces ekstrakcyjnej rafinacji złota – the Minataur™ Process –

przeznaczony do przerobu materiałów wsadowych o zawartości Au 50-99%. Proces ten wykorzystywany jest m.in. w RPA, Algierii i Dubaju [6].

Ekstrakcja palladu zwykle przebiega z zastosowaniem siarczków organicznych jako ekstrahentów, które selektywnie wydzielają Pd wobec innych metali szlachetnych, z wyjątkiem złota, które może być współekstrahowane razem z palladem; z tego względu ekstrakcja Au zwykle stanowi pierwszy etap schematu technologicznego. Proces ekstrakcji Pd charakteryzuje się bardzo wysoką wydajnością – w rafinacji pozostaje zwykle mniej niż 1 mg/l Pd [1, 5, 6].

Odzysk Pd z ekstraktów przebiega z zastosowaniem amoniaku, z którego, po uprzednim zakwaszeniu HCl, wytrącany jest $\text{Pd}(\text{NH}_3)_2\text{Cl}_2$. Końcowym etapem prowadzącym do uzyskania Pd w postaci metalicznej jest zwykle kalcynacja. Redukcję można również przeprowadzić z zastosowaniem HCOONa lub HCOOH . Dane literaturowe wskazują, iż światowe rafinerie, tj. INCO oraz Degussa wykorzystują do ekstrakcji Pd siarczki dioktylu oraz diheksylu; w przypadku Johnson Matthey i Anglo Platinum zastosowanie znalazł β -hydroksyoksym z dodatkiem amin jako katalizatorów przeniesienia fazowego [1, 4-6].

Najszerzej stosowanymi ekstrahentami do odzysku platyny są natomiast fosforan tributylu oraz ekstrahenty aminowe. Rafineria Johnson Matthey wykorzystuje trioktyloaminę do wydzielania Pt z roztworu zubożonego w złoto i pallad. Podczas ekstrakcji Pt niezwykle ważna jest kontrola potencjału redukcyjnego ze względu na możliwość współekstrahowania Ir(IV) [6].

Na rysunku 3 przedstawiono przykładowy schemat technologiczny rozdziału PGM.

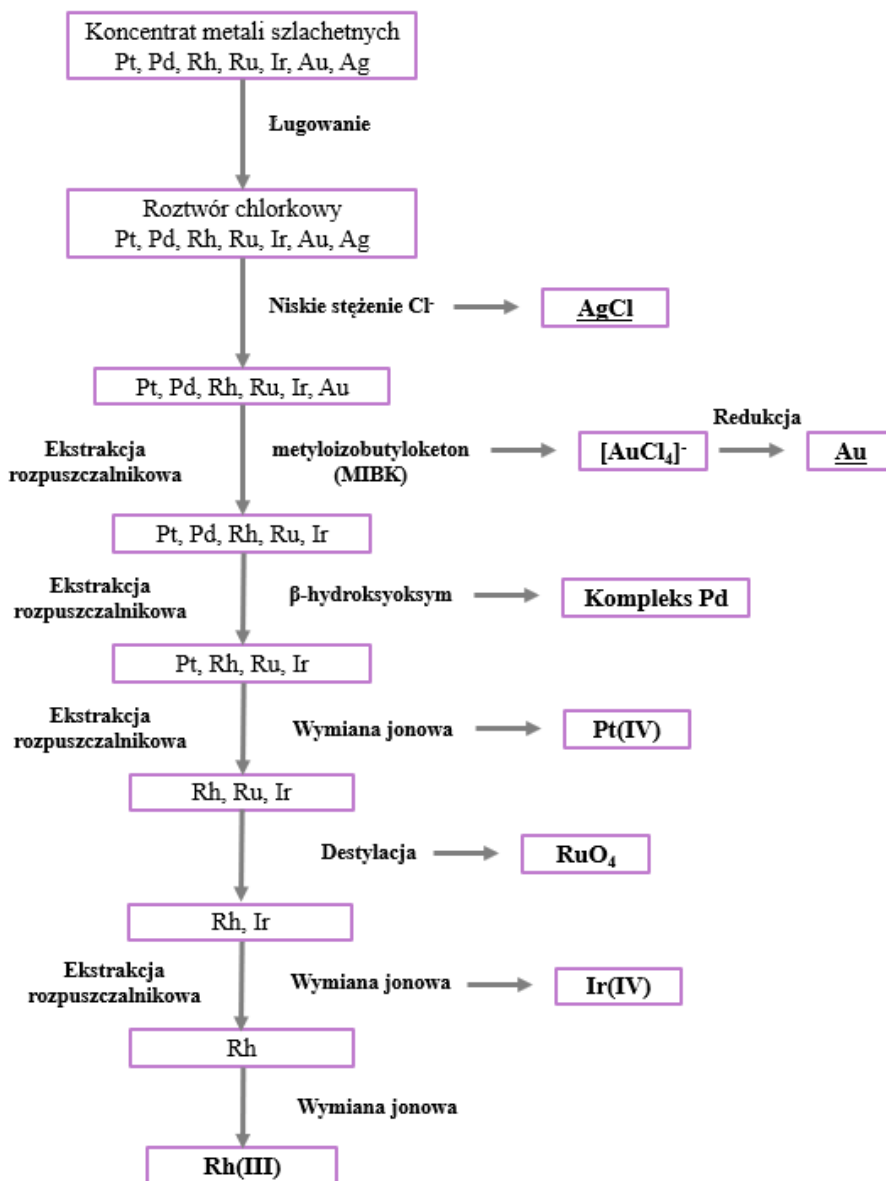
4. Ekstrakcja rozpuszczalnikowa

4.1. Podstawa teoretyczna

Ekstrakcja rozpuszczalnikowa jest skuteczną metodą rozdziału platynowców opartą na chemii chlorokompleksów, przy czym mechanizm ekstrakcji jest podobny zarówno w przypadku standardowych ekstrahentów organicznych, jak i cieczy jonowych. Sama technika ekstrakcji rozpuszczalnikowej, znanej również pod nazwą ekstrakcji cieczowej bądź ekstrakcji typu ciecz-ciecz, polega na przeniesieniu składników z jednej fazy, zwykle fazy wodnej, do drugiej, niemieszającej się z fazą pierwszą – fazy organicznej. W wyniku kontaktowania obu faz chlorokompleksy PGM migrują z fazy wodnej do fazy organicznej, co zwykle jest synergistycznym efektem przyciągania elektrostatycznego oraz procesów molekularnych [5, 15, 16].

Ekstrakcja chlorokompleksów platynowców może zachodzić według trzech mechanizmów:

- Wymiany anionów – polegającej na tworzeniu elektrycznie obojętnych par jonowych pomiędzy anionami chlorokompleksów MCl_x^{n-} a dodatnio naładowanymi, zasadowymi cząsteczkami związku organicznego;
- Solwatacji – gdzie otaczające aniony metali cząsteczki wody zastępowane są przez cząsteczki ekstrahenta;
- Tworzenia związku – mechanizm specyficzny dla Pd, zachodzący z ekstrahentami chelatującymi [16].



Rysunek 3. Schemat technologiczny rozdzielu PGM. Źródło: opracowano na podstawie [5]

Standardowe ekstrahenty cząsteczkowe, w zależności od rodzaju związku mogą wydzielać platynowce z roztworów wodnych poprzez każdy z powyższych mechanizmów. Amonowe lub fosforanowe ekstrahenty działają zwykle przez mechanizm anionowy-mienny bądź solwujący [13, 16].

W przypadku cieczy jonowych ekstrakcja może odbywać się zgodnie z mechanizmem jonowymiennym, gdzie grupa anionowa cieczy jonowej zostaje zastąpiona przez anion chlorokompleksowy, ale również poprzez koordynowanie platynowca,

asocjacje czy inne reakcje oparte na wymianie poszczególnych jonów. W praktyce, w przypadku układów faza wodna-ciecz jonowa reakcje ekstrakcji mają przebieg bardziej skomplikowany niż w przypadku klasycznych ekstrahentów. Poniżej przedstawiono reakcje wyjaśniające przebieg każdego z omówionych mechanizmów [5, 13, 16, 15].

Tabela 1. Mechanizmy ekstrakcji chlorokompleksów

Solwatacja	$MCl_y^{n-} + nH^+ + 2[HA] \rightarrow [H_nMCl_y] 2HA$
Wymiana jonowa	$MCl_y^{n-} + n[R_3NHCl] \rightarrow [(R_3NH)_n(MCl_y^{n-})] + nCl^-$ $MCl_y^{n-} + nH^+ + n[C][A] \rightarrow [C_nMCl_y] + nHA$
Tworzenie związku	$PdCl_4^{2-} + 2[L] \rightarrow [PdCl_2L_2] + 2 Cl^-$

(C – część kationowa; A – część anionowa; L – ligand)

Zródło: opracowano na podstawie [15].

Reaktywność chlorokompleksów jest ściśle związana z gęstością ładunku, liczbą koordynacyjną i geometrią cząstek. Na szybkość i skuteczność ekstrakcji duży wpływ mają również warunki prowadzenia procesu, takie jak pH, stężenie ekstrahowanego jonu w fazie wodnej, obecność innych jonów w roztworze, stosunek objętościowy poszczególnych faz, temperatura, intensywność mieszania, jak również stężenie ekstrahenta w fazie organicznej oraz lepkość, która jest kluczowym czynnikiem wpływającym na właściwości hydrodynamiczne układów mieszanych. Wysoka lepkość nie tylko hamuje transfer masy i szybkość ekstrakcji, ale również utrudnia proces ekstrakcji ciągłej [5, 13, 15, 16].

Pomimo dużej popularności technik ekstrakcyjnych i ich wieloletniej historii w zakresie selektywnego rozdziału platynowców, technika ta wciąż wzbudza szerokie zainteresowanie. Kluczowym aspektem są nie tylko wysoka selektywność procesu, ale również czystość otrzymywanych metali i ich związków, względnie krótki czas rafinacji, możliwość zawracania i regeneracji ekstrahentów oraz niewielka ilość generowanych odpadów [1-6]. Ponadto ekstrakcja znajduje zastosowanie zarówno w przypadku roztworów o względnie wysokich stężeniach PGM, jak i niskich. Kilkustopniowa ekstrakcja, jak również zachowanie odpowiedniego stosunku fazowego umożliwia wielokrotne załadowanie PGM w fazie organicznej, przy jednoczesnym niemal ilościowym wydzielaniu platynowca z roztworu macierzystego. Odpowiednie nasycenie fazy organicznej stanowi duże ułatwienie na dalszych etapach przerobu, takich jak otrzymywanie czystych, stężonych reekstraktów PGM, a następnie ich redukcji celem wydzielenia platynowców w postaci metalicznej. W niektórych przypadkach możliwa jest również redukcja PGM bezpośrednio z fazy organicznej [1-6].

4.2. Popularne ekstrahenty platynowców

Spośród dostępnych komercyjnie, skutecznych ekstrahentów platynowców można wyróżnić trzy główne grupy rozpuszczalników:

- ekstrahenty fosforoorganiczne;
- ekstrahenty aminowe;
- ekstrahenty na bazie siarki.

Do najpopularniejszych ekstrahentów na bazie fosforu można zaliczyć m.in. wspomniany już fosforan tributylu, czy tlenki trioktylofosfiny o rozgałęzionej bądź liniowej

budowie łańcucha alkilowego, szeroko znanymi pod nazwami handlowymi Cyanex 921 oraz Cyanex 925. Fosforan tributylu znajduje ponadto zastosowanie jako modyfikator w procesie ekstrakcji, gdzie ułatwia separację faz w układach mieszanych oraz zapobiega tworzeniu się trzeciej fazy [15, 16, 22].

Drugą dużą grupą ekstrahentów są ekstrahenty aminowe, które ze względu na budowę można podzielić na aminy pierwszo-, drugo-, trzeciorzędowe oraz czwartorzędowe sole amoniowe, gdzie jednym z najczęściej wykorzystywanych ekstrahentów jest trioktyloamina (Alamine 336, Alamine 300) i jej pochodne [5, 6, 15-22].

Ekstrahenty siarkowe, takie jak siarczek dioktylu, znajdują zastosowanie w selektywnej ekstrakcji Pd. Ponadto cechują się dużą stabilnością w silnie kwaśnym środowisku [15, 16, 22].

W ostatnich latach zaobserwowano znaczący wzrost popularności cieczy jonowych, takich jak Cyphos IL 101 czy Aliquat 336, jako ekstrahentów do wydzielania platynowców z roztworów wodnych. Pogłębiający się trend w kierunku ochrony środowiska, powoduje, że ciecze jonowe, uważane za „zielone rozpuszczalniki”, są coraz chętniej wykorzystywane. Właściwości, takie jak niepalność, niska lotność, zdolność rozpuszczania wielu substancji czy możliwość regeneracji, umożliwiły zastosowanie ich w wielu procesach, m.in. syntezie organicznej, produkcji leków, reakcjach katalitycznych, chromatografii cieczowej i elektrochemii. Duża popularność tego typu substancji chemicznych powoduje, że związki te cechują się dużą różnorodnością budowy i dostępnością na rynku [15-18, 31-35].

Pod względem chemicznym ciecze jonowe to stopione sole, zachowujące postać cieczy w temperaturze poniżej 100°C, składające się zwykle z dużych kationów organicznych oraz nieorganicznych anionów. Ze względu na budowę części kationowej można wyróżnić m.in. ciecze amoniowe, imidazolowe, fosfoniowe, pirydynowe, piperynowe, pirolidynowe, betainowe czy guanidynowe. Część anionową zwykle stanowią proste grupy nieorganiczne, tj. SO_4^{2-} , Cl^- , Br^- , I^- , NO_3^- bądź fluoroaniony – PF_6^- , NTf_2^- , BF_4^- , F^- . Ponieważ klasyczne ciecze jonowe nie zawsze wykazują selektywność względem niektórych kompleksów PGM, w ostatnich latach często podejmowano próby zastosowania anionów zawierających atomy P, N, S i O lub ich kombinacji w celu zwiększenia selektywności [15-18, 31-35].

Większość badań nad wykorzystaniem cieczy jonowych do ekstrakcji cieczowej platynowców skupia się na zastosowaniu ich rozcieńczonych roztworów organicznych, w których rolę rozpuszczalnika pełnią popularne niepolarne bądź półpolarne węglowodory alifatyczne i/lub aromatyczne. Ze względu na unikalne właściwości fizykochemiczne, dodatek cieczy jonowych do klasycznych rozpuszczalników organicznych stwarza idealne środowisko do selektywnego wydzielania metali szlachetnych. Naładowane cząsteczki cieczy jonowych zapewniają silniejsze oddziaływania pomiędzy aktywnymi ligandami zawartymi w fazie organicznej a cząsteczkami obecnymi w roztworze wodnym zwiększając w konsekwencji wydajność wydzielania chlorokompleksów platynowców. W przeszłości podejmowane były również badania nad zastosowaniem nierozcieńczonych cieczy jonowych, jednak ze względu na wysoką lepkość tych substancji oraz utrudnione mieszanie skutkujące powolnym przenoszeniem masy, zastosowanie tego typu układów, szczególnie w warunkach przemysłowych jest ograniczone [15-18, 31-35].

Jedną z rzadziej wspomnianych, ale równie ważną zaletą wynikającą z zastosowania cieczy jonowych w układach ekstrakcyjnych jest ich działanie powierzchniowo-

czynne, pozwalające na obniżenie napięcia powierzchniowego i zwiększenie kontaktu pomiędzy fazami [16].

W tabeli 3 przedstawiono zestawienie najczęściej stosowanych ekstrahentów organicznych, w tym również cieczy jonowych, znajdujących zastosowanie w rozdziale metali z grupy platynowców.

Tabela 2. Popularne ekstrahenty platynowców

Ekstrahent	Roztwór wyjściowy	Wydajność ekstrakcji	Reekstrahent	Wydajność reekstrakcji	Źródła
Trioktyloamina Nazwy handlowe: Alamine 336; Alamine 300; Alamine 308	Syntetyczny roztwór chlorkowy, zawierający Pt i Rh	97% Pt 21% Rh	8 M HNO ₃ (50°C)	~100% Pt < 1% Rh	[23]
	Syntetyczny roztwór chlorkowy, zawierający Pt i Pd	>99% Pt >99% Pd	1) 0,1 M NaSCN 2) 0,1 M TU w 0,5 M HCl	1) 99,9% Pt 2) 99,5% Pd	[24]
	Sole Pt i Rh otrzymane w wyniku ługowania materiałów odpadowych, rozpuszczone w glikolu etylenowym	~100% Pt	-	-	[25]
fosforan tributylu	Syntetyczny roztwór chlorkowy, zawierający Pt, Pd, Mn, Fe, Cr	99,9% Pd	0,5 M TU w 0,1 M HCl	99,8%	[26]
siarczek tri(izobutylo)-fosfiny Nazwy handlowe: Cyanex 471X	Roztwór chlorkowy po ługowaniu katalizatorów	~100%	0,1 M TU w 5% HCl	46% Pd	[27]
oksym 5,8-dietylo-7-hydrokso-6-dodekanu Nazwy handlowe: LIX63	Roztwór chlorkowy, zawierający Pd, Pt, Ir i Rh	99,9% Pd	1) 0,5 M TU 2) 0,5 M NaSCN	1) 99,9% 2) 90,1%	[28]
oksym 2-hydrokso-5-nonyloacetofenonu Nazwy handlowe: LIX 84I	Syntetyczny roztwór chlorkowy, zawierający Pd	> 95% Pd	-	-	[27]

Zastosowanie ekstrakcji cieczowej w technologii odzysku i rozdzielenia metali z grupy platynowców

Mieszanka: 5-nonylsalicylaldehyd i oksym 2-hydroksy-5-nonyloacetofenonu	Syntetyczny roztwór chlorkowy, zawierający Pd	> 97% Pd	-	-	[27]
Nazwy handlowe: LIX 984					
mieszanka: chlorek trioktylometyloamoniowy; chlorek tridecyloamoniowy	Syntetyczny roztwór chlorkowy, zawierający Pt, Rh, Al, Mg, Fe	99,97% Pt	0,1-0,5 M TU w 0,1-0,5 M HCl	97-100%	[29]
Nazwy handlowe: Aliquat 336					
	Syntetyczny roztwór chlorkowy, zawierający Pt, Pd, Mn, Fe, Cr	99,83% Pt	0,5 M TU w 0,5 M HCl	99,9%	[30]
	Roztwór po ługowaniu katalizatorów w wodzie królewskiej	>99% Pt	Na ₂ S ₂ O ₃	>99,9%	[31]
	Sole Pt i Rh otrzymane w wyniku ługowania materiałów odpadowych, rozpuszczone w glikolu etylenowym	~100% Pt	1,0 M TU	~100%	[25]
	Syntetyczny roztwór chlorkowy zawierający Pt, Pd, Ru i Rh	>95% Pt/Pd ~ 60% Ru <15% Rh	1) 0,1 M TU w 0,5 M HCl 2) 1,0 M HNO ₃	1) >90% Pd 2) >65% Pt	[32]
chlorek triheksylo(tetracylo)fosfoniowy	Roztwór chlorkowy po ługowaniu katalizatorów	~100%	1,0 M HNO ₃ (zastosowany do wymycia innych metali nieszlachetnych, gł. Fe, Pb, Zn i Mg)	Pt 18,6%	[33]
Nazwy handlowe: Cyphos IL 101					

	Syntetyczny roztwór chlorkowy zawierający Pt, Pd, Rh	99,9% Pt 98,1% Pd <2% Rh	1) 0,02 M TU w 5% HCl 2) 0,1 M NaSCN 3) 0,01 M NH ₄ OH	1) ~100% Pd 10,5% Pt 2) 92,8% Pt <2% Pd 3) 94,6% Pt 81,9% Pd	[34]
	Syntetyczny roztwór chlorkowy, zawierający Rh i Pd	99,9%- 89,7% Pd	-	-	[35]
	Sole Pt i Rh otrzymane w wyniku ługowania materiałów odpadowych, rozpuszczone w glikolu etylenowym	~100% Pt	1,0 M TU	~100%	[25]
bromek triheksylo(tetradecylo)fosfoniowy Nazwy handlowe: Cyphos IL 102	Syntetyczny roztwór, zawierający Pd	>80% ~100% Pd	0,5 M NH ₄ OH	84-90%	[36]
bis(2,4,4-trimetylopentylo)fosfinian triheksylo(tetradecylo)fosfoniowy Nazwy handlowe: Cyphos IL 104	Syntetyczny roztwór chlorkowy, zawierający Pd	52-96% Pd	0,5 M NH ₄ OH	~90%	[37]
chlorek trioktylododecylofosfoniowy (P ₈₈₈₁₂ Cl)	Syntetyczny roztwór chlorkowy, zawierający Pt, Pd i Rh	Pt, Pd: 99,9% Rh: 10,0- 90.0%	1) 5 M HNO ₃ 2) 1,0 M TU 3) 5,0 M HCl	1) 74,9% Pt 2) 91,2% Pd 3) 73,7% Rh	[38]
	Syntetyczny roztwór chlorkowy, zawierający Pd i Rh oraz zanieczyszczenia, m.in. Al, Mg, Ce, Ba	~100% Pd 80% Rh	1) 1M CS(NH ₂) ₂ 2) 5M HCl	1) >90% Pd 2) >75% Rh	[39]

(TU – tiomocznik).

5. Podsumowanie

Przytoczone dane literaturowe jednoznacznie wskazują, że ekstrakcja rozpuszczalnikowa jest bardzo skuteczną i selektywną metodą wydzielenia platynowców z wieloskładnikowych roztworów, dzięki czemu technika ta od lat stanowi stały element schematów technologicznych rafinerii metali szlachetnych na całym świecie. Kluczowym aspektem jest nie tylko wysoka selektywność procesu, ale również czystość otrzymanych metali i ich związków, względnie krótki czas rafinacji, możliwość zawracania i regeneracji ekstrahentów oraz niewielka ilość generowanych odpadów.

Pomimo dużej popularności technik ekstrakcyjnych i ich wieloletniej historii w zakresie selektywnego rozdzielu platynowców, technika ta nadal wzbudza szerokie zainteresowanie, czego dowodem są wciąż pojawiające się, nowe publikacje naukowe. Co więcej, analiza danych zamieszczonych w tabeli 3 wskazuje, że obecnie największą uwagę badaczy skupia wykorzystanie cieczy jonowych w ekstrakcyjnym rozdziale PGM. Dobrze poznane mechanizmy procesu ekstrakcji oraz chemii chlorokompleksów pozwalają na projektowanie oraz syntezę nowych, coraz skutecznych klas związków, mogących znaleźć zastosowanie w selektywnym odzysku tych cennych metali. Większość dostępnych badań naukowych skupia się jednak na rozdziale tych metali z roztworów syntetycznych, które ze względu na prosty skład i brak innych zanieczyszczeń nie są w stanie w pełni odwzorować warunków panujących w roztworach technologicznych, będących układami wieloskładnikowymi, charakteryzującymi się wysokimi stężeniami innych metali nieszlachetnych.

Z drugiej strony ciągły rozwój gospodarczy wraz z rosnącymi restrykcjami w zakresie ochrony środowiska wymuszają konieczność odzysku platynowców ze źródeł wtórnych, gdzie niezwykle cennym materiałem są zużyte katalizatory samochodowe oraz przemysłowe. Recykling tego typu odpadów pozwala na odzysk platynowców przy stosunkowo niewielkim nakładzie energetycznym oraz niewielkiej ilości potrzebnego materiału, gdzie w porównaniu do źródeł naturalnych, w celu otrzymania zaledwie kilku gramów PGM, nierzadko konieczne jest przerobienie kilku ton rudy. Otrzymywanie platynowców na drodze przerobu materiałów odpadowych jest więc niezwykle korzystne ze względów środowiskowych i gospodarczych. Należy jednak pamiętać, że w wyniku ługowania tego typu materiałów powstają roztwory, charakteryzujące się wysoką zawartością metali towarzyszących, nieszlachetnych. Mimo to odpowiednia optymalizacja parametrów procesu ekstrakcji i reekstrakcji, umożliwi skuteczną wydzielenie metali szlachetnych w postaci czystej, nawet z układów wieloskładnikowych. Co więcej, wiele korporacji opracowało już stosowane komercyjnie procesy odzysku platynowców ze źródeł wtórnych, gdzie ekstrakcja cieczowa jest kluczową techniką prowadzącą do szybkiego i wydajnego otrzymania tych metali, z jednoczesnym zachowaniem wysokiej czystości produktu końcowego.

Podziękowania

Serdeczne podziękowania dla wszystkich osób, które umożliwiły publikację tego rozdziału i pomogły w jego tworzeniu.

Praca jest częścią projektu doktorskiego wykonywanego w ramach V edycji programu „Doktorat Wdrożeniowy” finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego. Praca została zrealizowana również w części w ramach środków subwencyjnych Sieci Badawczej Łuksiewicz – Instytut Metali Nieżelaznych (sprawozdanie nr 8321/G/2022; kod pracy: 0334111002).

Literatura

1. Panda R., Jha M.K., Pathak D.D., *Commercial processes for the extraction of platinum group metals (PGMs)*, Minerals, Metals and Materials Series. Springer International Publishing, 2018, s. 119-130.
2. Nose K., Okabe T.H., *Treatise on Process Metallurgy*, Stockholm 2014, s. 1071-1097.
3. Bielański A., *Podstawy chemii nieorganicznej*, PWN, Warszawa 1994, s. 930-939.
4. Sinisalo P., Lundström M., *Refining Approaches in the Platinum Group Metal Processing Value Chain – A Review*, Metals, 8(4), 2018, s. 203.
5. Bernardis F.L., Grant R.A., Sherrington D.C., *A review of methods of separation of the platinum-group metals through their chloro-complexes*, Reactive and Functional Polymers, 65(3), 2005, s. 205-217.
6. Sole K., *Solvent extraction in the hydrometallurgical processing and purification of metals: process design and selected application*, Solvent Extraction and Liquid Membranes: Fundamentals and Applications in New Materials, 2008, s. 141-200.
7. <https://matthey.com/products-and-markets/pgms-and-circularity/pgm-management> [data dostępu: 15.03.2023].
8. <https://www.statista.com/statistics/273624/platinum-metal-reserves-by-country/> [data dostępu: 10.12.2022].
9. Oszczepalski S., Chmielewski A., *Zasoby przewidywane surowców metalicznych Polski na mapie w skali 1 : 200 000 – miedź, srebro, złoto, platyna i pallad w utworach cechsztyńskiej serii miedzionośnej*, Przegląd Geologiczny, 63(9), 2015, s. 534-545.
10. Chmielewski T., *Odzyskiwanie złota i platynowców z rudy pozabilansowej LGOM na drodze ługowania cyjankowego*, Fizykochemiczne Problemy Mineralurgii, 32, 1998, s. 43-56.
11. <https://kgm.com/pl/biznes/produkty/metale-szlachetne> [data dostępu: 10.12.2022].
12. Saguru C., Ndlovu S., Moropeng D., *A review of recent studies into hydrometallurgical methods for recovering PGMs from used catalytic converters*, Hydrometallurgy, 182, 2018, s. 44-56.
13. Zheng H., Ding Y., Wen Q., Liu B., Zhang S., *Separation and purification of platinum group metals from aqueous solution: Recent developments and industrial applications*, Resources, Conservation and Recycling, 167, 2021, s. 105417.
14. Yakoumis I., Panou M., Moschovi A.M., Pnias D., *Recovery of platinum group metals from spent automotive catalysts: A review*, Cleaner Engineering and Technology, 3, 2021, s. 100112.
15. Jha M.K., Gupta D., Lee J., Kumar V., Jeong J., *Solvent extraction of platinum using amine based extractants in different solutions: A review*, Hydrometallurgy, 142, 2014, s. 60-69.
16. Jae-chun Lee, Kurniawan Kurniawan, Sookyoung Kim, Viet Tu Nguyen, Banshi D. Pandey, *Ionic Liquids-Assisted Solvent Extraction of Precious Metals from Chloride Solutions*, Separation & Purification Reviews, 2022.
17. Sanuki S., Matsumoto, Y., Majima, H., *Preparation of ammonium chloroplatinate by a precipitation stripping of Pt(IV)-loaded Alamine 336 or TBP*, Metallurgical and Materials Transactions B, 30, 1999, s. 197-203.
18. Regel-Rosocka M., Cieszyńska A., Wiśniewski M., *Extraction of palladium (II) ions from chloride solutions with phosphonium ionic liquid CyphosŽIL101*, Polish Journal of Chemical Technology, 9, 2007, s. 99-101.
19. Nguyen T.H., Sonu C.H., Lee M.S., *Separation of Pt(IV), Pd(II), Rh(III) and Ir(IV) from concentrated hydrochloric acid solutions by solvent extraction*, Hydrometallurgy, 164, 2016, s. 164.

21. Nguyen T.H., Sonu, C.H., Lee, M.S., *Separation of platinum(IV) and palladium(II) from concentrated hydrochloric acid solutions by mixtures of amines with neutral extractants*, Journal of Industrial and Engineering Chemistry, 32, 2015, s. 238-245.
22. Firmansyah M.L., Kubota F., Goto M., *Solvent extraction of Pt(IV), Pd(II), and Rh(III) with the ionic liquid trioctyl(dodecyl) phosphonium chloride*, Journal of Chemical Technology & Biotechnology, 93, 2018.
23. Paiva A., *Recycling of Palladium from Spent Catalysts Using Solvent Extraction – Some Critical Points*, Metals, 7(11), 2017, s. 505.
24. Jaree A., Khunphakdee N., *Separation of concentrated platinum(IV) and rhodium(III) in acidic chloride solution via liquid–liquid extraction using tri octylamine*, Journal of Industrial and Engineering Chemistry, 17, 2011, s. 243-247.
25. Swain B., Jeong J., Kim S., Lee J., *Separation of platinum and palladium from chloride solution by solvent extraction using Alamine 300*, Hydrometallurgy, 104(1), 2010, s. 1-7.
26. Nguyen V.T., Riaño S., Aktan E., Deferm C., Fransaeer J., Binnemans K., *Solvometallurgical Recovery of Platinum Group Metals from Spent Automotive Catalysts*, ACS Sustainable Chemistry & Engineering, 9(1), 2020, s. 337-350.
27. Lee J.Y., Raju B., Kumar B.N., Kumar J.R., Park H.K., Reddy B.R., *Solvent extraction separation and recovery of palladium and platinum from chloride leach liquors of spent automobile catalyst*, Separation and Purification Technology, 73(2), 2010, s. 213-218.
28. Rane M.V., *PGM ore processing: LIX reagents for palladium extraction & platinum stripping from Alamine 336 using NaOH-NaCl*, Minerals Engineering, 138, 2019, s. 119-124.
29. Nguyen T.H., Sonu C.H., Lee M.S., *Separation of Pt(IV), Pd(II), Rh(III) and Ir(IV) from concentrated hydrochloric acid solutions by solvent extraction*, Hydrometallurgy, 164, 2016, s. 71-77.
30. Raju B., Kumar J.R., Lee J.-Y., Kwonc H.-S., Kantam M.L., Reddy B.R., *Separation of platinum and rhodium from chloride solutions containing aluminum, magnesium and iron using solvent extraction and precipitation methods*, Journal of Hazardous Materials, 227-228, 2012, s. 142-147.
31. Lee J.Y., Raju B., Kumar B.N., Kumar J.R., Park H.K., Reddy B.R., *Solvent extraction separation and recovery of palladium and platinum from chloride leach liquors of spent automobile catalyst*, Separation and Purification Technology, 73(2), 2010, s. 213-218.
32. Marinho R.S., Afonso J.C., da Cunha J.W.S.D., *Recovery of platinum from spent catalysts by liquid–liquid extraction in chloride medium*, Journal of Hazardous Materials, 179(1-3), 2010, s. 488-494.
33. Rzelewska-Piekut M., Regel-Rosocka M., *Separation of Pt(IV), Pd(II), Ru(III) and Rh(III) from model chloride solutions by liquid-liquid extraction with phosphonium ionic liquids*, Separation and Purification Technology, 2018.
34. Rzelewska-Piekut M., Paukszta D., Regel-Rosocka M., *Hydrometallurgical recovery of platinum group metals from spent automotive converters*, Physicochemical Problems of Mineral Processing, 57(2), 2021, s. 84-95.
35. Nguyen V.T., Lee J., Chagnes A., Kim M., Jeong J., Cote G., *Highly selective separation of individual platinum group metals (Pd, Pt, Rh) from acidic chloride media using phosphonium-based ionic liquid in aromatic diluent*, RSC Advances, 6(67), 2016, s. 62717-62728.
36. Svecova L., Papaiconomou N., Billard I., *Quantitative extraction of Rh(III) using ionic liquids and its simple separation from Pd(II)*, Dalton Transactions, 45(38), 2016, s. 15162-15169.
37. Regel-Rosocka M., Rzelewska M., Baczynska M., Janus M., Wisniewski M., *Removal of palladium(II) from aqueous chloride solutions with Cyphos phosphonium ionic liquids as metal ion carriers for liquid-liquid extraction and transport across polymer inclusion membranes*, Physicochemical Problems of Mineral Processing, 51(2), 2015, s. 621-631.

38. Cieszynska A., Wiśniewski M., *Extractive recovery of palladium(II) from hydrochloric acid solutions with Cyphos®IL 104*, Hydrometallurgy, 113(114), 2012, s. 79-85.
39. Firmansyah M.L., Kubota F., Goto M., *Solvent extraction of Pt(IV), Pd(II), and Rh(III) with the ionic liquid trioctyl(dodecyl) phosphonium chloride*, Journal of Chemical Technology & Biotechnology, 93(6), 2018, s. 1714-1721.
40. Firmansyah M.L., Kubota F., Yoshida W., Goto M., *Application of a Novel Phosphonium-based Ionic Liquid to the Separation of Platinum Group Metals from Automobile Catalyst Leach Liquor*, Industrial & Engineering Chemistry Research, 58(9), 2019, 3845-3852.

Zastosowanie ekstrakcji cieczowej w technologii odzysku i rozdziału metali z grupy platynowców

Streszczenie

Platynowce (PGM) stanowią grupę metali o szczególnej wartości rynkowej i niskim rozpowszechnieniu w skorupie ziemskiej. Wraz z rozwojem gospodarczym zapotrzebowanie na metale szlachetne sukcesywnie wzrasta, czemu towarzyszy stopniowe ubożenie naturalnych rezerw metali szlachetnych, a w konsekwencji poszukiwanie nowych dróg pozyskania tych cennych metali. Roztwory otrzymywane zarówno na drodze przerobu źródeł naturalnych, jak również materiałów odpadowych charakteryzują się zwykle niską koncentracją metali szlachetnych, przy jednoczesnej dużej ilości zanieczyszczeń, w tym miedzi, żelaza, cynku czy niklu. Co więcej, podobieństwo właściwości oraz duża odporność chemiczna platynowców powoduje, że metale te w stanie czystym otrzymać można wyłącznie metodami hydrometalurgicznymi, takimi jak metody strąceniowe, wymiana jonowa czy ekstrakcja rozpuszczalnikowa. Spośród dostępnych metod, na przestrzeni lat szczególne znaczenie zyskała ekstrakcja rozpuszczalnikowa, która obecnie stanowi główną metodę rozdziału PGM w ciągach technologicznych rafinerii na całym świecie. Ponadto szereg zalet, takich jak wysoka selektywność procesu, duża szybkość reakcji, wysoka czystość otrzymywanych metali i ich związków czy możliwość regeneracji i zwracania ekstrahentów, powoduje, iż mimo wieloletniej historii, technika ta nadal wzbudza szerokie zainteresowanie badaczy.

Niniejszy przegląd ma na celu przybliżenie tematyki rozdziału metali szlachetnych metodą ekstrakcji cieczowej, a także dokładną charakterystykę oraz ocenę przydatności wybranych ekstrahentów, w oparciu o najnowsze doniesienia literaturowe.

Słowa kluczowe: platynowce, metale szlachetne, ekstrakcja rozpuszczalnikowa, odzysk platynowców

Application of solvent extraction in the technology of recovery and separation of platinum group metals

Abstract

Platinum group metals (PGM) are a group of metals with a special market value and low distribution in the earth's crust. Along with the economic development, the demand for precious metals is gradually increasing, which is accompanied by the gradual depletion of natural reserves of precious metals, and consequently the search for new ways to obtain these precious metals. Solutions obtained both by processing natural sources and waste materials are usually characterized by a low concentration of precious metals, with a large number of impurities, including copper, iron, zinc or nickel. What's more, the similarity of properties and high chemical resistance of the platinum group metals mean that these metals in a pure state can only be obtained by hydrometallurgical methods, such as precipitation methods, ion exchange or solvent extraction. Among the available methods, over the years, solvent extraction has gained particular importance, which is currently the main method of PGM separation in technological lines of refineries around the world. In addition, several advantages, such as high selectivity of the process, high reaction rate, high purity of the obtained metals and their compounds, or the possibility of regeneration and recycling of extractants, cause that, despite many years of history, this technique still arouses wide interest of researchers.

This review is aimed at presenting the subject of the separation of precious metals by liquid extraction, as well as a detailed characterization and assessment of the usefulness of selected extractants, based on the latest literature reports.

Keywords: PGM, noble metals, solvent extraction, PGMs recovery

Indeks Autorów

Benke G.	184
Borkowski M.	157
Deniziak S.	157
Dzyundzja O.	107
Goc K.	184
Jońca J.	40
Jurczak E.	131
Kaczmarek D.	157
Kluczek J.	184
Koniew M.	173
Krajewski M.	74
Królak A.	157
Kuligowska K.	173
Kwietniewska-Śmietana A.	85
Leszczyńska-Sejda K.	184
Majkut L.	7
Malarz J.	184
Ochmański M.	184
Olszewski R.	7
Opaliński A.	85
Palimaka P.	18
Pianowska K.	184
Picheta P.	157
Pietrzyk S.	18
Pięta P.	157
Przywara M.	61
Przywara R.	61
Rasulov R.	107
Regulski K.	85
Sitkowski P.	85
Stanusch M.	173
Stryhuniwska O.	107
Swadźba J.	85
Urzenitzok P.	120
Wąsowicz P.	85
Wieczorek M.	95
Wiśniewski M.	157
Zapała W.	61