



**POLITECHNIKA
GDAŃSKA**

WYDZIAŁ ELEKTRONIKI,
TELEKOMUNIKACJI I INFORMATYKI



Imię i nazwisko autora rozprawy: Grzegorz Brodny
Dyscyplina naukowa: Informatyka

ROZPRAWA DOKTORSKA

Tytuł rozprawy w języku polskim: Problemy niepewności i integracji w przetwarzaniu danych o stanie emocjonalnym użytkownika komputera

Tytuł rozprawy w języku angielskim: Uncertainty and data integration in emotion recognition of emotional state of a computer user

Promotor

podpis

dr hab. inż. Agnieszka Landowska

Gdańsk, 2022

PODZIĘKOWANIA

Składam serdeczne podziękowania Pani dr hab. inż. Agnieszce Landowskiej za nieocenioną pomoc udzieloną w trakcie przygotowania pracy doktorskiej, wprowadzenie do zagadnienia automatycznego przetwarzania emocji – już od początku drugiego stopnia studiów, a także kontynuację w czasie doktoratu – ponad 8 lat współpracy. Szczególne podziękowania za cierpliwość, poświęcony czas, liczne rozmowy będące inspiracją oraz podnoszeniem na duchu w momentach wahania – umożliwiło mi to ukończenie pracy. Dziękuję za inspirację do podjęcia się pracy doktorskiej.

Szczególne słowa podziękowania składam mojej żonie Natalii za cierpliwość, wyrozumiałość i olbrzymie wsparcie oraz wiarę we mnie. Gdyby nie Ty, nie starczyłoby mi sił na ukończenie rozprawy.

Dziękuję także całemu zespołowi grupy badawczej EMORG za pomoc w realizacji wielu wspólnych badań eksperymentalnych, projektów i publikacji.

Dziękuję całej Katedrze Inżynierii Oprogramowania za wsparcie i celne uwagi podczas prezentacji seminaryjnych oraz rozmowach kulturalowych.

Chciałbym również podziękować rodzinie oraz wszystkim przyjaciołom, od których otrzymałem i czułem słowa otuchy. Dziękuję za wiarę we mnie oraz Wasze wsparcie.

STRESZCZENIE

Istnieje wiele programów automatycznego rozpoznawania emocji z pojedynczych modalności. Wykorzystują one różnorodne modele reprezentacji emocji, a także metody przetwarzania. Obecnie wyraźnie widoczny jest trend do wykorzystania analizy wielomodalnej oraz wielokanałowej, mającej na celu poprawę skuteczności rozpoznawania emocji oraz poprawę niezawodności. W dziedzinie automatycznego rozpoznawania emocji nie określono typowego modelu reprezentacji emocji ani typowej metody analizy wielomodalnej czy wielokanałowej.

Celem pracy doktorskiej było zaproponowanie metody integracji danych o stanie emocjonalnym użytkowników komputerów, która uwzględni różne programy, kanały dostępu oraz metody reprezentacji emocji, a także kontekstową niepewność pomiarów symptomów emocji.

Autor pracy zaproponował nową metodę integracji składającą się z wzorca architektonicznego Scoreboard, metody kwantyfikacji niepewności oszacowania emocji oraz algorytmu integracji informacji o stanie emocjonalnym, opartego o miarę niepewności. Zaproponowane rozwiązanie zapewnia modularność i adaptowalność poprzez zastosowane paradygmaty obiektowe, pozwala na łatwą integrację umożliwiającą wykorzystanie rozwiązań napisanych w różnych technologiach, cechuje się odpornością na nieprawidłowe dane wejściowe oraz nieprawidłowe działania komponentów, dzięki zastosowaniu ich dynamicznej oceny w czasie działania programu. Zaproponowana metoda kwantyfikacji pozwala na liczbową reprezentację niepewności. Z kolei zaproponowana metoda integracji informacji o stanie emocjonalnym wyznacza poziom wiarygodności odpowiadający jakości danych wejściowych, a także zgodności danych z poszczególnych kanałów obserwacji takich jak zgodności informacji o stanie emocjonalnym z dwóch kamer.

W ramach ewaluacji zaproponowanego rozwiązania wykorzystano trzy metody badawcze: eksperymenty na zaetykietowanym emocjonalnie zbiorze danych SEMAINE, symulacje przypadków brzegowych przygotowane przez autora pracy, eksperymenty oraz studium przypadku z wykorzystaniem danych z badania przeprowadzonego na Politechnice Gdańskiej na stanowisku Monitora Emocji. Przeprowadzono dziewięć eksperymentów, pięć symulacji oraz studium przypadku Monitora Emocji, co pozwoliło na walidację postawionych tez. Wykorzystywanymi w badaniach modalnościami były analiza mimiki twarzy, analiza prozodii głosu oraz analiza ładunku emocjonalnego tekstu. Za pomocą zaproponowanej wielomodalnej analizy uzyskano poprawę niezawodności.

Słowa kluczowe: przetwarzanie emocji, rozpoznawanie emocji, modele reprezentacji emocji, PAD, wczesna i późna fuzja, niepewność, integracja danych.

Dziedzina nauki i techniki, zgodnie z wymogami OECD: nauki przyrodnicze, nauki o komputerach i informatyka,

ABSTRACT

Nowadays, affective computing has provided many solutions for emotion recognition from a single modality. Most of them are based on a single input channel only. At the same time appropriate data integration might be the most efficient way to get more precise results. However, there is no typical integration method in emotion recognition present up to date.

The main goals of this PhD thesis were to propose and validate new integration method of emotional data, which supports multiple algorithms, observation channels and models of emotion representation, as well as uncertainty of measuring. The method includes: Scoreboard architectural pattern, method for quantifying the uncertainty of measuring emotion symptoms and an integration algorithm based on the measure of uncertainty. The proposed solution provides modularity and adaptability through object-oriented paradigms, allows for easy integration of solutions written in various technologies. It is has also robustness feature due to dynamic evaluation of components in runtime. The proposed quantification method allows for a numerical representation of uncertainty. The proposed method of integration determines the level of reliability corresponding to the observed convergence of results of multiple algorithms of emotion recognition.

In this PhD dissertation, three research method were used: the experiments with affect-labelled SEMAINE set, simulations for margin conditions and a case study conducted at Emotion Monitor stand at the Gdańsk University of Technology. Nine experiments, five simulations, and a case study were conducted in total to support the PhD thesis. The modalities used in the research were the analysis of facial expressions, the analysis of voice prosody and sentiment analysis of the text. In most cases, the proposed method obtained higher or no worse accuracy of emotion recognition than unimodal solutions. The proposed multimodal analysis achieved an improvement in reliability.

Keywords: Affective computing, Emotions Recognition, Models of Emotion, PAD, early and late fusion, uncertainty, data integration

OECD field of science and technology: Natural sciences, Computer and information sciences

SPIS TREŚCI

PODZIĘKOWANIA.....	2
SPIS TREŚCI.....	5
1. WPROWADZENIE.....	8
1.1. Motywacja podjęcia pracy.....	8
1.2. Cele pracy.....	9
1.3. Tezy rozprawy.....	10
1.4. Opis zawartości pracy.....	10
1.5. Słownik pojęć.....	11
1.6. Konwencje notacyjne.....	15
2. PRZETWARZANIE EMOCJI – ANALIZA ISTNIEJĄCEGO STANU WIEDZY.....	16
2.1. Stan emocjonalny i jego obserwacje.....	17
2.2. Modele reprezentacji emocji na potrzeby aplikacji komputerowych.....	20
2.3. Odzworowania pomiędzy modelami reprezentacji emocji.....	24
2.4. Rozpoznawanie emocji w podziale na modalności.....	26
2.5. Zastosowania przetwarzania emocji.....	30
2.6. Wyzwania rozpoznawania emocji.....	32
3. METODY INTEGRACJI WIELOMODALNEJ.....	33
3.1. Zestawienie stosowanych metod integracji.....	33
3.2. Scenariusz bazowy – rozwiązanie jednomodalne i jednokanałowe.....	35
3.3. Scenariusz wczesnej fuzji (ang. early/feature–level fusion).....	35
3.4. Scenariusz późnej fuzji – wiele klasyfikatorów.....	37
3.5. Scenariusz fuzji hybrydowej (ang. hybrid fusion).....	39
3.6. Podsumowanie metod integracji w rozpoznawaniu emocji.....	40
3.7. Rodzaje metod integracji w późnej fuzji.....	40
3.8. Podsumowanie metod integracji wielomodalnej.....	49
4. PROBLEM BADAWCZY.....	50
4.1. Wyzwanie: niekompletność danych.....	51
4.2. Wyzwanie: różny zakres stosowalności modalności.....	52
4.3. Wyzwanie: określenie niepewności związanej z pomiarem symptomów emocji oraz wyniku klasyfikacji.....	54
4.4. Wyzwanie: rozbieżności hipotez automatycznego rozpoznawania emocji w tym samym kontekście użycia.....	55
4.5. Wyzwanie: synchronizacja czasowa danych wejściowych oraz programów automatycznego rozpoznawania emocji.....	56
4.6. Podsumowanie analizy stanu wiedzy.....	57
5. PROPONOWANE ROZWIĄZANIE.....	58
5.1. Konceptyjny model rozwiązania.....	58
5.2. Architektura rozwiązania oparta o wzorzec projektowy Scoreboard.....	60
5.3. Przykładowa implementacja rozwiązania – Monitor Emocji.....	66
5.4. Algorytm integracji bazujący na współczynnikach wiarygodności.....	68
5.5. Funkcja oceny programów automatycznego rozpoznawania emocji.....	78
5.6. Propozycja sposobu wyliczenia współczynnika niepewności dla wybranych modalności.....	80
5.7. Zestawienie problemów i propozycji rozwiązania – podsumowanie.....	83



5.8. Tezy	86
6. OCENA ARCHITEKTURY ZAPROPONOWANEGO ROZWIĄZANIA Z WYKORZYSTANIEM WZORCA ARCHITEKTONICZNEGO SCOREBOARD88	
6.1. Projekt badania.....	89
6.2. Eksperyment 1. – integracja danych wielokanałowych w oparciu o analizę ładunku emocjonalnego tekstu.....	91
6.3. Symulacja 1. – weryfikacja miary odporności na podstawie atrap programów rozpoznawania (ang. mock-up)	95
6.4. Studium przypadku Monitora Emocji.....	96
6.5. Podsumowanie wyników	101
7. EWALUACJA ZAPROPONOWANEJ METODY KWANTYFIKACJI INFORMACJI UMOŻLIWIAJĄCEJ REPREZENTACJĘ NIEPEWNOŚCI ...	103
7.1. Konwencje notacyjne.....	104
7.2. Wybór zbioru danych SEMAINE	106
7.3. Wykorzystane programy automatycznego rozpoznawania emocji	107
7.4. Analiza danych SEMAINE	107
7.5. Eksperyment 2. Walidacja metody wyznaczania współczynnika wiarygodności	111
7.6. Eksperyment 3. Dobór współczynnika wiarygodności w przypadku jego braku	114
7.7. Podsumowanie walidacji zaproponowanej metody kwantyfikacji informacji	121
8. EWALUACJA METODY INTEGRACJI DANYCH Z WYKORZYSTANIEM NIEPEWNOŚCI.....	122
8.1. Opis rozwiązań referencyjnych.....	123
8.2. Eksperyment 4. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na zbiorze danych SEMAINE	123
8.3. Eksperyment 5. Odpowiednia konfiguracja funkcji integracji – dobór odpowiednich parametrów.....	127
8.4. Eksperyment 6. Porównanie skuteczności integracji zaproponowanej metody z rozwiązaniami jednomodalnymi oraz referencyjnymi na zbiorze danych SEMAINE	136
8.5. Eksperyment 7. Porównanie poprawności wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania przez zaproponowaną metodę oraz rozwiązania referencyjne	141
8.6. Podsumowanie eksperymentów na zbiorze danych SEMAINE	146
9. SYMULACJE SYTUACJI GRANICZNYCH	148
9.1. Symulacja 2. Analiza zmiany wartości współczynnika wiarygodności dla rozwiązań zbieżnych i rozbieżnych.....	148
9.2. Symulacja 3. Współczynnik wiarygodności uzgodnionego rozwiązania dla zmiennej wartości współczynnika wiarygodności hipotez.....	151
9.3. Symulacja 4. Analiza wpływu wartości wiarygodności rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania	155
9.4. Symulacja 5. Analiza wpływu odległości rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania	157
9.5. Podsumowanie symulacji	158
10. EWALUACJA ZAPROPONOWANEGO ROZWIĄZANIA Z WYKORZYSTANIEM BADANIA Z CZTEREMA KAMERAMI.....	160
10.1. Opis przebiegu badania z czterema kamerami	160
10.2. Eksperyment 8. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na danych z eksperymentu z 4 kamerami....	162
10.3. Określenie poziomu zbieżności danych dla eksperymentu z Politechniki Gdańskiej	

dla czterech kamer	163
10.4. Eksperyment 9. Porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez	166
10.5. Podsumowanie eksperymentów	171
11. PODSUMOWANIE	172
11.1. Teza pomocnicza 1	173
11.2. Teza pomocnicza 2	174
11.3. Teza pomocnicza 3	175
11.4. Teza główna oraz osiągnięcia pracy	177
11.5. Kierunki dalszych prac.	180
LITERATURA	182
WYKAZ RYSUNKÓW	190
WYKAZ TABEL	191
WYKAZ ZAŁĄCZNIKÓW	194

1. WPROWADZENIE

1.1. *Motywacja podjęcia pracy*

Przetwarzanie emocji (ang. affective computing) to obszar informatyki zajmujący się metodami rozpoznawania, reprezentacji oraz przetwarzania informacji o stanach emocjonalnych użytkowników komputerów. Dziedzina ta ma wiele zastosowań, takich jak terapia osób dotkniętych zaburzeniami (m.in. autyzmem), monitorowanie chorych (szpitale, rehabilitacja), pomiar użyteczności produktów i skuteczności reklam, poprawa narzędzi oraz materiałów nauczania na odległość, a także zastosowania w sektorze bezpieczeństwa (np. automatyczne wykrywanie agresywnych zachowań na podstawie analizy ładunku emocjonalnego tekstu czy obrazu wideo).

Współcześnie dziedzina ta wypracowała wiele programów rozpoznających stany emocjonalne człowieka na podstawie różnych sygnałów, z których wydobywa się informacje o symptomach emocji. Sygnały te są określane mianem modalności i są to: mimika twarzy, mowa ciała, ton i barwa głosu, parametry fizjologiczne (temperatura, przewodność skóry, tętno) i wiele innych. Jednym z coraz częściej stosowanych podejść jest rozpoznawanie wielomodalne (wykorzystujące wiele różnych sygnałów). W podejściu tym pojawia się wyzwanie dotyczące integracji danych, zarówno z poszczególnych kanałów obserwacji, jak i danych o rozpoznanych stanach emocjonalnych.

Aktualny stan wiedzy w dziedzinie przetwarzania emocji, będący podstawą niniejszej rozprawy, można podsumować następująco:

- istnieje wiele programów automatycznego rozpoznawania stanów emocjonalnych człowieka, wykorzystujących różne kanały obserwacji, o różnej dokładności (ang. accuracy);
- istnieje wiele modeli reprezentacji emocji wykorzystywanych na potrzeby przetwarzania komputerowego, w tym dyskretne modele reprezentacji emocji (pojedyncze emocje odróżnialne od siebie) oraz ciągłe modele reprezentacji emocji;
- istnieją metody odwzorowania między modelami reprezentacji emocji, z zastrzeżeniem, że nie dla wszystkich par modeli istnieją takie rozwiązania;
- istnieje wiele metod integracji bazujących zarówno na integracji danych bezpośrednio z kanałów wejściowych czy na wynikach innych programów – na danych częściowo przetworzonych, jak i podejścia mieszane.

Stan wiedzy w zakresie wielomodalnego rozpoznawania emocji przedstawiono szerzej w rozdziale 2. niniejszej rozprawy.

Analiza aktualnego stanu wiedzy oraz praktycznych problemów przy zastosowaniach rozpoznawania emocji pozwoliła na zaobserwowanie problemu badawczego.

W pracach eksperymentalnych prowadzonych na Wydziale ETI Politechniki Gdańskiej wielokrotnie zaobserwowano, że programy automatycznego rozpoznawania emocji dla tej samej sytuacji, osoby i czasu zwracają niespójne wyniki. Jest to związane z różnicami w dostępności (ang. availability), dokładności (ang. accuracy) oraz próbkowaniu poszczególnych kanałów obserwacji. Niespójności te zostały także potwierdzone w badaniach opisanych w literaturze przedmiotu.

Dodatkowo wyzwaniem jest reprezentacja stanów emocjonalnych na potrzeby ich przetwarzania, która odbywa się z użyciem różnych modeli reprezentacji emocji, na różnym poziomie granulacji oraz z różną częstotliwością.

Niespójność informacji o stanie emocjonalnym człowieka jest wyzwaniem dla integracji i uzyskania wiarygodnej informacji o stanie emocjonalnym. Ten zaobserwowany problem badawczy stał się podstawą niniejszej rozprawy doktorskiej, której celem stało się opracowanie metody integracji w wielokanałowej lub wielomodalnej obserwacji symptomów emocji, uwzględniającej niespójność i wiarygodność (ang. plausibility) integrowanych danych.

Problem badawczy, jakiemu jest poświęcona niniejsza rozprawa, przedstawiono szerzej w rozdziale 4. niniejszej rozprawy.

Zagadnienie integracji w obszarze wielomodalnego rozpoznawania emocji jest powiązane z kilkoma lukami badawczymi, takimi jak:

- reprezentacja niepewności związanej z niedoskonałością metod obserwacji symptomów emocji;
- integracja uwzględniająca okresową niedostępność poszczególnych kanałów obserwacji i niepewność obserwacji;
- rozwiązania pozwalające na synchronizację strumieni symptomów (z kanałów obserwacji) i estymat stanu emocjonalnego (z poszczególnych programów);
- zunifikowany model reprezentacji emocji umożliwiający integrację przy różnorodności zastosowanych technik i w różnych kontekstach zastosowań.

1.2. **Cele pracy**

Celem pracy doktorskiej jest zaproponowanie metody integracji danych o stanie emocjonalnym użytkowników komputerów, która uwzględni różne programy, kanały obserwacji oraz modele reprezentacji emocji, a także kontekstową niepewność pomiarów symptomów emocji. Cel nadrzędny można zdekomponować następująco:

- zaproponowanie metod wyznaczania niepewności pomiarowej związanej z kanałem wejściowym na podstawie danych dostępnych w czasie trwania pomiaru;
- zaproponowanie metody integracji bazującej na współczynniku niepewności wyznaczanym dynamicznie w czasie działania programu;
- przeprowadzenie analizy zaproponowanej metody integracji oraz wyników z poszczególnych programów automatycznego rozpoznawania emocji.

1.3. *Tezy rozprawy*

Teza niniejszej rozprawy została zatytułowana w następujący sposób:

Zaproponowana metoda integracji danych o stanie emocjonalnym umożliwia poprawę wybranych atrybutów jakości wielomodalnego rozwiązania automatycznego rozpoznawania emocji.

Wraz z tezą główną zostały także przygotowane trzy tezy pomocnicze, sformułowane następująco:

- ▶ Teza pomocnicza 1: Zastosowanie wzorca architektonicznego Scoreboard pozwala na poprawę modularności, łatwości integracji, adaptowalności i odporności w wielokanałowym rozpoznawaniu emocji.
- ▶ Teza pomocnicza 2: Zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwia reprezentację niepewności oszacowania stanu emocjonalnego.
- ▶ Teza pomocnicza 3: Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwia poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

1.4. *Opis zawartości pracy*

W rozdziale drugim przedstawiony został stan wiedzy dotyczący automatycznego przetwarzania emocji. Zostały w nim przedstawione modele reprezentacji emocji, ich odwzorowania, a także przykłady programów automatycznego rozpoznawania emocji podzielone na poszczególne modalności. Na koniec rozdziału zostały przedstawione zastosowania przetwarzania emocji oraz problemy badawcze związane z rozpoznawaniem emocji.

W kolejnym rozdziale (trzecim) opisano stan wiedzy o metodach integracji wielomodalnej, w szczególności model wczesnej i późnej fuzji. W rozdziale przedstawiono ich wady oraz zalety. Następnie, na podstawie aktualnej literatury, zostały przedstawione rodzaje metod integracji oraz znane rozwiązania fuzji w rozpoznawaniu emocji.

Rozdział czwarty zawiera opis aktualnych problemów badawczych dotyczących tematyki integracji w rozpoznawaniu emocji. Bazuje on częściowo na publikacjach [1] [2], których współautorem jest autor pracy. Opisane w nim zostały wyzwania związane z brakami w danych, różnym zakresem stosowalności kanałów wejściowych, rozbieżnościami wyników pomiędzy różnymi programami, brakiem wskazania poziomu wiarygodności dla hipotez o stanie emocjonalnym, a także kwestie odwzorowania pomiędzy różnymi modelami reprezentacji emocji.

Piąty rozdział pracy prezentuje proponowane przez autora pracy rozwiązanie. Bazuje on na publikacjach [1] [2] [3] [4] oraz dokumencie technicznym [5], których współautorem jest autor pracy. W rozdziale został przedstawiony koncepcyjny model rozwiązania, a także architektura rozwiązania oparta o wzorzec Scoreboard, wraz z przykładową implementacją wykonaną przez autora pracy. Następnie zaprezentowana została metoda integracji CFIEE oraz propozycja sposobu wyliczenia współczynnika niepewności dla różnych modalności. W podsumowaniu rozdziału przygotowano zestawienie problemów i propozycję ich rozwiązań. Na końcu rozdziału została powtórzona teza pracy doktorskiej wraz z listą planowanych eksperymentów.

Szósty rozdział bazuje na publikacji [4], której współautorem jest autor pracy. Zawiera on plan badań, a także wyniki przeprowadzonej symulacji, eksperymentu oraz studium przypadku Monitora Emocji. Zbadano w nim cztery zaproponowane miary jakości dla rozwiązania wykorzystującego wzorzec architektoniczny Scoreboard. Rozdział kończy się potwierdzeniem poprawności pierwszej tezy pomocniczej.

Siódmy rozdział zawiera opis przeprowadzonych analiz zaproponowanej metody kwantyfikacji informacji o danym kanale obserwacji. Opisano w nim wykorzystany zaetykietowany emocjonalnie zbiór danych SEMAINE oraz wykorzystywane programy automatycznego rozpoznawania emocji. Rozdział zawiera plan oraz przebieg dwóch eksperymentów. Rozdział kończy się potwierdzeniem poprawności drugiej tezy pomocniczej.

W ósmym rozdziale opisano plany, przebieg oraz analizę eksperymentów przeprowadzonych na zaetykietowanym emocjonalnie zbiorze danych SEMAINE z perspektywy zagadnień trzeciej tezy pomocniczej. Omówiono w nim kwestie niezawodności oraz wiarygodności zaproponowanej metody integracji z wykorzystaniem reprezentacji niepewności pomiarowej. Rozdział dziewiąty kontynuuje analizę poprzez sprawdzenie warunków brzegowych poprzez cztery symulacje.

Dziesiąty rozdział również adresuje problematykę podjętą w trzeciej tezie pomocniczej. Przeprowadzone w nim eksperymenty bazują na badaniu przeprowadzonym na Politechnice Gdańskiej z wykorzystaniem czterech kamer. W badaniu analizowane są przede wszystkim wzrost niezawodności oraz powiązanie zaproponowanego sposobu wyznaczania współczynnika wiarygodności z poziomem zbieżności danych. Rozdział kończy się potwierdzeniem poprawności trzeciej tezy pomocniczej, a więc i głównej tezy rozprawy.

Ostatni rozdział zawiera podsumowanie osiągniętych rezultatów pracy oraz kierunki dalszego rozwoju.

1.5. **Słownik pojęć**

W poniższym podrozdziale zostały przedstawione najważniejsze pojęcia wykorzystywane w niniejszej pracy. Kolejność pojęć nie jest alfabetyczna, a została pogrupowana i ułożona od najbardziej ogólnych do najbardziej specjalistycznych.



Przetwarzanie emocji (ang. affective computing) – nazywane także informatyką afektywną to „dział informatyki, zajmujący się metodami i narzędziami rozpoznawania, analizy, interpretacji i symulacji stanów emocjonalnych użytkowników komputerów” [6]. Jest to dziedzina funkcjonująca na przecięciu informatyki, psychologii, fizjologii oraz nauk kognitywnych [7].

Emocje – są to silne i nietrwałe stany afektywne, wywołane najczęściej konkretnym bodźcem np. zdarzeniem lub refleksją. Najczęściej znikają szybko po ustaniu bodźca. Emocje zależą od stanu organizmu i mogą zmieniać się nawet w ciągu kilku minut. Przykłady emocji to np. radość, złość, strach. Emocji nie należy utożsamiać z uczuciami – tymi dziedziną przetwarzania emocji przez komputery się nie zajmuje [7].

Nastrój – stan afektywny o średnim czasie trwania, dłuższym niż dla emocji, ale o słabszym działaniu. Wyzwalany jest on większymi wydarzeniami takimi jak traumatyczne przeżycie czy osiągnięcie jednego z życiowych celów. Nastrojem może być drażliwość (łatwe wpadanie w gniew) czy obniżony nastrój (z poczuciem smutku).

Osobowość – nazywana także temperamentem. Osobowością psychologia nazywa całą mentalną działalność człowieka, w tym perspektywę i poglądy. W kontekście emocji osobowość wyraża się w skłonności do reagowania określonymi emocjami na określone bodźce. Jest to afekt o najsłabszym, ale i najdłuższym działaniu – osobowość jest stabilna w okresie wielu lat, a czasem nie zmienia się przez całe życie.

Prawdziwy stan emocjonalny (ang. ground truth) – rzeczywisty stan emocjonalny człowieka. W literaturze wyróżnia się kilka możliwości określania „prawdziwego” stanu emocjonalnego (ang. ground truth): raportowanie własnych emocji przez badanego (ang. self-reports), manualne oznaczanie obserwacji lub danych fizjologicznych przez ekspertów oraz analiza wielomodalna wraz ze sprawdzeniem spójności [8] [9].

Rozpoznawanie emocji – proces nazywania emocji na podstawie ich zewnętrznych symptomów. Manualny (etykietowanie przez ludzi) lub automatyczny (rozpoznanie przez programy) [7].

Automatyczne rozpoznawanie emocji – „Dział informatyki zajmujący się tworzeniem algorytmów rozpoznających stany emocjonalne na podstawie zewnętrznych symptomów, pozyskanych za pomocą różnych urządzeń i programów monitorujących. Często stosuje techniki uczenia maszynowego” [7].

Program automatycznego rozpoznawania emocji – program komputerowy, który na podstawie obserwacji jedno lub wielomodalnych, jedno lub wielokanałowych, zwraca hipotezę stanu emocjonalnego w konkretnym modelu reprezentacji emocji. Autor jest świadomy różnicy pomiędzy definicją programu i algorytmu, jednakże na potrzeby tej pracy określenia te będą używane zamiennie.

Facial Action Coding System (FACS) – jest to taksonomia ruchów mięśni twarzy (ang. action units — AU). Emocje są rozpoznawane jako kombinacja tych ruchów.

Prozodia mowy – ogół brzmieniowych właściwości języka, które różnicują, postacują i członują głoskowy, sylabiczny i wyrazowy strumień mowy [10]. Należą do nich akcent, intonacja i iloczas.

Czujnik – lub sensor, jest to urządzenie służące do pozyskania sygnału np. termometr, kamera, mikrofon, opaska biometryczna czy czujnik nacisku [7]

Kanał obserwacji (ang. observation channel) – medium rejestracji sygnału zawierającego informację o obserwowalnym symptomie stanu emocjonalnego. Przykładem jest np. strumień wideo z konkretnej kamery oraz strumień danych z konkretnego czujnika [7]. System bazujący tylko na jednym kanale nazywamy jednokanałowym, a na wielu kanałach – wielokanałowym.

Modalność (ang. modality) – rodzaj informacji dotyczącej konkretnego obserwowanego symptomu emocji, która jest analizowana w celu oszacowania stanu emocjonalnego. Modalnością jest np. mimika twarzy, prozodia mowy, wzorce gestów, rytm serca [7]. System bazujący tylko na jednej modalności nazywamy jednomodalnym (ang. unimodal), a na kilku – wielomodalnym [11]

Model reprezentacji emocji – jest to ustrukturyzowany sposób zapisu stanu emocjonalnego na potrzeby przetwarzania przez aplikacje komputerowe lub człowieka [7].

Model emocji podstawowych Ekmana [12] – jest jednym z najczęściej stosowanych modeli emocji podstawowych. Przyjmuje on zbiór sześciu emocji, za pomocą których opisywane są pozostałe emocje. Zbiór ten składa się z emocji: **radości** (ang. happy), **złości** (ang. anger), **wstrętu** (ang. disgust), **strachu** (ang. fear), **smutku** (ang. sadness), oraz **zaskoczenia** (ang. surprise). Model ten nazywany jest czasem skrótowo **modelem Ekmana**. Na potrzeby niniejszej pracy nazwy te są używane zamiennie.

Model 2D – dwuwymiarowy model reprezentacji emocji, opisujący stan emocjonalny za pomocą wymiaru **wartościowości** (ang. valence lub pleasantness, rozróżnia emocje pozytywne od negatywnych) oraz wymiaru **pobudzenia** (ang. arousal, określa natężenie emocji).

Model PAD – trójwymiarowy model opisu temperamentu, nastroju oraz emocji stworzony przez Mehrabian'a i Russell'a [13]. Nazywany również modelem 3D opisujący stan emocjonalny za pomocą wymiarów: **wartościowość–pobudzenie–dominacja** (ang. Pleasure–Arousal–Dominance), przy czym wymiar dominacji określa stopień reakcji walki lub ucieczki.

Model 4D – czterowymiarowy model reprezentacji emocji, który oprócz wymiarów modelu PAD, dodawany jest czwarty wymiar, przy czym w literaturze zaproponowano różne wymiary dodatkowe np. wymiar nieprzewidywalności (ang. unpredictability) [14].

Pojęcia związane z integracją

Metoda fuzji – jest to scenariusz integracji danych z wielu modalności lub wielu kanałów obserwacji, który pozwala uzyskać jedną odpowiedź systemu na temat szacowanego stanu emocjonalnego.

Metoda wczesnej fuzji (ang. early fusion) – jest to klasa metod fuzji, w których dane integrowane są na poziomie danych z kanałów obserwacji, tworząc wspólny wektor cech.

Metoda późnej fuzji – jest to klasa metod fuzji, w których dane integrowane są na poziomie wyników klasyfikacji (hipotez o stanie emocjonalnym), otrzymanych z programów automatycznego rozpoznawania emocji. Nazywana jest również fuzją na poziomie decyzyjnym.

Ekspert – w tej pracy przyjęto zgodnie z literaturą [15] [16] [17] [18] [19]: klasyfikator lub inny program automatycznego rozpoznawania emocji, wykorzystywany w trakcie metody późnej fuzji lub fuzji hybrydowej, który na podstawie danych wejściowych o symptomach emocji zwraca hipotezę o stanie emocjonalnym.

Hipoteza o stanie emocjonalnym, Estymata, estymowany stan emocjonalny, uzyskany rezultat klasyfikacji, hipoteza o rezultacie klasyfikacji (ang. obtained classification result, classification hypothesis, obtained/estimated emotional state) – rezultat działania programu automatycznego rozpoznawania stanu emocjonalnego. W kontekście integracji danych będzie nazywany skrótowo *hipotezą* programu automatycznego rozpoznawania emocji lub *hipotezą eksperta*. Dla programów klasyfikujących wynik działania to hipoteza (klasa), natomiast dla programów estymacji jest to wartość liczbową w pewnej skali (estymata), jednak w przetwarzaniu emocji często stosuje się te dwa terminy zamiennie [18] [19].

Uzgodnione rozwiązanie – rezultat działania fuzji, działający na podstawie cząstkowych informacji (danych lub hipotez o stanie emocjonalnym).

Pojęcia związane z miarami sprawności i klasyfikatorów

Ze względu na potrzebę precyzji językowej oraz zapewnienie dobrego zrozumienia, poniżej zostały przedstawione najważniejsze miary wraz z ich angielskimi odpowiednikami.

Granulacja (ang. granularity) – również ziarnistość, poziom szczegółowości reprezentacji informacji [7] np. precyzja podawanych estymat.

Miary skuteczności klasyfikatorów (ang. classifiers effectiveness measures) – miary skuteczności klasyfikacji, odpowiadają na pytanie: „Czy klasyfikator robi to, co powinien robić?”, czyli czy skutecznie odróżnia klasy.

Miary sprawności klasyfikatorów (ang. classifiers efficiency measures) – miary sprawności odpowiadają na pytanie: „Czy klasyfikator wykonuje swoją pracę sprawnie/wydajnie (performance)?”

Niezawodność (ang. reliability) – cecha jakości (np. programu) polegająca na ciągłości dostarczenia jakiegoś wyniku (bez odzwierciedlenia jego jakości) [20].

Klasyfikacja (ang. classification) – Problem polegający na rozróżnieniu obiektów świata na podstawie różnic między nimi poprzez przypisanie ich do grup określających ich cechy (klas).

Macierz pomyłek (ang. confusion matrix) – Macierz zawierająca liczbowe ujęcie przypadków: poprawnie i niepoprawnie sklasyfikowanych z podziałem na klasy.

Estymacja (ang. estimation) – problem polegający na oszacowaniu wartości pewnej zmiennej.

Estymator (ang. estimator) – program dokonujący estymacji.

Miary jakości:

Modularność (ang. modularization) – cecha architektury pozwalająca na odseparowanie logicznych jednostek rozwiązania (modułów, komponentów). Literatura wskazuje na wysoką spójność (ang. cohesion) i niski poziom powiązań (ang. coupling) jako standardowe kryterium ewaluacji modularności [21].

Integracja (ang. integration) – cecha architektury reprezentująca zdolność do połączenia wyników z różnorodnych rozwiązań (tu: programów automatycznego rozpoznawania emocji). To kryterium powinno być rozumiane jako zdolność do integracji oraz otrzymania wyników nie gorszych od integrowanych rozwiązań.

Adaptowalność (ang. adaptability) – cecha architektury reprezentująca zdolność do łatwego dołączania/odłączania/podmiany ekspertów (tu: programów automatycznego rozpoznawania emocji).

Odporność (ang. robustness) – cecha programu automatycznego rozpoznawania emocji, wykorzystującego metodę integracji. Jeżeli do integracji zostanie wykorzystany niepoprawny ekspert (program zwracający niepoprawne wyniki) – powinien zostać wykryty, a jego wpływ na uzgodnione rozwiązanie powinien być ograniczony.

Niepewność (ang. uncertainty) – cecha zjawiska, mówiąca o tym, że nie można przewidzieć stanu zjawiska ze względu na zmienność jego samego lub brak wiedzy czy informacji. Każdy pomiar wiąże się z niepewnością [22]. W języku technicznym stosuje się wiele terminów w celu określenia problemów związanych z niepewnością i wiarygodnością pomiaru m.in.: niedokładność (ang. inaccuracy), nieprecyzyjność (ang. imprecision), błąd oraz ich przeciwieństwa: dokładność (ang. accuracy), pewność (ang. confidence), wiarygodność (ang. plausibility), zaufanie (ang. trust) itp. [20].

Wiarygodność (ang. plausibility) – W dziedzinie rozpoznawania emocji jednym z kluczowych ryzyk jest uznanie niepewnego wyniku za wiarygodny (ryzyko wiarygodności) [9]. Ryzyko wiarygodności można ograniczyć poprzez podanie wyniku (oszacowanego stanu emocjonalnego) wraz z pewnością tego wyniku. **W tej pracy** miara pewności wyniku rozpoznawania emocji jest określana w uproszczeniu jako jego wiarygodność.

Zaproponowane miary:

Współczynnik D_{max} (ang. distance treshold) – progowa wartość odległości, dla której hipoteza o stanie emocjonalnym (rezultat działania programu automatycznego rozpoznawania stanu emocjonalnego) uznana jest za zbieżną z inną hipotezą.

1.6. **Konwencje notacyjne**

Sposób redakcji poszczególnych elementów jest zgodny z rozporządzeniem Rektora PG nr 49/2014 z 5 grudnia 2014 r. Fragmenty kodów źródłowych oznaczone są w postaci Listingów, pisanych czcionką Courier New, rozmiaru 9.

Szczegółowe konwencje notacyjne dotyczące tabel z wynikami podano na początku rozdziału 7.

2. PRZETWARZANIE EMOCJI – ANALIZA ISTNIEJĄCEGO STANU WIEDZY

Niniejsza praca badawcza bazuje na dotychczasowych osiągnięciach naukowych w zakresie przetwarzania emocji przez aplikacje komputerowe oraz integracji danych. W tym rozdziale przedstawiono zestawienie badań i informacji, które były fundamentem, pozwalającym na przeprowadzenie prac nad integracją przetwarzania stanu emocjonalnego człowieka.

Przetwarzanie emocji (ang. affective computing) – nazywane także informatyką afektywną to „dział informatyki, zajmujący się metodami i narzędziami rozpoznawania, analizy, interpretacji i symulacji stanów emocjonalnych użytkowników komputerów” [6]. Jest to dziedzina funkcjonująca na przecięciu informatyki i psychologii.

Przetwarzanie emocji obejmuje [6]:

- metody rozpoznawania stanów emocjonalnych użytkowników komputerów i urządzeń mobilnych na podstawie różnorodnych symptomów (mimika twarzy, postawa ciała, parametry fizjologiczne, głos, tekst, wzorce behawioralne, analiza standardowego wejścia – mysz i klawiatura);
- modele reprezentacji stanów emocjonalnych na potrzeby przetwarzania komputerowego;
- tworzenie aplikacji afektywnych (ang. affective) i postrzegających emocje (ang. affect-aware) [23];
- modele interwencji afektywnych;
- analizę ładunku emocjonalnego tekstu (ang. sentiment analysis);
- symulacje stanów emocjonalnych na potrzeby wirtualnych postaci i robotów [24].

Rozpoznawanie emocji przez aplikacje komputerowe jest zagadnieniem złożonym. Również dla ludzi jest to trudne zadanie – wymaga dobrze rozwiniętej empatii i niesprzecznych symptomów. Ludzie często próbują (świadomie lub nieświadomie) ukryć swoje emocje, przykładowo poprzez sztuczny uśmiech (sugerujący radość), co może oznaczać w rzeczywistości zakłopotanie.

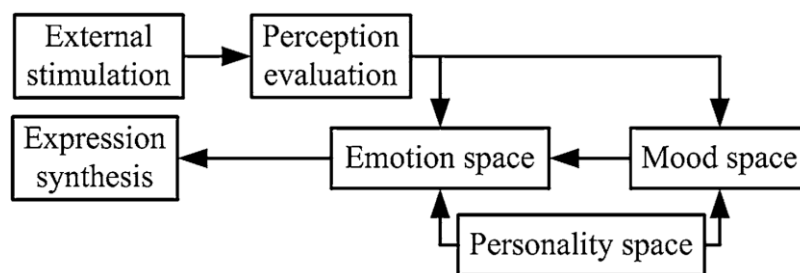
Innym wyzwaniem rozpoznawania emocji jest poprawne przedstawienie wykrytych stanów poprzez wybór odpowiedniego modelu reprezentacji emocji, w tym dobór miary i skali. Współcześnie w literaturze psychologicznej istnieje kilka różnych modeli, które zostały zaadaptowane na potrzeby przetwarzania emocji przez aplikacje komputerowe – modele te przybliżono w rozdziale 2.2.

2.1. Stan emocjonalny i jego obserwacje

Problemem poruszonym w tej pracy jest automatyczne rozpoznawanie stanów emocjonalnych przez systemy komputerowe. Warto podkreślić, że ten obszar badawczy wpisujący się w dyscyplinę informatyki ma interdyscyplinarny charakter i bazuje na osiągnięciach psychologii.

Stan emocjonalny ma w literaturze różne definicje, jednak rozróżnia się trzy pojęcia o różnym czasie trwania [25]: emocja (trwa najkrócej), nastrój (bardziej rozciągnięty w czasie) oraz osobowość. Zależności między tymi terminami pokazano na Rys 2.1. Osobowość wpływa zarówno na emocje, jak i na nastrój, a nastrój wpływa na emocje.

Emocje są to silne i nietrwałe stany afektywne, wywołane najczęściej konkretnym bodźcem np. zdarzeniem lub refleksją. Najczęściej znikają szybko po ustaniu bodźca. Emocje zależą od stanu organizmu i mogą zmieniać się nawet w ciągu kilku minut. Przykłady emocji są: radość, złość, strach. Emocji nie należy utożsamiać z uczuciami – tymi dziedziną przetwarzania emocji przez komputery się nie zajmuje [7].



Rys 2.1 Struktura emocji w procesie podejmowania decyzji [31]

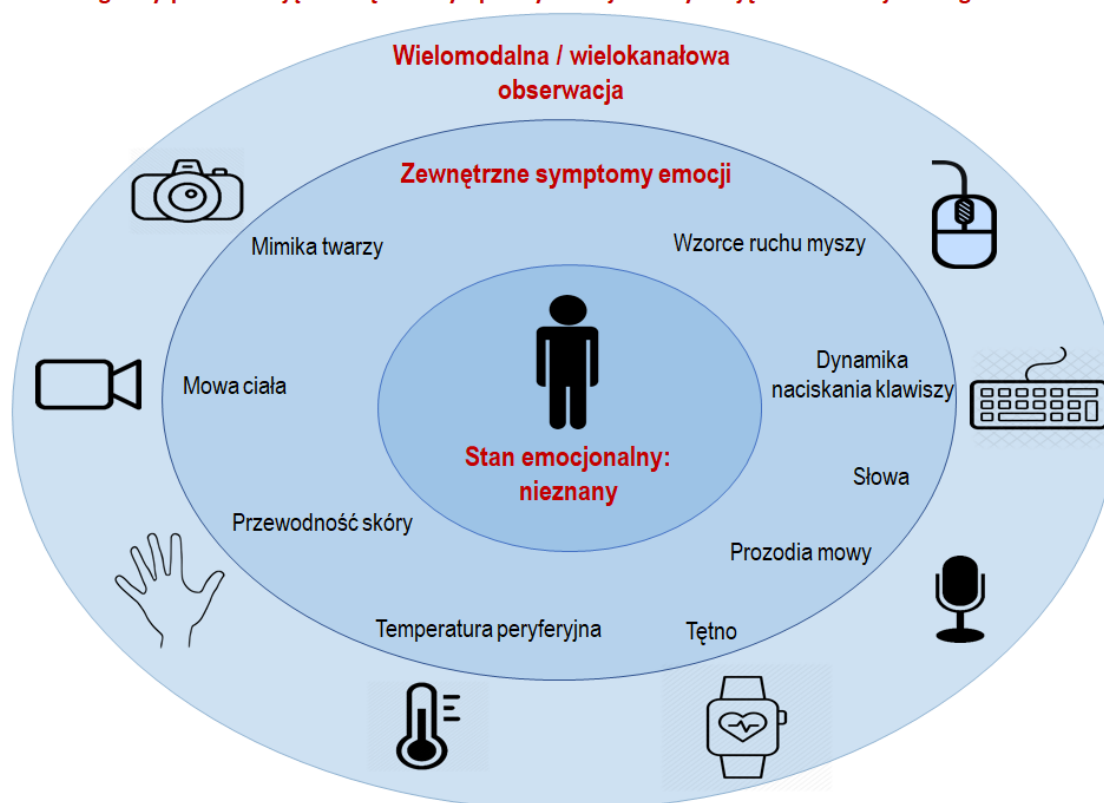
Nastrojem nazywamy stan afektywny o średnim czasie trwania (dłuższym niż dla emocji), ale słabszym działaniu. Wyzwalany jest on większymi wydarzeniami takimi jak traumatyczne przeżycie lub osiągnięcie jednego z życiowych celów. Nastrojem może być przykładowo drażliwość (łatwe wpadanie w gniew) czy obniżony nastrój (z poczuciem smutku).

Osobowością, nazywaną także temperamentem, psychologia nazywa całą mentalną działalność człowieka – w tym perspektywę i poglądy. W kontekście emocji osobowość wyraża się w skłonności do reagowania emocjami na określone bodźce. Jest to afekt o najsłabszym, ale i najdłuższym działaniu. Jest stabilna w okresie wielu lat, a czasem nie zmienia się przez całe życie – dlatego w kontekście rozpoznawania emocji często się przyjmuje ją jako stałą.

Zarówno ludzkie, jak i komputerowe wykrywanie stanów emocjonalnych bazuje jedynie na zewnętrznych symptomach emocji, takich jak mimika twarzy, mowa ciała, prozodia mowy, treść tekstu, dane fizjologiczne, takie jak temperatura czy przewodność skóry [26]. Te zewnętrzne symptomy stanów emocjonalnych można rejestrować za pomocą różnych urządzeń pomiarowych, a na podstawie wyodrębnionych informacji oszacowywać stan emocjonalny osoby. Kontekst przetwarzania emocji człowieka został pokazany na Rys 2.2. Program komputerowy na podstawie obserwacji zewnętrznych symptomów przetwarza je w oszacowanie stanu emocjonalnego. Dla usystematyzowania słownictwa, poniżej przedstawiono tłumaczenia na język polski kluczowych określeń:

- **Prawdziwy stan emocjonalny** (ang. ground truth) – rzeczywisty stan emocjonalny człowieka. W literaturze wyróżnia się kilka możliwości określania „prawdziwego” stanu emocjonalnego (ang. ground truth): raportowanie własnych emocji przez badanego (ang. self-reports), manualne oznaczanie obserwacji lub danych fizjologicznych przez ekspertów oraz analiza wielomodalna wraz ze sprawdzeniem spójności [8] [9]. Problematyka prawdziwego stanu emocjonalnego została szerzej poruszona w rozdziale 2.6.
- **Rozpoznawanie emocji** – proces nazywania emocji na podstawie ich zewnętrznych symptomów. Manualny (etykietowanie przez ludzi) lub automatyczny (rozpoznanie przez programy) [7].
- **Automatyczne rozpoznawanie emocji** – „Dział informatyki zajmujący się tworzeniem algorytmów rozpoznających stany emocjonalne na podstawie zewnętrznych symptomów, pozyskanych za pomocą różnych urządzeń i programów monitorujących. Często stosuje techniki uczenia maszynowego” [7].
- **Program automatycznego rozpoznawania emocji** – program komputerowy, który na podstawie obserwacji jedno lub wielomodalnych, jedno lub wielokanałowych, zwraca hipotezę stanu emocjonalnego w konkretnym modelu reprezentacji emocji.
- **Kanał obserwacji** (ang. observation channel) – medium rejestracji sygnału zawierającego informację o obserwowalnym symptomie stanu emocjonalnego. Przykładem jest m.in. strumień wideo z konkretnej kamery, strumień danych z konkretnego czujnika [7]. System bazujący tylko na jednym kanale nazywamy jednokanałowym, a na wielu kanałach – wielokanałowym.
- **Modalność** (ang. modality) – rodzaj informacji dotyczącej konkretnego obserwowanego symptomu emocji, która jest analizowana w celu oszacowania stanu emocjonalnego. Modalnością jest np. mimika twarzy, prozodia mowy, wzorce gestów, rytm serca [7]. System bazujący tylko na jednej modalności, nazywamy jednomodalnym (ang. unimodal), a na wielu informacjach – wielomodalnym [11].

Programy przetwarzają zewnętrzne symptomy emocji w estymację stanu emocjonalnego



Rys 2.2 Kontekst rozpoznawania emocji [opracowanie własne]

W praktyce istotne jest rozróżnienie między modalnością a kanałem obserwacji, zwłaszcza że w literaturze pojęcia te często są stosowane zamiennie. W Tabeli 2.1 pokazano różnicę między obserwacją wielokanałową a wielomodalną. Obserwacja wykorzystująca analizę mimiki twarzy z dwóch kamer to obserwacja dwukanałowa, ale jednomodalna. Jeżeli dla sygnału wideo z jednej kamery analizowane są symptomy emocji z mimiki twarzy, prozodii oraz mowy ciała, jest to analiza jednokanałowa, ale wielomodalna. Obserwacja wykorzystująca jedną kamerę i tekst pisany to obserwacja wielomodalna i wielokanałowa.

Tabela 2.1 Przykłady obserwacji jednomodalnych, wielomodalnych oraz jednokanałowych i wielokanałowych [opracowanie własne].

Obserwacja	Jednomodalna	Wielomodalna
Jednokanałowa	Analiza ekspresji mimicznej, z jednej kamery.	Analiza ekspresji mimicznej, gestów i prozodii mowy na podstawie sygnału z jednej kamery.
Wielokanałowa	Analiza ekspresji mimicznej, z dwóch kamer.	Wiele sygnałów (np. wiele kamer lub wiele czujników) i wiele analizowanych modalności (np. rytm serca, ekspresja mimiczna i wzorce gestów).

2.2 Modele reprezentacji emocji na potrzeby aplikacji komputerowych

W tym rozdziale przedstawiono kilka najczęściej wykorzystywanych modeli reprezentacji emocji w przetwarzaniu emocji – z racji mnogości i różnorodności modeli zostały przedstawione tylko wybrane, do których odwołuje się praktyczna część niniejszej pracy. Modele można podzielić ze względu na charakterystykę zmiennych na modele: dyskretne, modele emocji podstawowych, komponentowe oraz ciągłe, w których warto zwrócić uwagę na modele wielowymiarowe.

Modelem reprezentacji emocji jest nazywany ustrukturyzowany sposób zapisu stanu emocjonalnego pozwalający zapisać zaobserwowany lub rozpoznany stan emocjonalny.

Modele dyskretne reprezentacji emocji

Modele dyskretne charakteryzują się skończonym zbiorem stanów określających emocje. Najczęściej służą one do zapisu wykrytych emocji, określając czy emocja wystąpiła, czy nie. Jest to najprostszy sposób zapisu stanu emocjonalnego. Określa on zazwyczaj jednym lub dwoma słowami stan emocjonalny np. szczęśliwy, wesoły, smutny, przygnębiony. Jest to najbardziej intuicyjny sposób określenia emocji dla człowieka. Emocje w ujęciu dyskretnym najczęściej określane są poprzez etykietę przypisaną do okna czasowego, tworząc ciąg etykiet jako zapis stanu emocjonalnego osoby. W rozpoznawaniu emocji przy takim ujęciu sygnały są dzielone na próbki o przypisanych etykietach, a klasyfikatory są używane do rozpoznania tych dyskretnych emocji.

Do wad takiej reprezentacji emocji stanu emocjonalnego należy zaliczyć: brak możliwości obliczenia odległości jednej emocji od drugiej, trudność przedstawienia zmienności emocji w czasie, porównania i reprezentacji złożonych emocji w tym modelu. Dla potrzeb reprezentacji złożonych emocji psychologia definiuje pewne reguły, które najczęściej są wyprowadzane z eksperymentów psychologicznych np. połączenie emocji gniewu i zniesmaczenia to pogarda [27]. Zestaw reguł nie jest kompletny, co utrudnia jego zastosowanie w przetwarzaniu emocji przez aplikacje.

Modele emocji podstawowych

Jednym z podejść do definiowania stanów emocjonalnych jest model emocji podstawowych. Przyjmuje on, że wszystkie emocje da się opisać poprzez złożenie podstawowych emocji. Zestawienie kilku istniejących modeli emocji podstawowych zostało zaprezentowane w Tabeli 2.1. Najczęściej wykorzystywanym z nich jest model sześciu emocji podstawowych Ekmana.

Model emocji podstawowych Ekmana [12] przyjmuje zbiór sześciu emocji, za pomocą których opisywane są pozostałe emocje. Zbiór ten składa się z emocji: **radości** (ang. happy), **złości** (ang. anger), **wstrętu** (ang. disgust), **strachu** (ang. fear), **smutku** (ang. sadness) oraz **zaskoczenia** (ang. surprise). Model ten nazywany jest czasem skrótowo modelem Ekmana – na potrzeby niniejsze pracy nazwy te są używane zamiennie.

Tabela 2.2 Wybrane modele reprezentacji emocji, bazujące na emocjach podstawowych [opracowanie własne].

Model	Liczba emocji	Emocje podstawowe
Ekman [12]	6	anger, disgust, fear, joy, sadness, surprise
Parrot [28]	6	anger, fear, joy, love, sadness, surprise
Frijda [28]	6	desire, happiness, interest, surprise, wonder, sorrow
Plutchik [29] [30]	8	acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Tomkins [28]	9	desire, happiness, interest, surprise, wonder, sorrow
Matsumoto [28]	22	joy, anticipation, anger, disgust, sadness, surprise, fear, acceptance, shy, pride, appreciate, calmness, admire, contempt, love, happiness, exciting, regret, ease, discomfort, respect, like

W badaniach ten zbiór podstawowy bywa rozszerzany o inne emocje, przykładowo w robocie WE-3RV dodano emocje: neutralną, pijaną (ang. drunk) oraz nieśmiałości (ang. shy) [31]. W celu uproszczenia modelu czasem pomijane są emocje zaskoczenia oraz wstrętu [31].

Reprezentacja stanów emocjonalnych człowieka za pomocą zbioru emocji podstawowych może być oznaczana binarnie (emocja jest/nie ma). W zaawansowanych modelach reprezentacji dla każdej z emocji podstawowych określana jest jej intensywność w sposób ciągły w skali od 0 do 1, co ułatwia integrację wyników z różnych programów automatycznego rozpoznawania emocji. Model emocji podstawowych Ekmana jest powszechnie wykorzystywany zwłaszcza podczas analizy mimiki twarzy, ze względu na opracowane przypisanie akcji mięśni mimicznych do stanów emocjonalnych (ang. Facial Action Coding System – FACS).

Modele ciągłe

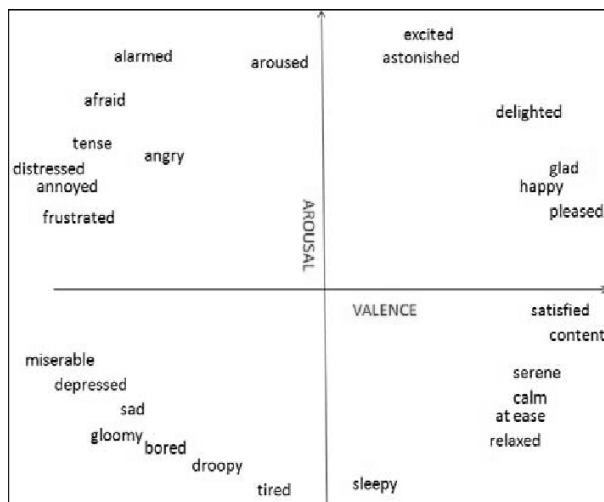
Dane o stanie emocjonalnym w modelach ciągłych mają charakter zmiennych ciągłych, dzięki czemu można prowadzić na nich operacje matematyczne takie jak dodawanie, odejmowanie, obliczanie odległości czy średniej. Wśród modeli ciągłych warto zwrócić uwagę na modele wielowymiarowe, w których stan emocjonalny jest opisywany przez kilka niezależnych zmiennych.

Modele wielowymiarowe opisują stany emocjonalne na podstawie wymiarów, które są elementami opisu stanów emocjonalnych, ale same nie reprezentują wprost emocji. Stan emocjonalny w modelu wielowymiarowym reprezentowany jest jako punkt w wielowymiarowej przestrzeni. Z powodów teoretycznych, jak i praktycznych coraz więcej badaczy sięga po modele wielowymiarowe [28]. Wadą tych modeli jest możliwość opisanie tylko jednej emocji na raz.

Model dwuwymiarowy reprezentuje emocje jako punkty na dwuwymiarowej powierzchni, opisując je za pomocą dwóch zmiennych, które są różnie nazywane w literaturze:

- **wartościowość** (ang. valence), inaczej: przyjemność (ang. pleasure) lub ocena (ang. evaluation), która może być pozytywna albo negatywna,
- **pobudzenie** (ang. arousal), inaczej: aktywność (ang. activation) lub intensywność (ang. intensity), która może być wysoka albo niska.

Model dwuwymiarowy czasem jest określany nazwą modelu Russela, koła Whissel (ang. whissel wheel) [32] lub w skrócie modelem 2D. Wizualizację rozmieszczenia emocji w modelu 2D przedstawia Rys 2.3 [33]. Na osi pionowej reprezentowany jest wymiar pobudzenia, a na osi poziomej wartościowości. Wykres przedstawia orientacyjne położenie emocji (etykiety) na płaszczyźnie modelu. Opis za pomocą dwóch wymiarów pozwala na wyliczenie odległości pomiędzy stanami emocjonalnymi. Jest to jeden z najczęściej wykorzystywanych modeli wielowymiarowych. Wadą tego modelu jest brak rozróżnienia niektórych emocji takich jak gniew i strach, które znajdują się blisko siebie – obie są negatywne i mają wysokie pobudzenie.



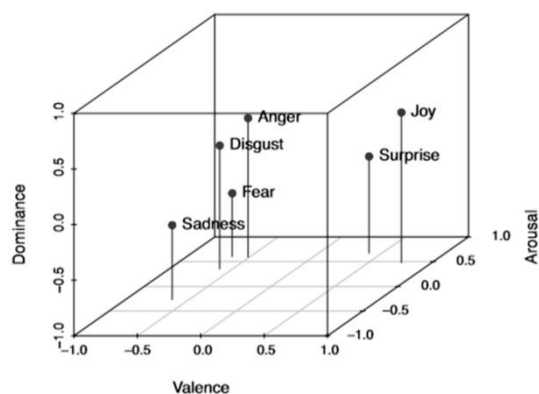
Rys 2.3 Model przestrzeni wartość–pobudzenie [33]

Model trójwymiarowy PAD to model opisu temperamentu, nastroju oraz emocji stworzony przez Mehrabian'a i Russell'a [13], nazywany również modelem 3D. W tym modelu stan emocjonalny jest opisywany za pomocą wymiarów: **przyjemności** (ang. pleasure, wartości pozytywne i negatywne), **pobudzenia** (ang. arousal, określa natężenie emocji) oraz **dominacji** (ang. dominance, określa stopień wpływu na otoczenie – reakcja walki lub ucieczki).

Poszczególne wymiary miewają także inne nazwy [13] [34]:

- przyjemności (ang. pleasure) odpowiadają także ocena (ang. evaluation) lub wartościowość (ang. valence);
- pobudzeniu (ang. arousal) odpowiada aktywacja (ang. activation) lub intensywność (ang. intensity);
- dominacji odpowiada czasami moc (ang. power).

Model ten, podobnie jak model 2D, reprezentuje stan emocjonalny jako punkt w przestrzeni trójwymiarowej, co zostało zobrazowane na Rys 2.4 [35]. Model ten jest kompatybilny z modelem 2D, przy przyjęciu założenia zerowej wartości wymiaru dominacji. W niektórych badaniach z dziedziny psychologii proponuje się rozszerzenie modelu 3D o dodatkowy wymiar, np. wymiar nieprzewidywalności (ang. unpredictability) [14].



Rys 2.4 Emocje podstawowe z modelu reprezentacji emocji podstawowych Ekmana w przestrzeni modelu reprezentacji emocji PAD [35]

Modele komponentowe

Model OCC (Ortony, Clore, & Collins) koncentruje się na procesie powstawania emocji. Zakłada, że emocje powstają w odpowiedzi na zdarzenia w środowisku oraz interakcje między przedmiotami i osobami [36]. Wyróżnia 22 emocje. Przyczyny emocji dzieli na trzy kategorie: wyniki zdarzeń, reakcje agenta oraz postrzeganie obiektu. Jest to powszechnie znany model, który doczekał się wielu publikacji i implementacji [31]. Ze względu na złożoność modelu, porównywanie stanów emocjonalnych w nim jest utrudnione. Model ten jest rzadko wykorzystywany w rozwiązaniach analizy wielomodalnej.

Modele reprezentacji emocji – podsumowanie

Istnieje jeszcze wiele innych modeli reprezentacji emocji, takich jak model pięcioczynnikowy (ang. FFM – Five Factor Model), zwany także modelem wielkiej piątki (ang. big five) [37], europejska przestrzeń bazowych emocji sztucznej inteligencji (ang. european space based artificial emotion model) [31], model wielu składowych (ang. multi-componential model) [31] oraz model klepsydry (ang. hourglass model) [28]. Ze względu na jednostkowe lub tylko teoretyczne zastosowania w przetwarzaniu emocji nie zostały one tutaj opisane.

Pomimo tego, że istnieje wiele modeli reprezentacji emocji, żadnego z nich nie można wskazać jako typowe rozwiązanie dla wszystkich rodzajów zastosowań w dziedzinie rozpoznawania emocji. Wybór modelu zależy od zastosowania, co stanowi wyzwanie przy porównywaniu wyników pochodzących z różnych programów, stosujących odmienne modele reprezentacji emocji. Z tego powodu potrzebne są odwzorowania pomiędzy modelami reprezentacji emocji, które zostały opisane w kolejnym podrozdziale.

2.3. Odzworowania pomiędzy modelami reprezentacji emocji

Dziedzina przetwarzania emocji nie stosuje jednego modelu reprezentacji emocji, dlatego opracowane zostały metody odzworowania wybranych modeli w inne modele.

Istotnym problemem dotyczącym wszystkich tych przekształceń jest błąd odzworowania, który nawarstwia się przy wykorzystaniu każdego kolejnego odzworowania.

Znanych jest szereg odzworowań m.in. etykiet i modeli dyskretnych na modele wielowymiarowe lub model emocji podstawowych Ekmana, odzworowanie modelu emocji podstawowych Ekmana w model PAD, jak również odzworowania modelu OCC oraz modeli osobowości w model PAD [13] [31] [38] [39]. W niniejszej pracy dwa ostatnie zostały pominięte, a szerzej opisane zostały odzworowania wykorzystywane w dalszej części pracy.

Odzworowanie etykiet i modeli dyskretnych w modele wielowymiarowe

Najprostszym i najczęściej stosowanym sposobem odzworowania emocji z modeli dyskretnych jest przypisanie etykietom (słowom) odpowiednich punktów przestrzeni w modelu wielowymiarowym. Przykładowo, Hupont i inni [40] skorzystali z afektywnie zaetykietowanej bazy 9 tysięcy słów z języka angielskiego. Do każdego słowa baza posiadała przypisaną wartość ładunku emocjonalnego w modelu 2D.

Największą wadą tego rodzaju odzworowań jest ich jednokierunkowość. O ile można łatwo zamienić etykiety na punkty opisane zmiennymi ciągłymi, to odwrotna transformacja jest jedynie przybliżona – jeżeli wartości nie będą zgodne z referencyjnymi, możliwy jest tylko wybór najbliższego zaetykietowanego punktu.

Odzworowanie etykiet w model emocji podstawowych Ekmana

Zgodnie z badaniami z dziedziny psychologii, emocje podstawowe mogą posłużyć do reprezentacji bardziej złożonych stanów emocjonalnych tzw. emocji drugo- i trzeciorzędowych [27]. Emocje złożone są reprezentowane jako połączenie etykiet emocji podstawowych np. z połączenia emocji podstawowych złości (ang. anger) oraz niesmaku (ang. disgust) otrzymamy drugorzędową emocję pogardy (ang. contempt) [27]. Możliwe jest również przypisanie do emocji podstawowych intensywności i za pomocą kombinacji natężenia emocji podstawowych, przedstawienie danego złożonego stanu emocjonalnego. W praktyce jednak takie odzworowanie jest obarczone niedokładnością i wymaga podparcia teorią lub doświadczeniem z dziedziny psychologii.

Odzworowanie etykiet w model Ekmana można wykonać poprzez przypisanie wartości każdej z emocji podstawowej (ze stopniem intensywności lub bez). Rozszerzeniem tego przypisania może być wykorzystanie emocji drugo- i trzeciorzędnych, a także wykorzystanie synonimów. Należy jednak pamiętać, że takie odzworowanie jest obarczone dużą niedokładnością.

Odwzorowanie modelu emocji podstawowych Ekman w model trójwymiarowy

Dwoma z najczęściej wykorzystywanych modeli reprezentacji emocji są modele Ekmana oraz PAD, dlatego też odwzorowanie między tymi modelami jest istotne. Zespół badawczy z Chin [31] [38] zaproponował odwzorowanie modelu PAD w model Ekmana, z pominięciem emocji zaskoczenia. Dla każdego wymiaru modelu PAD przypisane są współczynniki transformacji dla każdego wymiaru modelu Ekmana, które są określone jako: PADanger (współczynnik dla P, A oraz D), PADdisgust, PADfear, PADhappiness, PADsadness. Współczynniki te zostały przedstawione w postaci macierzy we wzorze 2.1.

$$\begin{aligned} PAD &= [PADanger, PADdisgust, PADfear, PADhappiness, PADsadness] = \\ &= \begin{pmatrix} -0,51 & -0,40 & -0,64 & 0,40 & -0,40 \\ 0,59 & 0,20 & 0,60 & 0,20 & -0,20 \\ 0,25 & 0,10 & -0,43 & 0,15 & -0,50 \end{pmatrix} \quad (2.1) \end{aligned}$$

Aby obliczyć wartość w modelu Ekmana należy skorzystać ze wzoru 2.2, gdzie PAD oznacza macierz opisaną wzorem 2.1, M – stan emocjonalny w modelu PAD, $D = [d_1, d_2, d_3, d_4, d_5]$ oznacza wymiary modelu Ekmana, d_i oznacza kolejne wymiary modelu Ekmana, i – kolejne emocje z modelu Ekmana.

$$E = f(M, PAD) = \frac{D}{\sum_{i=1}^5 d_i} \quad (2.2)$$

W analogiczny sposób można rozszerzyć tę metodą o brakujący szósty wymiar modelu Ekmana (zaskoczenie) – o ile wyznaczona zostanie odpowiednia macierz transformacji. Takim przykładem są macierze transformacji zaproponowane przez Landowską [41] dla modelu PAD (wzór 2.3) oraz PA (wzór 2.4), w których dodatkowo pojawia się współczynnik wolny. Macierze zostały wyznaczone przy pomocy regresji liniowej, wykorzystującej cztery leksykony języka angielskiego, w których słowa zostały oznakowane wartościami zmiennych z modelu Ekmana lub PAD.

$$\begin{aligned} PAD &= [Happy, Sad, Angry, Scared, Disgusted, Surprised, 1] = \\ &= \begin{pmatrix} 0,46 & -0,30 & -0,29 & -0,19 & -0,14 & 0,24 & 0,52 \\ 0,07 & -0,11 & 0,19 & 0,14 & -0,08 & 0,15 & 0,53 \\ 0,19 & -0,18 & -0,02 & -0,10 & -0,02 & 0,08 & 0,50 \end{pmatrix} \quad (2.3) \\ PA &= [Happy, Sad, Angry, Scared, Disgusted, Surprised, 1] = \\ &= \begin{pmatrix} 0,54 & -0,14 & -0,21 & -0,06 & -0,16 & 0,00 & 0,46 \\ 0,50 & 0,06 & 0,37 & 0,36 & 0,12 & 0,00 & -0,01 \end{pmatrix} \quad (2.4) \end{aligned}$$

Ukazane przykłady pokazują, że odwzorowanie pomiędzy modelami jest zagadnieniem trudnym – wiele transformacji jest jednokierunkowych, a dodatkowo każde odwzorowanie wprowadza pewną niedokładność.

Te wyzwania w połączeniu z mnogością wykorzystywanych modeli reprezentacji emocji utrudniają proces integracji rozwiązań rozpoznawania emocji.

2.4. *Rozpoznawanie emocji w podziale na modalności*

Programy przetwarzające zewnętrzne symptomy emocji są zróżnicowane. Głównym czynnikiem różnicującym są modalności, będące źródłem informacji. Najczęściej wykorzystywanymi rodzajami modalności są:

- modalność wizualna (mimika oraz mowa ciała) [42],
- prozodia mowy [43],
- tekst pisany lub mówiony [44],
- wzorce behawioralne [45] oraz
- sygnały fizjologiczne człowieka [26].

Zgodnie z przeglądem literatury z 2017 roku, na modalnościach wykorzystujących dane audio i wideo bazuje około 90% programów automatycznego rozpoznawania emocji [34].

W tym podrozdziale zostały opisane programy w podziale na wykorzystywane przez nie modalności wraz z informacją o wykorzystywanych modelach reprezentacji emocji. Istotne w kontekście niniejszej rozprawy jest to, że programy bazujące na poszczególnych modalnościach podają wyniki swojego działania w różny sposób i z różną częstotliwością, co ma decydujący wpływ na mechanizmy integracji informacji o stanie emocjonalnym.

Programy automatycznego rozpoznawania emocji z ekspresji mimicznej

Do rozpoznawania stanu emocjonalnego na podstawie analizy mimiki, najczęściej wykorzystywany jest model sześciu podstawowych emocji Ekmana [46]. Większość z programów opartych o tę modalność bazuje na wyodrębnionych akcjach mięśni twarzy, w szczególności na systemie kodowania akcji mięśni twarzy FACS (ang. Facial Action Coding System) [47]. **FACS** jest taksonomią ruchów mięśni twarzy (ang. action units – AU), a emocje są rozpoznawane jako kombinacja tych ruchów. Wiele komercyjnych rozwiązań rozpoznających emocje bazuje na FACS, a do przedstawienia rezultatów wykorzystuje model Ekmana (systemy w kolejności alfabetycznej):

- Affdex, Emotient (od firmy Affectiva) – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, a także jest rozszerzona o dodatkowe emocje: pogardy, frustracji oraz zmieszania. Bazuje na FACS i jest w stanie wykryć 19 AU;
- EmoVu [48] – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, a także emocję neutralną. Program wykorzystuje głębokie sieci neuronowe [48];
- InSight – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, a także emocję neutralną [49];
- Face Reader – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, a także emocję neutralną oraz pogardę. Potrafi także zwrócić wynik w modelu 2D. Bazuje na FACS i jest w stanie wykryć 20 AU [49] [50];
- nViso – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, bazuje na FACS. Jest to rozwiązanie chmurowe [51];
- Xpress Engine (od firmy QuantumLab, Polska) – rozpoznaje sześć emocji z modelu emocji podstawowych Ekmana, a także emocję neutralną. Bazuje na FACS.

Także autorskie rozwiązania naukowe często wykorzystują model emocji podstawowych Ekmana rozszerzony o emocję neutralną. Programy te często wykorzystują akcje mięśni twarzy na podstawie odległości i kątów pomiędzy kluczowymi częściami twarzy, takimi jak brwi, oczy czy usta [40], ale znane są również zastosowania głębokich sieci neuronowych, którym jest podawany nie wektor cech, a odpowiednio przygotowany obraz twarzy.

Programy rozpoznawania emocji z mowy ciała

W ostatnim czasie wykorzystanie analizy mowy ciała zyskuje większe zainteresowanie [42] [52] [53] [54] [55] [56]. Analiza mowy ciała jest często wykorzystywana w parze z analizą twarzy oraz analizy mowy, głównie jako modalność wspomagająca ze względu na zbyt dużą niejednoznaczność wyrażania emocji [42] [57]. Modalność ta może wykorzystywać ruch głowy, rąk lub gesty dłoni [57] czy też pełną mowę ciała w tańcu [58].

Dla przypadków gdzie mowa ciała jest modalnością wspomagającą, często jest ona odwzorowywana na model Ekmana – niepełny, pełny lub rozszerzony o inne emocje [52] [53] [54] [59] takie jak nerwowość czy frustrację. Odwzorowanie to następuje najczęściej poprzez określenie pewnych gestów jako przypisanych do danej emocji [56]. Niekiedy dla mowy ciała mierzony jest także poziom pobudzenia (ang. arousal) [55], wymiar wartościowości (ang. valence) [58] lub oba łącznie (model dwuwymiarowy) [56] [60].

Analiza prozodii mowy

Prozodią mowy jest ogół brzmieniowych właściwości języka, które różnicują, postaciują i członują głoskowy, sylabiczny i wyrazowy strumień mowy [10]. Należą do nich akcent, intonacja i iloczas. Dane dotyczące prozodii niosą dużo informacji na temat stanu emocjonalnego. Są one coraz częściej przedmiotem badań naukowców [43] [61] [62]. Podobnie jak wiele innych modalności, wymaga ona zgody użytkownika, przez co jej wykorzystanie jest ograniczone [63] [64].

W analizie prozodii mowy rozwiązania często wykorzystują model emocji podstawowych Ekmana lub jego część (w formie binarnej lub ciągłej) [43] [59] [65] [66]. Modalność związana z mową podobnie jak dane fizjologiczne jest dobrym nośnikiem informacji dotyczącej wymiaru pobudzenia [59], ale również i wymiaru wartości [61] [62] [67].

Podczas jednego z badań rozpoznawano trzy emocje: radość, smutek, gniew oraz stan neutralny ze skutecznością na poziomie 97% [59]. W innym badaniu program analizujący prozodię mowy osiągnął skuteczność na poziomie 57% dla emocji gniewu, smutku, radości, komfortu oraz spokoju, przy 63% skuteczności rozpoznawania przez ludzi [59].

Programy rozpoznawania ładunku emocjonalnego z tekstu

W odniesieniu do analizy tekstu dziedzina posługuje się pojęciem ładunku emocjonalnego (ang. sentiment analysis), a nie samych emocji. W tych programach automatycznego rozpoznawania kluczową rolę odgrywa zazwyczaj słownik, który odwzorowuje słowo na odpowiadające mu nasycenie w danym modelu reprezentacji emocji. Słowniki wykorzystują między innymi modele emocji podstawowych, takie jak modele Ekmana czy Plutchika [44] lub wielowymiarowe 2D oraz 3D.

Dla modeli podstawowych emocji następuje przypisanie konkretnej emocji do danego słowa w sposób binarny [44] lub ciągły. Dla modeli wielowymiarowych słowu przypisywany jest najczęściej punkt w przestrzeni modelu (przypisanie wartości dla każdego z wymiarów) [40].

Istnieje kilka leksykonów przypisujących słowom punkt w przestrzeni modelu PAD, takich jak leksykon opracowany przez Mehrabiana oraz Russela [68], a także leksykon ANEW [69]. Dla modelu emocji podstawowych Ekmana opracowano takie leksykony jak Synesketch [70]. Leksykon NAWL (Nencki AffectiveWord List) [71], [72] z kolei posiada odwzorowanie w model Ekmana (bez emocji zaskoczenia) oraz model 2D.

Najczęściej ładunek emocjonalny jest rozpoznawany poprzez analizę występowania słów kluczowych nacechowanych emocjonalnie (słowa ze słowników). Ta metoda ma jednak szereg wyzwań takich jak niejednoznaczność w definicji emocjonalnych słów kluczowych czy problem ze zdaniem bez słów kluczowych [73].

Osobnym zagadnieniem jest przypisanie wartości ładunku emocjonalnego do wykorzystywanych w tekście emotikon i znaków interpunkcyjnych jako celowego nacechowania emocjonalnego treści, który można uznać za intencjonalne wyrażenie emocji i przypisać tym samym nawet większą wagę niż słowom [40]. Odwzorowanie emotikon w model reprezentacji emocji następuje analogicznie jak dla tekstu – najczęściej na podstawie słowników.

Dane fizjologiczne w rozpoznawaniu stanów emocjonalnych

Dane fizjologiczne, takie jak tętno, przewodność skóry, rozmiar źrenicy, rytm oddechu, są wynikiem pobudzenia autonomicznego układu nerwowego (ang. ANS Autonomous Nervous System) niosącego informację dotyczącą mimowolnych reakcji. Są one dobrym nośnikiem informacji o poziomie pobudzenia (ang. arousal) [1] [26] [59].

Rozpoznawanie emocji z sygnałów fizjologicznych jest zróżnicowane. Zaproponowane programy różną się używanymi kanałami wejściowymi, techniką rozpoznawania, dokładnością, jak i wykorzystywanym modelem reprezentacji emocji. W przeglądzie literaturowym [14] zestawiono 17 programów, spośród których:

- 7 programów wykorzystuje podzbiór emocji podstawowych modelu Ekmana lub modelu Plutchicka,
- 5 programów zwraca własny, niepowtarzalny zbiór stanów emocjonalnych,
- 3 programy opisują jedynie poziom stresu,
- 2 programy opisują poziom przyjemności, który jest jednym z wymiarów modeli wielowymiarowych,
- jeden program zwraca dane w modelu dwuwymiarowym,
- jeden program opisuje emocje muzyczne,
- jeden program opisuje poziom zabawy.

Programy rozpoznawania a modele reprezentacji emocji – podsumowanie

Podsumowując analizę poszczególnych modalności, brakuje powszechnie przyjętego modelu reprezentacji emocji, który byłby uniwersalny dla wszystkich zastosowań. Z tego też powodu często programy automatycznego rozpoznawania emocji korzystają z odmiennych modeli, co jest problematyczne w procesie integracji, gdyż wymaga to przygotowania odwzorowania między modelami. Istniejące metody takich odwzorowań zostały opisane w podrozdziale 2.3. W procesie integracji istotny jest wybór takiego modelu docelowego, który pozwoli połączyć hipotezy o stanie emocjonalnym wyrażone w różnorodnych modelach początkowych.

Zestawienie modalności, na podstawie których działają programy rozpoznawania emocji wraz z zastosowanym modelem reprezentacji emocji w dotychczasowych badaniach, zostało zaprezentowane w Tabeli 2.3.

Tabela 2.3 Zestawienie par: modalności i modelu reprezentacji emocji stosowane przez różnych autorów [opracowanie własne].

Modalność	Pojedyncze etykiety/ emocje	Model emocji podstawowych Ekmana	Model emocji Plutchika	2D (wartościowość i pobudzenie)	Model PAD
Ekspresja mimiczna		++ [74] [49] [48] [50] [51] [34]		+	
Mowa ciała	+	+ [52][53] [54] [59]		+ [55] [56] [58] [60]	Plan [54]
Tekst	+ [40]	+ [70][71], [72] [34]	+ [44]	+ [71], [72]	+ [40] [68]
Mowa	+ [59]	+ [59] [65] [34]		+ [59][61] [62] [67] [34]	
Sygnały fizjologiczne	++ [75]	* [32]	* [32]	++ [59]	

Oznaczenie + oznacza wystąpienie, ++ częste wystąpienie lub dominacja modelu; * model niepełny lub zbliżony.

Prawie w każdej modalności istnieje dominujący model reprezentacji emocji, zgodnie z którym podaje wyniki większość programów automatycznego rozpoznawania emocji. Model podstawowych emocji Ekmana ze względu na powiązanie z FACS jest najczęściej wykorzystywanym w rozpoznawaniu emocji z mimiki twarzy. Modele wielowymiarowe są używane w programach bazujących na analizie sygnałów fizjologicznych i analizie ładunku emocjonalnego tekstu. Etykiety i pojedyncze emocje (modele własne) są również wykorzystywane w programach bazujących na każdej z modalności. Różnorodność stosowania modeli reprezentacji emocji oraz tworzenie własnych modeli jest jedną z przyczyn problemów z integracją danych o stanie emocjonalnym człowieka.

Zauważalna jest tendencja do zmiany modeli korzystających z emocji podstawowych na modele wielowymiarowe [76], które mają większe możliwości dotyczące integracji i przetwarzania komputerowego niż modele emocji podstawowych. Model 2D jest częściej wykorzystywany niż 3D [77], jednakże model 3D jest młodszy, a także w pełni kompatybilny z modelem 2D, gdyż jest jego rozwinięciem o wymiar dominacji.

2.5. Zastosowania przetwarzania emocji

Zakres stosowalności przetwarzania emocji jest szeroki, a potencjalnych zastosowań przyrasta wraz z rozwojem metod przetwarzania i rozpoznawania. Wiele z tych zastosowań zaproponowała Rosalind W. Picard w swojej książce „Affective Computing” [78]. Zgodnie z opinią autorki książki, aby maszynę można było nazwać w pełni inteligentną oraz by mogła komunikować się w sposób naturalny z ludźmi, należy nadać komputerom nie tylko możliwość rozpoznawania emocji, ale także rozumienia, a nawet posiadania i wyrażania własnych emocji.

Przetwarzanie emocji już jest stosowane w pomiarze użyteczności produktu [79] obok takich metod jak śledzenie aktywności czy śledzenie wzroku (ang. eye tracking). Pozwala ono na poszerzenie wiedzy o doświadczeniu użytkownika (ang. user experience), na które w dużym stopniu wpływają emocje.

Innym komercyjnym przykładem wykorzystania jest zastosowanie rozpoznawania emocji jako wsparcia dla działu HR w monitorowaniu pracownika [80] lub w rozmowie kwalifikacyjnej przeprowadzanej online poprzez komunikatory [81] [82] [83], ale także w tradycyjnej rozmowie [84]. Taki system może być również pomocny dla pracownika, pozwalając mu się lepiej się przygotować do rozmowy, a także opanować stres. Przykładowo gra Tardis symuluje proces rekrutacyjny, w którym wykorzystywane są realistyczne interakcje z wykorzystaniem emocji, pomagając w ten sposób w rozwoju umiejętności miękkich [85].

Kolejnym przykładem zastosowania rozpoznawania emocji są systemy rekomendacyjne, w których informacja o stanie emocjonalnym pozyskana w mało inwazyjny sposób pomogłoby określić, które utwory muzyczne są preferowane przez użytkownika w zależności od jego nastroju [86]. Program mógłby także sugerować inne produkty w jego guście, takie jak meble czy obrazy. Takie mechanizmy są opracowywane przez platformy streamingowe takie jak Netflix [87] czy Spotify [88], które dostarczają filtry oparte na emocjach.

Jedną z pierwszych i największych grup zastosowań rozpoznawania emocji jest wsparcie procesów e-nauczania. Liczne badania wskazują na istotny wpływ emocji na efektywność procesu uczenia [78] [89], zwłaszcza stan skupienia (ang. flow) [90] [91]. Zostało to potwierdzone w badaniu z wykorzystaniem inteligentnego systemu nauczania AutoTutor [89]. Jednocześnie rozwój przetwarzania emocji doprowadził do powstania nowej generacji inteligentnych systemów nauczania (ang. intelligent tutoring systems) – afektywnych systemów nauczania (ang. affective tutoring systems). Pozwalają one na adaptacyjne dostosowanie się nie tylko do wiedzy ucznia, ale także do jego stanu emocjonalnego, aby poprawić wyniki w nauce, zwiększyć motywację, pewność siebie uczniów, a także zapobiegać rezygnacji czy przerywania procesu nauki [92]. Systemy te mogą stanowić także wsparcie nauczycieli w czasie nauki zdalnej podczas egzaminów [93].

Wykorzystanie metod rozpoznawania emocji może być przydatne w terapii i monitorowaniu stanów psychicznych, takich jak stres [94], myśli samobójcze [95] czy depresja [61]. Na przykład DeVault i inni opracowali SimSensei Kiosk, wirtualną ankietkę o imieniu Ellie do automatycznej oceny wskaźników zagrożenia związanych z depresją, lękiem oraz zespołem stresu pourazowego [96]. Inny zespół z kolei opracował program monitorujący stan depresji poprzez analizę drżenia głosu [97].

Przetwarzanie emocji niesie także możliwość wsparcia osób z zaburzeniami ze spektrum autyzmu [98], które często mają problem w nawiązywaniu relacji społecznych, a także odczytywania oraz okazywania emocji [99]. Znane są badania, w których rozpoznawanie emocji na podstawie sygnałów fizjologicznych wspomaga terapeutę w pracy z dziećmi z autyzmem [100].

Coraz częściej wykorzystywani są inteligentni asystenci (ang. spoken dialogue systems), którzy często są jednak ograniczeni do prostych komunikatów np. w systemach nawigacyjnych, systemy informacji dla podróżnych [73]. Przykładem bardziej zaawansowanego zastosowania jest robot mobilny „Florence” – prototyp pielęgniarskiego asystenta domowego [101]. Rozwój innych zaawansowanych robotów wspierających chorych może być wartościowy w przypadku osób samotnych, wymagających wsparcia oraz możliwości komunikacji „z drugim człowiekiem”, co jest szczególnie istotne w przypadku izolacji osób chorych na COVID-19 [9]. Rozwinięcie rozumienia oraz mowy samego robota poprzez informację afektywną zwiększa naturalność rozmowy, a więc i wrażenie rozmowy z drugim człowiekiem, a nie maszyną.

Także obszar rozrywki znalazł swoje zastosowanie dla przetwarzania emocji w postaci wsparcia gier komputerowych. Najprostszym przykładem zastosowania jest regulacja poziomu trudności na podstawie poziomu znużenia lub pobudzenia użytkownika [102] [103]. Gry mogłyby w ten sposób skuteczniej redukować stres albo indukować oczekiwane przez użytkownika pobudzenie.

Wymienione zastosowania pokazują możliwości rozpoznawania emocji przez aplikacje komputerowe, a także istotność ich skuteczności i niezawodności, które mogą zostać poprawione dzięki mechanizmom integracji opracowanym w ramach niniejszej pracy doktorskiej.

2.6. *Wyzwania rozpoznawania emocji*

Wyzwania w dziedzinie rozpoznawania emocji można podzielić na dwie grupy: wyzwania programów automatycznego rozpoznawania emocji oraz wyzwania związane z integracją wielokałową i wielomodalną. Problematyka integracji wielomodalnej została opisana w kolejnym rozdziale (rozdział 3).

Głównymi wyzwaniami automatycznego rozpoznawania emocji są:

1. dokładność klasyfikatora (ang. accuracy) i wydajność (ang. performance) [34],
2. odporność programu na zakłócenia [34],
3. czasowa niedostępność danych w modalności,
4. nieznaną prawdziwy stan emocjonalny (ang. ground truth).

Wyzwanie dokładności i wydajności pojedynczych programów nie jest przedmiotem niniejszej pracy – rozprawa nie koncentruje się na skuteczności rozpoznawania tylko na wyzwaniu drugim oraz trzecim. Podejście wielomodalne pozwala zająć się problemami odporności na zakłócenia i okresowej niedostępności sygnału w kanale obserwacji, co zostało zaadresowane w dalszej części pracy.

W literaturze [8] wyróżnia się kilka możliwości określania „prawdziwego” stanu emocjonalnego (ang. Ground truth): raportowanie własnych emocji przez badanego (ang. self-reports), manualne oznaczanie obserwacji czy danych fizjologicznych przez ekspertów oraz analiza wielomodalna wraz ze sprawdzeniem spójności [9]. Raportowanie stanów emocjonalnych przez samego badanego wydaje się rozwiązaniem najprostszym, ale stwarza ono również szereg problemów związanych z metodyką badania, gdyż każda z metod i kwestionariuszy ma swoje wady i ograniczenia. Trudnością jest odpowiedź na pytanie, jak często i w jaki sposób zadawać pytania użytkownikowi, aby jednocześnie nie przeszkadzać w zadaniu lub interakcji, którą chcemy obserwować [1]. Badany często nie zna modeli reprezentacji emocji, przez co posługuje się etykietami lub wskazuje wartość na skali, a obie te metody są nieporównywalne dla dwóch osób.

Ekspertci zazwyczaj znają skale, jednakże posługują się jedynie symptomami emocji, bez możliwości wniknięcia w wewnętrzne procesy myślowe człowieka. Dodatkowo zewnętrzne symptomy emocji mogą być przez człowieka umyślnie lub nieumyślnie zafałszowane. Przykładowo, jeżeli okazywanie jakiejś emocji nie jest dobrze postrzegane w danej kulturze (tabu) np. okazywanie strachu przez mężczyzn, to okazywane reakcje mogą być nieadekwatne do prawdziwego stanu emocjonalnego. Z tych względów rozpoznawanie emocji zwraca się w kierunku analizy wielomodalnej, która może poprawić skuteczność rozpoznawania.

3. METODY INTEGRACJI WIELOMODALNEJ

W poprzednim rozdziale przedstawiono mnogość obecnie stosowanych modeli reprezentacji emocji, metod ich odwzorowania, modalności źródłowych oraz przykłady zastosowań, a także wyzwania związane z rozpoznawaniem emocji. Dodatkowo wskazano na tendencję do wykorzystywania analiz wielomodalnych, co ukazuje potrzebę integracji danych z wielu kanałów i modalności [20] [64]. Podobnie ludzie podczas odczytywania emocji innych ludzi wykorzystują analizę wielomodalną [40] [104]. Według badań psychologów ludzkie komunikaty są przekazywane i odbierane drogą werbalną, intonacji oraz mowy ciała [105].

Współcześnie prowadzone badania wskazują na zdolność rozwiązań wielomodalnych do przewyższenia rozwiązań jednomodalnych z perspektywy kryteriów takich jak dostępność (ang. availability) oraz dokładność (ang. accuracy) [77] [106]. Dla 86% systemów (26 z 30 badanych) analizy wielomodalne przyniosły poprawę dokładności, która wyniosła średnio 9,83% (mediana: 6,6%) w stosunku do najlepszego z integrowanych rozwiązań jednomodalnych [77]. Wiele pozycji literatury wskazuje na poprawę wyników w analizie wielomodalnej [107].

Jednocześnie analiza wielomodalna pozwala na zwiększenie ciągłości rozpoznawania emocji, gdyż przykładowo program jednomodalny analizujący dane mowy nie zwraca wyniku, gdy badany akurat nie mówi, a program jednomodalny analizujący dane z twarzy nie zwróci wyniku, gdy twarz zniknie z obszaru obejmowanego rejestracją za pomocą kamery. Zastosowanie wielu kanałów obserwacji i integracja wyników może zwiększyć dostępność oszacowania stanu emocjonalnego nawet przy okresowych brakach informacji z określonego kanału.

3.1. *Zestawienie stosowanych metod integracji*

Metoda fuzji jest to scenariusz integracji danych z wielu modalności lub wielu kanałów obserwacji, który pozwala uzyskać jedną odpowiedź systemu na temat szacowanego stanu emocjonalnego.

W tym podrozdziale zostały przedstawione obecnie stosowane scenariusze/poziomy integracji oszacowań stanów emocjonalnych, są to:

- **pojedynczy program** (brak integracji) – rozwiązanie jednomodalne, jednokanałowe;
- **metoda wczesnej fuzji** (ang. early fusion) – jest to klasa metod fuzji, w których dane integrowane są na poziomie danych z kanałów obserwacji, tworząc wspólny wektor cech;
- **metoda późnej fuzji** – jest to klasa metod fuzji, w których dane integrowane są na poziomie wyników klasyfikacji (hipotez o stanie emocjonalnym), otrzymanych z programów automatycznego rozpoznawania emocji (klasyfikatorów); nazywana jest również fuzją na poziomie decyzyjnym; czasem z późnej fuzji wyodrębniana jest **fuzja na poziomie oceny** (ang. score-level fusion), jako rodzaj fuzji uwzględniający prawdopodobieństwo hipotez [77];

- **fuzja hybrydowa** (ang. hybrid fusion) – połączenie wczesnej i późnej fuzji. Czasem z fuzji hybrydowej w literaturze wyodrębniana jest **fuzja na poziomie modelu** (ang. model-level fusion), jako rodzaj fuzji dodatkowo biorący pod uwagę powiązania między modalnościami.

Ważnymi z perspektywy integracji pojęciami są:

- **ekspert** – w tej pracy przyjęto zgodnie z literaturą [15] [16] [17] [18] [19]: klasyfikator lub inny program automatycznego rozpoznawania emocji, wykorzystywany w trakcie metody późnej fuzji lub fuzji hybrydowej, który na podstawie danych wejściowych o symptomach emocji zwraca hipotezę stanu emocjonalnego badanego użytkownika komputera,
- **hipoteza o stanie emocjonalnym** – rezultat działania programu automatycznego rozpoznawania stanu emocjonalnego (eksperta). W kontekście integracji danych będzie nazywany skrótowo *hipotezą* programu automatycznego rozpoznawania emocji lub hipotezą eksperta,
- **uzgodnione rozwiązanie** – rezultat działania fuzji, działający na podstawie cząstkowych informacji (danych lub hipotez o stanie emocjonalnym).

Tabela 3.1 przedstawia statystyki wykorzystania różnych metod fuzji oraz wykorzystanych modalności [77]. W 87% przypadków wykorzystano 2 modalności, a dla 13% przypadków 3 modalności. Najczęściej wykorzystywanymi kanałami obserwacji były kanał wideo oraz kanał audio. Najbardziej popularnymi metodami fuzji są metoda wczesnej i późnej fuzji po ok. 35–38% każda, na trzecim miejscu jest fuzja na poziomie modelu (rodzaj fuzji hybrydowej). Jednocześnie 35% rozwiązań wykorzystuje wielowymiarowe modele reprezentacji emocji, a 65% modele dyskretne. Szczegółowa analiza znajduje się w artykule przeglądowym [77].

Tabela 3.1 Statystyki dotyczące metod fuzji [77]

Dimension	Prop.	Dimension	Prop.
Data type		Measure. model	
Acted	0.522	Disc.	0.644
Induced	0.278	Dim.	0.356
Natural	0.200		
Detection model		Affect detected	
Classification	0.922	Disc. basic	0.367
Regression	0.078	Disc. nonbasic	0.078
		Disc. mixed	0.178
No. of modalities		Dim. simple	0.278
Bimodal	.867	Dim. complex	0.100
Trimodal	.133		
Modality		Fusion method	
Face	0.767	Feature	0.389
Voice	0.822	Decision	0.356
Text	0.167	Hybrid	0.056
Body	0.133	Model	0.200
Eye Gaze	0.011	Validation method	
Peri. physio.	0.111	Person indep.	0.378
Central physio.	0.056	Person dep.	0.622
Content	0.067		

Notes: Prop. = Proportion; Peri = Peripheral; Physio. = Physiology; Content = Content/Context; Measure. = Measurement; Disc. = Discrete; Dim. = Dimensional; Indep. = Independent; Dep. = Dependent.

Dla każdego scenariusza integracji w rozdziałach 3.2–3.5 przedstawiono charakterystykę, przebieg realizacji, a także ich zalety i wady. Opisy te stanowią opracowanie własne i są istotne ze względu na cel niniejszej pracy, jakim jest zaproponowanie ulepszenia w zakresie metod integracji informacji o stanie emocjonalnym.

3.2. **Scenariusz bazowy – rozwiązanie jednomodalne i jednokanałowe**

Obecnie większość rozwiązań stosuje wykrywanie jednej konkretnej emocji [40] lub podzbiór modelu reprezentacji emocji, co zostało opisane w punkcie 0. Rozwiązania te korzystają tylko z jednej modalności [34] co upraszcza rozwiązanie z powodu braku potrzeby integracji.

Przebieg scenariusza bazowego wygląda następująco:

1. Rejestrowane są dane wejściowe z pojedynczego sensora.
2. Program przetwarza je do postaci wektorów cech, klasyfikuje i podaje końcowy stan emocjonalny.

Zalety:

- najprostszy model, znane dane i sposób działania,
- zazwyczaj duża dokładność rozwiązania,
- jeden klasyfikator, brak integracji,
- pełna dowolność modelu reprezentacji emocji.

Wady:

- ograniczone obszary zastosowań,
- możliwość występowania nieciągłości informacji o stanie emocjonalnym ze względu na okresową niedostępność danych z kanału wejściowego,
- zazwyczaj wykrywanie pojedynczych, dyskretnych emocji,
- skuteczność rozpoznawania zależna od jakości sygnału w pojedynczym kanale obserwacji.

3.3. **Scenariusz wczesnej fuzji (ang. *early/feature-level fusion*)**

W metodzie wczesnej fuzji, w której wykorzystywane jest wiele modalności (np. mimika twarzy i głos) lub wiele kanałów (np. kilka kamer) integracja następuje na poziomie danych z poszczególnych źródeł. Zarejestrowane i przetworzone sygnały łączone są w jeden wspólny wektor cech [40] [42]. Im więcej jest kanałów oraz modalności, tym więcej danych, co powoduje konieczność selekcji części parametrów (problem zbyt dużego wektora cech) [42].

Rozwiązania bazujące na wczesnej fuzji często mają wysoką dokładność (ang. accuracy) – zazwyczaj lepszą od rozwiązań późnej fuzji [108]. Jednakże metoda wczesnej fuzji jest bardziej złożona. Dodanie nowego modułu często wymaga przebudowy całego rozwiązania oraz ponownego uczenia klasyfikatora. Większym wyzwaniem jest jednak synchronizacja czasowa, gdyż dane z różnych modalności przychodzą z różną częstotliwością, a uruchomienie analizy na niepełnym wektorze może pogorszyć wynik, gdy klasyfikator był uczony na pełnym wektorze cech.

Przebieg działania:

1. Zbieranie i przetwarzanie danych wejściowych z każdego z kanałów.
2. Łączenie danych we wspólny wektor cech.
3. Opcjonalnie: zastosowanie metod redukcji liczby cech wektora.
4. Działanie klasyfikatora, określającego stan emocjonalny na podstawie zbiorczego wektora cech.

Zalety:

- jeden klasyfikator,
- przeważnie wyższa skuteczność niż rozwiązania bazujące na późnej fuzji,
- wyższa skuteczność rozwiązania od rozwiązań jednomodalnych, jednokanałowych, nawet w przypadku częściowo sprzecznych danych.

Wady:

- problem synchronizacji czasowej – dane z modułów mogą pojawić się z różną częstotliwością, co powoduje otrzymanie niepełnego wektora cech, gdy podczas trenowania zazwyczaj są pełne dane [34],
- wrażliwy na całkowitą lub okresową niedostępność danych z kanałów obserwacji – awaria kamery lub zaszumiony sygnał z sensorów powoduje, że niektóre cechy mają wartości zerowe lub puste, co wpływa na skuteczność klasyfikacji;
- mała skalowalność:
 - dołączenie nowego kanału obserwacji lub nowej modalności wymaga ponownego uczenia oraz testowania całego klasyfikatora – złożony proces,
 - wektor cech może przybrać duże rozmiary,
 - im więcej modalności/kanałów, tym bardziej rośnie złożoność rozwiązania.

Przykładem wczesnej fuzji jest wykorzystanie modelu regresji liniowej do integracji danych z różnych modalności. W projekcie AVEC 2016 [61] prowadzono badania nad identyfikacją symptomów depresji. Wykorzystano trzy modalności: kanał wideo (analiza twarzy), kanał audio (mowa) oraz dane fizjologiczne. Do ich integracji wykorzystano kilka modeli regresji liniowej, osobno dla kilku wariantów:

- audio–ECG–EDA–wideo – dane ze wszystkich modalności,
- audio–ECG–EDA – dane dla wszystkich modalności poza wideo,
- audio–ECG – dla braku modalności wideo oraz EDA.

Dane dla każdej modalności były walidowane. Przykładowo, gdy na nagraniu nie wykryto twarzy, ramka zostawała odrzucana. Modelem reprezentacji był model 2D. Dla rozwiązań jednomodalnych poziom pobudzenia był najskuteczniej rozpoznawany przez modalność mowy, a dla poziomu wartościowości przez ekspresję mimiki twarzy. W przypadku braku jednej z modalności, rozwiązanie przełączało się na inny model regresji liniowej. Badania potwierdziły, że metoda integracji okazała się skuteczniejsza od rozwiązań jednomodalnych. Do rozwiązania problemu synchronizacji wykorzystano okna czasowe, których rozmiar wahał się między 4 a 14 s z przesunięciem o 2s dla kolejnych okien [61].

3.4. *Scenariusz późnej fuzji – wiele klasyfikatorów*

Późna fuzja, nazywana także fuzją na poziomie decyzyjnym lub systemem opartym na zespołach (ang. ensemble-based systems) jest alternatywną metodą do wczesnej fuzji. Wykorzystuje ona wiele modalności (np. audio i wideo) lub wiele kanałów (np. kilka kamer). Integracja następuje na poziomie wyników klasyfikacji. Późna fuzja została zaproponowana z powodu problemu synchronizacji czasowej we wczesnej fuzji [40] [42], a także problemu czasowej niedostępności danych z części kanałów [16]. Jej celem jest redukcja współczynnika błędów klasyfikacji przez strategiczne łączenie klasyfikatorów.

Metoda ta polega na niezależnym przetwarzaniu i wstępnej klasyfikacji danych z każdej modalności/kanału osobno tworząc zbiór cząstkowych hipotez o stanie emocjonalnym człowieka w danej chwili. Cząstkowe wyniki klasyfikacji łączone są dopiero na etapie decyzyjnym, poprzez funkcję integrującą. Może być ona sumą, średnią, wartością minimalną, maksymalną, medianą, wynikiem głosowania większościowego lub dowolną inną funkcją. Funkcje metod integracji zostały opisane w podrozdziale 3.7.

Dokonywanie zmian w sposobie przetwarzania poszczególnych modalności przez dany program automatycznego rozpoznawania emocji/klasifikator jest niezależne od pozostałych programów i nie wymaga zmiany całego rozwiązania – co zwiększa elastyczność metody późnej fuzji. Nie wymaga też znajomości technicznych aspektów każdego z programów automatycznego rozpoznawania emocji (w późnej fuzji nazywanych także ekspertami). Późna fuzja pozwala na niezależność struktur danych, jednakże wymaga ujednoczenia modelu reprezentacji emocji na etapie decyzyjnym.

Przebieg rozwiązania wygląda następująco:

1. Każdy z ekspertów pobiera dane ze swoich prywatnych kanałów wejściowych.
2. Każdy moduł (ekspert) przetwarza dane, klasyfikuje i przygotowuje hipotezę o stanie emocjonalnym.
3. Hipotezy ekspertów są przekształcane do jednego modelu reprezentacji emocji za pomocą odwzorowań (np. opisanych w rozdziale 2.3) – jeśli wyjściowe modele reprezentacji poszczególnych ekspertów są kompatybilne, można pominąć ten krok.
4. Integracja wyników według odpowiednio przyjętego kryterium/funkcji integracji.



Charakterystyka scenariusza opartego o późną fuzję przedstawia się następująco:

- jednocześnie działa wiele klasyfikatorów – każdy ekspert posiada swój,
- integracja następuje na końcu procesu,
- programy lub klasyfikatory są uczone/specjalizowane w swoim wąskim zakresie cech – zazwyczaj jedna modalność [34], choć możliwe jest stosowanie różnych programów wielomodalnych.

Zalety:

- bardziej wydajny trening klasyfikatorów, krótszy czas trenowania, a także mniejsze wektory cech niż w przypadku wczesnej fuzji [16],
- redukcja złożoności dzięki metodzie „dziel i zwyciężaj” – bardziej złożony problem zostaje rozbity na kilka mniejszych podproblemów (np. dla każdego kanału osobny klasyfikator), które są prostsze do rozwiązania [16],
- możliwość wykorzystania jednego modelu (sieci neuronowej lub innego modelu statystycznego) na różnych danych, przykładowo dla każdej modalności osobno, w celu optymalizacji efektu uczenia [16],
- wytwarzanie modułów oraz funkcji integrującej bez konieczności znajomości całości rozwiązania:
 - pozwala na zrównoleglenie pracy ludzkiej,
 - pozwala na wykorzystanie komercyjnych rozwiązań, w których brak dostępu do informacji o wewnętrznej budowie – gdy możliwe jest tylko pobranie wyniku klasyfikacji przez API,
 - brak zależności między modułami – pozwala na asynchroniczną integrację,
- błędy klasyfikatorów nakładają się w mniejszym stopniu niż w przypadku wczesnej fuzji,
- skalowalność:
 - mały narzut spowodowany dołączaniem kolejnych modułów,
 - nowy program lub modyfikacja nie wymaga przebudowy całego rozwiązania,
- odporność na okresową niedostępność wyników lub braku decyzji pojedynczych modułów – póki co najmniej jeden z ekspertów zwraca wynik, uzgodniony rozwiązanie jest możliwe do otrzymania.

Wady:

- brak typowych funkcji integracji i kryteriów ich oceny,
- wymaga sprowadzenia wszystkich wyników do jednego, spójnego modelu reprezentacji emocji, a ich odwzorowanie może wprowadzić dodatkowy błąd do rozwiązania,
- niższa skuteczność rozpoznawania niż dla wczesnej fuzji przy kompletnych danych.

Wyzwania:

- znane są w literaturze rozwiązania integracji dla dwóch i trzech kanałów, jednak brak typowych rozwiązań dla większej ich liczby [34] [77],
- wiele rozwiązań fuzji jest proponowana ad-hoc, co skutkuje brakiem możliwości dodawania nowej funkcjonalności bez przebudowy całego systemu [109],
- rozwiązania często nie adaptują się do jakości dostarczanych danych wejściowych [16] [17] [40] [109],
- problem synchronizacji czasowej (ze względu na różny czas działania programów), jednak mniej problematyczny niż w przypadku wczesnej fuzji,
- przypadki rozbieżnych oszacowań stanu emocjonalnego dla tego samego momentu w czasie [2].

3.5. **Scenariusz fuzji hybrydowej (ang. hybrid fusion)**

Aby pogodzić koncepcje wczesnej oraz późnej fuzji powstają metody starające się połączyć zalety ich obu. Mają one na celu zachowanie elastyczność późnej fuzji poprzez utrzymywanie wielu klasyfikatorów, pozwalając jednak na współdzielenie danych z każdego kanału między klasyfikatorami [40].

Takie rozwiązanie pozwala zachować przewagę rozwiązań wczesnej fuzji w przypadkach, gdy charakterystyki czasowe kanałów są zbliżone. Nie wymagają one jednak przebudowy całego systemu, co miałyby miejsce podczas wczesnej fuzji.

Przebieg rozwiązania:

1. Zebranie danych wejściowych z wielu kanałów.
2. Przygotowanie hipotezy o stanie emocjonalnym przez każdego z ekspertów
3. Sprowadzenie hipotez do jednego modelu reprezentacji emocji.
4. Integracja wyników według odpowiednio przyjętego kryterium.

Zalety:

- większość zalet późnej fuzji (w zależności od sposobu integracji),
- każdy z modułów ma do dyspozycji wszystkie kanały, na podstawie których może przygotować hipotezę stanu emocjonalnego,
- lepsze wyniki, gdy istnieją zależności między kanałami, czego brakuje w rozwiązaniu opartym na fuzji decyzyjnej.

Wady:

- wykorzystuje wiele klasyfikatorów,
- dodanie nowego kanału może powodować przebudowę kilku ekspertów.

3.6. *Podsumowanie metod integracji w rozpoznawaniu emocji*

Według wielu badań rozwiązania bazujące na wczesnej fuzji mają wyższą dokładność (ang. accuracy), jednakże różnice są nieznaczne [108]. Dla obu typów fuzji raportowana jest widoczna poprawa skuteczności w stosunku do najlepszego rozwiązania jednomodalnego [58] [110] [111].

Rozwiązania wykorzystujące późną fuzję, mogą być obarczone większym błędem, gdyż bazują na przetworzonych danych, przez co tracą informację o powiązaniach poszczególnych symptomów, które mogą zostać uwzględnione w klasyfikatorach wczesnej fuzji. Jednocześnie dodawanie wielu modalności do wczesnej fuzji powoduje duże wektory cech oraz wyzwania związane z okresowo występującymi brakami w tym wektorze.

Zarówno metoda wczesnej, jak i późnej fuzji, mają swoje zalety i wady, a ich wykorzystanie zależy od kontekstu użycia. Na podstawie analizy dotychczasowych badań można określić, że prawdopodobnie metoda wczesnej fuzji jest lepsza dla rozwiązań, gdzie sensory przesyłają dane w podobnych interwałach czasowych oraz z podobną częstotliwością np. dla danych z jednej modalności (kilka kamer) lub dane wielu modalności z jednej kamery. Dla przypadków bardziej zróżnicowanych lub potencjalnie rozszerzanych o nowe czujniki (np. dane fizjologiczne) lepszym rozwiązaniem wydaje się późna lub hybrydowa fuzja [34] [17].

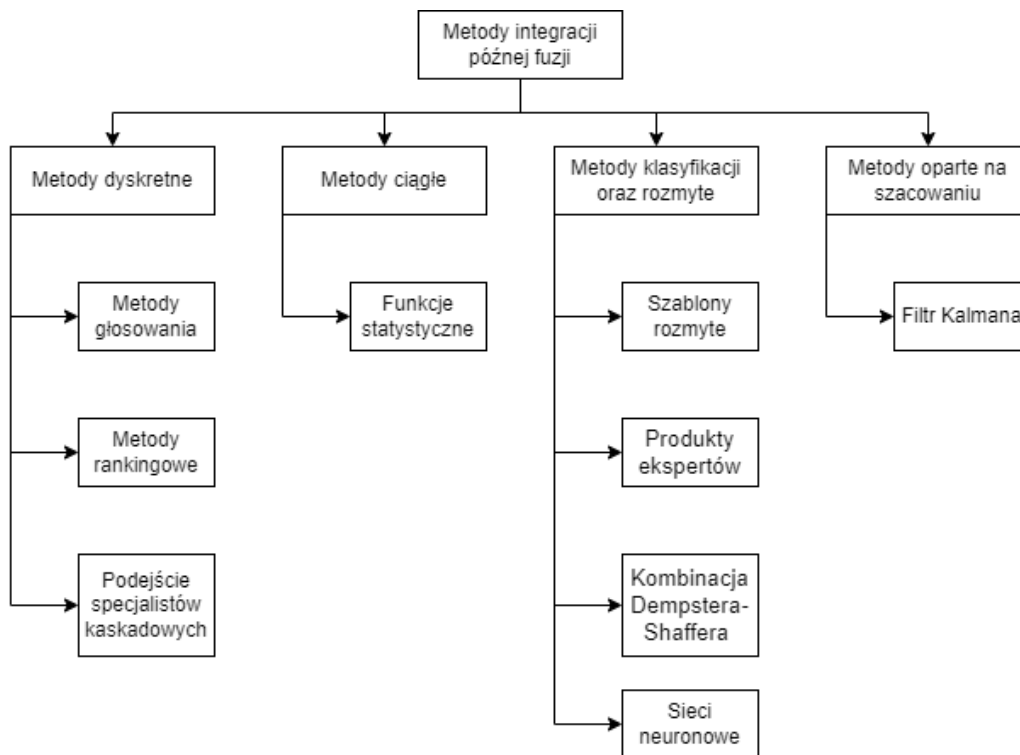
Istotną wadą rozwiązania bazującego na wczesnej fuzji jest nieodporność klasyfikatora, który był uczony na kompletnych danych, na wartości puste w wektorze cech. Rozwiązanie to jest ściśle zależne od dostępności sygnałów z poszczególnych kanałów obserwacji [16].

Według Poria i innych [34] można zauważyć rosnącą tendencję do wykorzystywania później fuzji dla rozwiązań, w których wykorzystywane są moduły audio i wideo. Dla modalności analizy ładunku emocjonalnego tekstu nie widać wyraźnej tendencji do żadnej z metod integracji.

3.7. *Rodzaje metod integracji w późnej fuzji*

Kluczowym elementem późnej fuzji jest metoda integracji, która może być zrealizowana poprzez dodatkowy klasyfikator, funkcję czy ranking. W literaturze można znaleźć wiele rozwiązań takich jak funkcje statystyczne, metody rankingowe, klasyfikatory czy też różne formy głosowania [16]. Występuje wiele sposobów klasyfikacji rozwiązań [16] [18] [34], na podstawie których autor pracy opracował własną klasyfikację, która została zaprezentowana na Rys 3.1.

Autor niniejszej pracy przedstawił najważniejsze z metod, podając dla nich zalety, wady oraz ograniczenia, które stanowią opracowanie własne.



Rys 3.1 Klasyfikacja metod integracji późnej fuzji [opracowanie własne]

Cechą wspólną dla wielu rozwiązań znalezionych w literaturze jest stosowanie wag dla programów automatycznego rozpoznawania emocji (ekspertów). Istotną luką wielu rozwiązań jest brak uwzględniania wiedzy na temat jakości danych, które są przetwarzane, a określanie wag jedynie na podstawie danych treningowych, które mogą się znacząco różnić od rzeczywistych danych [16].

Zgodnie z najlepszą wiedzą autora i stanem literatury wagi dla programów najczęściej określane są w sposób statyczny na podstawie wyników otrzymanych na etapie uczenia klasyfikatorów [16] [109] [111] [115] lub wyliczana na koniec badania (w wersji offline). W ten sposób wyznaczone wagi są przypisywane do całego programu automatycznego rozpoznawania emocji (eksperta) zamiast dynamicznego przypisania do fragmentów danych (np. ramek czasowych) w trakcie przeprowadzanych badań (ang. run-time). Część z tych metod stosowanych statycznie nie jest możliwa do realizacji dla rozwiązań czasu rzeczywistego np. przypisanie współczynnika po analizie całego pliku. Jest to istotna wada ze względu na możliwość pogorszenia się jakości danych w trakcie pomiaru, takiej jak zmiana kąta kamery, oświetlenia czy szumu spowodowanego zmiennym otoczeniem. Zmienne otoczenia takie jak ruch uliczny, mogą wpłynąć na skuteczność rozpoznawania emocji. Dodatkowo wyznaczone wagi często są używane w sposób liniowy – tak jak to jest w przypadku średniej ważonej, co jest prostym, lecz nie zawsze jest najbardziej adekwatnym rozwiązaniem.

3.7.1 Metody dyskretne

Metody te bazują na dyskretnych modelach reprezentacji emocji, takich jak binarne modele – emocja jest lub jej nie ma, emocja jest pozytywna lub negatywna, poziom pobudzenia jest wysoki lub niski itd. Przestrzeń rozwiązań może być także złożeniem kilku binarnych decyzji lub podprzestrzenią modelu wielowymiarowego [16] np. wysokie pobudzenie pozytywne, wysokie pobudzenie negatywne itd. lub skala: wysokie, średnie, niskie i inne warianty.

Metody głosowania

Metoda głosowania możliwa jest do zastosowania w rozwiązaniach z modelem dyskretnym, w którym istnieje skończona liczba decyzji [18]. Głosowanie większościowe zgodnie z nazwą zakłada wybór rozwiązania, które zostaje wybrane większością głosów – dla problemów dwuklasowych i nieparzystej liczby ekspertów – zawsze możliwe jest podanie uzgodnionego rozwiązania. W wariacie ważonym głos każdego z ekspertów otrzymuje inną wagę [16], co zdejmuje ograniczenie nieparzystej liczby ekspertów.

Lanze i inni w swoich badaniach integrowali dane z kanału wideo (mimika twarzy) oraz audio (prozodia) z wykorzystaniem głosowania większościowego [109]. Wykorzystali model emocji podstawowych Ekmana oraz klasyfikatory binarne takie jak smutny vs szczęśliwy, smutny vs rozgniewany itd., gdzie każda binarna odpowiedź była traktowana jako jeden głos. Współczynnik zaufania został wyliczony jako: liczba głosów za daną opcją (np. smutny) podzielona przez wszystkie głosy (np. smutny i niesmutny). Powyższy współczynnik zaufania został użyty jako waga w średniej ważonej, która posłużyła jako funkcja integracji dla obu modalności. Ich program osiągnął najlepsze wyniki dla wag: 0,75 dla kanału wideo oraz 0,25 dla kanału audio.

Caridakis i inni [111] w swoich badaniach integrowali dane z kilku modalności: mimiki twarzy, mowy ciała, prozodii mowy oraz gestów. Wykorzystany przez nich model zawierał osiem emocji, których granulacja ograniczona była do wyboru jednej z nich, bez wskazania intensywności. Wykorzystali dwa podejścia do funkcji integracji w późnej fuzji. Pierwsza wskazywała emocję z najwyższym prawdopodobieństwem, druga zaś przewidywała głosowanie większościowe, a w przypadku braku rozstrzygnięcia (3 różne emocje) – wybierano emocje z największym prawdopodobieństwem (powrót do metody nr 1). Nieznacznie lepsze wyniki osiągnięto dla głosowania większościowego.

Zaletami metod głosowania jest promowanie zbieżności oraz prosty mechanizm, który często jest wystarczający dla specjalistycznych aplikacji, ukierunkowanych na konkretne zadanie, gdzie wystarczającym jest odpowiedź na pytanie, czy dana emocja została wykryta. Z drugiej strony rozwiązania te często potrzebują wsparcia innych metod w przypadku braku rozstrzygnięcia. Sposób wyznaczania wiarygodności rozwiązania jest ograniczony do procentu poparcia hipotezy, co jednak nie uwzględnia w żaden sposób jakości danych, gdyż prawdopodobnie przy gorszej jakości danych, częściej będą padać odpowiedzi „nie wiadomo” oraz błędne. Stosowanie wagi dla programów są często przypisane statycznie, uwzględniając uśrednioną skuteczność, co nie odzwierciedla jakości danych spowodowanych słabym oświetleniem czy dużym szumem.

Metody rankingowe

Metody rankingowe bazują na wytypowaniu klasy rozwiązań poprzez:

1. redukcję klas – odrzucenie klas rozwiązań osiągane jest poprzez wyznaczenie progu na etapie trenowania danych, poniżej którego odpowiedzi klasyfikatora są odrzucane – wybrane zostają pozostałe klasy,
2. uporządkowanie kolejności klas od najbardziej prawdopodobnej do najmniej prawdopodobnej. Uporządkowanie osiągane jest m.in. w metodach Bordy i regresji logistycznej [18].

W metodzie Bordy każdy klasyfikator tworzy własny ranking klas rozwiązania, a następnie na tej podstawie wybierany jest najbardziej prawdopodobne rozwiązanie. W przypadku binarnych decyzji metoda to sprowadza się do głosowania większościowego. Domyślnie nie uwzględnia ona jakości danych ani współczynnika zaufania klasyfikatora, jednakże można metodę rozszerzyć poprzez wykorzystanie średniej ważonej.

Do zalet metod rankingowych można zaliczyć mechanizm odrzucania powtarzających się oraz słabszych klasyfikatorów przy zachowaniu promowania zbieżności rozwiązań. Podobnie jak metody głosowania nadają się one do prostych specjalistycznych aplikacji. Do wad tej metody należy ograniczony sposób przypisania poziomu wiarygodności, a także niska granularność wyniku.

Podejście specjalistów kaskadowych (ang. *cascading specialists*)

Kim oraz Lingenfelser, po skorzystaniu z metod głosowania większościowego oraz metody Bordy, utworzyli własne podejście „specjalistów kaskadowych” (CS) [16] [17]. Autorzy publikacji wskazują na problem metod ważonych spowodowany przecenieniem niektórych klas rozwiązań z powodu ważenia. Metoda specjalistów kaskadowych wybiera ekspertów dla pojedynczych klas i sortuje ich według znaczenia.

Na początku specjaliści (programy) są wybierani na podstawie współczynnika dodatnie pozytywnego (ang. true positive) dla wszystkich klas znalezionych w fazie testowania. Następnie wybierany jest najlepszy ekspert oraz kolejny ekspert. Jeżeli ich hipotezy są zgodne – odpowiedź jest wybierana. W przeciwnym razie odpytywany jest drugi i trzeci ekspert i ich hipotezy są porównywane itd. W przypadku braku zgody, wybierana jest hipoteza pierwszego klasyfikatora (z najwyższą skutecznością).

Autorzy rozwiązania wskazują na potencjalnie bardziej równomierne wyniki klasyfikacji i dokładniejsze wyniki w porównaniu do standardowych metod ze względu na priorytetowe traktowanie słabo rozpoznanych przypadków zamiast klasyfikacji dominującej klasy rozwiązań. Niewątpliwą zaletą rozwiązania jest promowanie dokładności programów przy jednoczesnym wsparciu zgodności co najmniej dwóch rozwiązań. Wadą rozwiązania są wskazanie poziomu zaufania ograniczone do wskazania kroku wyboru oraz możliwość przegłosowania programu przez słabsze programy o zbliżonym sposobie działania.

Wspólną wadą wszystkich rozwiązań dyskretnych jest brak możliwości wskazania poziomu intensywności emocji.

3.7.2 Metody ciągłe

W metodach ciągłych hipotezy programów automatycznego rozpoznawania emocji (ekspertów) są agregowane za pomocą funkcji statystycznych, takich jak maksimum, minimum, mediana, średnia czy średnia ważona [16] [110].

Jednym z przykładów jest kombinacja liniowa wykorzystująca funkcje statystyczne do łączenia danych z modalności (w przypadku wczesnej fuzji) lub hipotez (w przypadku późnej fuzji). W ważonej kombinacji liniowej poszczególnym programom przypisywane są wagi, gdzie 0 oznacza zignorowanie, a 1 oznacza wzięcie pod uwagę tylko danego programu [77] [86] [112]. Wagi są wyznaczone na podstawie danych treningowych tak, by zmaksymalizować skuteczność rozpoznawania.

Metody ciągłe pozwalają na szybką agregację dowolnej liczby hipotez, także wyliczenie odległości uzgodnionego rozwiązania od hipotez początkowych. Informację o jakości danych mogą zostać wykorzystane jako wagi w rozwiązaniu. Do wad można zaliczyć wrażliwość na wartości odstające, które zaburzają wynik [34]. Rozwiązania te zazwyczaj nie mają możliwości promowania zbieżności rozwiązania. Podobnie jak w przypadku metod dyskretnych – stosowane wagi często przypisywane są statycznie.

3.7.3 Metody klasyfikacji i metody rozmyte jako funkcja integracji w późnej fuzji

Metody klasyfikacji i metody rozmyte wykorzystywane są jako metody integracji w rozwiązaniach opartych o późną lub hybrydową fuzję. Metody rozmyte zwracają wartości rozwiązania w przedziale ciągłym $[0,1]$, które mogą być związane z takimi miarami dowodów jak: prawdopodobieństwo, możliwość (ang. possibility), zaufanie (ang. belief), wiarygodność (ang. plausibility) oraz konieczność (ang. necessity). Metody klasyfikacji w tym zastosowaniu wykorzystują hipotezy ekspertów (programów automatycznego rozpoznawania emocji) jako podstawę do zbudowania wejściowego wektora cech dla klasyfikatora.

Najczęściej stosowane metody obejmują [18] [19] [113]:

- maszyna wektorów nośnych (ang. support vector machine) [34],
- całki rozmyte (ang. fuzzy integrals) [18],
- szablony rozmyte (ang. fuzzy templates) [18] [113] [114],
- drzewa decyzyjne [113],
- metoda fuzji Bayesa (ang. Bayesian inference fusion) [18] [34] [102] [114],
- kombinacja Dempstera–Shaffera (ang. Dempster–Shaffer combination) [18],
- dynamiczne sieci Bayesa (ang. dynamic Bayesian network) [34],
- ukryte modele Markova (ang. hidden markov model) [115],
- produkty ekspertów (ang. product of experts) [18],
- uogólniona macierz odwrócona (ang. generalised inverse matrix) [114],
- sztuczne sieci neuronowe (ang. artificial neural networks) [18] [114] [116],
- sieci BLSTM (ang. bidirectional long short–term memory) [76],
- modele maksymalnej entropii (ang. maximum entropy models) [34].

Technika **szablonów rozmytych** przyjmuje hipotezy cząstkowe jako wejście do klasyfikatora, a następnie wyznacza poziom wsparcia każdej hipotezy przez poszczególne klasyfikatory. Zwraca macierz wyników, której wartości można zagregować za pomocą funkcji statystycznych (opisanych wcześniej w tym podrozdziale). Metoda jest stosunkowo prosta i elastyczna. Pozwala na wykorzystanie bardziej złożonych metod do agregacji.

Kolejną metodą jest wykorzystanie **drzewa decyzyjnego** (ang. ODT – Ordinary Decision Tree) lub meta–decyzyjnego (ang. MDT – Meta Decision Tree). W drzewach ODT przewidywany jest poziom prawdopodobieństwa stanu emocjonalnego. W drzewach MDT liście drzewa decydują o tym, który klasyfikator powinien zostać wykorzystany [117]. Drzewa MDT zostały wykorzystane przez Wu i Lianga [73] podczas integracji danych akustycznych z ładunkiem emocjonalnym z treści. Wyniki otrzymane z drzewa zostały uśrednione przy pomocy modelu maksymalnej entropii, który posłużył jako wagi.

Wyróżnia się dwie metody **fuzji Bayesa**: prostą średnią Bayesowską, a także metodę wprowadzającą miarę zaufania wyliczaną dla każdej hipotezy klasyfikatora. Pierwsza z nich polega na uśrednieniu prawdopodobieństwa wartości hipotez oraz ich współczynników zaufania. Druga wprowadza miarę zaufania wyliczaną dla każdej hipotezy klasyfikatora. Możliwa jest eliminacja hipotez, jeżeli współczynnik zaufania jest poniżej przyjętego progu. Metoda pozwala na wyznaczenie współczynnika wiarygodności na podstawie funkcji zaufania hipotez wejściowych. Aby skorzystać z fuzji Bayesa, hipotezy muszą mieć współczynnik zaufania wyrażony w prawdopodobieństwie a posteriori.

Przykładem wykorzystania klasyfikatora Bayesa jest integracja danych z ekspresji mimiki twarzy, mowy ciała, gestów oraz prozodii dla rozpoznania ośmiu emocji. Wykorzystano platformę EyesWeb do śledzenia ruchu, której głównymi estymatami są wykrywanie ruchu sylwetki oraz płynność. W mowie skupiono się na intensywności, wysokości dźwięku, a także pauzach. Dokładność pojedynczych modalności wynosiła 55%, a wynik fuzji hybrydowej to 78%. Dla tych samych danych wykonano także integrację metodą wczesnej fuzji, dla której wynik ten nie różnił się znacząco [108].

Teoria ewidencji Dempstera–Shaffera generalizuje Bayesowską teorię prawdopodobieństwa. Skupia się na fuzji niezależnych hipotez w jedną klasę, eliminując wadę metody Bayesowskiej, polegającą na wymogu współczynnika zaufania wyrażonego w postaci prawdopodobieństwa a posteriori [34]. Wykorzystywane są miary niepewności (ang. uncertainty) oraz nieprecyzyjności (ang. imprecision) poprzez definicję miar zaufania (ang. belief) oraz wiarygodności (ang. plausibility). Wybierana jest hipoteza o najwyższej wartości miary zaufania, która wyrażona jest w zakresie (0,1). Możliwe jest dodanie progu tolerancji dla odrzucenia hipotez o niskim poziomie zaufania, jak również odrzuceniu hipotez o poziomie zaufania równym 1, gdyż poziom zaufania równy 1 oznacza absolutną pewność oznaczającą, że inne hipotezy nie są potrzebne [18].

Dynamiczne sieci neuronowe są rozszerzeniem metody fuzji Bayesa do sieci grafów, gdzie węzły reprezentują różne modalności, a krawędzie oznaczają ich zależności probabilistyczne. Zaletą tej metody jest możliwość łatwej integracji czasowej danych wielomodalnych. Popularną formą DBN są ukryte sieci Markova (ang. HMM - Hidden Markov Model) [34].

Ukryte modele Markova (HMM) są statystyczną metodą klasyfikacji sekwencji zdarzeń. Wykorzystują one prawdopodobieństwo hipotez, które mogą reprezentować poziom zaufania programu do hipotezy. Rozszerzeniem HMM jest CHMM (ang. Conditioned Hidden Markov Model), który pozwala na wskazanie docelowego poziomu zaufania do uzgodnionej hipotezy [118].

Przy zastosowaniu **sieci neuronowych** danymi wejściowymi mogą być cechy pochodzące z różnych modalności (dla rozwiązań wczesnej fuzji) lub hipotezy ekspertów (dla późnej fuzji). Ukryta warstwa neuronów zawiera wagi, które są wyznaczone na etapie uczenia, aż do osiągnięcia zamierzonego poziomu dokładności.

Sieci syntezy Markova (ang. Markov fusion networks) zostały zaprojektowane w celu połączenia decyzji z wielu źródeł i są silnie powiązane z losowymi polami Markova (ang. Markov random fields), wykorzystywanymi w dziedzinie przetwarzania obrazu [119] [120]. Ostateczny poziom zaufania do wypracowanej metody został uzyskany na podstawie odchylenia standardowego hipotez (im bardziej zróżnicowane, tym mniejsze zaufanie do odpowiedzi) [119].

Zaletami metod klasyfikacji i metod rozmytych jako funkcji integracji w późnej fuzji jest działanie na danych ciągłych, a także możliwość wykorzystania jakości danych i współczynnika wiarygodności hipotez jako wag rozwiązania. Wszystkie powyższe metody są niezależne od kanałów wejściowych, a także pozwalają na wskazanie współczynnika zaufania dla uzgodnionego wyniku. Niektóre z metod pozwalają na eliminację hipotez o zbyt niskim poziomie zaufania poprzez stosowanie minimalnych progów.

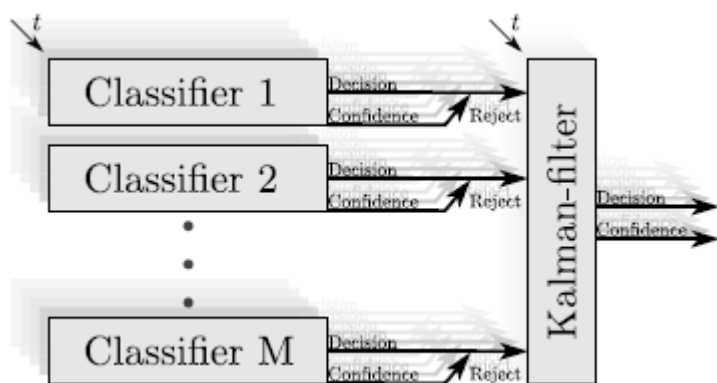
Rozwiązania te są zdecydowanie bardziej złożone od metod dyskretnych i ciągłych, m.in. wymagają wskazania współczynnika zaufania. Metody oparte na klasyfikatorach są wrażliwe na brak danych dla jednego z kanałów (podobnie jak w metodach wczesnej fuzji), a także ich dokładność jest zależna od dobranych danych trenujących. Z kolei metody bazujące na modelach Markova wymagają sekwencyjnego uruchomienia klasyfikatorów.

3.7.4 Metody oparte na szacowaniu

Rozwiązania tej kategorii stosowane są często do szacowania pozycji ruchomych obiektów, wykorzystując przy tym wielomodalną informację zwłaszcza audio i wideo. Przykładami tych metod są: filtr Kalmana, rozszerzony filtr Kalmana, a także filtr cząsteczkowy (ang. particle filter) [34], z czego warto szerzej opisać filtr Kalmana ze względu na możliwość zastosowania w integracji w wielomodalnym rozpoznawaniu emocji.

Filtr Kalmana (ang. Kalman filter) jest używany szeroko w nawigacji [121] i śledzeniu obiektów [122]. Celem filtru jest zmniejszenie szumu poprzez uwzględnienie wielu pomiarów i najnowszych oszacowań [123]. Algorytm ten zakłada, że wszystkie pomiary są niepewne. Brakujące wyniki klasyfikatora są rozwiązywane poprzez zwiększenie niepewności.

Oszacowanie podzielone jest na dwa kroki: predykcja oraz aktualizacja (nazywana również krokiem korekty). Aktualizacja bierze pod uwagę ostatnie n pomiarów ze wszystkich wykorzystanych klasyfikatorów. Wartość zaufania służy od odrzucania decyzji z niskim poziomem zaufania. Filtr Kalmana określa przybliżone oszacowanie na podstawie bieżących i przeszłych decyzji [123]. Schemat działania został zaprezentowany na Rys 3.2. Filtr Kalmana może zostać zastosowany dla sekwencyjnej struktury ekspertów oraz jeżeli spełnione zostaną założenia Markova. Filtr Kalmana jest ograniczony do układów liniowych, dlatego dla systemów nie-liniowych stosuje się rozszerzony model Kalmana [34].



Rys 3.2 Schemat działania filtra Kalmana [34]

Filtr Kalmana poprzez prognozowanie stanu emocjonalnego łagodzi skutki braku danych z kanałów informacyjnych [40], a także dużych skoków pomiędzy kolejnymi stanami emocjonalnymi (duży dystans na wykresie) korzystając z założenia – że gwałtowne przejścia z jednego stanu emocjonalnego do skrajnie przeciwnego są rzadkie. Inną zaletą filtra Kalmana jest branie pod uwagę historycznych stanów, a także odrzucenie hipotez z niskim poziomem zaufania oraz programów, które mają zbyt mało danych.

Do wad rozwiązania możemy zaliczyć wymaganie układu liniowego, które jednak można obejść poprzez zastosowanie rozszerzenia metody dla układów liniowych. Filtr Kalmana jednak nie zwraca końcowego współczynnika zaufania – jest on wykorzystywany do eliminacji hipotez.

Jednym z przykładów zastosowania filtra Kalmana jest projekt komunikatora czasu rzeczywistego [40]. Hupont i inni wykorzystali w nim trzy modalności:

- tekstu napisanego przez użytkowników,
- emotikon znajdujące się w tekście,
- analizy mimiki twarzy rozmówców.

Wykorzystano fuzję na poziomie decyzyjnym ze względu na różny czas otrzymywania danych – jedynie dane dotyczące mimiki były dostarczane regularnie. Emotikony zostały wykorzystane jako dodatkowe źródło informacji, które miało pomóc w interpretacji wypowiedzi sarkastycznych i ironicznych.

Kanały wejściowe otrzymały statycznie przypisany współczynnik wiarygodności:

- tekst – 65%, gdyż tyle wyniosła skuteczność na danych testowych,
- mimika twarzy – 87%, gdyż tyle wyniosła skuteczność na danych testowych,
- emotikony – 100%, gdyż założono, że użytkownik świadomie wyraża w ten sposób emocje.

Każdy z kanałów miał swój własny model reprezentacji emocji:

- treść – 24 etykiet emocji,
- mimika twarzy – model emocji podstawowych Ekmana, rozszerzony o emocję neutralną,
- emotikony – 16 etykiet emocji.

Na wynikowy model został wybrany model 2D ze względu na wielowymiarowy charakter. Sprowadzenie hipotez do tego modelu nastąpiło poprzez odwzorowanie etykiet do przypisanych im współrzędnych modelu 2D. Hipotezy zostały zintegrowane za pomocą opracowanej przez zespół funkcji integracji. Na końcu wynik poddawany jest „filtrowaniu kinematyki emocjonalnej” za pomocą filtru Kalmana.

3.8. Podsumowanie metod integracji wielomodalnej

Podczas analizy wielomodalnej, możliwe są trzy warianty fuzji: fuzja wczesna, późna lub fuzja hybrydowa.

Przykładem wczesnej fuzji są badania przeprowadzone w projekcie AVEC [61]. Wykorzystano w nim kilka modeli regresji liniowych – odpowiednio dla każdej kombinacji dostępnych modalności. W przypadku braku danych w jednej z modalności rozwiązanie przełączało się na inny model regresji. Każdy model był trenowany na innym zbiorze danych i cechował się inną skutecznością.

Komunikator czasu rzeczywistego opracowany przez Hupont i innych [40] jest przykładem zastosowania późnej fuzji. Dane z każdej modalności były analizowane indywidualnie, sprowadzane do wspólnego modelu reprezentacji (2D), a następnie łączone poprzez funkcję integracji, a na końcu wynik poddawany został „filtrowaniu kinematyki emocjonalnej” za pomocą filtru Kalmana.

Z kolei zastosowanie klasyfikatora Bayesa na danych z każdej modalności, a następnie poddanie wyników klasyfikacji, jest przykładem wykorzystania funkcji hybrydowej [108].

Na podstawie przytoczonej w rozdziale klasyfikacji oraz powyższych przykładach, autor pracy zaproponował i zaimplementował rozwiązanie umożliwiające wykorzystanie każdego z trzech rodzajów fuzji, ze wskazaniem dla późnej fuzji – ze względu na większą elastyczność rozwiązania. Poniżej przedstawiono elementy wcześniejszych badań wykorzystane w przez autora rozprawy.

Z metod dyskretnych autor pracy zaczerpnął dwa główne kryteria dla funkcji integracji: zbieżność hipotez z metod głosowania oraz preferencję najbardziej wiarygodnej hipotezy – z podejścia specjalistów kaskadowych, gdzie najwyższą skuteczność zastąpiono poziomem wiarygodności. Z metod rankingowych zaczerpnięto możliwość wprowadzania progu odcięcia dla hipotez o niskiej skuteczności.

Metody ciągle oparte o średnią oraz średnią ważoną zostały wykorzystane w dalszej części jako rozwiązania referencyjne.

Rozmycie hipotez z metod rozmytych zostało wykorzystane przez autora pracy w zaproponowanej metodzie integracji. Z kolei z teorii ewidencji Dempstera–Shaffera, zaczerpnięte zostały miary niepewności, zaufania i wiarygodności oraz wybór hipotezy o najwyższej wartości miary wiarygodności.

Z wykorzystania filtru Kalmana, autor przyjął założenie o niepewności każdego pomiaru oraz zmniejszenie zaufania do rozwiązań, które mają zbyt mało danych, a więc niższą jakość danych.

4. PROBLEM BADAWCZY

Wyzwania w dziedzinie rozpoznawania emocji można podzielić na dwie grupy: wyzwania programów automatycznego rozpoznawania emocji oraz wyzwania związane z integracją wielokałową i wielomodalną. Wyzwania związane z programami automatycznego rozpoznawania emocji zostały omówione w rozdziale 2.6. Podejście wielomodalne pozwala zająć się problemami odporności na zakłócenia i okresowej niedostępności sygnału w kanale obserwacji.

Do głównych wyzwań integracji w wielomodalnym rozpoznawaniu emocji należy zaliczyć:

- **niekompletność danych** – w przypadku czasowej lub kontekstowej niedostępności kanału [77] [124],
- **różny zakres stosowalności modalności** – niedostępność spowodowana kontekstem wykorzystania [1],
- odpowiedni **dobór modalności** do problematyki zadania [77]
- właściwe przypisanie **wag modalnościom** [124],
- **określenie niepewności** związanej z pomiarem symptomów emocji oraz wynikiem klasyfikacji [3],
- **różnorodność modeli reprezentacji emocji** – różna skala, precyzja oraz struktura danych [125];
- **rozbieżności** wyników **oszacowania stanu emocjonalnego** pomiędzy różnymi klasyfikatorami w tym samym kontekście [2] [63];
- **synchronizacja czasowa** danych wejściowych oraz programów automatycznego rozpoznawania emocji – różny czas i częstotliwość otrzymywania danych z różnych modalności [77] [124] [126],
- **adaptowalność** – rozszerzalność o nową modalność/kanał, często wymaga przebudowy całego systemu;
- **modularność** – łatwość wprowadzania zmian np. rozwój klasyfikatorów niezależnie od całej architektury.

Niekompletność danych, różny zakres stosowalności modalności, określenie niepewności pomiarowej, problem rozbieżności wyników oszacowania stanu emocjonalnego między różnymi klasyfikatorami, synchronizacja czasowa danych wejściowych, adaptowalność oraz modularność są obszarem badań opisanych w niniejszej rozprawie doktorskiej.

Zaproponowana metoda integracji pozwala na poprawę w wybranych obszarach wyzwań integracji. Te wybrane wyzwania zostały szerzej opisane w dalszej części tego rozdziału, z wyjątkiem adaptowalności i modularności, które zostały opisane przy okazji zaproponowanej metody – w rozdziale 5.

Rozdział ten bazuje częściowo na publikacjach [1] [2], których współautorem jest autor pracy.

4.1. *Wyzwanie: niekompletność danych*

Jednym z wyzwań związanych z automatycznym rozpoznawaniem emocji jest okresowy brak danych dla modalności. Zamyka on drogę do rozpoznania dla programów jednomodalnych, a także jest jednym z powodów dla powstania rozpoznawania wielomodalnego.

Przyczyny braku danych wejściowych z kanałów obserwacji mogą być różnorodne:

- awaria urządzenia;
- odłączenie się czujnika/mikrofonu;
- przesunięcie kamery – przy poprawnym działaniu kamery śledzona twarz może być poza zasięgiem kamery albo może być pod zbyt dużym kątem, co obniża skuteczność rozpoznawania emocji;
- jakość przesyłanego sygnału może być niska z powodu szumu otoczenia;
- okresowy brak istotnych informacji dla analizy modalności w kanale obserwacji np. gdy kanał audio jest rejestrowany poprawnie, ale użytkownik nic nie mówi, ale także gdy mówi jednocześnie więcej osób lub gdy w tle jest szum.

Podczas jednego z badań użyteczności oprogramowania z wykorzystaniem mechanizmów rozpoznawania emocji dostępność danych fizjologicznych wynosiła 100% czasu (poza jednym przypadkiem, gdy czujnik się odłączył), a dane dotyczące rozpoznawania emocji z twarzy były dostępne przez 42–100% czasu – w zależności od uczestnika badania [63].

W wielu naukowych badaniach zakłada się optymistyczny wariant, że dane dla każdego kanału obserwacji są zawsze dostępne – co jest łatwo osiągalne w warunkach eksperymentalnych, gdzie fragmenty z brakującymi danymi można wykluczyć z analizy. W przypadkach zastosowania czasu rzeczywistego oznaczałoby to wyłączenie modułu rozpoznawania emocji w przypadku awarii jednego z czujników. Z tego powodu, istotnym jest znalezienie sposobu na poprawę dostępności rozpoznawania [16].

Istotną wadą rozwiązań bazującego na wczesnej fuzji jest nieodporność klasyfikatora na brak części danych, gdy był on uczony na kompletnych danych. Skuteczność klasyfikacji we wczesnej fuzji jest ściśle związana z kompletnością wektora cech. W literaturze są znane sposoby radzenia sobie z brakiem danych dla metod wczesnej fuzji, takie jak predykcja na podstawie poprzednich próbek [127], przełączanie programu między różne klasyfikatory [61]. Z kolei dla późnej fuzji, znaną metodą jest ignorowanie brakujących danych aby nie zakłócać procesu decyzyjnego [128]. Często jest to zapewnione poprzez ustawienie progów wiarygodności (gdy są dostępne), poniżej których dane te nie są brane pod uwagę np. w przypadku głosowania większościowego [16]. W przypadku statystycznych metod, w których wykorzystywane są wagi, poziom zaufania do zaszumionego kanału jest stosownie obniżany, czasem nawet do zera [19]. Rozwiązania bazujące na późnej fuzji lepiej radzą sobie z brakiem danych w stosunku do metod wczesnej fuzji.

Proponowanym w niniejszej pracy podejściem jest uwzględnienie dynamicznej oceny dostępności kanałów obserwacji i dynamiczne dopasowanie poziomów zaufania do wyników działania modułów rozpoznawania.

4.2. **Wyzwanie: różny zakres stosowalności modalności**

Zakres stosowalności kanałów wejściowych jest zależny zarówno od badanego użytkownika (ang. user-dependent), jak i od kontekstu użycia (ang. task-dependent). Zgoda użytkownika na przetwarzanie jego danych w celu przetwarzania emocji, a także poinformowanie o przetwarzaniu tych danych to dwa z czterech głównych czynników regulujących ryzyko w dziedzinie automatycznego rozpoznawania emocji [9].

Zakres stosowalności zależny od użytkownika:

- użytkownik aplikacji internetowej może nie wyrazić zgody na udostępnienie danych z kamery, co będzie skutkowało brakiem dostępności tego kanału,
- okrycia twarzy – takie jak okulary, wąsy, zarost, twarz zasłonięta dłonią są często utrudnieniem dla programów automatycznego rozpoznawania emocji z analizy mimiki [2], [63], [120], [129],
- twarz poza kadrem kamery, ale także zbyt blisko ekranu np. podczas czytania [2], [120]
- analiza gestów – dostępna tylko, gdy użytkownik gestykuje, a jego dłonie są w zasięgu kamery [120],
- analiza mowy – jest dostępna tylko wtedy, gdy użytkownik mówi – brak danych do analizy, gdy użytkownik milczy, ale także, gdy mówi jednocześnie więcej osób lub gdy w tle jest szum [120].

Sam sposób ekspresji jest też zależny od użytkownika – jego temperamentu, tła kulturowego, płci, stanu zdrowia czy też życiowych doświadczenia, które wpływają na poziom i sposób ekspresji emocji [120].

Zakres stosowania modalności może być też podyktowany kontekstem użycia:

- brak kanału audio, jeżeli zadanie polega na pisaniu treści lub wykonania zadania przy użyciu myszki i klawiatury [103],
- brak kanału tekstowego np. w przypadku rozmowy wideo, zdalnego ustnego egzaminu, oglądania tutoriala [1],
- brak dostępu do danych fizjologicznych, jeżeli użytkownik łączy się z domu za pomocą własnego laptopa, a nie w sali laboratoryjnej, gdzie dostępny jest specjalistyczny sprzęt [26],
- w zależności od aktywności w zadaniu np. mimika twarzy zmienia się najczęściej przy interakcji z człowiekiem niż z komputerem [1],
- zakłócenia danych fizjologicznych – jeżeli czujniki są przyłączone do miejsc, które są ruchome w czasie badania, takie jak palce przy zadaniach wykorzystujących urządzenia peryferyjne [26].

W Tabeli 4.1 [1] zostało przedstawione zestawienie dostępności modalności w zależności od kontekstu e-nauczania. Oznaczenia: AV – modalność dostępna (ang. available), UN – niedostępna (ang. unavailable), TD – dostępność zależna od zadania (ang. task dependency), UD – dostępność zależna od użytkownika (ang. user dependency).

Tabela 4.1 Dostępność modalności w zależności od kontekstu e-nauczania [1].

Kanał obserwacji	Kontekst				
	Domowy, model asynchroniczny	Domowy, wirtualna klasa	Laboratorium komputerowe	Nauczanie mobilne	Warunki eksperymentalne
Raportowanie własnych emocji przez badanego	AV	AV	AV	AV	AV
Ekspresja mimiczna	UD	AV	AV	UD	AV
Wzorce wykorzystania urządzeń peryferyjnych	TD, UD	TD, UD	TD	TD, UD	TD
Analiza prozodii mowy	TD, UD	AV	TD	TD	TD
Rozpoznawanie ładunku emocjonalnego z tekstu	TD	TD	TD	UN	TD
Dane fizjologiczne	UN	UN	UN*	UN*	AV

Warte uwagi jest też inne zestawienie z tej samej publikacji, mówiące o poziomie dostosowania kanału do kontekstu edukacyjnego, które zostało przedstawione w Tabeli 4.2 [1]. Pokazuje ono, że żadna modalność nie jest wolna od wad. Pozornie dobrą modalnością jest raportowanie własnych emocji przez badanego ze względu na najwyższą odporność na zakłócenia, jednakże charakteryzuje się ono jednocześnie niską niezależnością od woli człowieka. Z kolei analiza prozodii mowy, dane fizjologiczne oraz mimika twarzy są w dużym stopniu niezależne od woli, ale mocno podatne na zakłócenia. Analiza danych fizjologicznych jest problematyczna pod względem wygody w nauczaniu zdalnym, a analiza mimiki pod kątem kompatybilności z nauczaniem zdalnym ze względu na najczęściej wykorzystywany model reprezentacji emocji, jakim jest model emocji podstawowych Ekmana. Emocje podstawowe takie jak radość, złość, gniew są rzadkie w kontekście edukacyjnym, a najbardziej wartościowymi są stany takie jak poziom zaangażowania oraz skupienia [90] [91].

Tabela 4.2 Porównanie modalności w kontekście edukacyjnym. Wartość Wysoka – oznacza lepsze dostosowanie, Niska – oznacza gorsze dopasowanie [1].

Modalności	Odporność na zakłócenia	Kompatybilność z nauczaniem zdalnym	Niezależność od woli człowieka	Wygoda w nauczaniu zdalnym
Raportowanie własnych emocji przez badanego	Wysoka	Średnia	Niska	Średnia
Ekspresja mimiczna	Niska	Niska	Średnia	Wysoka
Wzorce wykorzystania urządzeń peryferyjnych	Średnia	Średnia	Wysoka	Wysoka
Analiza prozodii mowy	Niska	Wysoka	Wysoka	Średnia
Rozpoznawanie ładunku emocjonalnego z tekstu	Średnia	Średnia	Niska	Wysoka
Dane fizjologiczne	Niska	Wysoka	Wysoka	Niska

Zgodnie z literaturą, czujniki dotykowe – takie jak czuła na nacisk mysz czy krzesło powodują mniejszy dyskomfort i uczestnicy badania ich nie zauważają, jednakże wykrywają tylko część ruchów [71]. Modalności bezdotykowe takie jak analizy mimiki twarzy czy analiza prozodii mowy nie są fizycznie uciążliwe, ale sprawiają wrażenie dyskomfortu poprzez świadomość bycia obserwowanym [71] i mogą powodować zaburzenie wyników – znane jako efekt Hawthorne [130].

4.3. Wyzwanie: określenie niepewności związanej z pomiarem symptomów emocji oraz wyniku klasyfikacji

Zgodnie z założeniami filtra Kalmana, opisanego w rozdziale 3.7, wszystkie badania obarczone są niepewnością pomiarową, a wskazania ze zbyt małą ilością danych lub niską ich jakością powinny zostać oznaczone wyższą niepewnością.

Zgodnie z literaturą przedstawioną w rozdziale 3.7, w rozwiązaniach wielomodalnego rozpoznawania emocji miary skuteczności są wykorzystywane jako wagi w funkcjach integracji. Wartości te są zazwyczaj wyznaczane na etapie uczenia klasyfikatorów, ignorując jakość danych podczas pomiaru.

Rozwiązaniem problemu określenia niepewności jest uwzględnienie w ocenie wskaźnika jakości informacji dla analizowanego fragmentu danych. Pozwala to wykryć i odpowiednio uwzględnić sytuację nagłego pogorszenia się warunków badania np. w postaci zmiany oświetlenia, ruchu głowy czy zwiększenia szumu w kanale obserwacji [3]. Ma to szczególne znaczenie, gdy w jednym z kanałów występuje pogorszenie, a drugi ma nadal prawidłowe warunki.

Osobną kwestią jest brak określenia niepewności przez rozwiązania fuzji. W języku technicznym stosuje się wiele terminów w celu określenia problemów związanych z niepewnością i wiarygodnością pomiaru m.in.: niedokładność (ang. inaccuracy), nieprecyzyjność (ang. imprecision), błąd oraz ich przeciwieństwa: dokładność (ang. accuracy), pewność (ang. confidence), wiarygodność (ang. plausibility), zaufanie (ang. trust) itp. [20]

Sposób wyznaczenia niepewności jest różny w zależności od funkcji integracji – możliwości wyznaczenia współczynnika zaufania zostały zaprezentowane w rozdziale 3.7 przy poszczególnych metodach integracji. Są to między innymi: poziom poparcia hipotezy, wyrażony w stosunku liczby głosów za odpowiedzią do wszystkich głosów (metody głosowania) czy wartości miar: prawdopodobieństwa, możliwości, zaufania, wiarygodności, konieczności czy niepewności.

Zgodnie z metodami integracji późnej fuzji znanymi z literatury (rozdziale 3.7), dwoma ważnymi kryteriami dla funkcji integracji są: promowanie zbieżności oraz preferencja najbardziej wiarygodnej hipotezy. Stąd następujące założenia dla metod wyznaczania niepewności pomiarowej w rozpoznawaniu emocji:

- promowanie zbieżności – im więcej programów popiera hipotezę, tym poziom wiarygodności powinien być wyższy (niższa niepewność pomiaru) i odwrotnie, im więcej rozbieżnych hipotez, tym poziom zaufania powinien być niższy (większa niepewność pomiaru) [19],
- im większe różnice w hipotezach, tym większa niepewność do uzgodnionego rozwiązania [19].

Niepewność pomiarowa może być reprezentowana jako współczynnik wiarygodności lub poprzez przedziały ufności [131].

4.4. **Wyzwanie: rozbieżności hipotez automatycznego rozpoznawania emocji w tym samym kontekście użycia**

Kolejnym z wyzwań związanych z rozwiązaniami późnej fuzji są przypadki dużych rozbieżności pomiędzy programami, w tym samym czasie i kontekście użycia.

Podczas badania użyteczności w grze edukacyjnej GraPM mierzono estymaty emocjonalne za pomocą dwóch modalności: analizy mimiki twarzy oraz ankiet opisu stanu emocjonalnego wypełnianej przez badanego. Co najbardziej interesujące, znaleziono duże rozbieżności pomiędzy nimi. Raportowane emocje były pozytywne, a rozpoznane przez program były negatywne – jedne i drugie z dodatnim pobudzeniem, jednak te zaraportowane przez użytkownika przeważnie charakteryzowały się wyższym jego poziomem [63]. Rozbieżności te sugerują popełnienie błędu w badaniu, jednakże manualna weryfikacja wskazała, że program pomylił skupienie z emocją gniewu, prawdopodobnie przez lokalizację kamery. W tego typu badaniach zawsze wpływ może mieć kwestia ukrywania emocji lub zredukowanie ekspresji przy użytkownaniu komputera, które powoduje sprzeczność pomiędzy różnymi kanałami informacji – jednakże jest to kwestia nierozwiązywalna z punktu widzenia maszyny [16].

Innym przykładem są badania, w których zauważono znaczące różnice pomiędzy wynikami podczas tego samego pomiaru z zachowaniem niezmienności użytkownika, modalności, a także programu. Jediną różnicą było umiejscowienie kamery. Dolna kamera przeszacowywała emocję zaskoczenia (prawdopodobnie dlatego, że od dołu brwi wydają się bardziej uniesione niż w pozycji na wprost), górna kamera emocję gniewu (prawdopodobnie dlatego, że brwi wydają się bardziej obniżone niż w pozycji na wprost). Spójność emocji w modelu emocji podstawowych Ekmana, jak i dla wymiaru wartościowości była niska. Wysoka spójność wystąpiła jedynie dla pobudzenia [2]. Badanie te pokazują, jak duże znaczenie ma umiejscowienie kamery, co potwierdza obserwację, że jakość danych wejściowych ma duży wpływ na wynik rozpoznawania.

4.5. **Wyzwanie: synchronizacja czasowa danych wejściowych oraz programów automatycznego rozpoznawania emocji**

Dane z różnych modalności są dostarczane w różnym czasie i z różną częstotliwością, a także czas działania różnych programów może się różnić. Przykładem równomiernego otrzymywania informacji są modalności wideo oraz dane fizjologiczne, dla których otrzymywane informacje otrzymywane są nieprzerwanie i regularnie, o ile dysponujemy odpowiednio skonfigurowanymi czujnikami np. kamery o podobnych parametrach. Dostępność kanału (kwestie niekompletności danych zostały opisane w podrozdziale 4.1) nie oznacza otrzymania estymat stanu emocjonalnego, na co wskazują badania, w których czas dostępności twarzy w kamerze wynosił od 40 do 100% (w zależności od badanego) [63].

Wyzwaniem synchronizacji czasowej jest fuzja danych z modalności o różnej częstotliwości napływania sygnałów, jak ma to miejsce w przypadku wykorzystania komunikatora czasu rzeczywistego [40], w którym wykorzystano dane z modalności mimiki twarzy, głosu oraz danych tekstowych. Dane z tych modalności mają naturę nieciągłą, ograniczoną odpowiednio do: momentów uchwycenia twarzy w kamerze, wypowiedzi badanego bez przerywania przez rozmówcę oraz do momentów pisania wiadomości.

Najczęstszym sposobem radzenia sobie z wyżej opisanymi trudnościami jest wyznaczenie okna czasowego [40], w którym zbierane są dane ze wszystkich kanałów. Preferowane jest zbieranie informacji na poziomie zdarzenia (ang. event-level) lub na poziomie segmentów odpowiedniej długości. Analiza na poziomie zdarzenia, jakim może być cała wypowiedź użytkownika, jest preferowana w przypadku analizy ładunku emocjonalnego z modalności mowy [116].

Rzędem wielkości analizowanych segmentów powinny być sekundy ze względu na długość trwania reakcji emocjonalnej, który trwa od kilku do kilkunastu sekund [77]. Zdania wśród psychologów są jednak podzielone odnośnie do długości ich trwania: większość wskazuje, że emocje trwają od 0,5 do 4s. Część psychologów wskazuje jednak, że emocje mogą trwać nawet od kilku minut do nawet godzin [25], inni zaś uznają te dłuższe stany za nastrój, a nie emocje [89].

W projekcie AVEC 2016 wykorzystano okna czasowe o długości od 4 do 14s, z przesunięciem o 2s dla kolejnych okien [61]. W swoich badaniach wykorzystali cztery modalności: wideo (mimika twarzy), audio (prozodia mowy), a także dwie modalności fizjologiczne – ECG oraz EDA. Inna grupa badawcza również wykorzystwała mechanizm okna czasowego dla integracji danych, które wynosiło 8–10 s, z przesunięciem dla kolejnych okien 4s [132]. W swoim badaniu wykorzystali tylko modalność audio, jednakże przy wykorzystaniu dwóch różnych programów.

4.6. Podsumowanie analizy stanu wiedzy

Autor pracy w zaproponowanym rozwiązaniu bazuje na następujących elementach analizy literatury:

- Istniejący model reprezentacji emocji – model emocji podstawowych Ekmana oraz model PAD (3D) wraz z odwzorowaniami, których wykorzystanie zależy jednak od wykorzystywanych programów automatycznego rozpoznawania emocji. Zarówno modele, jak i odwzorowania zostały opisane w rozdziale 2;
- Metodę późnej fuzji, wraz ze wsparciem metod wczesnej oraz hybrydowej fuzji, które zostały opisane w rozdziale 3;
- Dwa główne kryteria oceny funkcji integracji – zbieżność hipotez oraz preferencję najbardziej skutecznej/wiarygodnej hipotezy, wraz z możliwością wprowadzenia proggu odcięcia dla hipotez o niskiej skuteczności/wiarygodności – zaczerpnięte z literaturowych przykładów metod integracji późnej fuzji, opisanych w rozdziale 3.7;
- Metody ciągłe integracji (opisane w rozdziale 3.7.2) zostały wykorzystane jako metody referencyjne;
- Rozmycie hipotez z metod rozmytych zostało wykorzystane przez autora pracy w zaproponowanej metodzie integracji, opisanych w rozdziale 3.7.3;
- Z kolei z teorii ewidencji Dempstera–Shaffera (rozdział 3.7.3), zaczerpnięte zostały miary niepewności, zaufania i wiarygodności oraz wybór hipotezy o najwyższej wartości miary wiarygodności;
- Z wykorzystania filtra Kalmana (rozdział 3.7.4), autor przyjął założenie o niepewności każdego pomiaru oraz zmniejszenie zaufania do rozwiązań, które mają zbyt mało danych, a więc niższą jakość danych;
- Ze sposobu reprezentacji niepewności pomiarowej, omówionej w rozdziale 4;
- Z wykorzystania innych metod integracji zaczerpnięte zostało także słownictwo:
 - Eksperta – rozumianego jako program automatycznego rozpoznawania emocji zgodnie z literaturą [15] [16] [17] [18] [19],
 - Hipotezy – rozumianej jako rezultat działania programu automatycznego rozpoznawania emocji (eksperta). Dla programów klasyfikujących wynik działania to hipoteza (klasa), natomiast dla programów estymacji jest to wartość liczbowa w pewnej skali (estymata), jednak w przetwarzaniu emocji często stosuje się te dwa terminy zamiennie [18] [19].

5. PROPONOWANE ROZWIĄZANIE

Rozdział ten został poświęcony rozwiązaniu opracowanemu i zaimplementowanemu przez autora niniejszej rozprawy. Na proponowane rozwiązanie, czyli metodę integracji wyników w wielokanałowym lub wielomodalnym rozpoznawaniu emocji składa się:

- architektura oparta o wzorzec architektoniczny Scoreboard, umożliwiająca integrację wielu programów automatycznego rozpoznawania emocji;
- propozycja sposobu wyliczania współczynnika niepewności dla programów automatycznego rozpoznawania emocji;
- metoda dynamicznej oceny działania programów automatycznego rozpoznawania emocji w czasie działania aplikacji;
- algorytm integracji oparty o współczynniki niepewności i zbieżność hipotez o stanie emocjonalnym użytkownika, bazujący na miarach niepewności, zaufania oraz wiarygodności zmodyfikowanej teorii ewidencji Dempstera–Shaffera.

Rozwiązanie to odpowiada na wybrane wyzwania wielomodalnego rozpoznawania emocji – wskazane na podstawie przeglądu literatury w rozdziale 4. Elementy metody zaczerpnięte z literatury zostały opisane w rozdziale 4.6.

Rozdział ten bazuje częściowo na publikacjach [1] [2] [3] [4] oraz dokumencie technicznym [5], których współautorem jest autor pracy.

5.1. *Koncepcyjny model rozwiązania*

Głównym problemem badawczym postawionym w niniejszej rozprawie jest integracja danych o symptomach stanów emocjonalnych człowieka z wielu źródeł, a także wielu programów.

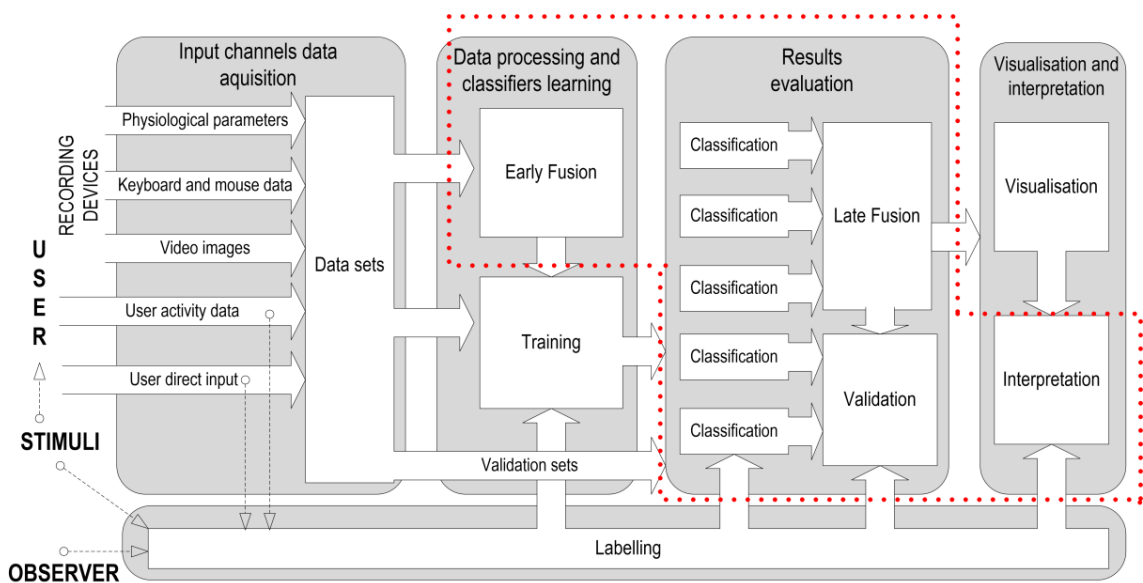
Zgodnie z analizą literatury przeprowadzoną w rozdziale 3, istnieją dwie główne kategorie rozwiązań wielomodalnych: wczesna fuzja oraz późna fuzja, a także połączenie obu metod w postaci fuzji hybrydowej. Liczne publikacje (cytowane w rozdziale 3) wskazują na różny zakres stosowalności obu podejść w zależności od kontekstu użycia. Z reguły im większa różnorodność rozwiązań (różne modalności, różne czasy otrzymania informacji), tym większy potencjał ma rozwiązanie późnej fuzji. Z tego względu w niniejszym rozwiązaniu najwięcej uwagi jest poświęcone późnej fuzji, jednak fuzja wczesna oraz hybrydowa także są możliwe do realizacji w zaproponowanej architekturze.

W rozwiązaniu tym możliwa jest:

- wczesna fuzja, jeżeli zostanie wykorzystany jeden klasyfikator, wykorzystujący wiele kanałów informacji,
- późna fuzja, jeżeli każdy z programów będzie wykorzystywał tylko prywatne kanały informacji.
- fuzja hybrydowa, jeżeli każdy z programów będzie wykorzystywał więcej niż jeden współdzielony kanał informacji.

Koncepcyjny model integracji w wielomodalnym rozpoznawaniu emocji został zilustrowany na Rys 5.1 [3]. Zagadnienia integracji, czyli zakres, którym zajmuje się niniejsza praca, znaczone na rysunku wykropkowaną czerwoną linią. Istnieje szereg zagadnień takich jak trenowanie klasyfikatorów, etykietowanie danych czy wizualizacja wyników, które to obszary, choć ważne badawczo, nie stanowią jednak przedmiotu badań przedstawionych w niniejszej rozprawie. W szczególności zaproponowane rozwiązanie nie wkracza w problematykę algorytmów uczenia maszynowego służących automatycznemu rozpoznawaniu emocji ani w problematykę przekształcania sygnału z kanału obserwacji do postaci wektora cech. W pewnym sensie programy automatycznego rozpoznawania emocji są traktowane zgodnie z podejściem czarnej skrzynki (ang. black-box), co pozwala na koncentrację na kwestiach integracji.

Zagadnienia będące przedmiotem tej pracy obejmują: integrację danych z wielu kanałów i modalności (fuzja); uruchomienie programów automatycznego rozpoznawania emocji na tych danych (klasyfikacja), w celu uzyskania hipotez o stanie emocjonalnym użytkownika, sprowadzenie hipotez do wspólnego modelu reprezentacji w celu ich porównania i walidacji, integracja wielu niezależnych hipotez o stanie emocjonalnym (późna fuzja), a także reprezentację informacji o niepewności związanej z uzyskanym wynikiem rozpoznawania na potrzeby walidacji i interpretacji wyniku.



Rys 5.1 Koncepcyjny model integracji w wielomodalnym rozpoznawaniu emocji [3].

Przedstawiona koncepcja posiada następujące założenia:

- kanały wejściowe są nagrywane przez pewną, określoną liczbę urządzeń wejściowych (np. kamery, mikrofony),
- kanały wejściowe niosą informację o symptomach stanów emocjonalnych (np. mimika twarzy, rytm serca),
- dane z każdego kanału są współdzielone pomiędzy wszystkie programy automatycznego rozpoznawania emocji (klasyfikatory), które mogą, ale nie muszą z nich korzystać,

- pośrednie wyniki (np. wektory cech) oraz rezultaty działania poszczególnych programów są współdzielone, co pozwala realizować wariant wczesnej, jak i późnej fuzji oraz metaanalizy działania programów automatycznego rozpoznawania emocji,
- możliwe jest wykorzystanie zamkniętego kodu (zgodnie z podejściem czarnej skrzynki, ang. black box) dla rozwiązań późnej i hybrydowej fuzji,
- możliwa jest synchronizacja czasowa dla kanałów wejściowych (poprzez odpytanie za zadany okres) oraz programów automatycznego rozpoznawania emocji,
- dla podejść późnej i hybrydowej fuzji, uzgodnione rozwiązanie określane jest poprzez funkcję integracji uwzględniającą niepewność związaną z kanałami obserwacji.

5.2. *Architektura rozwiązania oparta o wzorzec projektowy Scoreboard*

Dla problemu integracji istotną kwestią jest wybór odpowiednich programów oraz przypisanie im odpowiednich wag, ale także eliminacja wpływu błędnych wyników spowodowanych np. niską jakością danych. W odpowiedzi na te problemy, a także w zgodzie z przedstawioną na Rys 5.1 koncepcją, autor pracy zaimplementował wzorzec architektoniczny Tabeli Ocen (ang. scoreboard) [4]. Spełnia on wymienione wcześniej założenia i pozwala na elastyczne zarządzanie zestawem programów automatycznego rozpoznawania emocji.

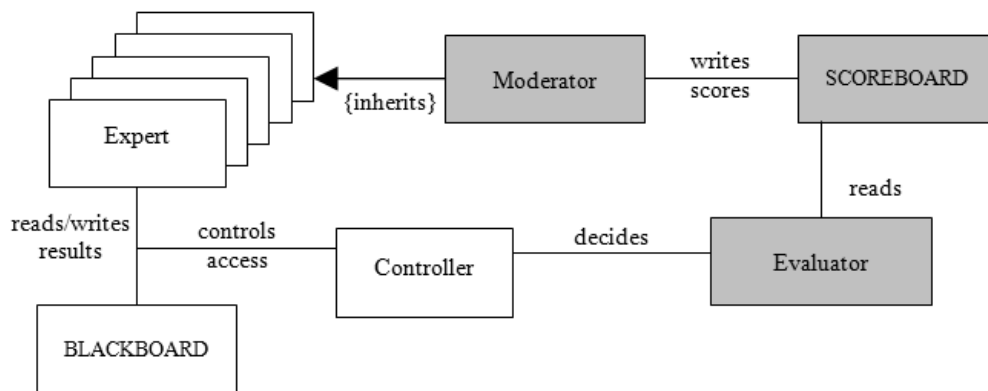
Wzorzec Tabeli ocen jest rozwinięciem wzorca tablicy (ang. blackboard). Wzorzec Tablicy (ang. blackboard) składa się z następujących klas:

- **tablicy** (ang. blackboard – od niej pochodzi nazwa wzorca), która jest współdzielona przez komponenty aplikacji. Są na niej zapisywane dane wejściowe, jak i dane pośrednie,
- programów wykonujących zadania (tu: rozpoznające emocje), w tym wzorcu nazywane są **ekspertami**, oraz
- **kontroler**, sterujący przepływem działania aplikacji.

Wzorzec Scoreboard rozszerza wzorzec Blackboard o 3 dodatkowe klasy:

- **tabeli ocen**, na której zapisywane są oceny poszczególnych *ekspertów* oraz liczniki dotyczące działania aplikacji,
- **moderatora**, który jest podklasą *eksperta*, oceniającą pozostałych *ekspertów* i zapisujący wyniki w *tabeli ocen*; odpowiada on także za realizację funkcji integracji hipotez o stanie emocjonalnym – pełni rolę końcowego klasyfikatora (późna fuzja),
- **arbitra** (ang. evaluator), który jest modulem odpowiedzialnym za wybór *ekspertów*, którzy mają zostać wywołani z kolejną iteracją na podstawie *tabeli ocen*; jego działanie jest wywoływane przez moduł *kontrolera*.

Diagram przedstawiający klasy wzorca projektowego **Tabela ocen** (Scoreboard) został przedstawiony na Rys 5.2. Na biało widoczne są elementy wzorca **Tablicy** (Blackboard), na szaro elementy rozszerzające ten wzorzec.



Rys 5.2 Diagram klas wzorca projektowego Scoreboard [4]

Wzorec ten jest dedykowany rozwiązaniom niedeterministycznym z potencjalnie dużą przestrzenią rozwiązań wymagających integracji wielu złożonych programów, zwłaszcza gdy programy będą często zmieniane i aktualizowane – wzorec nie wymaga wtedy wielu zmian w kodzie. Jednocześnie ma on wbudowany mechanizm ciągłej oceny działania aplikacji i programów automatycznego rozpoznawania emocji (*ekspertów*), co może być wartościowe, gdy nie jest możliwa wystarczająca weryfikacja w środowisku testowym. Wzorec Scoreboard został utworzony z myślą o automatycznym przetwarzaniu emocji, ale może zostać użyty również dla innych zastosowań.

Dla zastosowania automatycznego rozpoznawania emocji klasy reprezentują odpowiednio:

- **Ekspert** – program automatycznego rozpoznawania emocji (klasyfikator),
- **Tablica** – współdzielone zasób, na którym zapisywane są dane z poszczególnych kanałów obserwacji, a także wektory cech i wyniki klasyfikacji,
- **Moderator** – klasa implementująca funkcję integracji późnej fuzji oraz funkcję oceny, która może zostać wykorzystana do określenia poziomu zaufania dla programów automatycznego rozpoznawania emocji (*ekspertów*),
- **Kontroler** – klasa odpowiedzialna za przepływ sterowania programem,
- **Arbiter** – klasa odpowiedzialna za wybór programów automatycznego rozpoznawania emocji (*ekspertów*) do wywołania,
- **Tabela ocen** – zasób, na którym przechowywane są oceny programów automatycznego rozpoznawania emocji (*ekspertów*).

Wzorec jest elastyczny w kwestii tworzenia przebiegu działania programu, który jest obsługiwany przez klasę *kontrolera*. Działanie programu jest podzielone na fazy definiowane przez twórcę programu, a *eksperci* mogą brać udział w każdej lub tylko w wybranych fazach. Przykładowo działanie aplikacji można podzielić na fazę inicjalną, w której w klasie *tablicy* są zapisywane współdzielone dane z czujników; fazę wstępnych obliczeń, gdzie *eksperci* przetwarzają surowe dane w wektory cech, a następnie wprowadzają go na *tablicę*; fazę przetwarzania emocji; oraz fazę moderacji, w której następuje uruchomienie funkcji integracji wyników (przez klasę *moderatora*). Faza moderacji (oceny) może być wywołana zarówno jako osobna faza, jak i etap końcowy każdej z faz.

Typowy przebieg faz wygląda następująco:

1. *Kontroler* wykonuje żądanie do *Arbitra* o zwrócenie listy *ekspertów* do wywołania;
2. *Arbiter* wybiera określoną liczbę *ekspertów* gotowych do działania w danej fazie, zgodnie z zadaną konfiguracją;
3. *Kontroler* po otrzymaniu listy *ekspertów* wywołuje ich w losowej kolejności;
4. Obiekty *ekspertów* czytają dane zapisane w klasie *tablicy* i wpisują swoje wyniki z powrotem na klasę *tablicy*;
5. Po działaniach *ekspertów* może być wywołany *moderator* do integracji wyników.

Faza analizy może być ograniczona liczbą iteracji lub trwać aż żaden z *ekspertów* nie będzie miał „nic do dodania”. Przebieg sterowania może zostać dostosowany przez programistę implementującego wzorzec wedle potrzeb. Fragment kodu klasy *kontrolera* został przedstawiony na Listingu 5.1. Opisuje on metodę odpowiedzialną za pętlę sterowania. Klasa *Kontrolera* iteruje po wszystkich fazach programu (zdefiniowanych w pliku konfiguracyjnym). W każdej fazie odpytuje klasę *arbitra* o listę *ekspertów* (funkcja: *GetActiveExperts*), a następnie ich uruchamia. Po wywołaniu wszystkich *ekspertów* – następuje zmiana fazy programu.

```
override public void run()
{
    bool nothingAdded; //logical field (flag for stage stop criteria)

    foreach (String stage in stages) //iterate stages
    {
        do
        {
            nothingAdded = true; //setting initial stage flag value

            /*iterate list of Experts provided by Arbiter class instance (limited no), if required
            number of Experts is less than available, Arbiter chooses Experts based on algorithm
            informed by Scoreboard entries*/
            foreach (var expert in arbiter.GetActiveExperts
                (stage,MaxExpertsNumber,MinExpertsNumber))
            {
                /*call expert - expert reads Blackboard and might add entry to Blackboard (then
                stage flag is set to false) or report that he has nothing to add*/

                tryLaunchExpert(expert, ref nothingAdded);
            }

            //if any Expert added something change turn and repeat stage
            blackboard.ChangeTurn();

            //if there is a Moderator class instance for this stage
            if (moderator.processesPhase(stage))
                moderator.DoWork(); //launch Moderator (subclass of Expert)
        }
        /*repeat until turn resulted in any new Blackboard entry or limit of turns was reached*/
        while (!nothingAdded && blackboard.GetCurrentTurn() < MaxRoundInPhaseNumber);

        blackboard.ChangePhase(); //change stage
    }
}
```

Listing 5.1 Przykładowy fragment kodu klasy kontrolera we wzorcu Scoreboard zaimplementowanego w języku C#

Moderator posiada dwie funkcje: ustala uzgodnione rozwiązanie poprzez wywołanie funkcji integracji na podstawie hipotez zapisanych w klasie *tablicy*, określając przy tym poziom zaufania do uzgodnionego rozwiązania, a także wywołuje funkcję oceny *ekspertów*. Funkcja ta wyznacza ocenę na podstawie odległości ich hipotez od uzgodnionego rozwiązania, a następnie zapisuje ją na *tabeli ocen*. Funkcja oceny może zostać oczywiście nadpisana wedle potrzeb programisty.

Klasa *Arbitra* wybiera *ekspertów* na podstawie zadanych kryteriów, takich jak minimalna i maksymalna liczba *ekspertów*, faza, dla których ma zostać wywołany, a także informacji czy wywołanie jest obligatoryjne, czy opcjonalne. Dla przypadków opcjonalnych wywołanie *ekspertów* następuje w sposób niedeterministyczny z prawdopodobieństwem odpowiadającym ocenie *eksperta* na *tabeli ocen*. Pozwala to na eliminację słabszych *ekspertów* oraz tych, dla których warunki w danym badaniu nie są optymalne. Oczywiście, programista implementujący klasę *arbitra* może dostosować algorytm do swoich potrzeb. Główną zasadą jednak jest, to aby każdy, nawet „najsłabszy” *ekspert*, miał szansę na wywołanie. Fragment interfejsu *arbitra* został przedstawiony na Listingu 5.2. Przedstawia on metody zwracające listy *ekspertów* do wywołania, które muszą zostać zaimplementowane (są wywoływane przez klasę *kontrolera*).

```
public interface IArbiter
{
    /*returns all experts*/
    List<IExpert> GetActiveExperts();

    /*returns all experts, taking into account current phase */
    List<IExpert> GetActiveExperts(String type);

    /*returns all experts, taking into account current phase and in a given range */
    List<IExpert> GetActiveExperts(String type, int max=int.MaxValue, int min = 0);
}
```

Listing 5.2 Definicja interfejsu IArbiter w języku C#

Wpisy w instancji *tablicy* z wzorca **Blackboard** zawierają takie informacje jak: wartość hipotezy; nazwa *eksperta*, który ją wpisał; typ danych, czyli poziom abstrakcji np. dane wejściowe, hipoteza, element pośredni oraz znacznik czasowy. Fragment kodu interfejsu *tablicy* został przedstawiony na Listingu 5.3. Wzorzec **Scoreboard** rozszerza te wpisy o informację na temat poziomu pewności hipotezy, który powinien bazować na jakości danych. Z kolei w klasie *Tabeli ocen* zebrane są informacje o ocenie działania *ekspertów*, która może być określona jako odległość od uzgodnionego rozwiązania. Może również zawierać inne statystyki takie jak liczba wywołań lub okresy niedostępności *eksperta*, które mogą posłużyć do bardziej zaawansowanych funkcji *arbitra*.

```

public interface IBlackboard
{
    //adding a new entry (of any type T) to Blackboard
    void AddBlackboardEntry(T entry);

    //reading all entries (of any type) currently at Blackboard
    List<T> GetEntries();

    //reading all entries of specific type T
    List<BlackboardEntry>GetEntries<T>(IBlackboardEntryType<T> type);

    //reading all entries added by specific Expert
    List<BlackboardEntry> GetEntries<T>(string moduleName, IBlackboardEntryType<T> type);

    //getting turn count (processing is divided into stages and turns)
    int GetCurrentTurn();

    //getting stage count (processing has stages and turns)
    int GetCurrentPhase();

    //Blackboard clearing (removes all entries)
    void ClearEntry();

    //removing single Blackboard entry
    void RemoveEntry(BlackboardEntry entry);

    void ChangeTurn();           //increment turn count
    void ChangePhase();         //increment stage count

    PhaseTurnNumber GetPhaseAndTurn(); //getting processing status (stage and turn)
}

```

Listing 5.3 Definicja interfejsu IBlackboard w języku C#

Dzięki zastosowaniu wstrzykiwania zależności (ang. dependency injection) możliwe jest przygotowanie programów automatycznego przetwarzania emocji (*ekspertów*) bez zmiany kodu sterowania programem. Klasa *eksperta* musi realizować interfejs *IExpert*, wraz z wymaganymi metodami: *DoExpertWork* – odpowiedzialnej za wykonanie oszacowania stanu emocjonalnego; *GetExpertWorkCondition* – odpowiedzialnej za zwrócenie informacji, czy *ekspert* ma coś do dodania; *ProcessingPhase* – zwracającą informację, czy *ekspert* wykonuje się w tej fazie; a także *GetExpertName* – zwracającą nazwę *eksperta*. Kod interfejsu został zaprezentowany na Listingu 5.4.

```

public interface IExpert : IComparable<IExpert>
{
    /*reading data from Blackboard, processing (expert's work), adding a new entry*/
    void DoExpertWork();
    //checking the criteria for the expert to be called
    IWorkCondition GetExpertWorkCondition();
    //getting information about processing stage
    bool ProcessingPhase(string phase);
    string GetExpertName(); //getting Expert name
}

```

Listing 5.4 Definicja interfejsu IExpert w języku C#

Taki mechanizm pozwala na wywołanie programu na różnych kombinacjach *ekspertów* bez potrzeby ponownej kompilacji programu, a jedynie poprzez zmiany w zewnętrznych plikach konfiguracyjnych, które zawierają listę ekspertów (programów automatycznego rozpoznawania emocji) do wywołania.

W przypadku zewnętrznych programów lub innego języka programowania niż C# możliwe jest wykorzystanie klasy opakowującej (ang. wrapper), która dostarczy dane z instancji *tablicy* do zewnętrznego programu i przekaże od niego wynik na instancję *tablicy*. Biblioteka wzorca została zaimplementowana w języku C#, a jej kod źródłowy wraz z dokumentacją, instrukcją użytkownika i opisem parametrów konfiguracyjnych jest dostępny pod adresem <https://bitbucket.org/pgscoreboard/scoreboard/wiki/GetStarted>. Jest on udostępniony na licencji BSD (OpenSource). Na Listingach 5.1 – 5.4 znajdują się fragmenty kodu biblioteki Scoreboard.

Istotne cechy wzorca z punktu widzenia integracji danych dla wielomodalnego rozpoznawania emocji:

- możliwość wykorzystania programów w zaimplementowanych w różnych technologiach (wykorzystanie jako czarnej skrzynki), wymaga jedynie klasy opakowującej (ang. wrapper) implementującej interfejs *eksperta (IExpert)*,
- możliwość modyfikacji *ekspertów* bez zmiany całego rozwiązania dzięki separacji – brak wzajemnych zależności i powiązań pomiędzy modułami (programami automatycznego rozpoznawania emocji, funkcji integracji itp.),
- realizacja paradygmatów wstrzykiwania zależności (ang. dependency injection) oraz odwrócenia sterowania (ang. Inversion of Control), w których przebieg działania aplikacji jest odseparowany od funkcjonalności. Ich realizacja pozwala na łatwiejszą rekonfigurację reguł wywołania *ekspertów* bez przebudowy całej aplikacji, podczas gdy kontrolę nad sterowaniem sprawuje klasa *Kontrolera* wraz z klasą *Arbitra*, niezależnie od wybranych programów automatycznego rozpoznawania emocji,
- posiada wbudowaną dynamiczną ocenę programów automatycznego rozpoznawania emocji (*ekspertów*) w trakcie działania aplikacji, pozwalającą na ich dynamiczną selekcję,
- wsparcie dla oceny oraz wyznaczania współczynnika zaufania,
- wszystkie implementacje dla wzorca **Blackboard** mogą zostać wykorzystane we wzorcu **Scoreboard**,
- możliwość wykorzystania wczesnej, późnej oraz hybrydowej fuzji,
- odporność na brak danych z jednego kanału, czasową niedostępność oraz awarię jednego z programów automatycznego rozpoznawania emocji,
- pozwala na łatwe zapisywanie przebiegu pracy, wyniku oraz ocen,
- pozwala na określenie maksymalnej liczby wykorzystywanych modułów, przy dostępnej dużej ich liczbie, bez arbitralnego wykluczania jakiegokolwiek z programów – wybór *ekspertów* w każdej iteracji.

Ograniczenia wzorca:

- trudność testowania, zwłaszcza gdy wykorzystanych jest wielu *ekspertów* (ograniczenie odziedziczone ze wzorca **Blackboard**), typowe dla rozwiązań wykorzystujących wstrzykiwanie zależności
- nie ma gwarancji uzyskania satysfakcjonującego rozwiązania – cecha charakterystyczna dla rozwiązań niedeterministycznych,
- wykorzystanie wzorca może spowodować pogorszenie wydajności, zwłaszcza gdy kryteria ograniczające liczbę wywoływanych *ekspertów* nie zostaną zastosowane,
- podstawowa implementacja zakłada działanie sekwencyjne.

Szczegółowy opis wzorca został przedstawiony w publikacji [4]. Znajduje się w niej m.in. szczegółowa analiza wzorca pod kątem pięciu kryteriów jakości: modularyzacji (ang. modularization), integracji (ang. integration), adaptowalności (ang. adaptability), odporności (ang. robustness) oraz wydajności (ang. efficiency).

5.3. **Przykładowa implementacja rozwiązania – Monitor Emocji**

Wzorzec Scorebard został zaimplementowany, a także jest wykorzystywany w czasie badań przeprowadzanych na Politechnice Gdańskiej, na wydziale Elektroniki, Telekomunikacji i Informatyki na stanowisku w laboratorium EA01. Stanowisko to nosi nazwę Monitora Emocji [4] [133]. Pierwsza implementacja wzorca powstała w ramach pracy magisterskiej autora niniejszej pracy [134].

Stanowisko badawcze pozwala na wielokanałową i wielomodalną obserwację użytkownika komputera. Jest ono w pełni konfigurowalne. Składa się z trzech komputerów, specjalistycznego oświetlenia, kilku kamer oraz zestawu czujników fizjologicznych. Wykorzystywane oprogramowanie:

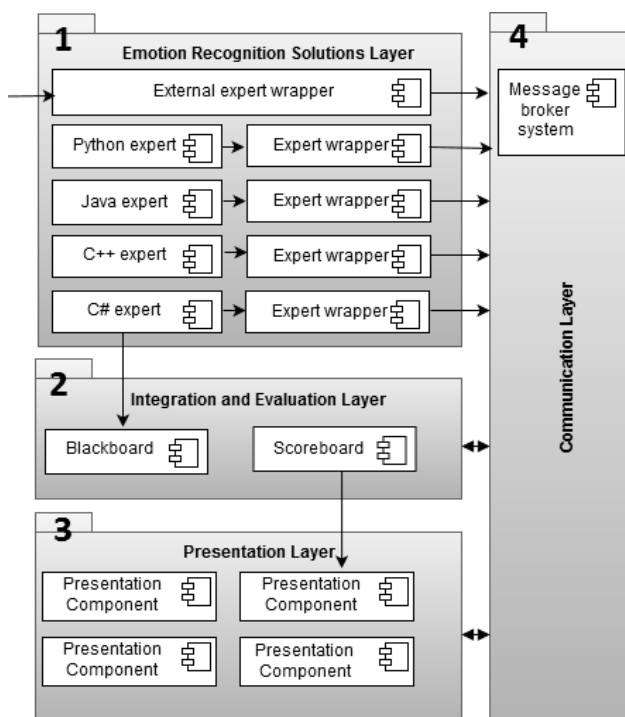
- Morae Recorder i Morae Observer – oprogramowanie do obserwacji reakcji użytkownika oraz zapisywania stanów czujników fizjologicznych,
- aplikacja do synchronizacji czasowej kamer (równoczesne włączenie, wyłączenie oraz wspólna linia czasowa),
- aplikacja do monitorowania ruchu myszą oraz dynamiki naciskania klawiszy klawiatury,
- Noldus Face Reader – program automatycznego rozpoznawania emocji na podstawie modalności mimiki twarzy.

Obserwator na stanowisku Monitora Emocji jest oddzielony od badanego ścianą tak, iż badany go nie widzi. Monitorowanie odbywa się za pomocą dwóch komputerów. Konfiguracja jest możliwa do modyfikacji w zależności od potrzeb badawczych.

Techniczne wyzwania związane ze stanowiskiem:

- dostępne oprogramowanie jest zmienne ze względu na aktualizacje, licencje ograniczone czasowe oraz nabywanie nowych rozwiązań,
- rozwiązania znacznie się różnią pod względem technologii, dostępnego API, a także wydajności,
- rozwiązania różnią się zastosowanymi modelami reprezentacji emocji,
- rozwiązania różnią się dostępnością dostarczanych hipotez, zwłaszcza przy czasowej niedostępności kanału.

W odpowiedzi na te wyzwania powstała implementacja Monitora Emocji wykorzystującego wzorzec Scoreboard [3]. Architektura Monitora emocji została podzielona na 4 warstwy: (1) warstwę programów automatycznego rozpoznawania emocji, (2) warstwę integracji i oceny, (3) warstwę prezentacji oraz (4) warstwę komunikacji – wymaganą ze względu na wykorzystanie różnych języków programowania. Graficzna prezentacja podziału na warstwy została zaprezentowana na diagramie komponentowym Rys 5.3 [3].



Rys 5.3 Diagram komponentowy monitora emocji [3]

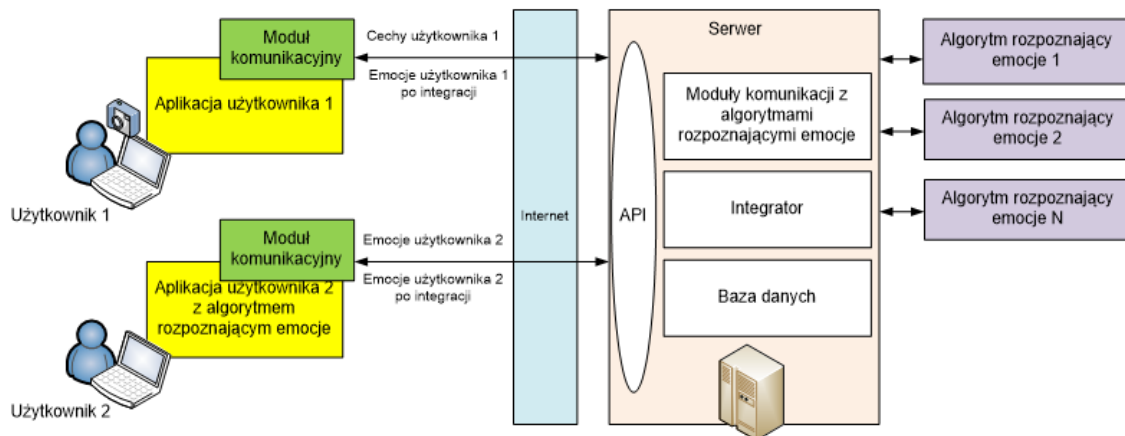
Warstwa rozpoznawania pozwala na implementację ekspertów bezpośrednio jako klasy w języku C# lub wykorzystanie klas opakowujących dla innych języków programowania oraz dla zamkniętego kodu, które trzeba wywołać z linii poleceń lub poprzez API.

Warstwa integracji wykorzystuje bezpośrednio mechanizmy wzorca Scoreboard (opisanego w 5.2), w szczególności funkcje integracji (opisaną w 5.4) oraz funkcję oceny (opisaną w 5.5).

Wydzielenie warstwy prezentacji pozwala na niezależną wizualizację w zewnętrznym programie lub programach, które nie muszą być zintegrowane bezpośrednio z samym Monitorem Emocji.

Ostatnią warstwą jest warstwa komunikacji, do której realizacji został wykorzystany system kolejkowy RabbitMQ. Komunikacja poprzez system kolejkowy pozwala na integrację rozwiązań z różnych technologii, a także separację, poszczególnych warstw, takich jak warstwa prezentacji.

Monitor emocji stanowi lokalną implementację wzorca. Istnieje także implementacja oparta o Web Serwisy, której architektura została przedstawiona na Rys 5.4 [5]. Rozwiązanie to pozwala na wykorzystanie Monitora Emocji równoległe przez kilku użytkowników. Wyniki analizy są zapisywane bezpośrednio do centralnej bazy danych. Wynik integracji jest zwracany do aplikacji (warstwa wizualizacji). Wewnętrzna architektura rozwiązania jest tożsama z lokalną, dlatego też nie została tutaj powtórzona.



Rys 5.4 Architektura Monitora Emocji zaimplementowanego z pomocą Web Serwisów [5]

5.4. Algorytm integracji bazujący na współczynnikach wiarygodności

Drugim komponentem zaproponowanej metody integracji, oprócz wzorca architektonicznego Scoarboard jest algorytm integracji bazujący na współczynnikach wiarygodności.

Podjęcie wielomodalne sprawdziło się już w wielu dziedzinach, posiada też znane zastosowania w rozpoznawaniu emocji [135]. Zaproponowany algorytm o nazwie CFIEE (ang. Confidence-based Fuzzy Integration of Emotion Estimates) został zaprojektowany z myślą o rozwiązaniach dotyczących rozpoznawania emocji, choć obszar jego zastosowań może być szerszy. Służy on do integracji danych z dwóch i więcej źródeł danych, które mogą być programami automatycznego rozpoznawania emocji lub jednym programem uruchomionym wielokrotnie np. programem wykorzystującym modalność mimiki twarzy, gdzie każda instancja bazuje na niezależnej kamerze (osobnym kanale).

Zaproponowany algorytm zakłada, że każdy *ekspert* (tu: program automatycznego rozpoznawania emocji) zwraca oprócz hipotezy o stanie emocjonalnym także współczynnik wiarygodności tej hipotezy. Współczynnik ten powinien być podawany wraz z hipotezą *eksperta* w odpowiedzi na każdą próbkę danych. Problem określenia współczynnika niepewności został szerzej opisany w rozdziale 4.3.

Jeżeli *ekspert* sam nie wyznacza współczynnika wiarygodności, może on zostać wyznaczony na podstawie miar jakości zwróconych przez program, takich jak np. kąt ustawienia płaszczyzny twarzy w stosunku do kamery czy jakość sygnału dźwiękowego. Jeżeli program sam nie zwraca takich miar, mogą być one wyznaczone przez program integrujący dane lub poprzez klasę opakowującą (ang. wrapper) – stosownie do wykorzystanych modalności. W przypadku, gdy *ekspert* nie dostarcza żadnych informacji poza stanem emocjonalnym i nie jest możliwe dynamiczne wyznaczenie współczynnika wiarygodności, można przyjąć jako współczynnik obliczoną statycznie skuteczność programu (ang. accuracy) na testowej bazie danych. Przyjęcie takiej statycznej miary skuteczności nie zapewnia jednak dynamicznego dopasowania do zmieniających się warunków pozyskiwania sygnałów wejściowych, ale jest korzystniejsze niż brak informacji o wiarygodności.

Algorytm CFIEE wyznacza uzgodnione rozwiązanie na podstawie hipotez *ekspertów*, które muszą zostać sprowadzone do wspólnego modelu reprezentacji emocji (jeżeli wykorzystują różne modele) za pomocą odwzorowań, które zostały opisane w rozdziale 2.3. Do reprezentacji uzgodnionego rozwiązania można zastosować różne modele reprezentacji emocji. Każdy model ciągły, który da się zaprezentować w postaci współczynników w przedziale $\langle -1, 1 \rangle$ lub $\langle 0, 1 \rangle$, a jego wymiary mogą być analizowane niezależnie, może zostać wykorzystany.

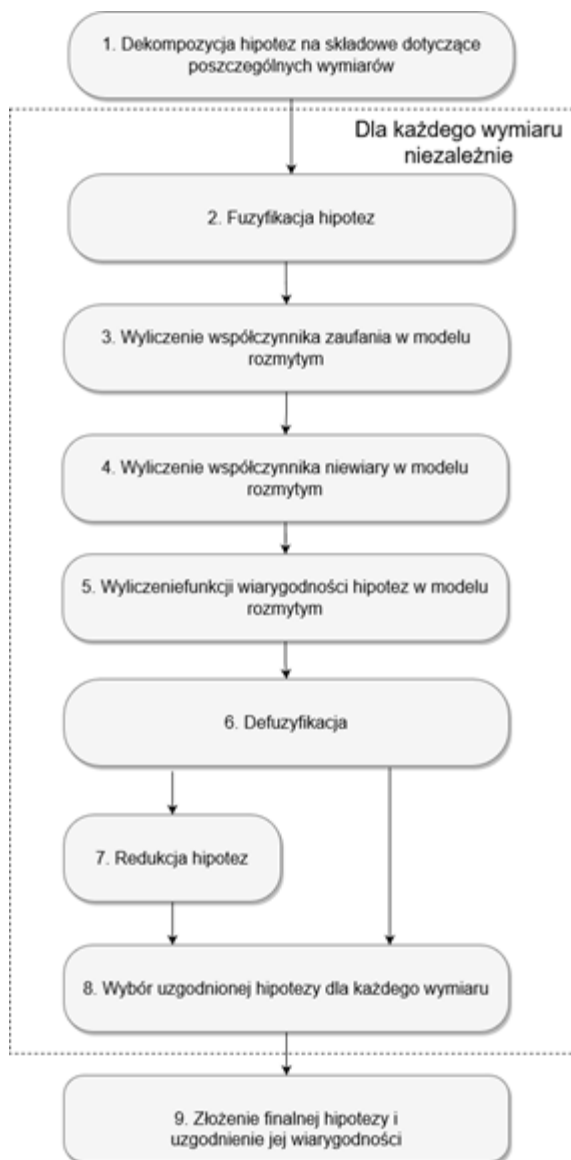
Algorytm CFIEE analizuje każdy z wymiarów niezależnie, co jest istotne dla programów wykorzystujących niekompletne modele np. jedynie wymiar wartościowości z modelu 2D, 3D lub 4D, gdzie wskazanie wartości zerowej dla brakujących wymiarów byłoby błędnym założeniem.

Algorytm CFIEE przyjmuje indywidualny współczynnik wiarygodności dla każdego wymiaru. Takie rozwiązanie zostało przyjęte, ponieważ kanały obserwacji różnią się w zakresie skuteczności rozpoznawania poszczególnych wymiarów. Dla przykładu dane fizjologiczne dostarczają więcej informacji o pobudzeniu (ang. arousal) człowieka niż o wartościowości (ang. valence). Jest to związane z pracą układu nerwowego, który ma za zadanie stymulować pracę organizmu w sytuacjach wymagających działania niezależnie od jego charakteru.

Kolejne ze zidentyfikowanych wyzwań dotyczy sytuacji, w której dla tej samej osoby, miejsca i czasu, programy automatycznego rozpoznawania emocji podają rozbieżne wyniki oszacowania stanu afektywnego [2], [63], co zostało szerzej opisane w rozdziale 4.4. W algorytmie CFIEE zostało zaproponowane rozwiązanie dla tego wyzwania poprzez wyróżnienie dwóch kryteriów: zbieżności hipotez oraz wiarygodności hipotez, zaczerpniętych z istniejących rozwiązań opisanych w rozdziale 3.7.



Algorytm integracji CFIEE składa się z 9 kroków, które pokazano na rysunku Rys 5.5. W pierwszym kroku hipotezy zostają zdekomponowane na poszczególne wymiary, a następnie w krokach od drugiego do ósmego są wykonywane dla każdego wymiaru osobno, by ostatecznie w kroku 9 zostać złączone w uzgodnione rozwiązanie.



Rys 5.5 Schemat blokowy algorytmu CFIEE [Opracowanie własne]

W drugim kroku algorytmu następuje rozmycie hipotez, które zostało zaczerpnięte z metod opisanych w rozdziale 3.7. Rozmycie hipotez zostało wykorzystane z powodu natury rozpoznawania stanów emocjonalnych oraz niemożności określania „prawdziwego” stanu emocjonalnego (ang. ground truth), co zostało szerzej omówione w rozdziale 2.6. Hipoteza odległa o małą wartość od wskazanej hipotezy (np. 0,01) jest tylko niewiele mniej prawdopodobna dlatego, aby umożliwić interpolację wyników na pobliskie wartości, zastosowano metody rozmyte.

W krokach trzecim, czwartym oraz piątym wykorzystano zmodyfikowaną teorię ewidencji Dempstera–Shafera [136], której wykorzystanie, jak i zarys koncepcji został zaproponowany w raporcie technicznym [137]. Wykorzystano z niej funkcję:

- zaufania (ang. belief) [krok 3 zaproponowanego algorytmu] – rozumianej jako poziom wsparcia hipotezy przez obserwację – może do tego posłużyć np. funkcja prawdopodobieństwa,
- niewiary (ang. disbelief) [krok 4 zaproponowanego algorytmu] – rozumianej jako brak zaufania do hipotezy przez obserwację – przekonanie o fałszywości hipotezy.

Rozbicie to jest spowodowane sytuacją braku kompletnych informacji. Jeżeli w obserwacji brakuje dostatecznej informacji do poparcia, jak również odrzucenia hipotezy – obie wartości mogą być niskie lub równe zero. Złączenie tych informacji – odjęcie wartości funkcji niewiary od wartości zaufania zostało wykonane jako 5 krok metody (funkcja wiarygodności).

Bezpośrednie zastosowanie Dempstera–Shafera teorii nie jest możliwe, ponieważ:

- teoria Dempstera–Shafera dotyczy zmiennych o wartościach dyskretnych, które się wzajemnie wykluczają (np. ktoś jest chory – nie jest chory), a w tym przypadku mamy do czynienia z wartościami ciągłymi, które wykluczają się dopiero po przekroczeniu pewnych wartości granicznych,
- w teorii Dempstera–Shafera preferowana jest przede wszystkim zgodność dowodów, a nie ich wiarygodność (jeżeli są dwa dowody wspierające o wiarygodności 0,01, przyjęta hipoteza będzie miała 100% wiarygodności mimo niskiej wiarygodności dowodów).

Ogólny schemat działania algorytmu:

Poniżej został przedstawiony ogólny schemat działania każdego z kroków algorytmu. Założenia te nie powinny być zmieniane ze względu na poprawność działania rozwiązania. W dalszej części pracy został opisany szczegółowy schemat zawierający przyjęte przez autora pracy założenia, które mogą być modyfikowane przez osoby implementujące metodę.

1. Dekompozycja hipotez na składowe dotyczące poszczególnych wymiarów

W tym kroku, hipotezy muszą zostać rozbite na pojedyncze składowe dotyczące każdego z wymiarów.

2. Fuzyfikacja hipotez (ang. Fuzzyfication)

Dla każdego wymiaru hipotezy stosowane jest rozmycie w postaci rozkładu statystycznego w taki sposób, aby maksimum funkcji znajdowało się w wartości hipotezy i wynosiło tyle, ile wiarygodność hipotezy. Rozkład ten powinien być symetryczny, tak aby dodatnie, jak i ujemne odchylenia były tak samo prawdopodobne. Im punkt jest mocniej oddalony od hipotezy – tym wartość wiarygodności maleje. Krok ten jest wykonywany dla każdej hipotezy.

3. Funkcja zaufania w modelu rozmytym (ang. belief function)

W trzecim kroku obliczana jest całkowita funkcja zaufania $b(x)$, która określa uzgodniony poziom zaufania dla każdej wartości x z przedziału $\langle -1, 1 \rangle$ lub $\langle 0, 1 \rangle$ (w zależności od przyjętego modelu reprezentacji emocji) – dla wszystkich hipotez łącznie. Wartość funkcji obliczana w punkcie x oznacza poziom zaufania w punkcie x . Współczynnik zaufania powinien spełniać kilka kryteriów:

- przyjmować wartości z zakresu $(0, 1)$,
- jeżeli dla danej wartości x nie ma poparcia żadnej z hipotez (częstkowe funkcje zaufania, wyznaczone w kroku nr 2, wynoszą 0), współczynnik zaufania powinien wynosić 0,
- jeżeli dla danej wartości x jest poparcie choćby jednej hipotezy (wszystkie poza jedną częstkowe funkcje zaufania wynoszą 0), wówczas powinna być przyjęta wartość niezerowej cząstkowej funkcji dla tego punktu,
- jeżeli dla danej wartości x więcej niż jedna cząstkowa funkcja zaufania przyjmuje wartość niezerową, całkowite zaufanie powinno być zwiększone proporcjonalnie do cząstkowych funkcji zaufania i przekraczać wartość większej z nich (maksymalnie 1).

4. Obliczanie współczynnika niewiary w modelu rozmytym (ang. disbelief function)

W czwartym kroku obliczana jest całkowita funkcja niewiary, która reprezentuje odległość i wiarygodność hipotez niezgodnych z daną wartością x – dla wszystkich hipotez łącznie.

Hipotezy nadają badanemu punktowi niewiarę – im punkt bardziej oddalony od hipotezy, tym większy poziom niewiary. Funkcja niewiary ma charakter ciągły i może przyjmować wartości większe niż 1. Wartości z funkcją niewiary większą niż 1 nie mogą zostać uznane za wiarygodny wynik.

5. Wyliczenie funkcji wiarygodności hipotez w modelu rozmytym (ang. plausibility function)

W piątym kroku obliczana jest funkcja wiarygodności, która jest różnicą wyznaczonej funkcji zaufania i niewiary dla każdej wartości x , przy czym wartości poniżej zera, przyjmują wartość 0.

6. Defuzyfikacja hipotez

W kroku szóstym następuje defuzyfikacja hipotez poprzez znalezienie lokalnych maksimów funkcji wiarygodności. Wynikiem jest zbiór hipotez alternatywnych M o liczności m , które mogą, ale nie muszą pokrywać się ze zbiorem hipotez początkowych N .

7. Redukcja liczby hipotez (opcjonalny)

W kroku siódmym można przeprowadzić dodatkową redukcję liczby hipotez na podstawie ich zbieżności.



8. Wybór uzgodnionej hipotezy dla każdego wymiaru

W przedostatnim kroku wybierana jest uzgodniona hipoteza na podstawie hipotez końcowych otrzymanych po defuzyfikacji i poddanych ewentualnej redukcji.

9. Złożenie finalnej hipotezy i uzgodnienie jej wiarygodności

Kroki 2–8 wykonywane są niezależnie dla każdego z wymiarów. W ostatnim kroku wyniki cząstkowe z kroku nr 8 są składane w jedną odpowiedź – uzgodnione rozwiązanie. Współczynnik wiarygodności powinien zostać podany dla każdego wymiaru osobno – zgodnie z otrzymanymi wartościami z kroku 5. Jeżeli przyjęty model reprezentacji emocji wymaga pojedynczego współczynnika wiarygodności, należy wyliczyć średnią arytmetyczną, jednakże nie jest to zalecane ze względu na charakterystykę modeli. Przykładowo w modelu 3D, oczekuje się, że wartości dla wymiaru dominacji będzie mniejsza, ponieważ jest to najmniej zbadany wymiar i może być pominięty w programach automatycznego rozpoznawania emocji – dla programów wykorzystujących model 2D.

Szczegóły zastosowanego rozwiązania

Poniżej przedstawione zostały szczegóły zastosowania algorytmu integracji wraz z przykładowymi wartościami. Elementy te mogą zostać dostosowane wedle potrzeb do konkretnego zastosowania.

1. Dekompozycja hipotez na składowe dotyczące poszczególnych wymiarów

Wartości należy poddać normalizacji do wspólnej skali $\langle -1,1 \rangle$ lub $\langle 0,1 \rangle$ (w zależności od modelu) dla danych oraz $\langle 0,1 \rangle$ dla współczynnika wiarygodności.

Przykład

Docelowym modelem jest model PAD (3D), składowe zostały zaprezentowane w trzech różnych modelach, co zostało zestawione w Tabeli 5.1. W kolejnych kolumnach przedstawione są: numer porządkowy hipotezy, źródłowy model reprezentacji i jego skala, wartość w modelu źródłowym – hipotezy i jej wiarygodności oraz wartości po normalizacji – wartości i jej wiarygodności.

Tabela 5.1 Przykładowe wartości dla normalizacji danych z różnych modeli reprezentacji emocji [opracowanie własne].

Lp	Model i skala	Bazowe		Po normalizacji	
		Wartość	Wiarygodność	Wartość	Wiarygodność
1	Model: PAD Skala (-5,5) Skala wiarygodności: (0,1)	(2; 2,5;1)	(0,8;0,7;0,6)	(0,4;0,5;0,2)	(0,8;0,7;0,6)
2	Model: 2D Skala (-1,1) Skala wiarygodności: (0,1)	(0,5; -0,5)	(0,7)	(0,5; -0,5;-)	(0,7;0,7;-),
3	Model: 2D Skala (-2,2), Skala wiarygodności: (0,1)	(0,8; -0,6)	(0,6;0,8)	(0,4; -0,3;-)	(0,6;0,8;-),

Dla wymiaru D mamy jedynie jedną wartość, także nie uruchamiamy dalszych kroków integracji dla tego wymiaru. Dalszy kroki wykonujemy jedynie dla wymiarów P oraz A.

2. Fuzyfikacja hipotez (ang. fuzzyfication)

W niniejszej pracy zaproponowane zostało wykorzystanie rozmycia hipotez z wykorzystaniem rozkładu normalnego o średniej odpowiadającej jej wartości i wariancji dobranej tak, aby maksimum funkcji znajdowało się w wartości hipotezy i wynosiło tyle, ile wiarygodność hipotezy cząstkowej podanej przez źródłowy program automatycznego rozpoznawania emocji.

W rezultacie otrzymywany jest zbiór cząstkowy funkcji zaufania (ang. belief) w postaci rozkładów normalnych o następujących charakterystykach: wartość średnia (5.1), wariancja (5.2), rozkład normalny (5.3).

$$\mu_i = h_i \quad (5.1)$$

$$\sigma_i = \frac{1}{\sqrt{2\pi * \omega(h_i)}} * c \quad (5.2)$$

$$b_i(x) = \frac{1}{\sqrt{2\pi * \sigma_i}} * \exp\left(-\frac{(x-\mu_i)^2}{2 * \sigma_i}\right) * d \quad (5.3)$$

Rozmycie hipotez zgodnie z rozkładem normalnym znajduje swoje uzasadnienie dzięki następującym założeniom [138]:

- problem dotyczy przypadku ogólnego, nieskalibrowanego do przypadku pojedynczego osobnika, a więc dotyczy całej populacji ludzi – rozkład normalny jest jednym z najczęściej występujących rozkładów w naturze,
- występuje silna tendencja do przyjmowania wartości położonych blisko środka rozkładu – gdyż skrajne, silne emocje występują stosunkowo rzadko;
- dodatnie i ujemne odchylenia od środka rozkładu są jednakowo prawdopodobne – dlatego też rozkład powinien być symetryczny. Każda hipoteza ma swoją precyzję, co oznacza, że jeżeli moduł podała wartość P jako 0,8 to wartość 0,79 oraz 0,81 także można przyjąć za prawdziwe, z niewiele mniejszym prawdopodobieństwem,
- im punkt na wykresie jest bardziej oddalony od hipotezy eksperta, tym mniejsze powinno być prawdopodobieństwo jego prawdziwości,
- algorytm jest bezstanowy i nie bierze pod uwagę poprzednich wartości,
- liczność odchyłek gwałtownie spada wraz ze wzrostem ich wielkości,
- współczynniki c i d, które stanowią mnożniki rozkładu normalnego, zostały wprowadzone, żeby zapewnić odpowiednią stromość rozkładu normalnego dopasowaną do przyjętego przedziału ufności.

Stromość funkcji, a więc dobór parametrów, powinien zależeć od przyjętej wartości progowej odległości, dla której hipoteza uznana jest za zbieżną – w dalszej części pracy wartość ta będzie nazywana współczynnikiem D_{max} (ang. distance treshold). Współczynniki c i d muszą być równe, aby maksimum funkcji znajdowało się w punkcie hipotezy początkowej.

W celu powiązania wartości progowej odległości ze współczynnikami rozkładu normalnego autor pracy zaproponował, aby współczynniki były równe odchyleniu standardowemu dla wiarygodności równej 1, co sprowadza się to do wzoru 5.4. Upraszcza to wzór na odchylenie standardowe do postaci wzoru 5.5

$$c = d = \sqrt{2\pi} * D_{max} \quad (5.4)$$

$$\sigma_i = \frac{D_{max}}{\omega(h_i)} \quad (5.5)$$

Przykład

Zależność zmiany wartości progowej odległości od kształtu krzywej dla wartości współczynnika zaufania równego odpowiednio: 1;0,75;0,5;0,25 oraz wartości progowej D_{max} równej: 0,05; 0,1; 0,2; 0,5. została zilustrowana na Rys 5.6. Skala wykresu $\langle -2,2 \rangle$ została specjalnie dobrana, by lepiej zobrazować działanie funkcji, gdyż docelową skalą jest $\langle -1,1 \rangle$, a więc maksymalna odległość dwóch punktów na tej skali wynosi 2. Im większa wartość progowa odległości tym wykres jest bardziej wygładzony.

Wartość progową odległości należy dobrać do konkretnego zastosowania i do skuteczności programów. W miarę wzrostu skuteczności, wartość progowa może być obniżana.

Innymi potencjalnie odpowiednimi rozkładami do wykorzystania są rozkłady symetryczne, takie jak:

- rozkład Cauchy'ego,
- rozkład logarytmiczno-normalny,
- rozkład Studenta.

Jednakże ich działanie należałoby sprawdzić eksperymentalnie w dużej skali.

3. Funkcja zaufania w modelu rozmytym (ang. belief function)

Wariantem naiwnym tej funkcji jest wyliczenie średniej arytmetycznej, jednakże nie spełnia ona ostatniego kryterium opisanego w ogólnych krokach algorytmu, gdyż wartość wiarygodności nigdy nie będzie większa niż początkowa hipoteza (efekt zbieżności hipotez nie da dodatniego wpływu).

Funkcją zaufania spełniającą warunki postawione powyżej jest funkcja iloczynowa opisana wzorem 5.6

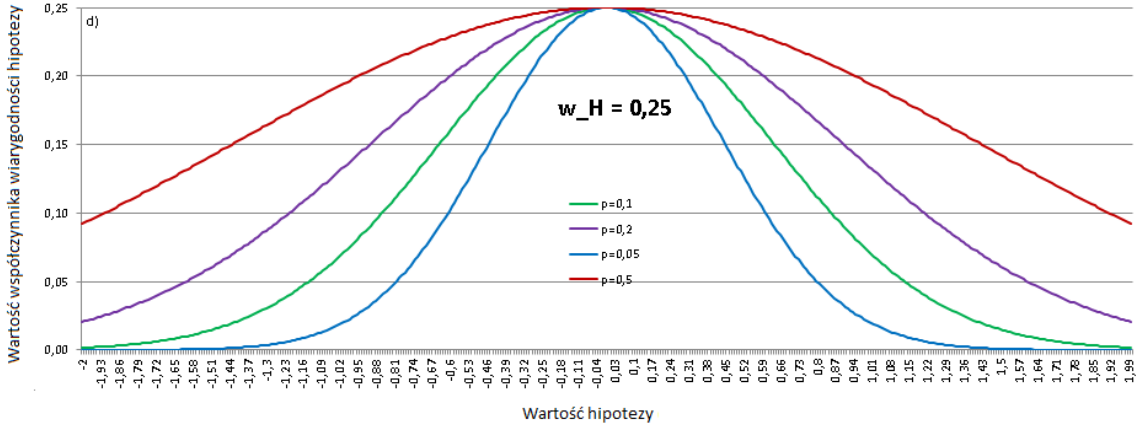
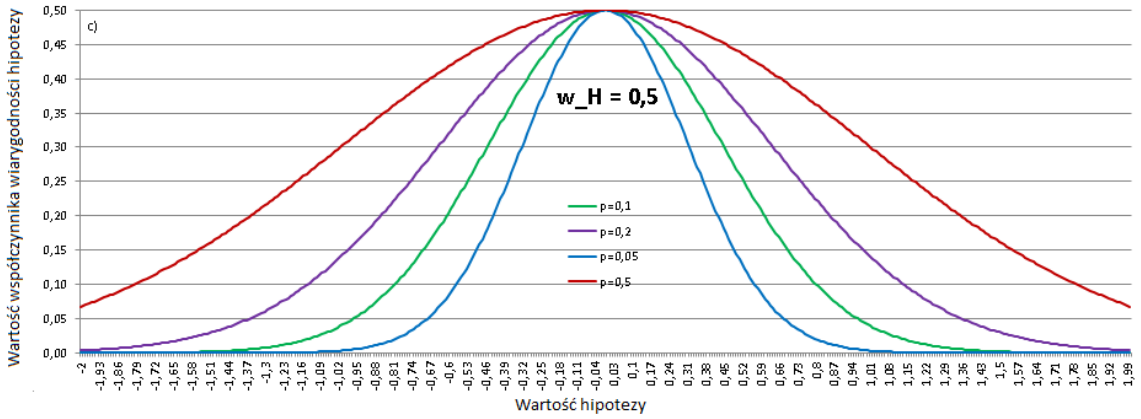
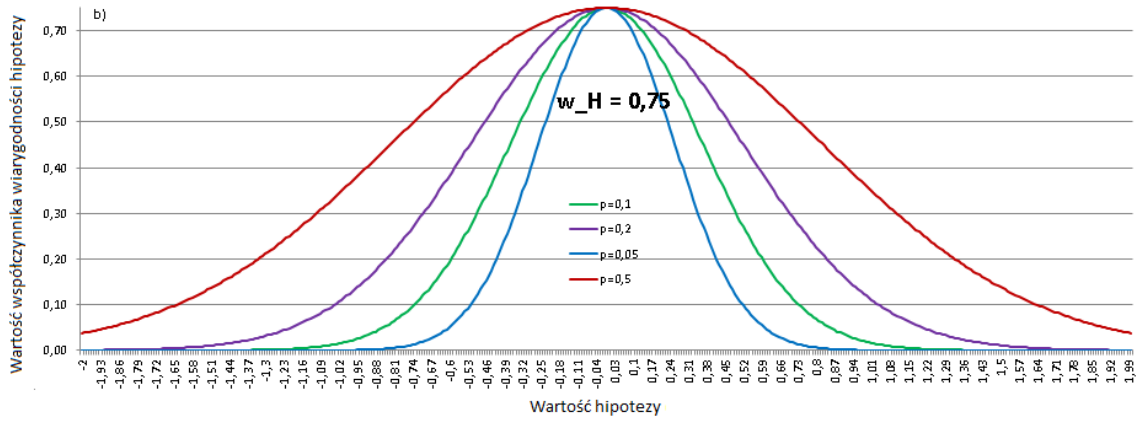
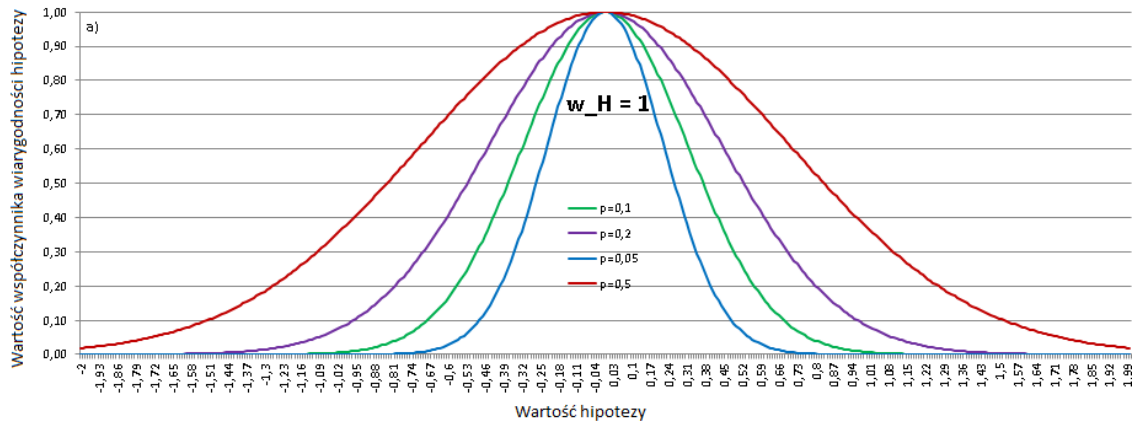
$$b(x) = 1 - \prod_i (1 - b_i(x)) \quad (5.6)$$

Powyższa funkcja jest znanym wzorem charakteryzującym między innymi niezawodność sprzętowych elementów równoległych. Przyjęcie tej funkcji jest więc rozwiązaniem sprawdzonym dla obliczania miar zaufania w różnych kontekstach.

Przykład

Dla wartości zaufania w punkcie x trzech hipotez wynoszących odpowiednio $b_1 = 0,9$; $b_2 = 0,8$; $b_3 = 0,7$ otrzymujemy: $1 - (0,1 * 0,2 * 0,3) = 0,994$. Dla przypadku, gdzie tylko dla jednej hipotezy wartość jest większa od 0 np. 0,1, otrzymamy wartość tej hipotezy $1 - (0,9) = 0,1$, co spełnia kryteria podane w ogólnym przypadku.





Rys 5.6 Wykres rozmycia hipotezy w punkcie 0, przy współczynniku wiarygodności równym: a) 1 b) 0,75 c) 0,5 d) 0,25 dla danych z przykładu kroku 2 [opracowanie własne]

4. Obliczanie współczynniki niewiary w modelu rozmytym (ang. disbelief function)

Propozycja autora na funkcję niewiary opisana jest wzorem 5.7. Wzór ten pochodzi z teorii ewidencji Demsptera–Shafera [139].

$$d(x) = \frac{\sum_j |x - H_j| * b(H_j)}{n-1} \quad (5.7)$$

5. Wyliczenie funkcji wiarygodności hipotez w modelu rozmytym (ang. plausibility function)

Propozycja autora na wyliczenie funkcji wiarygodności w modelu opisana jest wzorem 5.8. Wzór ten pochodzi z teorii ewidencji Demsptera–Shafera [139]. Jest to najprostszy wariant tej funkcji, polegający na wyliczeniu różnicy pomiędzy funkcją zaufania i niewiary. Funkcja ta może zostać zmodyfikowana, co może być przedmiotem dalszych badań.

$$c(x) = b(x) - d(x) \quad (5.8)$$

6. Defuzyfikacja hipotez

Praktyczną optymalizacją wyliczenia jest ograniczenie badanego zakresu do przedziału pomiędzy wartością minimalną oraz maksymalną dla hipotez początkowych, gdyż tylko w tym zakresie może znaleźć się uzgodnione rozwiązanie.

7. Redukcja liczby hipotez

Hipotezy uznane są za zbieżne, o ile ich odległość jest mniejsza niż zadana wartość progowa odległości D_{\max} np. przy dwóch hipotezach równych 0,81 oraz 0,8 i progu D_{\max} równym 0,02 hipotezy zostaną uznane za zbieżne, ale 0,83 i 0,8 już nie. Redukcję zbieżnych hipotez do jednej można przeprowadzić na kilka sposobów, poprzez:

- średnią ważoną (rozwiązanie wybrane przez autora pracy),
- wybór pierwszej ze zbieżnych hipotez (podejście naiwne),
- średnią arytmetyczną.

Wyznaczony w ten sposób nowy punkt może przyjmować różne wartości wiarygodności:

- wartość danego punktu wynikającego z funkcji wiarygodności z kroku 5 (rozwiązanie wybrane przez autora pracy),
- średnią z wartości wiarygodności łączonych hipotez,
- wartość maksymalną ze składowych.

8. Wybór uzgodnionej hipotezy dla każdego wymiaru

Jeżeli po wykonaniu wcześniejszych kroków pozostaje jedna hipoteza – jest ona zwracana jako rozwiązanie. Jeżeli jest ich więcej, sprawdzana jest wartość współczynnika wiarygodności i zwracana jest hipoteza o największej wiarygodności. Jeżeli wartość najwyższą osiągnęło więcej hipotez, wybór należy podjąć w zależności od typu aplikacji lub rodzaju zastosowaniu. Możliwe jest kilka rozwiązań:

- wybór pierwszej/ostatniej hipotezy (podejście naiwne),
- wybór rozwiązania dalszego od 0 – dla zastosowań aplikacji typu „affective” [23], w których minimalizowany jest błąd II rodzaju np. jeżeli lepsza jest nadmiarowa reakcja niż brak reakcji (rozwiązane domyślnie, zalecane przez autora pracy),
- wybór rozwiązania bliższego 0 – dla zastosowań aplikacji typu „affect-aware” [23], w których minimalizowany jest błąd I rodzaju np. jeżeli lepszy jest brak reakcji aplikacji niż niepotrzebna reakcja,
- wyliczenie średniej arytmetycznej z kandydatów, gdzie wiarygodność wyliczona jest jako wartość danego punktu wynikającego z funkcji wiarygodności z kroku 5.

5.5. Funkcja oceny programów automatycznego rozpoznawania emocji

Kolejnym komponentem zaproponowanej metody integracji, obok wzorca architektonicznego Scoarboard oraz algorytmu integracji, jest funkcja oceny programów automatycznego rozpoznawania emocji.

Funkcja oceny znajduje się w module *moderatora* i jest uruchamiana po funkcji integracji. Sposób jej działania może być modyfikowany niezależnie od pozostałych komponentów. Funkcja oceny porównuje hipotezę *eksperta* z wartością referencyjną, którą domyślnie jest uzgodnione rozwiązanie. W przypadku aplikacji testowych za wartość referencyjną może posłużyć wartość podana przez innego *eksperta* (referencyjnego lub posiadającego listę odpowiedzi). Zwrócona ocena powinna znajdować się w zakresie $\langle 0,1 \rangle$, gdzie wartość 1 to pełna zgodność z wartością referencyjną.

Ocena powinna uwzględniać współczynnik wiarygodności *eksperta*, który powinien odpowiadać jakości danych. Im większa pewność programu, tym większa powinna być kara za popełnienie błędu. Przykładowo, jeżeli program podaje skrajnie rozbieżną hipotezę z dużą pewnością, powinien otrzymać niską ocenę.

Funkcja oceny spełnia dwa zadania: służy moderatorowi jako wskazówka do wyboru kolejnego programu do wywołania, ale także ocenie skuteczności wykorzystywanych programów automatycznego rozpoznawania emocji. Autor zaproponował wzór 5.9 do wyliczenia oceny, gdzie n – liczba analizowanych wymiarów, i – kolejny wymiar np. P, A, D (dla modelu PAD), A_i – wartość hipotezy dla wymiaru i , W_{A_i} – wiarygodność hipotezy cząstkowej wymiaru i , R_i – wartość hipotezy uzgodnionej.

$$O = \frac{\sum_{i=1}^n 1 - (|A_i - R_i|) * W_{A_i}}{n} \quad (5.9)$$

Przykład

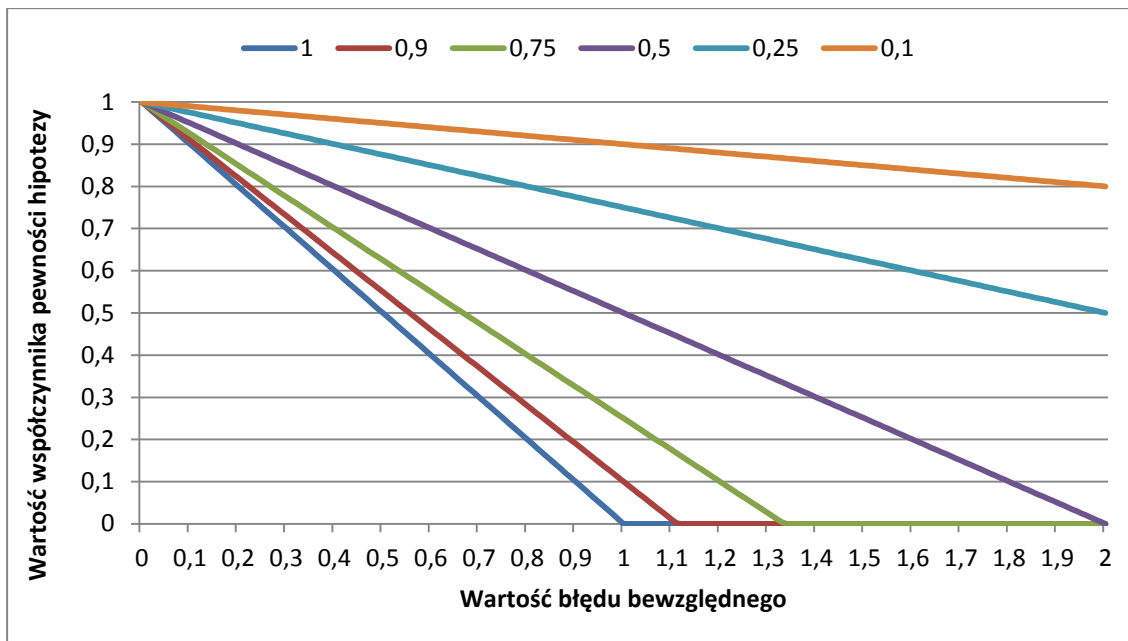
Kilka przykładowych wartości oceny zostało zestawionych w Tabeli 5.2. Wizualizacja oceny dla współczynników 0,1; 0,25; 0,5; 0,75; 0,9 oraz 1 znajduje się na Rys 5.7. Jeżeli ocena wyniesie mniej niż 0, jako ocena powinna zostać przypisana wartość 0. Wartość oceny równa 0 zostanie osiągnięta dla błędu (odległości od hipotezy) równego lub większego:

- 1 dla współczynnika wiarygodności równego 1,
- 1,11 dla współczynnika wiarygodności równego 0,1,
- 1,34 dla współczynnika wiarygodności równego 0,25,
- 2 dla współczynnika wiarygodności równego 0,5.

Tabela 5.2 Przykładowe wartości błędów i wyliczone przy nich oceny wyznaczone przez proponowaną funkcję oceny [opracowanie własne].

Wartość błędu bezwzględnego	Wartość współczynnika pewności	Wartość oceny
0,2	0,9	0,82
0,5	0,9	0,55
1	0,9	0,1
0,2	0,75	0,85
0,5	0,75	0,625
1	0,75	0,25
0,2	0,5	0,9
0,5	0,5	0,75
1	0,5	0,5
0,2	0,25	0,95
0,5	0,25	0,875
1	0,25	0,75
0,2	0,1	0,98
0,5	0,1	0,95
1	0,1	0,9
2	0,1	0,8

Warto zwrócić uwagę na to, że niski współczynnik wiarygodności (poniżej 50%) niejako gwarantuje wysoką ocenę – co jest zgodne z założeniem, by nie karać programu złą jakością danych. Może to jednak spowodować sztuczne zawyżenie oceny programu i częstsze jego wywołanie mimo niskiej wartości dodanej. W tym celu można rozważyć modyfikację szansy wywołania do wartości $\text{ocena}/2$, co ograniczy szansę wywołania programu do 50%.



Rys 5.7 Wartości funkcji oceny w zależności od współczynnika pewności hipotezy oraz wartości błędu bezwzględnego [opracowanie własne]

5.6. *Propozycja sposobu wyliczenia współczynnika niepewności dla wybranych modalności*

Ostatnim komponentem zaproponowanej metody integracji jest propozycja wyliczenia współczynnika niepewności dla wybranych modalności.

Problem określenia niepewności pomiarowej programów automatycznego rozpoznawania emocji został szerzej opisany w rozdziale 4.3. Zostało tam wskazane, że programy rzadko prezentują miary jakości, a współczynnik wiarygodności wykorzystywany na potrzeby integracji często jest wyznaczany poprzez wyliczenie skuteczności rozwiązania na danych testowych. Podejściem alternatywnym dla wskaźnika z programu jest wyliczenie wskaźnika poza nim przy pomocy otrzymanych od niego lub pobranych bezpośrednio z kanału obserwacji parametrów jakości danych.

W tym rozdziale zostały zaproponowane sposoby wyliczenia współczynnika niepewności dla wybranych modalności.

Wskaźnik niepewności dla analizy twarzy

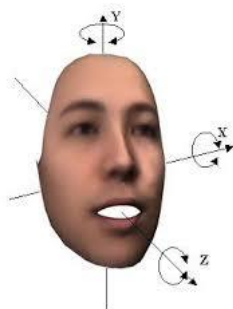
W trakcie analizy mimiki twarzy w literaturze często wskazuje się kilka czynników mówiących o jakości rozpoznawania emocji: (1) wykrycie twarzy, (2) oświetlenie oraz (3) kąt kamery.

(1) Wykrycie twarzy jest kwestią kluczową, gdyż jeżeli twarz nie zostanie wykryta, program nie zwróci wyniku, a więc wiarygodność oceny będzie równa 0 (brak oceny). Na wykrycie twarzy mają wpływ okrycia twarzy takie jak okulary, zarost czy dłoń na twarzy.

(2) Nierównomierne oświetlenie twarzy może być przyczyną pogorszenia jakości analizy emocji z twarzy oraz utrudnieniem dla samego wykrywania twarzy.

(3) Preferowanym ustawieniem kamery jest ustawienie na wprost twarzy, co jest zadaniem o tyle trudnym, że dla użytkownika komputera kamera powinna być zlokalizowana na środku monitora. Wpływ kąta kamery na wynik rozpoznawania emocji został opisany w rozdziale 4.4, a także w publikacjach [1] [2].

Ustawienie twarzy do kamery można analizować w trzech osiach – X,Y,Z, które zostały zilustrowane na Rys 5.8 [140]. Im twarz mocniej odbiega od preferowanego ustawienia (kąty: [0;0;0]), tym trudniejsza analiza obrazu twarzy. Przydatność analizy jest możliwa do pewnego kąta, po którym twarz nie jest możliwa do rozpoznania. Za punkt graniczny autor pracy przyjął kąt 90 stopni, dla którego w osi Y widoczny jest jedynie jeden profil.



Rys 5.8 Osie obrotów twarzy w przestrzeni 3D [140]

Autor pracy opracował metryki opartych o powyższe trzy kąty, opisane wzorami 5.10 – 5.14, które oznaczają odpowiednio:

- wzór (5.10) jest naiwnym podejściem polegającym na wyliczeniu błędu bezwzględnego jako sumę odchyień, dla którego punktem granicznym (dla którego wartość jest równa 0) to jeden z kątów równy 90 stopni,
- wzór (5.11) jest modyfikacją wzoru 1, dla którego punktem granicznym są 3 kąty o wartości 90 stopni. Opiera się on na zmodyfikowanym wzorze na odchylenie przeciętne, gdzie zamiast średniej użyto wartość oczekiwaną,
- wzór (5.12) jest znanym wzorem charakteryzującym między innymi niezawodność sprzętową elementów równoległych. Przyjęcie tej funkcji jest więc rozwiązaniem sprawdzonym dla obliczania miar zaufania w różnych kontekstach,
- wzór (5.13) bazuje na odchyleniu standardowym populacji, wykorzystując średnią kwadratową różnic między wartością graniczną a wartością kątów. Wzór ten ze względu na kwadrat różnic jest bardziej wyczulony na duże odchylenie w pojedynczym kącie,
- wzór (5.14) powstał z połączenia wzorów (5.12) i (5.13).

$$1 - \left| \frac{\alpha}{90} \right| - \left| \frac{\beta}{90} \right| - \left| \frac{\gamma}{90} \right| \quad (5.10)$$

$$1 - \frac{\left| \frac{\alpha}{90} \right| + \left| \frac{\beta}{90} \right| + \left| \frac{\gamma}{90} \right|}{3} \quad (5.11)$$

$$\left(1 - \left| \frac{\alpha}{90} \right| \right) * \left(1 - \left| \frac{\beta}{90} \right| \right) * \left(1 - \left| \frac{\gamma}{90} \right| \right) \quad (5.12)$$

$$1 - \sqrt{\frac{\left(\frac{\alpha}{90}\right)^2 + \left(\frac{\beta}{90}\right)^2 + \left(\frac{\gamma}{90}\right)^2}{3}} \quad (5.13)$$

$$\sqrt{\frac{\left(1 - \frac{\alpha}{90}\right)^2 + \left(1 - \frac{\beta}{90}\right)^2 + \left(1 - \frac{\gamma}{90}\right)^2}{3}} \quad (5.14)$$

Przykładowe wartości dla wzorów 5.10–5.14 zostały przedstawione w Tabeli 5.3. Dla wzorów 5.10 i 5.13 otrzymano wartości ujemne, jednakże w takim przypadku wartości te powinny być interpretowane jako 0.

Założenia wspólne dla przedstawionych wzorów:

- dla preferowanego ustawienia na wprost kamery współczynnik zaufania jest równy 100%,
- dla niewielkich odchyłeń (do 5 stopni) [140], współczynnik zaufania jest stosunkowo wysoki – powyżej 80%,
- dla jednego kąta powyżej 90 stopni, współczynnik zwraca wynik 0% – wyjątek stanowią wzory 5.11 oraz 5.14, dla których wynik równy 0% osiągany jest dla trzech kątów równych 90 stopni.
- Weryfikacja wzorów na istniejącej bazie została przeprowadzona w rozdziale 7.

Tabela 5.3 Wartości kątów odchylenia od twarzy na wprost kamery oraz wartości proponowanych miar niepewności [opracowanie własne]

α	β	γ	Wzór 5.10	Wzór 5.11	Wzór 5.12	Wzór 5.13	Wzór 5.14
0	0	0	100%	100%	100%	100%	100%
5	0	0	94%	98%	94%	94%	98%
5	5	5	83%	94%	84%	92%	94%
15	0	0	83%	94%	83%	83%	95%
30	0	0	67%	89%	67%	67%	90%
30	30	30	0%	67%	30%	49%	67%
60	0	0	33%	78%	33%	33%	84%
60	30	0	0%	67%	22%	25%	72%
90	0	0	0%	67%	0%	0%	82%
60	60	60	-100%	33%	4%	-2%	33%
90	90	90	-200%	0%	0%	-53%	0%

Wskaźnik niepewności dla analizy ładunku emocjonalnego tekstu

Jednym z najczęściej wykorzystywanych i najprostszych programów analizy ładunku emocjonalnego z tekstu jest analiza oparta na ładunku emocjonalnym słów w tekście. Najłabszym punktem tych metod są ograniczenia słowników afektywnych, które nie posiadają wszystkich możliwych słów. Co więcej, gdy użytkownik napisze wyraz niedbale (z błędem ortograficznym lub literówką), użyje zapożyczeń z innego języka lub nazw własnych, algorytm nie znajdzie przypisania ładunku w słowniku. Przy braku przypisania wartości afektywnej do słowa (choćby neutralnego), zwiększa się niepewność rozwiązania związanego z ryzykiem pominięcia istotnego słowa nacechowanego afektywnie, gdyż jedno słowo może zmienić znaczenie całego zdania, przykładowo „Jej ciało leżało na łóżku”, a „Jej **martwe** ciało leżało na łóżku” mają zupełnie inny wydźwięk emocjonalny.

W niniejszej pracy przyjęto uproszczony sposób wyznaczania współczynnika niepewności polegający na wykorzystaniu proporcji rozpoznanych słów (z przypisanym w słowniku ładunkiem emocjonalnym) i modyfikatorów zdań, takich jak znaki interpunkcyjne i emotikony, do wszystkich użytych słów i modyfikatorów w danym zdaniu. Zaproponowane rozwiązanie, określone wzorem (5.15), może być rozszerzone o kolejne współczynniki lub różne wagi dla różnych części mowy.

$$\frac{\text{Liczba rozpoznanych słów}}{\text{Liczba użytych słów}} * 100\% \quad (5.15)$$

Dla bardziej złożonych programów wartości rozpoznanych słów nie muszą być liczbami całkowitymi. Przykładowo przy zastosowaniu logiki rozmytej, binarna odpowiedź na temat znajomości może zostać zastąpiona poziomem trafności przypisania słowa, co byłoby odpowiedzią na problem literówek.

5.7. Zestawienie problemów i propozycji rozwiązania – podsumowanie

Podsumowując, zaproponowane rozwiązanie składa się z:

- wzorca architektonicznego Scoreboard,
- metody kwantyfikacji informacji o niepewności pomiarów w danym kanale obserwacji,
- metody integracji informacji o stanie emocjonalnym uwzględniającej niepewność i niedostępność kanałów obserwacji.

W tym rozdziale została przedstawiona koncepcja rozwiązania, które koresponduje z problemami badawczymi opisanymi w rozdziale 4. Dla podsumowania komponenty proponowanego rozwiązania wraz z ich konsekwencjami zostały zebrane w Tabeli 5.4.

Tabela 5.4 Proponowane komponenty rozwiązania i ich konsekwencje [opracowanie własne].

Komponent rozwiązania	Konsekwencje
<ul style="list-style-type: none"> Wzorzec architektoniczny Scoreboard; Paradygmaty odwrócenia sterowania (ang. inversion of control) Paradygmat wstrzykiwania zależności (ang. dependency injection) 	<p>Korzyści:</p> <ul style="list-style-type: none"> łatwe dodawanie, modyfikowanie oraz usuwanie <i>ekspertów</i> (tu: programów automatycznego rozpoznawania emocji), a także modyfikacja przebiegu sterowania w odizolowanym fragmencie kodu, możliwość wykorzystania programów napisanych w różnych technologiach, możliwość zastosowania zamkniętego kodu jako podejście czarnej skrzynki (ang. blackbox), niezależna implementacja poszczególnych modułów m.in. <i>ekspertów</i>, brak ograniczenia liczby <i>ekspertów</i>, rozwiązanie niezależne od zastosowanych modalności – jest to problem warstwy rozpoznawania emocji (programu automatycznego rozpoznawania emocji). <p>Ograniczenia:</p> <ul style="list-style-type: none"> wymusza znajomość i wykorzystanie wzorców projektowych oraz paradygmatów programowania – wyższy próg wejścia, większy poziom skomplikowania rozwiązania – dodatkowy narzut struktur, trudność testowania, zwłaszcza gdy wykorzystanych jest wielu ekspertów (ograniczenie odziedziczone ze wzorca Blackboard).
Wzorzec architektoniczny Scoreboard	<p>Korzyści:</p> <ul style="list-style-type: none"> możliwość wyboru metody fuzji: wczesna, późna, hybrydowa. <p>Ograniczenia:</p> <ul style="list-style-type: none"> wykorzystanie wzorca może spowodować pogorszenie wydajności, zwłaszcza gdy kryteria ograniczające liczbę <i>ekspertów</i> nie zostaną zastosowane.
Wzorzec architektoniczny Scoreboard, komponenty: <ul style="list-style-type: none"> Moderator oraz Tabela ocen 	<p>Korzyści:</p> <ul style="list-style-type: none"> możliwość dynamicznej oceny działania programów w czasie działania aplikacji, wbudowana możliwość zbierania statystyk całego przebiegu działania, możliwość przeprowadzenia porównania różnych wersji programu. <p>Ograniczenia:</p> <ul style="list-style-type: none"> narzut wydajnościowy związany z moderacją i oceną.
Wzorzec Scoreboard, komponenty: <ul style="list-style-type: none"> Arbiter 	<p>Korzyści:</p> <ul style="list-style-type: none"> pozwała na określenie maksymalnej liczby wykorzystanych <i>ekspertów</i> bez ich preselekcji – dynamiczne podejmowanie decyzji w każdej iteracji, dynamiczna selekcja wadliwych <i>ekspertów</i>, skalowalność. <p>Ograniczenia:</p> <ul style="list-style-type: none"> narzut wydajnościowy związany z wyborem ekspertów (programów) do wywołania.
Algorytm integracji CFIEE	<p>Korzyści:</p> <ul style="list-style-type: none"> brak ograniczeń liczby programów do integracji, nie wymaga zmiany programu przy dodaniu bądź usunięciu <i>eksperta</i> z wywołania, pozwała na wykorzystanie dowolnego ciągłego modelu reprezentacji emocji, program wykorzystuje współczynnik wiarygodności, a także promuje zbieżność hipotez, odporność na brak danych z kanału lub jego czasową niedostępność. <p>Ograniczenia:</p> <ul style="list-style-type: none"> wymaga niewielkiej zmiany kodu przy zmianie modelu reprezentacji emocji na inny, może wymagać wykorzystania odwzorowania modeli reprezentacji emocji, jeżeli <i>eksperci</i> zwracają wyniki w różnych modelach, wymaga wykorzystanie współczynnika wiarygodności hipotez, algorytm ma tendencje do uśredniania wyniku – preferuje stany bliższe 0 niż skrajne (1 oraz -1), brak gwarancji uzyskania satysfakcjonującego rozwiązania – cecha charakterystyczna dla rozwiązań niedeterministycznych.
Algorytm integracji CFIEE, <ul style="list-style-type: none"> Wykorzystanie współczynnika wiarygodności 	<p>Korzyści:</p> <ul style="list-style-type: none"> odpowiedź na problem rozbieżności hipotez ekspertów – wskazanie rozwiązania w przypadku rozbieżności hipotez, indywidualna analiza każdego z wymiarów i ocena wiarygodności każdego z osobna, oszacowanie poziomu wiarygodności uzgodnionego rozwiązania na podstawie zbieżności hipotez oraz poziomu zaufania do samych hipotez (np. poprzez określenie jakości danych kanału), uwzględnienie natury rozpoznawania stanów emocjonalnych oraz niemożności określania „prawdziwego” stanu emocjonalnego (ang. ground truth) poprzez rozmycie hipotez.
Zaproponowany współczynnik wiarygodności	<p>Korzyści:</p> <ul style="list-style-type: none"> Uwzględnia jakość danych, Możliwość oszacowania poziomu wiarygodności danych dla poszczególnych modalności.

Jednym z najważniejszych założeń rozwiązania jest przedstawienie elastycznego, konfigurowalnego rozwiązania, dlatego też rozwiązanie pozwala na dostosowanie funkcji integracji w zależności od potrzeb dla dwóch kryteriów:

- zbieżności wyników algorytmów,
- wiarygodności wyników algorytmów.

Zaproponowane rozwiązanie pozwala na różną konfigurację, poprzez:

- wybór metody fuzji: wczesna, późna, hybrydowa,
- wybór modelu reprezentacji emocji przy stosunkowo małej modyfikacji kodu,
- aktualizację, usunięcie lub dodanie nowego *eksperta* bez konieczności modyfikacji całego programu,
- ustalenie zestawu wykorzystywanych *ekspertów* (tu: algorytmów automatycznego rozpoznawania emocji) oraz parametrów przebiegu programu m.in. maksymalna liczba *ekspertów* w iteracji, liczba faz itp. poprzez zewnętrzny plik konfiguracyjny,
- wyznaczania *ekspertów* do wywołania poprzez metodę *arbitra*,
- modyfikację przebiegu programu poprzez modyfikację klasy *kontrolera*,
- modyfikację funkcji oceny programów poprzez modyfikację klasy *moderatora*,
- ustalenie wartości progowej odległości, dla której dwie hipotezy można uznać za zbieżne,
- zmiany współczynników rozkładu normalnego – regulujących stromość rozkładu, w zależności od przyjętego przedziału ufności,
- zmianę wykorzystywanego rozkładu statystycznego na inny,
- możliwość modyfikacji wzorów w poszczególnych krokach algorytmu integracji, w tym:
 - funkcję zaufania,
 - funkcję niewiary,
 - funkcję wiarygodności,
 - funkcję redukcji hipotez – wybór sposobu wyznaczania współczynnika wiarygodności dla redukowanych przypadków,
 - zmianę algorytmu wyboru uzgodnionej hipotezy w przypadku równej wartości współczynnik wiarygodności m.in. redukcja błędu I lub II rodzaju, w zależności od zastosowania.

5.8. Tezy

Badania przeprowadzone w ramach niniejszej pracy doktorskiej mają na celu udowodnienie postawionej tezy:

Zaproponowana metoda integracji danych o stanie emocjonalnym umożliwia poprawę wybranych atrybutów jakości wielomodalnego rozwiązania automatycznego rozpoznawania emocji

Wraz z tezą główną zostały przygotowane trzy tezy pomocnicze, których potwierdzenie potwierdza tezę główną. Tezy pomocnicze zostały sformułowane następująco:

- ▶ Teza pomocnicza 1: Zastosowanie wzorca architektonicznego Scoreboard pozwala na poprawę modularności, łatwości integracji, adaptowalności i odporności w wielokanałowym rozpoznawaniu emocji.
- ▶ Teza pomocnicza 2: Zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwia reprezentację niepewności oszacowania stanu emocjonalnego.
- ▶ Teza pomocnicza 3: Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwia poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

W celu udowodnienia pierwszej tezy pomocniczej zostały przygotowane:

- Eksperyment 1. Analiza ładunku emocjonalnego z tekstu,
- Symulacja 1. Weryfikacja miary odporności oraz
- Studium przypadku Monitora Emocji.

Zostały one opisane w rozdziale 6 – Ocena architektury zaproponowanego rozwiązania z wykorzystaniem wzorca architektonicznego Scoreboard.

W celu udowodnienia drugiej tezy pomocniczej zostały przygotowane:

- Eksperyment 2. Walidacja metody wyznaczania współczynnika wiarygodności oraz
- Eksperyment 3. Dobór współczynnika wiarygodności w przypadku jego braku.

Zostały one opisane w rozdziale 7 – Ewaluacja zaproponowanej metody kwantyfikacji informacji umożliwiającej reprezentację niepewności.

W celu udowodnienia trzeciej tezy pomocniczej zostały przygotowane eksperymenty, dotyczące poprawy niezawodności:

- Eksperyment 4. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na zbiorze danych SEMAINE (rozdział 8.2),
- Eksperyment 8. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na danych z eksperymentu z 4 kamerami (rozdział 10.2).

A także eksperymenty i symulacje do weryfikacji poprawności wyznaczania współczynnika wiarygodności:

- Eksperyment 5. Odpowiednia konfiguracja funkcji integracji – dobór odpowiednich parametrów (rozdział 8.3),
- Eksperyment 6. Porównanie skuteczności integracji zaproponowanej metody z rozwiązaniami jednomodalnymi oraz referencyjnymi na zbiorze danych SEMAINE (rozdział 8.4),
- Eksperyment 7. Porównanie poprawności wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania przez zaproponowaną metodę oraz rozwiązania referencyjne (rozdział 8.5),
- Eksperyment 9. Porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez (rozdział 10.4),
- Symulacja 2. Analiza zmiany wartości współczynnika wiarygodności dla rozwiązań zbieżnych i rozbieżnych (rozdział 9.1),
- Symulacja 3. Współczynnik wiarygodności uzgodnionego rozwiązania dla zmiennej wartości współczynnika wiarygodności hipotez (rozdział 9.2),
- Symulacja 4. Analiza wpływu wartości wiarygodności rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania (rozdział 9.3),
- Symulacja 5. Analiza wpływu odległości rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania (rozdział 9.4).

W kolejnych rozdziałach zostały użyte takie określenia jak **hipoteza**, rozumiana jako wynik zwrócony przez program automatycznego rozpoznawania emocji oraz **zbieżne hipotezy**, które oznaczają wyniki oddalone nie bardziej niż założona wartość progowa odległości D_{\max} (distance max) np. wartości 0,8 oraz 0,89 są zbieżne dla współczynnika równego 0,1, ale 0,8 i 0,91 już są rozbieżne.

Ze względu na specyfikę tematyki rozpoznawania emocji oraz stosowanie metod rozmytych, nie jest możliwe wykorzystanie klasycznej definicji miary skuteczności. Jest to spowodowane naturą rozpoznawania stanów emocjonalnych i niemożnością określania „prawdziwego” stanu emocjonalnego (ang. ground truth), co zostało szerzej przedyskutowane w rozdziale 2.6.

Zaproponowana miara skuteczności liczona jest jako liczba przypadków pozytywnego rozpoznania wobec liczby wszystkich prób, przy czym pozytywne rozpoznanie stanu emocjonalnego jest rozumiane jako wynik znajdujący się w odległości nie większej niż określona wartość progu odległości (D_{\max}) od wartości referencyjnej, przykładowo: jeżeli zadana wartość progu odległości wynosi 0,05, a wartość referencyjna to 0,9, to wszystkie hipotezy z przedziału 0,85–0,95 zostaną uznane za poprawne.

6. OCENA ARCHITEKTURY ZAPROPONOWANEGO ROZWIĄZANIA Z WYKORZYSTANIEM WZORCA ARCHITEKTONICZNEGO SCOREBOARD

Pierwszą tezę pomocniczą zatytułowano:

Zastosowanie wzorca architektonicznego Scoreboard pozwala na poprawę modularności, łatwości integracji, adaptowalności i odporności w wielokanałowym rozpoznawaniu emocji.

Aby zweryfikować tezę, autor pracy wykorzystał metodę GQM [141] (ang. Goal–Question–Metric) do ewaluacji atrybutów jakości z tezy [4]. Cel oraz pytania do GQM zostały przedstawione w Tabeli 6.1.

Tabela 6.1 Cel i pytania do metody GQM, na podstawie [141].

Cel	Analiza wzorca architektonicznego pod kątem jakości wielokanałowej integracji z punktu widzenia programisty w odniesieniu do wyzwań wielomodalnego rozpoznawania emocji.
Pytanie 1	Czy zaproponowane rozwiązanie jest dość modularne , by obsłużyć istniejące (różnorodne) programy automatycznego rozpoznawania emocji (<i>ekspertów</i>)?
Pytanie 2	Czy zaproponowane rozwiązanie pozwalana na integrację wyników programów automatycznego rozpoznawania emocji (<i>ekspertów</i>)?
Pytanie 3	Czy rozwiązanie dostosowuje się do zmian w trakcie działania (ang. runtime), takich jak tymczasowy brak dostępności wyników <i>ekspertów</i> ?
Pytanie 4	Czy rozwiązanie jest odporne na (czasowe lub stałe) błędne odpowiedzi <i>ekspertów</i> ?

Pytania zostały odwzorowane na odpowiednie miary jakości, których definicje zostały przytoczone poniżej:

Modularność (ang. modularization) – cecha architektury pozwalająca na odseparowanie logicznych jednostek rozwiązania (modułów, komponentów). Literatura wskazuje na wysoką spójność (ang. cohesion) i niski poziom powiązań (ang. coupling) jako standardowe kryterium ewaluacji modularności [21].

Integracja (ang. integration) – cecha architektury reprezentująca zdolność do połączenia wyników z różnorodnych rozwiązań (tu: programów automatycznego rozpoznawania emocji). To kryterium powinno być rozumiane jako zdolność do integracji oraz otrzymania wyników nie gorszych od integrowanych rozwiązań.

Adaptowalność (ang. adaptability) – cecha architektury reprezentująca zdolność do łatwego dołączania/odłączania/podmiany modułów (m.in. ekspertów, tu: programów automatycznego rozpoznawania emocji).

Odporność (ang. robustness) – cecha programu automatycznego rozpoznawania emocji, wykorzystującego metodę integracji. Jeżeli do integracji zostanie wykorzystany niepoprawny ekspert (program zwracający niepoprawne wyniki) – powinien zostać wykryty, a jego wpływ na uzgodnione rozwiązanie powinien być ograniczony.

Projekt badania oraz wykorzystane metryki dla powyższych miar jakości zostały opisane w kolejnej sekcji. Rozdział ten bazuje na publikacji [4], której współautorem jest autor pracy.

6.1. *Projekt badania*

W ramach badań nad udowodnieniem tezy został przygotowany i przeprowadzony odpowiedni eksperyment, symulacja oraz studium przypadku. Rzeczywiste dane wykorzystane w eksperymencie, nie pozwalają na przetestowanie wszystkich możliwych scenariuszy – dlatego w ramach uzupełnienia została przeprowadzona symulacja, w której dane zostały wygenerowane. W tym rozdziale opisano:

- Eksperyment 1 – Integracja danych wielokanałowych w oparciu o analizę ładunku emocjonalnego tekstu,
- Symulacja 1 – Scenariusz weryfikujący miarę odporności, na podstawie atrap (ang. mock-up) programów rozpoznawania,
- Studium przypadku Monitora Emocji.

Eksperyment, symulacja oraz studium przypadku pozwalają na obliczenie metryk ilościowych, jednocześnie dla studium przypadku Monitora Emocji przeprowadzono jakościową ewaluację architektury poprzez wykorzystanie metody ATAM [142] [143]. Jako rozwiązanie referencyjne przyjęto wzorzec architektoniczny Blackboard.

Dla każdej z miar jakości zostały zdefiniowane odpowiednie metryki.

Adaptowalność (ang. adaptability)

Adaptowalność została zmierzona za pomocą następujących metryk:

- (A1) LOC (liczba linii kodu) potrzebna do podpięcia programu do integracji – dodanie do pętli sterowania,
- (A2) LOC potrzebna do usunięcia programu z integracji,
- (A3) LOC potrzebna do modyfikacji interfejsu integracji,
- (A4) liczba klas wymagających modyfikacji dla zmiany przepływu sterowania program,
- (A5) liczba klas wymagających modyfikacji dla zmiany funkcji integracji,
- (A6) LOC potrzebna do dodania nowego programu do integracji,
- (A7) narzut kontroli przepływu – stosunek LOC dla mechanizmu kontroli przepływu do całkowitej liczby LOC,
- (A8) narzut komunikacji – stosunek LOC dla mechanizmu komunikacji do całkowitej liczby LOC.

Integracja (ang. integration)

Integracja jest miarą jakościową mówiąca o zdolności systemu do integracji różnorodnych rozwiązań. Różnorodność jest tutaj rozumiana jako wykorzystanie różnych metod, modalności, technologii itp. Integracja może zostać zdefiniowana za pomocą takich terminów jak: poprawność (ang. correctness), wydajność (ang. performance) oraz łatwość użycia (ang. ease-of-use). Łatwość użycia jest powiązana z **adaptowalnością** i na potrzeby tej pracy, jest rozumiana jako cecha wzorca architektonicznego, która określa łatwość podłączenia/odłączenia/zamiany istniejącego programu automatycznego rozpoznawania emocji. Wykorzystaną miarą integracji jest miara skuteczności rozwiązania (I1).

Odporność (ang. robustness)

Odporność rozwiązania jest rozumiana jako cecha ograniczająca wpływ awarii jednego z modułów *ekspertów* (programów automatycznego rozpoznawania emocji) na uzgodnione rozwiązanie. Cecha ta została zbadana poprzez wybór podzbioru programów automatycznego rozpoznawania emocji. Na potrzeby analizy miary jakości odporności, wykorzystano następujące metryki:

- (R1) – skuteczności rozpoznawania uzgodnionego rozwiązania,
- (R2) – liczby wywołań programu,
- (R3) – wynik programu uzyskany po wszystkich etapach.

Modularność (ang. modularization)

Na potrzeby analizy modularności, wykorzystano następujące metryki:

- (M1) unikalna liczba klas,
- (M2) liczba metod przypadająca na pojedynczą klasę,
- (M3) liczba linii kodu (ang. Line Of Code – LOC) na pojedynczą metodę,
- (M4) złożoność cyklometryczna metod, która powinna być mniejsza niż 10 dla każdej klasy (klasy z wartością powyżej 20 są podatne na błędy i należy rozważyć ich refaktoryzację),
- (M5) głębokość dziedziczenia – im niższa wartość, tym lepiej, jednakże wartość 1 oznacza brak abstrakcji, co można uznać za błąd,
- (M6) poziom powiązań (ang. coupling) klas oraz (M7) poziom powiązań metod – zasadniczo, im niższa wartość, tym lepiej.
- niski poziom powiązań (ang. coupling)

Poziom powiązań klas powinien być niższy niż 10–20 [144] [145]. Zgodnie z zaleceniami Oliveira i innych [146] wartość ta powinna być osiągnięta dla pewnego procenta klas np. wartość 10 dla 70% klas, 22 dla 80% [145] itd. Inne badania podają inne wartości – średnią 13,4 oraz medianę równą 8 [147].

Zestawienie

W Tabeli 6.2 zestawiono informację na temat miar jakości oraz metod badawczych odpowiadających na zagadnienia tych miar. W komórkach tej tabeli zostały wskazane tabele, w których opisane zostały dane dotyczące poszczególnych miar jakości. Symulacja i eksperyment prezentują dane ilościowe, a studium przypadku zarówno dane ilościowe, jak i jakościowe.

Tabela 6.2 Pokrycie miar jakości przez eksperymenty i symulacje [opracowanie własne].

Miary jakości	Metoda badawcza		
	Eksperyment 1	Symulacja 1	Studium przypadku (ilościowe)
Modularyzacja			Tabela 6.9
Integracja	Tabela 6.3		
Adaptowalność	Tabela 6.4 Tabela 6.5		Tabela 6.10
Odporność	Tabela 6.6 Tabela 6.7	Tabela 6.8	

Tabela 6.11
Tabela 6.12
Tabela 6.13

6.2. *Eksperyment 1. – integracja danych wielokanałowych w oparciu o analizę ładunku emocjonalnego tekstu*

Plan eksperymentu

Celem eksperymentu 1 było porównanie miary integracji, odporności, a także adaptowalności. Wykorzystano w nim siedem wersji programu analizującego ładunek emocjonalny tekstu na sześciu zbiorach po 40 sentencji. Działanie programów było oparte na regułach (ang. rule-based) oraz na różnych słownikach (ang. dictionary-based). Wykorzystano słownik opracowany na Politechnice Gdańskiej [148] [149] oraz słownik ANEW [69]. Słownikami w tym przypadku były leksykony z afektywnymi etykietami przypisanymi do słów.

W podejściu bazującym na wzorcu referencyjnym (Blackboard) – zawsze wywoływani są wszyscy *eksperci* (programy automatycznego rozpoznawania emocji), a w przypadku wzorca *Tabeli wyników* (ang. scoreboard) wybór *ekspertów* jest określany na podstawie wartości na *Tabeli wyników* (według ocen *ekspertów*).

W eksperymencie wykorzystano dwa scenariusze doboru ekspertów: 1) wybór spośród wszystkich wersji programu oraz 2) wybór z trzech dobrych wersji i jednej słabej. Pierwszy scenariusz pozwala ocenić skuteczność integracji, a drugi określa poziom odporności rozwiązania. Miara adaptowalności dotyczy kodu rozwiązania, który nie ulega zmianie przy zmianie scenariusza.

Wyniki eksperymentu I – Integracja

W Tabeli 6.3 zestawiono wyniki integracji, rozumianej w kontekście skuteczności. Wyniki przedstawiono w podziale na poszczególne wersje programu (S2–S2.3 AD) oraz rozwiązania zintegrowane (Blackboard oraz Scoreboard).

Tabela 6.3 Skuteczność rozpoznawania emocji w podziale na zbiory danych i wykorzystane rozwiązanie integracji dla wyników eksperymentu. Na podstawie [4].

Rozwiązanie	Skuteczność dla poszczególnych zbiorów					
	Zbiór 1	Zbiór 2	Zbiór 3	Zbiór 4	Zbiór 5	Zbiór 6
S 2.0	72,16	71,64	65,70	72,24	71,49	70,98
S 2.0.1	71,63	73,36	68,42	72,65	72,04	70,10
S 2.1	71,87	69,03	68,09	74,00	67,66	70,82
S 2.2	71,87	74,06	68,42	72,64	72,04	70,10
S 2.2.1	73,42	73,39	68,42	74,28	69,12	70,10
S 2.2.2	70,23	na	75,00	64,27	84,58	70,04
S 2.3 AD	67,61	63,95	74,33	68,64	70,00	67,63
Uzgodnione rozwiązanie (Blackboard)	84,39	82,91	79,10	80,92	89,04	74,09
Uzgodnione rozwiązanie (Scoreboard)	87,05	88,43	81,06	81,48	82,60	81,52

Wartości skuteczności dla rozwiązań integracji były wyższe od rozwiązań jednokanałowych, z wyjątkiem zbioru piątego, dla którego program „S 2.2.2” uzyskał wyższą skuteczność od architektury opartej o wzorzec Scoreard, ale niższą od rozwiązania realizującego wzorzec Blackboard. Jednocześnie, był to jedyny zbiór, dla którego wzorzec Blackboard uzyskał wyższą skuteczność niż wzorzec Scoreboard. Zatem dla większości przypadków, dynamiczny dobór ekspertów pozwolił na poprawę skuteczności rozwiązania.

Otrzymane wyniki pozwalają na potwierdzenie tezy, że rozwiązanie oparte na wzorcu architektonicznym Scoreboard zapewnia wyższą skuteczność integracji niż pojedyncze rozwiązanie.

Wyniki eksperymentu I – Adaptowalność

Adaptowalność została zmierzona metrykami wykorzystującymi rozmiar kodu (liczba linii kodu oraz liczba klas), które są potrzebne do zmiany/dodania, by wprowadzić zmianę. Im mniejsza jest to liczba, tym lepsza adaptowalność rozwiązania. W Tabeli 6.4 przedstawione zostały wartości miar adaptowalności dla przeprowadzonego eksperymentu – programu integrującego wyniki programów analizy ładunku emocjonalnego tekstu.

Tabela 6.4 Miary adaptowalności (A1–A6) dla rozwiązania z eksperymentu z programami do analizy ładunku emocjonalnego z tekstu. Na podstawie [4].

Miara adaptowalności	Blackboard	Scoreboard	Dodatkowe informacje
(A1) LOC dla dodania programu	1	1	Nie wymaga rekompilacji rozwiązania
(A2) LOC dla usunięcia programu	1	1	Nie wymaga rekompilacji rozwiązania
(A3) LOC dla podmiany programu	1	1	Nie wymaga rekompilacji rozwiązania
(A4) liczba klas do zmiany przepływu danych	1	1	Nadpisanie modułu kontrolera (jedna klasa)
(A5) liczba klas do zmiany funkcja oceny	1	1	Nadpisanie modułu oceny (jedna klasa)
(A6) LOC wymaganych do dodania nowego program automatycznego rozpoznawania emocji	87–158	87–158	Komunikacja z systemem komunikacji (ang. message broker) – jeden dla każdej technologii (możliwe ponowne wykorzystanie)
	33–111	33–111	Interfejs do kontroli przepływu – jeden dla każdej technologii (możliwe ponowne wykorzystanie)
	162	162	Klasa opakowująca eksperta (program automatycznego rozpoznawania emocji) – jeden dla każdej technologii (możliwe ponowne wykorzystanie)
	18	18	Zmiana parametrów – jeden dla każdej technologii (możliwe ponowne wykorzystanie)

Dzięki wykorzystaniu paradygmatów odwrócenia sterowania (ang. inversion of control) oraz wstrzykiwania zależności (ang. dependency injection) – wywoływane programy są definiowane w zewnętrznym pliku konfiguracyjnym i uruchamiane poprzez wstrzyknięcie kodu. Pozwoliło to na osiągnięcie małej liczbie zmian w kodzie – co jest widoczne w wartościach metryk – wymagana zmiana jednej linii kodu lub jednej klasy. W eksperymencie nie zauważono różnic między rozwiązaniem opartym o wzorzec Blackboard oraz Scoreboard.

W zaproponowanym rozwiązaniu wszystkie moduły muszą być zdefiniowane w pliku konfiguracyjnym z następującymi parametrami:

- **Services** – definiującym funkcję klasy – *ekspert*, *moderator* lub *arbiter*,
- **To** – definiujące konkretną klasę, która ma zostać wywołana,
- **Type** – typ wywołania – jedynie dla modułów ekspertów, czy wywołanie jest obligatoryjne, czy opcjonalne.

Każdy plik konfiguracyjny wymaga zdefiniowanie co najmniej jednego *eksperta* – dla wzorca **Blackboard** oraz dodatkowo, jedną klasę *moderatora* oraz *arbitra* dla wzorca **Scoreboard**. To podejście reprezentuje paradygmat późnego wiązania (ang. late binding paradigm) i jest preferowane dla kryterium adaptowalności.

Jeżeli program automatycznego rozpoznawania emocji jest wywołany zewnętrznie (został napisany w innej technologii, jest wywoływany poprzez API lub z poziomu linii komend), wymagane jest napisanie dla niego funkcji opakowującej, której statystyki są opisane w metryce A6. Informuje ona o bezwzględnej liczbie linii kodu wymaganej do napisania – bez odniesienia do linii kodu całego rozwiązania. Z tego też powodu, metryki A7 i A8 prezentują liczbę linii kodu w odniesieniu do całego rozwiązania (odpowiednio: narzut komunikacji i sterowania) dla programów automatycznego rozpoznawania emocji poprzez analizę ładunku emocjonalnego z tekstu. Zestawienie to zostało zaprezentowane w Tabeli 6.5. Dane w tabeli dotyczą pojedynczych programów (*ekspertów*), nie sumy linii kodu całego rozwiązania.

Tabela 6.5 Miary adaptowalności (narzuty – A7 i A8) dla eksperymentu z programami do analizy ładunku emocjonalnego z tekstu. Na podstawie [4].

Program automatycznego rozpoznawania emocji	Rozmiar kodu [LOC]			(A7) Narzut kontroli [%]	(A8) Narzut komunikacji [%]
	Suma	Komunikacja	Przepływ sterowania		
S 2.0 (C++)	608	158	55	9,1	26,0
S 2.0.1 (C++)	800	158	55	6,9	19,8
S 2.1 (C++)	713	158	55	7,7	22,2
S 2.2 (C++)	799	158	55	6,9	19,8
S 2.2.1 (C++)	806	158	55	6,8	19,6
S 2.2.2 (C++)	841	158	55	6,5	18,8
S 2.3 (C++)	908	158	55	6,1	17,4

Wyniki wskazują na narzut kodu odpowiedzialnego za kontrolę przepływu na poziomie 6–9% oraz narzut komunikacyjny na poziomie 17–26% kodu programu. Warto jednak zauważyć, że narzut ten jest jednorazowy i nie wymaga zmiany podczas modyfikacji logiki działania programu (stała liczba linii kodu). Dalszy ciąg analizy miary adaptowalności został przeprowadzony w studium przypadku Monitora Emocji.

Wyniki eksperymentu I – Odporność

Dla oceny **odporności** wykorzystano metrykę (R1) skuteczności rozpoznawania uzgodnionego rozwiązania. Tabela 6.6 oraz 6.7 przedstawiają dwa warianty doboru programów automatycznego rozpoznawania emocji, w których trzy z nich są „dobre” oraz jeden „słabszy”.

Tabela 6.6 Wariant I analizy odporności – trzy „dobre” i jeden „słaby” program. Skuteczność rozpoznawania emocji w podziale na zbiory danych z eksperymentu. Na podstawie [4].

Program automatycznego rozpoznawania emocji	Skuteczność w każdym zbiorze (R1)					
	Zbiór 1	Zbiór 2	Zbiór 3	Zbiór 4	Zbiór 5	Zbiór 6
S 2.0	72,16	71,64	65,70	72,24	71,49	70,98
S 2.1	71,87	69,03	68,09	74,00	67,66	70,82
S 2.2	71,87	74,06	68,42	72,64	72,04	70,10
S 2.3	67,61	63,95	74,33	68,64	70,00	67,63
Uzgodnione rozwiązanie	73,09	74,22	72,17	73,10	72,20	70,89

Wartym zauważenia jest, że skuteczności poszczególnych programów są zbliżone do siebie. Przykładowo dla wariantu I (Tabela 6.6), „słabszy” program z pierwszego zbioru (S2.3) – jest najskuteczniejszym dla zbioru trzeciego. Dla zbiorów 1, 2, 5 oraz 6 rozwiązanie zintegrowane jest skuteczniejsze lub tak samo skuteczne, jak najlepszy z programów. Dla zbiorów 3 oraz 4 zintegrowane rozwiązanie jest mniej skuteczne od najskuteczniejszego, ale wciąż lepsze od trzech pozostałych.

Tabela 6.7 Scenariusz II analizy odporności – trzy „dobre” i jeden „słaby” program. Skuteczność rozpoznawania emocji w podziale na zbiory danych z eksperymentu. Na podstawie [4].

Program automatycznego rozpoznawania emocji	Skuteczność w każdym zbiorze (R1)					
	Zbiór 1	Zbiór 2	Zbiór 3	Zbiór 4	Zbiór 5	Zbiór 6
S 2.2	71,87	74,06	68,42	72,64	72,04	70,10
S 2.2.1	73,42	73,39	68,42	74,28	69,12	70,10
S 2.2.2	70,23	Na	75,00	64,27	84,58	70,04
S 2.3	67,61	63,95	74,33	68,64	70,00	67,63
Uzgodnione rozwiązanie	73,27	73,22	74,99	73,67	77,80	70,42

Dla drugiego wariantu doboru programów (Tabela 6.7), dla zbiorów 1, 3 oraz 6 rozwiązanie zintegrowane jest skuteczniejsze lub tak samo skuteczne, jak najlepszy z programów. Dla zbiorów 2, 4 oraz 5 zintegrowane rozwiązanie jest mniej skuteczne od najskuteczniejszego, ale wciąż lepsze od trzech pozostałych wersji programu. Wartym zauważenia jest, że w przypadku braku danych dla jednego z programów (zbiór drugi), rozwiązanie jest w stanie zwrócić uzgodnione rozwiązanie. Pokazuje to zdolność rozwiązania do zwracania wyniku w przypadku tymczasowej niedostępności jednego z programów.

Ze względu na małe różnice pomiędzy wskazaniem „dobrego” oraz „słabego” programu rozpoznawania została przygotowana symulacja, w której różnice między wskazaniem są wyraźniejsze. Szczegóły symulacji opisano w kolejnej sekcji.

6.3. Symulacja 1. – weryfikacja miary odporności na podstawie atrap programów rozpoznawania (ang. mock-up)

Plan symulacji 1

Symulacja 1 została przeprowadzona ze względu na ograniczenia eksperymentu 1, w którym to wynik miary jakości odporności, nie był jednoznaczny dla scenariusza: silny–słaby program automatycznego rozpoznawania emocji.

Symulacja wykorzystuje atrapy programów (ang. mock-up) automatycznego rozpoznawania emocji. Zwracają one losowy wynik z określonego przedziału. Ze względu na to, że metoda integracji wspiera odpowiedzi spójne (gdyż prawdziwy stan emocjonalny nie jest znany), silne programy generują quasi–spójne wyniki. Dla dobrych programów odpowiedź zwracana jest z przedziału (0,2–0,8), a dla niepoprawnych z przedziału (–1,0).

Zmienną zależną symulacji jest architektura rozwiązania (Scoreboard oraz Blackboard), a zmienną niezależną jest **odporność** rozwiązania. Ze względu na brak znajomości poprawnej odpowiedzi (wykorzystanie atrap) nie można skorzystać z metryki (R1), dlatego autor pracy wykorzystał dwie inne metryki dla odporności:

- (R2) – liczba wywołań programu,
- (R3) – wynik programu uzyskany po wszystkich etapach.

Wyniki symulacji 1 – Odporność

W symulacji wykonano 100 iteracji – po każdej z nich następowała ocena oraz decyzja o wyborze programów do uruchomienia. Wadliwy program został szybko wykryty i otrzymał wynik 0,46 wobec wyników rzędu 0,99 dla pozostałych rozwiązań. Moduł moderatora wybrał go tylko 42 razy na 100 iteracji, gdy inne rozwiązania były wywoływane prawie za każdym razem (od 98 do 100 razy na 100).

Wartości metryk R2 (liczba wywołań) oraz R3 (wynik osiągnięty przez program) z symulacji zaprezentowano w Tabeli 6.8.

Tabela 6.8 Miary odporności dla symulacji ze atrapami (ang. mock-up) ekspertów. Na podstawie [4].

Nazwa	Typ programu	(R2) Liczba uruchomień (na 100)	(R3) Przyznana ocena
Program 1	Silny	98	0,989
Program 2	Silny	99	0,993
Program 3	Silny	98	0,993
Program 4	Referencyjny	100	0,990
Program 5	Słaby	42	0,465

Funkcja integracji bazuje w dużej mierze na spójności wyników, co generuje ograniczenia tej metody. Przykładowo, jeden dużo skuteczniejszy program może zostać przegłosowany przez kilka mniej skutecznych, ale spójnych programów. Można rozważyć zmianę funkcji integracji oraz oceny – co w przypadku rozwiązania Monitora Emocji wiąże się ze zmianą pojedynczej klasy.

Spostrzeżenia związane z odpornością można podsumować następująco:

- Uzgodnione rozwiązanie jest nieco skuteczniejsze od jednokanałowej analizy przy założeniu, że liczba „silnych” oraz „słabych” programów jest zbalansowana,
- Możliwe jest automatyczne wykrycie programów o niskiej skuteczności.

Powyższe obserwacje pozwalają przyjąć prawdziwość tezy, że zaproponowana architektura zapewnia pewną odporność na niską skuteczność rozpoznawania emocji poszczególnych modułów ekspertów.

6.4. *Studium przypadku Monitora Emocji*

Plan studium przypadku Monitora Emocji

Do studium przypadku zostało wykorzystane stanowisko Monitora Emocji na Politechnice Gdańskiej – opisane w rozdziale 5.3. Wykorzystano metodę ATAM do oceny jakościowej wzorca architektonicznego. Metoda ta została wybrana ze względu na możliwość uwzględnienia wielu atrybutów jakości i wielu architektonicznych podejść, a także przeprowadzenia analizy kompromisów (ang. trade-off).

Wyniki badań ilościowych – Modularność

Studium przypadku Monitora emocji pozwala na analizę złożonego, rzeczywistego rozwiązania. Opracowana architektura pozwala na:

- separację programów automatycznego rozpoznawania emocji od przepływu sterowania oraz komunikacji,
- niezależność programów automatycznego rozpoznawania emocji od pozostałych programów – nie ma potrzeby wiedzy o istnieniu innych modułów ani implementacji komunikacji między modułami w obrębie poszczególnych modułów (w szczególności *ekspertów*),
- separacji funkcji integracji od reszty kodu (w klasie *Moderatora*),
- separacji funkcji oceny modułów (*Ekspertów*) od reszty kodu (w klasie *Arbitra*).

W studium przypadku Monitora Emocji zmierzono wartości miar **modularyzacji**, które zestawiono w Tabeli 6.9. Ze względu na duże zróżnicowanie wartości, poza metryką (M1), przedstawiono wartości średnie, odchylenie standardowe (SD) oraz wartość maksymalną (max). Dla wszystkich metryk – im wartość jest niższa, tym lepiej. Architektura cechująca się modularnością ma niską poziomem powiązań (ang. coupling) (M6–M7). Miary te powinny mieć wartości poniżej 10, co zostało spełnione dla średniego przypadku, a jedynie dwie spośród 25 klas mają wartość wyższą niż 20. Maksymalny poziom powiązań (ang. coupling) klas (M6) ma klasa odpowiedzialna za komunikację, używająca wielu zewnętrznych bibliotek. 19 z 25 klas ma tylko jedną lub całkowity brak zależności.

Średnia liczba metod na klasę (M2) wyniosła 5,52, podczas gdy LOC na metodę (M3) to 22, przy wartości maksymalnej 188. Głębokość dziedziczenia (M5) nie przekracza wartości 4.



Tabela 6.9 Miary modularyzacji dla studium przypadku. Na podstawie [4].

Miary modularyzacji	Miara statystyczna	Emotion Monitor
(M1) unikalna liczba klas	N	25,00
(M2) liczba metod na klasę	Średnia (SD)	5,52 (6,86)
	Max	27,00
(M3) LOC na metodę	Średnia (SD)	22,16 (40,00)
	Max	188,00
(M4) złożoność cyklometryczna metod	Średnia (SD)	2,01 (3,55)
	Max	32,00
(M5) głębokość dziedziczenia	Średnia (SD)	2,64 (1,2)
	Max	4,00
(M6) poziom powiązań klas	Średnia (SD)	6,2 (6,97)
	Max	28
(M7) poziom powiązań metod	Średnia (SD)	1,88 (3,15)
	Max	21

Podsumowując powyższą analizę, rozwiązanie wykorzystujące wzorzec architektoniczny Scoreboard cechuje się modularnością i jest ona podobna do modularności rozwiązań opartych na wzorcu architektonicznym Blackboard.

Wyniki badań ilościowych – Adaptowalność

W przeprowadzonym eksperymencie wszystkie wykorzystane programy zostały zaimplementowane w C++. W studium przypadku Monitora emocji wykorzystano więcej rozwiązań, które były zaimplementowane w innych technologiach, takich jak Java czy C#. W Tabeli 6.10 przedstawiono miary adaptowalności ze studium przypadku Monitora Emocji dla różnych technologii: „S 2.3” program z eksperymentu z analizą danych ładunku emocjonalnego tekstu, Sam1 i Sam2 napisane w Java, FERT i FERT2 bazujące na rozpoznawaniu mimicznej ekspresji oraz dwóch zewnętrznych rozwiązaniach: FR – Noldus Face Reader – z wykorzystaniem API oraz XE – uruchamianym z linii komend. Wszystkie rozwiązania zostały z powodzeniem zintegrowane z Monitorem Emocji.

Tabela 6.10 Miary adaptowalności (narzuty – A7 i A8) dla studium przypadku Monitora emocji (zróżnicowane technologie). Na podstawie [4].

Program automatycznego rozpoznawania emocji	Rozmiar kodu [LOC]			(A7) Narzut kontroli [%]	(A8) Narzut komunikacji [%]
	Suma	Komunikacja	Przepływ sterowania		
S 2.3 (C++)	908	158	55	6,1	17,4
Sam1 (Java)	424	92	96	22,6	21,7
Sam2 (Java)	269	92	96	35,7	34,2
Sam3 (C#)	330	87	111	33,6	26,4
FERT (C#)	242	87	33	13,6	36,0
FERT2 (C#)	207	87	33	15,9	42,0
FR (C#)	413	87	45	10,9	21,1
XE (C#)	321	87	33	10,3	27,1

Narzut sterowania wahał się od 6 do 35%, a narzut komunikacji od 17 do 42%. Wykorzystane programy były dosyć proste, dlatego ich narzut w niektórych przypadkach może wydawać się znaczny – zwłaszcza dla rozwiązań o zamkniętym kodzie, gdzie wykorzystano jedynie klasę opakowującą (nie znamy rozmiaru kodu odpowiedzialnego za rozpoznawanie). Warty uwagi jest, że wartości liczby linii kodu są stałe lub zbliżone dla jednej technologii, a także możliwe do ponownego użycia – co poprawia skalowalność rozwiązania.

Powyższe analizy dotyczące adaptowalności można podsumować następująco:

- w zaproponowanej metodzie integracji, programy automatycznego rozpoznawania emocji, można w łatwy sposób podłączać, odłączać oraz modyfikować bez wpływu na pozostałe komponenty rozwiązania,
- złożoność adaptacji integracji programu automatycznego rozpoznawania emocji zależy głównie od technologii, w jakiej został napisany,
- funkcje integracji oraz oceny są zastępowalne bez wpływu na programy automatycznego rozpoznawania emocji – pełna izolacja.

Przedstawione dane pozwalają przyjąć tezę, że wybrana architektura rozwiązania jest adaptowalna. Adaptowalność architektury wykorzystującej wzorzec Scoreboard jest na tym samym poziomie co wykorzystująca wzorzec Blackboard, której użyto jako referencyjnej.

Wynik badań jakościowych – Analiza ATAM

Oprócz analizy ilościowej została przeprowadzona szczegółowa jakościowa analiza studium przypadku Monitora Emocji przy pomocy metody ATAM. Metoda składa się z trzech kroków:

1. Zdefiniowania scenariuszy do spełnienia przez rozwiązanie,
2. Zdefiniowanie podejść architektonicznych, które stanowią rozwiązanie,
3. Powiązanie podejścia architektonicznego ze scenariuszami w celu identyfikacji obsłużonych sytuacji (ang. non-risk), sytuacji potencjalnego zagrożenia (ang. risk) oraz wrażliwych punktów (ang. trade-offs).

Scenariusze dla Monitora Emocji zostały zaprezentowane w Tabeli 6.11. Zostały one przyporządkowane do wcześniej zdefiniowanych miar jakości w celu powiązania z analizą ilościową.

Tabela 6.11 Scenariusze ATAM dla studium przypadku Monitora Emocji. Na podstawie [4].

Kryteria	Scenariusz	Opis
Modularyzacja	Sc 1. Separacja ekspertów od sterowania	Programy automatycznego rozpoznawania emocji są niezależne od przepływu sterowania oraz procesu integracji.
	Sc 2. Separacja przepływu sterowania i integracji	Przebieg sterowania i moduł integracji są odseparowane od siebie oraz od innych modułów.
	Sc 3. Separacja ekspertów	Programy automatycznego rozpoznawania emocji są niezależne od siebie nawzajem – nie komunikują się ze sobą bezpośrednio.
Integracja	Sc 1. Integracja	Jest możliwa integracja dowolnej liczby programów automatycznego rozpoznawania emocji, różniących się modalnościami oraz wykorzystanymi algorytmami.
	Sc 2. Technologie	Możliwa jest integracja rozwiązania zaimplementowanych w różnych technologiach.
	Sc 3. Gotowe rozwiązania komercyjne (ang. off-the-shelf)	Możliwe jest integrowanie rozwiązania gotowego, pod warunkiem, że może zostać uruchomione z linii poleceń lub z API.
	Sc 4. Wczesna fuzja w wielomodalnym rozpoznawaniu emocji	Wszystkie kanały wejściowe są dostępne dla wszystkich ekspertów. Wczesna fuzja jest możliwa do implementacji w wielomodalnym rozpoznawaniu emocji.
	Sc 5. Późna fuzja w wielomodalnym rozpoznawaniu emocji	Wszystkie cząstkowe wyniki każdego z ekspertów są dostępne dla pozostałych. Późna fuzja jest możliwa do implementacji w wielomodalnym rozpoznawaniu emocji.
	Sc 6. Wielofazowość (ang. multistage)	Jest możliwe zdefiniowanie faz, w których wykonywane są wybrane moduły ekspertów np. przygotowanie danych, rozpoznawanie emocji, integracja, odwzorowanie modeli danych lub dowolna inna faza.
	Sc 7. Niezgodność	Niezgodność wyników programów automatycznego rozpoznawania emocji została zaadresowana w rozwiązaniu.
Adaptowalność	Sc 1. Podłączenie programu automatycznego rozpoznawania emocji	Moduły ekspertów są łatwo dołączane do rozwiązania. Brak konieczności rekompilacji rozwiązania.
	Sc 2. Odłączenie programu automatycznego rozpoznawania emocji	Moduły ekspertów są łatwo odłączane od rozwiązania. Brak konieczności rekompilacji rozwiązania.
	Sc 3. Zmiana programu automatycznego rozpoznawania emocji	Zmiana wewnętrznej implementacji programu eksperta nie wymaga zmian w innych modułach rozwiązania. Nie wymaga rekompilacji całego rozwiązania.
	Sc 4. Zmiana funkcji integracji	Funkcja integracji może być łatwo modyfikowana bez wpływu na pozostałe moduły. Nie wymaga rekompilacji całego rozwiązania.
	Sc 5. Łatwa modyfikowalność sekwencji wywołania programów	Zmiana listy programów rozpoznawania do wywołania może być zmieniana bez rekompilacji rozwiązania.
Odporność	Sc 1. Niska skuteczność	Możliwe jest wykrycie programu o niskiej skuteczności w czasie działania programu.
	Sc 2. Wpływ na rozwiązanie	Ograniczony jest wpływ rozwiązań o niskiej skuteczności na uzgodnione rozwiązanie.
	Sc 3. Zaufanie	Współczynnik zaufania do programu automatycznego rozpoznawania emocji jest budowany w trakcie działania rozwiązania.

Drugim krokiem metody ATAM jest zidentyfikowanie kluczowych architektonicznych założeń, które zostały zestawione w Tabeli 6.12. Zdefiniowano 14 założeń: pierwsze 5 (AA1–AA5) charakteryzuje oba wzorce architektoniczne (Blackboard oraz Scoreboard), kolejne dwa (AA6, AA7) opisują tylko wzorzec Blackboard, a pozostałe (AA8–AA14) są przypisane tylko do wzorca Scoreboard.

Tabela 6.12 Podejścia architektoniczne ATAM dla studium przypadku Monitora Emocji. Na podstawie [4].

Założenia architektoniczne	Blackboard	Scoreboard	Opis
AA1	X	X	Pojedyncza, odseparowana struktura do współdzielenia wyników pomiędzy ekspertami (Klasa <i>Tablicy</i>).
AA2	X	X	Pojedyncza klasa kontrolera, odseparowana od klas ekspertów.
AA3	X	X	Wielu ekspertów wywołujących te same zadania (rozpoznawanie emocji) z różnymi algorytmami. Podejście czarnej skrzynki (ang. <i>blackbox</i>) jest możliwe do realizacji.
AA4	X	X	<i>Eksperci</i> (programy automatycznego rozpoznawania emocji) komunikują się z <i>Tablicą</i> oraz są wywoływani przez <i>Kontrolera</i> . <i>Eksperci</i> nie komunikują się ze sobą bezpośrednio.
AA5	X	X	Programy zwracają wyniki jako parę: wartość stanu emocjonalnego oraz poziom współczynnika wiarygodności.
AA6	X		Wszyscy eksperci są wywoływani w każdej iteracji.
AA7	X		Losowa lub zdefiniowana regułą kolejność wywoływania ekspertów.
AA8		X	<i>Eksperci</i> mają możliwość oceny innych ekspertów (klasa <i>moderatora</i> , będąca rozszerzeniem klasy <i>Eksperta</i>).
AA9		X	Odseparowana struktura do przechowywania ocen ekspertów (klasa <i>Tabeli wyników</i>).
AA10		X	Grupowanie opcjonalnych ekspertów gotowych do uruchomienia i uruchomienie wybranych z nich na podstawie ocen.
AA11		X	Wydzielona klasa <i>moderatora</i> odpowiedzialna za ocenę i uspołnienie wyników w danej fazie.
AA12		X	Wykorzystanie paradygmatu odwrócenia sterowania i wstrzyknięcia zależności w klasie <i>kontrolera</i> .
AA13		X	Wykorzystanie zewnętrznych plików konfiguracyjnych z definicją listy <i>ekspertów</i> do uruchomienia.
AA14		X	Wykorzystanie klas opakowujących (ang. <i>wrappers</i>) dla zewnętrznych programów i modułów napisanych w innych technologiach.

Ostatnim krokiem metody ATAM jest powiązanie scenariuszy z założeniami architektury, które zostało przedstawione w Tabeli 6.13. Znak + oznacza brak ryzyka (ang. no risk), – oznacza ryzyko (ang. risk), a znakiem * oznaczone zostały „czułe punkty”. Jeżeli czułe punkty dotyczą więcej niż jednego kryterium – uznawane są za kompromisy (ang. trade-off).

Analiza powiązań pozwala na identyfikację wielu konsekwencji wykorzystania wzorca Scoreboard. Pierwszym z nich jest wsparcie dla **modularności** oraz **adaptowalności** – wsparcie dla wszystkich scenariuszy – zarówno dla wzorca architektonicznego Blackboard, jak i Scoreboard. Dla scenariuszy **integracji** przejście ze wzorca Blackboard na Scoreboard zaowocowało kilkoma korzyściami, między innymi w postaci lepszego wsparcia późnej fuzji oraz poradzenia sobie z niezgodnością, ale także lepsze wsparcie dla różnych technologii czy wykorzystania istniejących rozwiązań. Najwięcej zalet z przejścia na architekturę Scoreboard zauważono dla miary jakości **odporności** – wzorzec Scoreboard obsługuje wszystkie scenariusze, kiedy wzorzec Blackboard część z nich.

Tabela 6.13 Powiązanie scenariuszy ATAM z podejściami ATAM zaprojektowanymi dla studium przypadku Monitora Emocji. Na podstawie [4].

Criteria	Scenario	Architectural approaches													
		01	02	03	04	05	06	07	08	09	10	11	12	13	14
Modularyzacja	Sc 1. Separacja ekspertów od sterowania		+							+					
	Sc 2. Separacja przepływu sterowania i integracji		+							+					
	Sc 3. Separacja ekspertów	+		+	+					+		+			
Integracja	Sc 1. Integracja	+		+		+	+								+
	Sc 2. Technologie	*		+	+										+
	Sc 3. Gotowe rozwiązania komercyjne (ang. off-the-shelf)	*		+	+										+
	Sc 4. Wczesna fuzja w wielomodalnym rozpoznawaniu u emocji	+		*	*										
	Sc 5. Późna fuzja w wielomodalnym rozpoznawaniu u emocji				-	+						+			
	Sc 6. Wielofazowość (ang. multistage)	+		*	*			+							
	Sc 7. Niezgodność					+	-	-		+		+			
Adaptowalność	Sc 1. Podłączenie programu automatycznego rozpoznawania emocji		+										+	+	
	Sc 2. Odłączenie programu automatycznego rozpoznawania emocji		+										+	+	
	Sc 3. Zmiana programu automatycznego rozpoznawania emocji		+	+									+	+	+
	Sc 4. Zmiana funkcji integracji	+											+	+	
	Sc 5. Łatwa modyfikowalność sekwencji wywołania programów		+		+		+	+					+	+	
Odporność	Sc 1. Niska skuteczność						-	-	+			+			
	Sc 2. Wpływ		+				-	-	+			+			
	Sc 3. Zaufanie			*			-	-	+	+	+	+			

Jakościowa analiza stadium przypadku Monitora Emocji wspiera większość obserwacji z ilościowego eksperymentu oraz symulacji.

6.5. Podsumowanie wyników

W celu ewaluacji architektury zaproponowanego rozwiązania przeprowadzono eksperyment, symulację oraz studium przypadku Monitora Emocji. Ewaluację można podsumować następująco:

1. Rozwiązania bazujące na wzorcu Scoreboard są modułarne w sposób porównywalny ze wzorcem Blackboard, który został użyty jako wzorzec referencyjny. Możliwe jest wykorzystanie wielu różnych programów automatycznego rozpoznawania emocji dla tej samej fazy przetwarzania, równoległe uruchomienie programów oraz porównanie wyników.

2. Wykorzystanie wzorca Scoreboard poprawia skuteczność rozwiązania. Uzgodnione rozwiązanie w większości przypadków wykazuje nieco wyższą skuteczność niż rozwiązanie jednokanałowe oraz oparte na wzorcu Blackboard przy założeniu, że liczba programów „silnych” oraz „słabych” jest odpowiednio zbalansowana, a rozwiązania „silne” są zbieżne. Wzorzec Scoreboard wspiera zarówno metodę wczesnej, późnej, jak i hybrydowej fuzji.
3. Rozwiązanie oparte o wzorzec Scoreboard cechuje się adaptowalnością nie gorszą niż w przypadku wzorca Blackboard. Programy automatycznego rozpoznawania emocji są łatwo podłączane, odłączane oraz modyfikowane bez wpływu na pozostałe moduły. Złożoność adaptacji nowych programów zależy głównie od wybranej technologii. Dodatkowo modyfikacja funkcji integracji oraz oceny nie wymaga modyfikacji pozostałych modułów – zmiana dotyczy tylko jednej klasy.
4. Wiele programów może być ocenianych dynamicznie w czasie działania (ang. run-time). Możliwość oceny oraz adaptowalność do zmieniających się okoliczności (takich jak dostępność kanału, rozpoznawanie emocji w zależności od kontekstu) są wbudowane we wzorzec Scoreboard, który wspiera ciągłą ocenę programów. Rozwiązanie to pomaga budować zaufanie do wyniku rozwiązania i wykorzystywanych komponentów.
5. Wzorzec Scoreboard cechuje się odpornością na czasową niedostępność kanałów obserwacji oraz programów automatycznego rozpoznawania emocji. Możliwe jest automatyczne wykrycie programów o niskiej skuteczności i dostosowanie częstości ich wywołania w czasie działania aplikacji (dynamicznie). Odporność ta nie występuje w referencyjnym wzorcu Blackboard.

Dane ilościowe oraz jakościowe obserwacje pozwalają na zaakceptowanie pierwszej tezy pomocniczej mówiącej, że rozwiązanie oparte o wzorzec Scoreboard zapewnia modularyzację, łatwość integracji, odporność oraz adaptowalność dla rozwiązań automatycznego rozpoznawania emocji użytkownika komputera.

7. EWALUACJA ZAPROPONOWANEJ METODY KWANTYFIKACJI

INFORMACJI UMOŻLIWIAJĄCEJ REPREZENTACJĘ NIEPEWNOŚCI

W celu udowodnienia drugiej tezy pomocniczych autor pracy wykorzystał ogólnodostępny i uznany zbiór danych zaetykietowanych emocjonalnie SEMAINE [34] [150]. Zawiera ona dane wielomodalne i wielokanałowe z przypisanymi przez ekspertów stanami emocjonalnymi w ciągłym modelu reprezentacji emocji. Na danych z SEMAINE wykonano dogłębną analizę, która została opisana w rozdziale 7.4.

Druga teza pomocnicza została sformułowana następująco:

Zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwia reprezentację niepewności oszacowania stanu emocjonalnego

W celu jej udowodnienia przeprowadzono dwa eksperymenty:

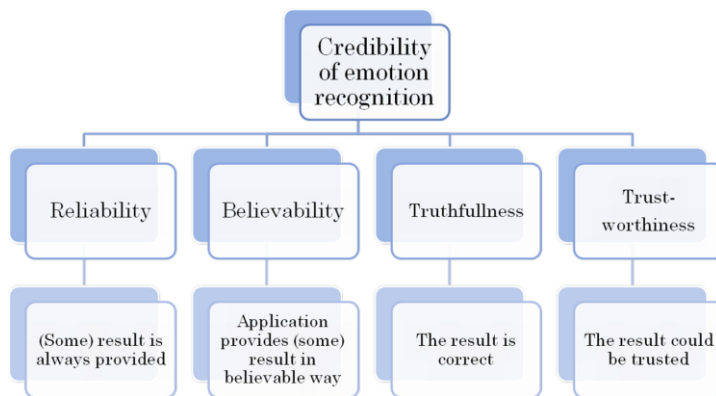
- Eksperyment 2. Walidacja metody wyznaczania współczynnika wiarygodności oraz
- Eksperyment 3. Dobór współczynnika wiarygodności w przypadku jego braku.

Niepewność jest to stan analizy, w którym nie można przewidzieć stanu zjawiska, ze względu na zmienność jego samego lub brak wiedzy czy informacji. Zgodnie z założeniami przyjętymi w filtrze Kalmana, każdy pomiar wiąże się z niepewnością [22] [123]. W języku technicznym stosuje się wiele terminów w celu określenia problemów związanych z niepewnością i wiarygodnością pomiaru m.in.: niedokładność (ang. inaccuracy), nieprecyzyjność (ang. imprecision), błąd oraz ich przeciwieństwa: dokładność (ang. accuracy), pewność (ang. confidence), wiarygodność (ang. plausibility), zaufanie (ang. trust) [20] itp. Jakość wyniku zazwyczaj określana jest miarą skuteczności, taką jak: precyzja (ang. precision) czy dokładność (ang. accuracy).

W przypadku niepewności związanej z rozpoznawaniem emocji niekiedy korzystniejsza dla danego obszaru zastosowań byłaby odpowiedź eksperta (programu) „Nie wiem”, zamiast podawania niepewnego wyniku.

W dziedzinie rozpoznawania emocji jednym z kluczowych ryzyk jest uznanie niepewnego wyniku za wiarygodny (ryzyko wiarygodności). Jest to szczególnie istotne ze względu na coraz częstsze wykorzystywanie tej technologii przez praktyków (niebędących ekspertami w dziedzinie), nie zawsze znających ograniczenia i założenia wykorzystywanych metod [9]. Z tego powodu, użytkownicy takich systemów nie są świadomi niepewności pomiaru i przyjmują rezultat jako pewny i wiarygodny.

Ryzyko wiarygodności można ograniczyć poprzez podanie wyniku (oszacowanego stanu emocjonalnego) wraz z pewnością tego wyniku. W niniejszej pracy miara pewności wyniku jest określana w uproszczeniu jako jego wiarygodnością. Dekompozycja wiarygodności w rozpoznawaniu emocji została przedstawiona na Rys 7.1.



Rys 7.1 Dekompozycja wiarygodności w rozpoznawaniu emocji [20]

7.1. **Konwencje notacyjne**

Ze względu na zaobserwowaną złożoność opisywanych eksperymentów, autor pracy podjął próbę uproszczenia analizy poprzez wprowadzenie konwencji notacyjnej dla tabel z wynikami eksperymentów w rozdziałach 7 oraz 8.

Jedną z zaobserwowanych trudności jest posługiwanie się metryką zwiększenia poprawności współczynnika wiarygodności.

Drugą trudnością jest rozróżnienie pomiędzy współczynnikiem wiarygodności dla oszacowanego stanu emocjonalnego pochodzącego od pojedynczego eksperta (programu automatycznego rozpoznawania emocji) a współczynnikiem wiarygodności końcowej hipotezy o stanie emocjonalnym będącej wynikiem funkcji integracji.

1. Kolorowanie komórek

Dla lepszej czytelności wartości procentowych w tabelach zostały one pokolorowane w trzystopniowej skali: od barwy czerwonej dla wartości równej 0% przez żółtą, dla wartości 50%, aż do barwy zielonej dla wartości 100%.

2. Tabele porównujące skuteczność

Tabele, w których zmienną zależną jest skuteczności automatycznego rozpoznawania emocji, utworzone są w następujący sposób:

- w kolejnych wierszach pokazano rezultat dla różnych rozwiązań (jednomodalnych oraz wielomodalnych) oraz pozostałych zmiennych takich jak wartość progowa odległości (dla której hipoteza jest uznana za zbieżną),
- w kolumnach prezentowane są wymiary modelu reprezentacji emocji (emocje), wraz z uśrednionym wynikiem na końcu tabeli (General),
- w komórkach znajduje się skuteczność rozpoznawania emocji, wyrażona w procentach, liczona jako liczba przypadków poprawnego rozpoznania (w ramach zadanego progu) podzielona przez liczbę wszystkich przypadków.

Poprawne rozwiązanie rozumiane jest jako odpowiedź znajdująca się w odległości nie większej niż progowa wartość odległości od wartości wskazanej przez ekspertów.

3. Tabele do weryfikacji poprawności wyznaczania współczynnika wiarygodności przez funkcję integracji

Tabele, w których zmienną zależną jest zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności, utworzone są w następujący sposób:

- w kolejnych wierszach pokazano rezultat dla różnych rozwiązań (jednomodalnych oraz wielomodalnych) oraz pozostałych zmiennych takich jak wartość progowa odległości (dla której hipoteza jest uznana za zbieżną),
- w kolumnach przedstawione są przedziały wynikowego współczynnika wiarygodności, zwróconego przez funkcję integracji,
- w komórkach znajduje się skuteczność rozpoznawania emocji, wyrażona w procentach, liczona jako liczba przypadków poprawnego rozpoznania (w ramach zadanego proggu), podzielona przez liczbę wszystkich przypadków, a także liczba przypadków.

Puste pola w kolumnie oznaczają brak wskazań wynikowych hipotez w danym przedziale współczynnika wiarygodności ($n = 0$).

4. Przedziały współczynnika wiarygodności

W tabelach do weryfikacji poprawności wyznaczania współczynnika wiarygodności, w kolumnach znajdują się przedziały wynikowego współczynnika wiarygodności uzgodnionego rozwiązania. Pożądanym efektem jest wyższy poziom skuteczności w wyższych przedziałach wiarygodności (bliżej 100%), a spadek współczynnika wraz z przechodzeniem do przedziałów bliższych zeru. Dane zostały podzielone na 9 przedziałów współczynnika wiarygodności uzgodnionego rozwiązania:

- od 0 do 20, oznaczony dalej jako „0–20”,
- od 20 do 40%, oznaczony dalej jako „20–40”,
- od 40 do 50%, oznaczony dalej jako „40–50”,
- od 50 do 60%, oznaczony dalej jako „50–60”,
- od 60 do 70%, oznaczony dalej jako „60–70”,
- od 70 do 80%, oznaczony dalej jako „70–80”,
- od 80 do 90%, oznaczony dalej jako „80–90”,
- od 90 do 95%, oznaczony dalej jako „90–95” oraz
- od 95 do 100%, oznaczony dalej jako „95–100”.

Przedziały bliższe wartości maksymalnej („90–95” oraz „95–100”) są mniejsze, ze względu na chęć dokładniejszej analizy wysoko ocenionych przypadków, z kolei przedziały poniżej 40% są większe (0–20, 20–40), gdyż ich istotność dla dziedziny jest ograniczona.

7.2. Wybór zbioru danych SEMAINE

Ze względu na zmniejszenie ryzyka wyboru niereprezentatywnego zbioru danych, autor pracy zdecydował się na wybór zweryfikowanego i ogólnie dostępnego, zaetykietowanego emocjonalnie przez ekspertów, zbioru SEMAINE [150]. Wyboru tego dokonano ze względu na zaetykietowanie danych m.in. dla modelu Ekmana oraz wymiarów pobudzenia (ang. arousal) i wartościowości (ang. valence). Dane te również pozwalają na analizę w dwóch modalnościach i pięciu kanałach – trzy kamery oraz dwa mikrofony. Analizowane były także inne zbiory danych, m. in. MAHNOB-HCI [151], który został jednak wykluczony ze względu na ścieżkę audio, w której połączone są dźwięki z prezentowanych materiałów wideo z reakcjami użytkownika – co uniemożliwiło wykorzystanie kanału audio do rozpoznawania emocji.

Wykorzystywanymi kanałami w zbiorze danych SEMAINE są dane audio oraz wideo. Dane audio zostały zebrane z dwóch mikrofonów – umieszczonego w pokoju oraz na słuchawkach badanego. Dane wideo zostały zarejestrowane poprzez kamerę przednią – w wersji kolorowej oraz monochromatycznej, a także poprzez kamerę boczną monochromatyczną.

Rozmowy z bazy SEMAINE zostały przeprowadzone zgodnie z paradygmatem SAL (ang. Sensitive Artificial Listener) [152], w którym to interakcja jest podzielona na człowieka (użytkownika) oraz operatora, którym może być człowiek lub maszyna. W tym przypadku obiema stronami byli ludzie, jednakże etykiety emocjonalne dotyczyły głównie użytkowników, dlatego badania w niniejszej pracy zostały ograniczone do symptomów emocji użytkownika.

Do oceny stanów emocjonalnych wykorzystano model pięciowymiarowy: wartość, aktywacja, moc (ang. power), przewidywanie/oczekiwanie (ang. anticipation/expectation) oraz intensywność (ang. intensity) – można więc z niego wyodrębnić wymiary z modelu 3D. Eksperci mieli także możliwość oceny za pomocą innych etykiet, w tym 5 z 6 podstawowych emocji modelu reprezentacji emocji Ekmana, a brakującą emocję zaskoczenia można uzyskać poprzez inwersję z emocji oczekiwania/przewidywania (ang. expectation/anticipation) [153].

Ludzcy eksperci wykorzystywali narzędzie FEEL-trace [154] do oceny stanów emocjonalnych za pomocą myszki komputerowej. Narzędzie to niezależnie od ocenianego wymiaru wykorzystywało skalę od -1 do 1 [155] [156], gdzie wartość -1 oznacza najmniejszą wskazaną wartość, a 1 najwyższą dla danego wymiaru [157]. Wyjątek stanowi wymiar oczekiwania/przewidywania, którego wartości w bazie zostały podane w skali <0,100>. Z tego powodu autor niniejszej pracy sprowadził dane do przedziału <0,1> dla emocji z modelu emocji podstawowych Ekmana, a wartości z modelu PAD pozostały w skali <-1,1>. Narzędzie FEEL-trace pozwoliło na zebranie informacji z dokładnością do 20 ms. Każdy wymiar był oceniany przez ekspertów niezależnie, także liczba ocenających, jak i zestaw ocenianych wymiarów/emocji różni się dla każdej sesji.

7.3. Wykorzystane programy automatycznego rozpoznawania emocji

Do analizy danych wideo został wykorzystany program Noldus FaceReader [50], który zwraca wynik w modelu 2D oraz modelu emocji podstawowych Ekmana. Do każdego wyniku dołączony jest współczynnik jakości danych oraz dane dotycząca ustawienia kamery względem twarzy (kąty na trzech osiach X, Y, Z). Granulacja danych wynosi 20 ms.

Do analizy danych audio wykorzystany został program wytworzony na Politechnice Gdańskiej [158] zwracający estymację stanu emocjonalnego w modelu emocji podstawowych Ekmana bez współczynnika wiarygodności. Przyjętą w programie skalą jest $(0,1>$, a granulacja danych wynosi 1s lub plik (cała wypowiedź).

Mimo tej samej granulacji danych dla Noldus FaceReadera oraz danych dla zaetykietowanego emocjonalnie zbioru danych SEMAINE (20 ms), występują przesunięcia w odpowiedziach Noldus FaceReadera (ok. 10 ms) – z tego też powodu, do analizy przyjęto uśrednione wyniki granulacji 100 ms jako metodę synchronizacji.

7.4. Analiza danych SEMAINE

W celu przeprowadzenia rzetelnej analizy zaproponowanej metody zostały przeanalizowane dane ze zbioru SEMAINE pod kątem poziomu skuteczności rozpoznawania emocji przez wykorzystane programy jednomodalne, jako wartości referencyjne do późniejszej analizy skuteczności rozpoznawania wielomodalnego.

Określenie skuteczności rozpoznawania emocji w kanale audio na zbiorze danych SEMAINE

Dane audio z bazy SEMAINE zostały podzielone na fragmenty odpowiadające werbalnej reakcji użytkownika, a następnie przeanalizowane przez program napisany na Politechnice Gdańskiej [158]. Wyodrębnienie głosu użytkownika zostało dokonane na podstawie zapisu transkrypcji, dlatego też fragmenty były analizowane jako całość – jedna odpowiedź programu oraz uśrednione wyniki z odpowiednich ramek czasowych z bazy SEMAINE. Do analizy wykorzystano dwa kanały audio – z mikrofonu słuchawek oraz mikrofonu z pokoju.

Do analizy wykorzystano 3983 próbki dla każdego kanału. Operacjonalizacja zmiennych została przedstawiona w Tabeli 7.1. Wartości zostały przedstawione w Tabeli 7.2., w podziale na poszczególne wymiary (tu: emocje) oraz zagregowane dla wierszy z etykietą „General”. Zauważono wyższą skuteczność dla mikrofonu zlokalizowanego bliżej użytkownika w porównaniu do danych z pomieszczenia, co może być spowodowane większą ilością zakłóceń w mikrofonie dalej oddalonym od badanego.

Tabela 7.1 Operacjonalizacja zmiennych dla określenia skuteczności rozpoznania kanału audio w bazie SEMAINE [opracowanie własne].

Zmienna niezależna	Lokalizacja mikrofonu: 1. Mikrofon w słuchawkach (ang. user headMounted), 2. Mikrofon w pokoju (ang. user room)
Zmienna zależna	Skuteczność rozpoznawania emocji (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold) (wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Różna liczba próbek dla różnych emocji jest spowodowana brakami w emocjonalnym zaetykietowaniu zbioru SEMAINE dla emocji z modelu Ekmana. Najwięcej etykiet było przypisanych do emocji radości (ang. happy) oraz zaskoczenia (ang. surprise), dla których to poprawność klasyfikacji była najwyższa. Dla wartości progowej odległości równej 0,3 rozpoznanie emocji radości i zaskoczenia przewyższają wartość 60%. Wyniki te wskazują jednoznacznie na istotność różnicowania poziomu zaufania dla różnych wymiarów, które są znacząco różne dla poszczególnych emocji. Niższe zaufanie do pojedynczego wymiaru może też być spowodowane różną liczbą próbek do uczenia (np. w przypadku użycia bazy SEMAINE jako zbiór uczący i testujący dla programu).

Tabela 7.2 Skuteczność (ang. accuracy) rozpoznania emocji poprzez program automatycznego rozpoznawania emocji bazujący na kanale audio z bazy SEMAINE na 3983 próbkach w podziale na poszczególne emocje [opracowanie własne].

		Distance treshold (D_{max})					
Location	Dimension	0,1	0,15	0,2	0,3	0,5	N
User HeadMounted	Happy	22,3%	34,6%	45,6%	67,4%	93,6%	1470
User Room		4,3%	7,4%	12,7%	28,7%	74,9%	
User HeadMounted	Sad	7,1%	12,2%	16,4%	26,2%	60,2%	646
User Room		0,0%	0,0%	0,0%	1,2%	21,4%	
User HeadMounted	Angry	1,9%	4,0%	7,4%	15,4%	48,2%	571
User Room		0,5%	1,4%	2,8%	10,0%	34,9%	
User HeadMounted	Scared	0,0%	0,0%	0,0%	1,9%	18,9%	53
User Room		0,0%	0,0%	0,0%	0,0%	3,8%	
User HeadMounted	Disgusted	4,5%	9,5%	11,4%	17,4%	40,3%	201
User Room		0,5%	1,5%	2,0%	8,0%	22,4%	
User HeadMounted	Surprised	11,5%	20,1%	32,2%	60,0%	98,3%	1042
User Room		12,5%	21,0%	32,4%	60,1%	98,1%	
User HeadMounted	General	12,90%	21,04%	29,53%	47,93%	79,21%	3983
User Room		4,95%	8,51%	13,66%	28,35%	62,94%	

*Oryginalna skala z bazy SEMAINE to $\langle -1,1 \rangle$, skala w tabeli to $\langle 0,1 \rangle$, która jest zgodna z wykorzystywaną przez użyty program oraz powszechnie stosowaną praktyką dla modelu Ekmana.

Co ciekawe, dla emocji zaskoczenia umiejscowienie urządzenia nagrywającego nie miało dużego wpływu na skuteczność, co z kolei było wyraźne dla pozostałych emocji (nawet 5 krotnie gorsze wyniki dla emocji radości), a nawet dla mikrofonu w pokoju wyniki dla emocji zaskoczenia były nieco lepsze. Różnica w rozpoznaniu dla emocji zaskoczenia mogła być spowodowana większym kontrastem pomiędzy ekspresją tej emocji a pozostałymi stanami i mniejszym wpływem szumu. Teza ta jednak wymaga dogłębniejszej analizy na większym zbiorze danych – wykracza to jednak poza ramy niniejszej pracy.

Określenie skuteczności rozpoznawania emocji w kanale wideo w bazie SEMAINE

Dane wideo z bazy SEMAINE zostały przeanalizowane za pomocą programu Noldus FaceReader [50], który analizuje mimikę twarzy. Ze względu na potrzebę synchronizacji danych, przyjęto próbkowanie na poziomie 100 ms – mimo granulacji danych SEMAINE oraz programu na poziomie 20 ms – występowały przesunięcia pomiędzy wskazaniem rzędu kilku milisekund.

Noldus FaceReader zwraca wynik w modelu 2D oraz modelu emocji podstawowych Ekmana. Do analizy wykorzystano trzy kanały: kamerę przednią kolorową (FC), przednią monochromatyczną (FBW) oraz profilową monochromatyczną (PBW).

Liczba próbek w analizie różni się w zależności od kanału i poszczególnych emocji. Operacjonalizacja zmiennych analizy została przedstawiona w Tabeli 7.3. Wartości zostały przedstawione w podziale na poszczególne wymiary (tu: emocje) oraz dane zagregowane dla wierszy etykietą „General” w Tabeli 7.4.

Tabela 7.3 Operacjonalizacja zmiennych dla określenia skuteczności rozpoznania kanału wideo w bazie SEMAINE [opracowanie własne].

Zmienna niezależna	Lokalizacja kamery: 1. Frontalna monochromatyczna (FBW) 2. Frontalna kolorowa (FC) 3. Profilowa monochromatyczna (PBW)
Zmienna zależna	Skuteczność rozpoznawania emocji (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold) (wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie, Pobudzenie, Wartościowość

Dla kamery monochromatycznej skuteczność rozpoznawania każdej z emocji jest najwyższa. Skuteczność dla kamery profilowej, dla której przez większość czasu nie wykryto twarzy, jest najniższa, jednakże nie odstaje ona znacząco od pozostałych kamer.

Najlepiej rozpoznaną emocją był strach, którego poprawność dla wartości progowej odległości równej 0,05 wyniósł 73% dla kamer przednich i 45% dla bocznej, co jest wynikiem 10 razy wyższym niż dla emocji zaskoczenia. Dla wartości progowej odległości równej 0,15 wynik powyżej 50% osiągnięty został dla emocji gniewu, zniesmaczenia oraz strachu. Najniższe wskazania dla poprawności rozpoznawania zostały odnotowane kolejno dla: emocji zaskoczenia, wymiaru pobudzenia, emocji radości oraz wymiaru wartościowości.



Rozbieżności wskazań dla poszczególnych emocji zarówno dla programu rozpoznającego dane z mimiki twarzy, jak i danych audio pokazują, jak istotne jest odrębne traktowanie poszczególnych wymiarów oraz przypisanie odpowiedniego współczynnika zaufania dla każdego z wymiarów. Przykładowo emocja zaskoczenia lepiej była rozpoznawana przez program bazujących na prozodii mowy, a gniew lepiej był rozpoznawany przez program bazujący na mimice twarzy.

Tabela 7.4 Skuteczność rozpoznawania emocji programu Noldus FaceReader na danych z trzech kamer z bazy SEMAINE. Analiza w podziale na lokalizację kamery, emocję oraz wartość progową odległości [opracowanie własne].

			Wartość progowa odległości (D_{max})					
Location	Dimension	N	0,05	0,1	0,15	0,2	0,3	0,5
Frontalna monochromatyczna (FBW)	Arousal	253 784	4,39%	8,87%	13,37%	19,07%	32,12%	62,57%
	Valence	253 784	8,94%	17,44%	25,53%	32,84%	46,10%	65,77%
	Angry	56 100	32,94%	43,15%	50,94%	58,79%	71,92%	90,15%
	Disgusted	24 256	48,43%	60,31%	67,14%	72,67%	82,20%	95,28%
	Happy	116 108	11,41%	21,64%	30,75%	40,19%	59,09%	91,15%
	Sad	69 698	28,93%	39,02%	46,62%	53,61%	69,28%	92,41%
	Scared	6 384	73,09%	75,88%	79,64%	80,50%	83,63%	88,61%
	Surprised	253 784	2,59%	5,20%	7,94%	10,82%	16,63%	32,93%
	General	253 784	10,51%	17,02%	22,92%	28,92%	40,93%	63,73%
Frontalna kolorowa (FC)	Arousal	242 917	4,37%	8,41%	12,60%	18,02%	30,63%	60,62%
	Valence	242 917	8,20%	16,25%	24,15%	31,25%	44,32%	64,81%
	Angry	54 054	30,91%	41,06%	49,79%	58,08%	71,53%	90,08%
	Disgusted	23 818	44,93%	57,18%	64,92%	70,88%	79,82%	94,20%
	Happy	109 017	11,63%	21,16%	30,70%	40,10%	58,28%	90,17%
	Sad	67 677	27,83%	38,02%	46,04%	53,22%	69,74%	92,24%
	Scared	6 408	73,11%	75,81%	79,59%	80,45%	83,55%	88,64%
	Surprised	242 917	3,03%	6,14%	9,37%	12,72%	19,84%	36,66%
	General	242 917	10,26%	16,60%	22,65%	28,67%	40,82%	63,78%
Profilowa monochromatyczna (PBW)	Arousal	9 927	7,32%	14,51%	22,50%	30,79%	45,95%	72,57%
	Valence	9 927	6,45%	13,13%	20,57%	29,33%	46,46%	73,66%
	Angry	2 687	45,74%	56,61%	64,01%	71,79%	79,90%	93,67%
	Disgusted	792	18,43%	31,44%	42,42%	57,45%	66,92%	84,72%
	Happy	4 549	5,01%	11,21%	16,09%	21,54%	39,85%	84,15%
	Sad	2 722	28,84%	38,46%	46,36%	53,38%	62,97%	87,77%
	Scared	649	45,61%	57,16%	62,40%	67,95%	77,04%	86,44%
	Surprised	9 927	0,72%	1,62%	2,43%	3,32%	5,60%	17,16%
	General	9 927	10,01%	16,03%	21,79%	28,06%	39,91%	63,59%

7.5. Eksperyment 2. Walidacja metody wyznaczania współczynnika wiarygodności

Celem eksperymentu nr 2 był dobór odpowiedniej metody reprezentacji niepewności pomiarowej. Propozycją sposobu walidacji poprawności zaproponowanej metody wyznaczania współczynnika wiarygodności jest spełnienie założeń:

- wraz z pogorszeniem się jakości danych wejściowych, współczynnik powinien maleć,
- wraz ze spadkiem współczynnika, skuteczność rozpoznawania nie powinna rosnąć.

W rozdziale 5.6 zaproponowano kilka metod wyznaczania współczynnika wiarygodności dla modalności mimiki twarzy. Zostały one określone wzorami 5.10–5.14, które zostały w tym miejscu powtórzone dla wygody czytania rozdziału. We wzorach α, β, γ oznaczają kąty, pod jakimi jest ustawiona płaszczyzna twarzy w stosunku do płaszczyzny obiektywu kamery.

$$1 - \left| \frac{\alpha}{90} \right| - \left| \frac{\beta}{90} \right| - \left| \frac{\gamma}{90} \right| \quad (5.10)$$

$$1 - \frac{\left| \frac{\alpha}{90} \right| + \left| \frac{\beta}{90} \right| + \left| \frac{\gamma}{90} \right|}{3} \quad (5.11)$$

$$\left(1 - \left| \frac{\alpha}{90} \right| \right) * \left(1 - \left| \frac{\beta}{90} \right| \right) * \left(1 - \left| \frac{\gamma}{90} \right| \right) \quad (5.12)$$

$$1 - \sqrt{\frac{\left(\frac{\alpha}{90}\right)^2 + \left(\frac{\beta}{90}\right)^2 + \left(\frac{\gamma}{90}\right)^2}{3}} \quad (5.13)$$

$$\sqrt{\frac{\left(1 - \frac{\alpha}{90}\right)^2 + \left(1 - \frac{\beta}{90}\right)^2 + \left(1 - \frac{\gamma}{90}\right)^2}{3}} \quad (5.14)$$

W weryfikacji zaproponowanych metod, wykorzystano program Noldus FaceReader [50], który zastosowano jako referencyjne rozwiązanie. Zwraca on wynik w modelu 2D (valence – arousal) oraz modelu emocji podstawowych Ekmana, a także współczynnik jakości danych.

Wyliczone współczynniki wiarygodności oraz współczynnik jakości programu Noldus FaceReader (oznaczony jako FR) w podziale na poszczególne kamery zostały przedstawione w Tabeli 7.5.

Tabela 7.5 Wartości zaproponowanych współczynników wiarygodności oraz wskaźnika jakości programu Noldus FaceReader w podziale na kanały dostępu [opracowanie własne].

Kamera	N	(5.10)	(5.11)	(5.12)	(5.13)	(5.14)	FR
FC	989 725	77,26%	91,74%	78,74%	90,18%	91,92%	91,77%
FBW	1 033 898	78,98%	92,26%	80,28%	90,94%	92,41%	91,84%
PBW	41 180	38,97%	45,35%	39,66%	44,61%	45,47%	37,35%

Zaproponowany przez autorów programu FaceReader współczynnik jakości zwrócił wysoką jakość danych dla kamer FC i FBW – powyżej 90%. Proponowane wskaźniki 5.11, 5.13 oraz 5.14 mimo swojej prostej budowy zachowują podobną charakterystykę do metody obliczanej przez autora komercyjnego programu, co pozwala przyjąć go za dobre uproszczenie, a także wskazuje na istotność ustawienia kamery również dla producenta oprogramowania.

Dla kamery profilowej (dla której ustawienie kamery jest gorsze) widoczne są większe rozbieżności wskazań zaproponowanych wskaźników od wskaźnika referencyjnego. Dla tej kamery, wartości wskaźników 5.10 oraz 5.12 są bliskie współczynnika jakości FaceReader.

Analogicznie jak dla poziomów skuteczności dla rozwiązań jednomodalnych, zaprezentowanych w rozdziale 7.4, wyższe wartości współczynnika wiarygodności są wskazywane dla kamery przedniej monochromatycznej (FBW) – co jest zachowaniem zgodnym z oczekiwaniami. Z drugiej strony, skuteczność kamery profilowej (PBW) jest niższa, ale zbliżona do pozostałych, a wartości wyznaczonego współczynnika wiarygodności czy jakości są ponad dwa razy niższe. Jest to prawdopodobnie spowodowane istotą działania algorytmu opartego o kąt kamery oraz gorszym kątem twarzy, ale wciąż rozpoznaną twarzą. Wysoka skuteczność może być też spowodowana małą próbką rozpoznanych emocji (wykrytej twarzy) lub wyraźniejszą reakcją emocjonalną – ruch głową mógł zostać wywołany działaniem emocji, a brak emocji lub emocja neutralna (nieoznaczona w zbiorze SEMAINE) częściej występować przy ustawieniu na wprost. To jednak wymagałoby dalszego dogłębnego badania, które wykracza poza zakres tej pracy.

W Tabeli 7.6 zaprezentowano zestawienie dla każdej z emocji – wartości współczynników zaufania dla sytuacji poprawnego wskazania wartości emocji (oznaczonej jako S) oraz dla błędnego (oznaczonej jako F). Dla danych w tabeli przyjęto wartość progową odległości równą 0,05. Dla kamer frontalnych (FBW i FC) zgodnie z oczekiwaniem zachowały się wskazania dla emocji gniewu oraz strachu – dla błędnej identyfikacji wartości jakości spadała. Odwrotne zachowanie zauważono dla wymiarów pobudzenia, zaskoczenia oraz smutku (słabszy efekt). Dla pozostałych emocji nie zauważono wyraźnych tendencji. Dla współczynnika jakości programu FaceReader nie zauważono podobnych tendencji.

Kamera boczna miała nieco inne wskazania – emocje zaskoczenia, radości, smutku oraz gniewu zachowały się zgodnie z oczekiwaniem, a odwrotnie wymiary: pobudzenia i wartościowości oraz emocje zniesmaczenia i strachu. Zachowanie to było analogiczne do zachowania współczynnika jakości FaceReader. Widoczna jest jednak różnica we wskazaniach współczynników wiarygodności dla poprawnie rozpoznanych stanów i dla niepoprawnie rozpoznanych dla emocji zaskoczenia – gdzie współczynnik jest o ponad 50% wyższy dla poprawnie rozpoznanych.

Analiza została powtórzona dla wyższych wartości progowych odległości (ang. distance threshold) aż do wartości 0,5 i tendencje wskaźników były analogiczne, dlatego też nie zostały tutaj przytoczone. Pełne zestawienie znajduje się w załączniku A. Otrzymane wyniki wskazują na podobieństwo zaproponowanych metod z metodą wyznaczania współczynnika jakości programu Noldus FaceReader.

Brak jednoznacznego powiązania poprawności rozpoznania oraz jakości danych w każdym z wymiarów może oznaczać niedoskonałość samego programu, afektywnego etykietowania bazy, bazowanie na szerszym kontekście przez ekspertów (prozodia mowy lub treść), jak również nieuwzględnienie jakiegoś istotnego parametru we wzorze wyznaczającym współczynnik wiarygodności. Weryfikacja powyższego wymaga jednak dalszych badań, które wykraczają poza zakres tej pracy.

Tabela 7.6 Wartości zaproponowanych współczynników wiarygodności w podziale na kanały, poszczególne emocje oraz przypadki rozpoznane poprawnie (oznaczone jako S) i nieprawidłowo (oznaczone jako F) dla wartości progowej odległości równej 0,05 [opracowanie własne].

			(5.10)		(5.11)		(5.12)		(5.13)		(5.14)		FR	
Location	Dimension	N	S	F	S	F	S	F	S	F	S	F	S	F
FBW	Arousal	253 784	76,12%	79,11%	90,14%	92,36%	77,46%	80,41%	88,60%	91,05%	90,33%	92,50%	88,99%	91,97%
	Valence	253 784	78,13%	79,06%	91,91%	92,30%	79,58%	80,35%	90,58%	90,97%	92,05%	92,44%	92,02%	91,82%
	Surprised	253 784	73,80%	79,11%	90,93%	92,30%	75,91%	80,40%	89,31%	90,98%	91,14%	92,44%	90,37%	91,88%
	Happy	116 108	77,51%	78,29%	91,82%	92,22%	79,07%	79,78%	90,55%	90,93%	91,96%	92,37%	91,75%	92,00%
	Sad	69 698	77,22%	79,62%	92,10%	92,81%	78,84%	80,95%	90,64%	91,67%	92,27%	92,93%	93,11%	93,25%
	Angry	56 100	79,90%	78,30%	92,64%	92,22%	81,14%	79,74%	91,59%	90,86%	92,74%	92,39%	91,40%	92,14%
	Disgusted	24 256	80,59%	81,08%	93,08%	93,20%	81,64%	82,14%	91,85%	92,16%	93,21%	93,29%	93,25%	93,44%
	Scared	6 384	83,29%	81,49%	94,34%	93,61%	84,09%	82,54%	93,10%	92,66%	94,46%	93,71%	93,77%	94,43%
FC	Arousal	242 917	72,79%	77,47%	88,95%	91,86%	74,45%	78,94%	86,94%	90,33%	89,23%	92,05%	88,76%	91,91%
	Valence	242 917	76,27%	77,35%	91,23%	91,78%	77,89%	78,82%	89,68%	90,23%	91,41%	91,97%	91,66%	91,78%
	Surprised	242 917	73,19%	77,39%	90,80%	91,76%	75,32%	78,85%	88,95%	90,22%	91,05%	91,95%	90,89%	91,80%
	Happy	109 017	75,79%	76,55%	91,39%	91,59%	77,51%	78,21%	89,85%	90,08%	91,58%	91,78%	92,25%	91,83%
	Sad	67 677	75,24%	77,03%	91,23%	91,92%	77,03%	78,58%	89,55%	90,47%	91,43%	92,10%	92,40%	92,83%
	Angry	54 054	77,99%	76,24%	92,02%	91,61%	79,42%	77,87%	90,81%	89,89%	92,15%	91,84%	91,80%	92,42%
	Disgusted	23 818	77,37%	78,24%	91,85%	92,24%	78,68%	79,57%	90,20%	90,99%	92,04%	92,37%	92,47%	92,28%
	Scared	6 408	80,41%	80,34%	93,35%	93,28%	81,48%	81,52%	92,06%	92,47%	93,49%	93,37%	93,75%	94,04%
PBW	Arousal	9 927	28,49%	39,79%	32,42%	46,37%	28,82%	40,52%	32,00%	45,61%	32,48%	46,50%	26,31%	38,23%
	Valence	9 927	29,99%	39,58%	35,18%	46,05%	30,54%	40,29%	34,51%	45,31%	35,30%	46,17%	28,61%	37,96%
	Surprised	9 927	64,72%	38,78%	78,66%	45,11%	66,32%	39,47%	77,18%	44,38%	78,87%	45,23%	67,78%	37,13%
	Happy	4 549	39,65%	38,67%	53,33%	45,52%	42,80%	39,46%	51,79%	44,73%	53,87%	45,65%	48,06%	37,68%
	Sad	2 722	39,95%	39,78%	46,74%	46,21%	40,70%	40,43%	46,00%	45,51%	46,85%	46,31%	38,34%	38,06%
	Angry	2 687	39,05%	37,91%	46,61%	45,45%	40,08%	39,02%	45,71%	44,58%	46,79%	45,63%	39,30%	38,36%
	Disgusted	792	31,21%	42,19%	36,25%	48,72%	31,63%	42,86%	35,57%	47,94%	36,32%	48,82%	29,47%	40,04%
	Scared	649	31,61%	43,09%	37,11%	48,28%	31,96%	43,39%	36,07%	47,50%	37,36%	48,34%	29,94%	38,65%

Wnioski z powyższej analizy pozwalają przyjąć poprawność zaproponowanej tezy pomocniczej nr 2, która mówi, że zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwi reprezentację niepewności oszacowania stanu emocjonalnego.

7.6. *Eksperyment 3. Dobór współczynnika wiarygodności w przypadku jego braku*

Jednym z istotnych wyzwań metod uwzględniających niepewność pomiarową jest określenie niepewności związanej z pomiarem, co zostało szerzej przedyskutowane w rozdziale 4.3.

Podczas analizy zbioru danych SEMAINE, wykorzystano program analizy danych audio (opisany w rozdziale 7.3), który nie zwraca współczynnika wiarygodności. Często spotykanym w literaturze przypadkiem jest statyczne wyznaczenie średniej skuteczności i przypisanie jej programowi [16] [109] [111].

W celu jak najlepszego wyznaczenia współczynnika wiarygodności autor pracy sprawdził kilka wariantów wyznaczenia współczynnika wiarygodności, a także czy lepiej wyznaczyć jeden ogólny współczynnik dla całego programu, czy indywidualny dla każdego z wymiarów.

Eksperyment 3a. Dobór współczynnika wiarygodności dla kanału audio

Dobór odpowiedniej wartości współczynnika wiarygodności jest pierwszą częścią eksperymentu 3. Analiza doboru współczynnika wiarygodności została przeprowadzona na przykładzie zbioru SEMAINE i została nazwana jako Eksperyment 3a, którego operacjonalizacja zmiennych znajduje się w Tabeli 7.7.

Tabela 7.7 Operacjonalizacja zmiennych dla eksperymentu 3a – dobór współczynnika zaufania [opracowanie własne].

Nazwa eksperymentu	Eksperyment 3a
Zmienna niezależna	Współczynnik zaufania, odpowiadający statycznej skuteczności programu dla wartości progowej odległości równej: (1) 0,1; (2) 0,15; (3) 0,2; (4) 0,3; (5) 0,5
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwany rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).
Pozostałe zmienne	<ul style="list-style-type: none">Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Dla programu analizującego dane audio [158] wyznaczono współczynnik wiarygodności dla każdej z sześciu analizowanych emocji w podziale na lokalizację kamery oraz przyjętą wartość progową odległości D_{max} . Wartości tych współczynników zostały przedstawione w Tabeli 7.2 (rozdział 7.4). Wykorzystano je do wyboru optymalnych wartości ze względu na poprawę skuteczności integracji, a także poprawność wyznacza współczynnika wiarygodności.

Wartości tych współczynników zostały przetestowane i porównane ze sobą dla różnych wartości progowych odległości D_{max} . Dla wartości równej 0,05; 0,1; 0,2 oraz 0,3 – im wartości współczynnika wiarygodności są niższe (wariant 1, 2 oraz 3) – tym skuteczność programu jest wyższa. Jest to szczególnie widoczne w spadku skuteczności dla wariantów 4 oraz 5. Dla wartości progowej odległości D_{max} równej 0,3 oraz 0,5 najlepszy rezultat został osiągnięty dla wariantu 5 współczynnika zaufania (zmiennej zależnej).

Tabela 7.8 przedstawia dane bez rozbicia na emocje dla wartości progowych odległości od 0,5 do 0,15. Pełne zestawienie znajduje się w załączniku B. Na uwagę zasługuje wariant 4, dla którego wartość skuteczności dla progów 95–100 jest najwyższa dla wartości progowej odległości D_{max} od 0,1 do 0,3.

Powyższa analiza nie wskazuje jednoznacznie, który wariant współczynnika wiarygodności i w jakich okolicznościach jest skuteczniejszy. Analiza była kontynuowana w podziale na poszczególne wymiary (emocje).

Tabela 7.8 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od doboru współczynnika wiarygodności rozwiązań jednomodalnych [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	22,99%	22,72%	26,87%	26,88%	23,70%	26,10%	28,18%	33,76%	41,90%	27,46%
		(N=12586)	(N=338190)	(N=353842)	(N=103744)	(N=22692)	(N=57584)	(N=271036)	(N=278566)	(N=155264)	(N=1593504)
	2	22,68%	23,06%	26,36%	28,17%	24,84%	24,81%	30,66%	35,77%	41,72%	27,69%
		(N=10436)	(N=303452)	(N=385436)	(N=110340)	(N=23814)	(N=68990)	(N=364296)	(N=264228)	(N=62512)	(N=1593504)
	3	21,51%	23,85%	26,86%	27,50%	25,80%	23,96%	32,75%	33,64%	40,43%	27,31%
		(N=9108)	(N=295476)	(N=391562)	(N=114212)	(N=26126)	(N=87692)	(N=437756)	(N=202340)	(N=29232)	(N=1593504)
	4	13,61%	23,95%	26,61%	28,55%	22,07%	27,13%	34,72%	22,71%	36,44%	25,70%
		(N=9090)	(N=302670)	(N=377544)	(N=124232)	(N=31204)	(N=167512)	(N=468532)	(N=104592)	(N=8128)	(N=1593504)
	5	20,36%	27,02%	31,97%	25,63%	22,88%	31,09%	27,27%	13,51%	22,64%	25,98%
		(N=12734)	(N=421400)	(N=269992)	(N=111196)	(N=128098)	(N=451826)	(N=185542)	(N=12364)	(N=352)	(N=1593504)
0.10	1	31,68%	30,41%	32,36%	32,04%	39,54%	33,96%	36,69%	42,33%	47,41%	35,30%
		(N=8784)	(N=323656)	(N=365184)	(N=111960)	(N=26104)	(N=57858)	(N=285638)	(N=316840)	(N=97480)	(N=1593504)
	2	27,31%	31,25%	32,62%	33,56%	39,14%	33,75%	39,38%	43,48%	45,53%	35,25%
		(N=6928)	(N=271760)	(N=411056)	(N=121560)	(N=26738)	(N=69606)	(N=396640)	(N=253952)	(N=35264)	(N=1593504)
	3	24,68%	31,46%	34,75%	34,13%	41,05%	33,11%	40,43%	39,90%	47,04%	35,26%
		(N=5834)	(N=245022)	(N=432968)	(N=128480)	(N=28438)	(N=89170)	(N=477584)	(N=169784)	(N=16224)	(N=1593504)
	4	24,23%	32,18%	35,83%	35,45%	33,89%	37,22%	41,72%	29,81%	67,79%	34,75%
		(N=4048)	(N=254312)	(N=417128)	(N=140776)	(N=32746)	(N=181714)	(N=486788)	(N=71960)	(N=4032)	(N=1593504)
	5	30,53%	38,84%	37,79%	32,93%	34,37%	38,80%	34,71%	25,07%	11,39%	34,48%
		(N=4658)	(N=406080)	(N=287922)	(N=119648)	(N=140974)	(N=480374)	(N=142976)	(N=10560)	(N=312)	(N=1593504)
0.15	1	41,78%	34,83%	36,40%	37,86%	39,68%	38,88%	43,60%	48,78%	55,45%	40,70%
		(N=7768)	(N=321320)	(N=365856)	(N=116450)	(N=27400)	(N=59422)	(N=311312)	(N=317176)	(N=66800)	(N=1593504)
	2	34,54%	37,05%	36,62%	39,51%	43,35%	38,74%	45,94%	50,67%	58,26%	40,94%
		(N=6248)	(N=260976)	(N=418706)	(N=127408)	(N=27640)	(N=72278)	(N=429688)	(N=224808)	(N=25752)	(N=1593504)
	3	33,62%	37,26%	39,62%	40,18%	43,73%	37,88%	47,11%	47,44%	57,47%	40,99%
		(N=5306)	(N=226190)	(N=446560)	(N=135928)	(N=29604)	(N=94772)	(N=495208)	(N=148152)	(N=11784)	(N=1593504)
	4	29,07%	37,58%	39,43%	43,04%	44,70%	43,08%	47,96%	34,37%	84,96%	41,21%
		(N=3520)	(N=208280)	(N=451840)	(N=154600)	(N=33732)	(N=194516)	(N=495864)	(N=48272)	(N=2880)	(N=1593504)
	5	35,48%	42,93%	45,73%	40,93%	42,25%	43,72%	42,03%	19,41%	20,00%	40,66%
		(N=3632)	(N=376736)	(N=304296)	(N=133560)	(N=153492)	(N=488452)	(N=124872)	(N=8256)	(N=208)	(N=1593504)

Dla emocji gniewu oraz strachu wartości z niższych wariantów (1 oraz 2) wykazują wyższą skuteczność rozpoznawania emocji. Dla emocji strachu najwyższa wartość średnia skuteczności jest osiągnięta dla wariantu 1 dla wartości progowej odległości D_{max} w przedziale od 0,05 do 0,3. Jednocześnie rozkład wartości według współczynnika wiarygodności jest zaburzony, zwłaszcza dla przedziałów współczynnika wiarygodności 0–20 oraz 40–60, gdzie pojawia się duży wzrost skuteczności. Dla emocji gniewu wyższa skuteczność rozpoznawania emocji jest widoczna dla wariantu 2, dla wartości progowej odległości D_{max} od 0,10 do 0,15, dla pozostałych wartości zaś dla wariantu 1.

Skuteczność dla emocji zniecierpliwienia jest najwyższa dla wariantu 4 dla wartości progowej odległości od 0,1 do 0,3. Wariant 2 jest skuteczniejszy dla wartości progowej odległości równej 0,05, a wariant 1 dla wartości progowej równej 0,5. Przypisanie współczynnika wiarygodności jest zbliżone do oczekiwanego – z wyjątkiem wysokiego poziomu skuteczności dla przedziałów 40–50 oraz 60–70 wobec niższej wartości dla przedziału 50–60 – dla wszystkich analizowanych wariantów. Dane dla emocji wstrętu zostały umieszczone w Tabeli 7.9 dla wartości progowych odległości od 0,05 do 0,15. Pełne zestawienie znajduje się w załączniku B.

Skuteczność dla emocji radości jest najwyższa dla wariantu 3, z wyjątkiem wartości progowej odległości D_{max} równej 0,5, dla której najwyższą skuteczność osiągnięto dla wariantu 1. Skuteczność rozpoznawania wahała się od 14 do 69,5%.

Dla emocji smutku najwyższą skuteczność osiągnięto odpowiednio: dla wartości progowej odległości D_{max} od 0,05 do 0,15 dla wariantu 5; 0,2 dla wariantu 2, 0,3 dla wariantu 1 oraz 0,5 dla wariantu 3. Przypisanie współczynnika wiarygodności nie jest odpowiednie dla żadnej z wariantów, brakuje widocznego trendu. Jego brak jest szczególnie widoczny dla wartości z przedziału współczynnika wiarygodności 0–20, dla którego większość wariantów otrzymała wyższą skuteczność.

W miarę zwiększania wariantu (wyższy numer wariantu – bliżej 5) – skuteczność rozpoznawania dla emocji zaskoczenia rośnie. Wzrost skuteczności jest widoczna dla wariantu 5, ale jedynie w przedziale wiarygodności 80–90, co jest zachowaniem oczekiwanym oraz dla przedziału 0–20, co jest przypadkiem błędnym. Skuteczności dla tej emocji waha się od 1 do 43% w zależności od wartości progowej odległości D_{max} . Wzrost skuteczności wraz ze zwiększeniem wariantu, a więc wyższa wartość współczynnika wiarygodności dla kanału audio, może być spowodowany większym wpływem kanału audio na rozwiązanie końcowe – kanały analizujące dane mimiki twarzy osiągały niższą skuteczność dla tej emocji. Niezależnie od wyboru wariantu dla analizy emocji zaskoczenia, współczynnik zaufania dla tej emocji nie jest przypisany zgodnie z oczekiwaniami.

Ze względu na duże różnice między wymiarami, autor pracy przygotował zestawienie emocji w podziale na wartości progowe odległości (D_{max}), dla których przedstawiono najlepsze warianty współczynnika. Zestawienie to przedstawiono w Tabeli 7.10 i zostało ono wykorzystane w dalszej części pracy.

Tabela 7.9 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od doboru współczynnika wiarygodności rozwiązań jednomodalnych. Tabela zawiera dane dla emocji wtrętu (ang. disgust) [opracowanie własne].

Distance Treshold (Dmax)	Variant	0-20	20-40	40-50	50-60	60-70	70-80	80-90	90-95	95-100	General	
0.05	1	0,00%	47,07%	71,76%	34,86%	43,52%	51,37%	46,67%	60,51%	71,67%	46,18%	
		(N=240)	(N=14184)	(N=17128)	(N=2416)	(N=1392)	(N=2592)	(N=10096)	(N=11312)	(N=8512)	(N=67872)	
	2	0,00%	46,42%	71,22%	39,34%	49,20%	50,38%	48,69%	64,51%	80,86%	48,20%	
		(N=208)	(N=13544)	(N=17712)	(N=2512)	(N=1496)	(N=2864)	(N=11992)	(N=11392)	(N=6152)	(N=67872)	
	3	0,00%	48,24%	70,01%	38,51%	47,58%	49,44%	52,00%	55,81%	74,48%	46,13%	
		(N=192)	(N=13416)	(N=17824)	(N=2632)	(N=1472)	(N=3112)	(N=14160)	(N=8976)	(N=6088)	(N=67872)	
	4	0,00%	52,49%	67,38%	45,45%	46,55%	51,86%	52,20%	51,06%	58,27%	47,45%	
		(N=176)	(N=13720)	(N=17440)	(N=2904)	(N=1600)	(N=4472)	(N=16384)	(N=5704)	(N=5472)	(N=67872)	
	5	24,00%	50,14%	61,64%	50,55%	48,99%	52,88%	40,33%	31,48%	100,00%	47,57%	
		(N=120)	(N=14064)	(N=16552)	(N=3752)	(N=2880)	(N=11808)	(N=13568)	(N=5120)	(N=8)	(N=67872)	
	0.10	1	8,33%	56,44%	82,38%	36,40%	89,18%	60,76%	57,86%	74,25%	76,82%	58,30%
			(N=104)	(N=13864)	(N=17400)	(N=2600)	(N=1600)	(N=2736)	(N=10288)	(N=13608)	(N=5672)	(N=67872)
		2	8,33%	56,01%	81,39%	41,61%	90,12%	60,57%	61,15%	76,27%	84,99%	59,99%
			(N=96)	(N=12904)	(N=18248)	(N=2728)	(N=1728)	(N=2936)	(N=13000)	(N=11368)	(N=4864)	(N=67872)
		3	8,33%	57,45%	80,23%	46,17%	90,20%	61,08%	62,44%	69,87%	79,99%	59,43%
			(N=80)	(N=12336)	(N=18760)	(N=2888)	(N=1720)	(N=3168)	(N=16288)	(N=7776)	(N=4856)	(N=67872)
		4	12,50%	62,15%	78,30%	53,63%	84,66%	64,46%	61,62%	56,45%	82,06%	62,07%
			(N=64)	(N=11232)	(N=19800)	(N=3144)	(N=1816)	(N=4728)	(N=16336)	(N=6784)	(N=3968)	(N=67872)
		5	40,00%	66,05%	70,20%	52,20%	75,86%	65,79%	50,67%	35,16%		57,73%
			(N=48)	(N=12760)	(N=17648)	(N=4232)	(N=3032)	(N=13976)	(N=12744)	(N=3432)	(N=0)	(N=67872)
0.15		1	8,33%	63,25%	87,52%	35,52%	92,48%	57,93%	67,86%	81,95%	79,93%	61,14%
			(N=56)	(N=13976)	(N=17280)	(N=2664)	(N=1840)	(N=2640)	(N=11408)	(N=13512)	(N=4496)	(N=67872)
		2	8,33%	62,66%	86,39%	40,53%	91,99%	59,15%	70,23%	82,61%	82,94%	61,84%
			(N=56)	(N=12872)	(N=18224)	(N=2832)	(N=1936)	(N=2936)	(N=14208)	(N=10920)	(N=3888)	(N=67872)
		3	8,33%	63,95%	85,33%	44,52%	92,40%	58,83%	71,03%	76,47%	82,94%	61,82%
			(N=56)	(N=12176)	(N=18840)	(N=3000)	(N=1912)	(N=3168)	(N=17304)	(N=7528)	(N=3888)	(N=67872)
		4	12,50%	66,42%	83,80%	55,56%	93,65%	65,85%	70,45%	83,04%	84,96%	67,34%
			(N=40)	(N=10424)	(N=20448)	(N=3392)	(N=1904)	(N=5088)	(N=16472)	(N=7224)	(N=2880)	(N=67872)
		5	40,00%	71,60%	77,15%	53,75%	80,65%	69,45%	57,41%	36,73%		61,36%
			(N=40)	(N=11968)	(N=18016)	(N=4704)	(N=3296)	(N=14112)	(N=13536)	(N=2200)	(N=0)	(N=67872)

Tabela 7.10 Zestawienie najlepszych wariantów współczynnika wiarygodności dla kanału audio w podziale na emocje oraz wartości progowych odległości (D_{max}) [opracowanie własne].

Dimension\Distance treshold (Dmax)	0,05	0,1	0,15	0,2	0,3	0,5	Accurate
Angry	1	2	2	1	1	1	33-90%
Disgusted	2	4	4	4	4	1	48-89%
Happy	3	3	3	3	1	5	14-95%
Sad	5	5	5	2	1	3	28-90%
Scared	1	1	1	1	3	5	60-93%
Surprised	5	5	5	5	5	5	4-43%

Na czerwono zostały oznaczone przypadki, dla których przeszacowanie wartości wiarygodności poprawiło skuteczność integracji, a na zielono przypadki, dla których niedoszacowanie wiarygodności poprawiło wynik integracji. Wartości współczynników wiarygodności po przełożeniu na wartości procentowe (użyte w dalszej części pracy) dla każdego z wymiarów i umiejscowienia czujników (mikrofonów) przedstawiono w Tabeli 7.11.

Tabela 7.11 Wartości współczynników wiarygodności dla poszczególnych emocji, dla kanału audio w podziale na wartości progowe odległości (D_{max}) [opracowanie własne].

Location	Distance treshold (D_{max})	Happy	Sad	Angry	Scared	Disgusted	Surprised
Head	0,05	45,60%	60,20%	1,90%	0,00%	9,50%	98,30%
Room	0,05	12,70%	21,04%	0,50%	0,00%	1,50%	98,10%
Head	0,1	45,60%	60,20%	4,00%	0,00%	17,40%	98,30%
Room	0,1	12,70%	21,04%	1,40%	0,00%	0,80%	98,10%
Head	0,15	45,60%	60,20%	4,00%	0,00%	17,40%	98,30%
Room	0,15	12,70%	21,04%	1,40%	0,00%	0,80%	98,10%
Head	0,2	45,60%	12,20%	1,90%	0,00%	17,40%	98,30%
Room	0,2	12,70%	0,00%	0,50%	0,00%	0,80%	98,10%
Head	0,3	22,30%	7,10%	1,90%	0,00%	17,40%	98,30%
Room	0,3	4,30%	0,00%	0,50%	0,00%	0,80%	98,10%
Head	0,5	93,60%	16,40%	1,90%	18,90%	4,50%	98,30%
Room	0,5	74,90%	0,00%	0,50%	3,80%	0,50%	98,10%

Podsumowując, pokazano, że zmiana wartości współczynnika wiarygodności hipotez modyfikuje skuteczność integracji w zależności od emocji. Zwiększenie współczynnika wiarygodności dla wartości małych (poniżej 25%) powoduje poprawę skuteczności integracji, co jest widoczne na przykładzie emocji zaskoczenia. Podobny wpływ nie jest widoczny dla przypadków ze skutecznością powyżej 75%, dla których przeszacowanie współczynnika wiarygodności obniża skuteczność. Dla przypadków z zakresu od 25 do 75% skuteczności, zarówno przeszacowanie, jak i niedoszacowanie współczynnika wiarygodności może mieć zarówno pozytywny, jak i negatywny wpływ, także jest szczególnie ważnym, by był on wyznaczony w sposób rzetelny.

Eksperyment 3b. Uśrednianie wartości współczynnika wiarygodności

Drugą częścią eksperymentu jest weryfikacja czy wyliczanie indywidualnego współczynnika wiarygodności przynosi korzyść w stosunku do uśrednionej wartości. Analiza została przeprowadzona na przykładzie bazy SEMAINE i została oznaczona jako Eksperyment 3b, którego operacjonalizacja znajduje się w Tabeli 7.12.

Tabela 7.12 Operacjonalizacja zmiennych dla eksperymentu 3b – uśrednianie wartości współczynnika wiarygodności [opracowanie własne].

Nazwa eksperymentu	Eksperyment 3b
Zmienna niezależna	Współczynnik wiarygodności: 1. Indywidualny dla każdego wymiaru, 2. Uśredniony dla wszystkich wymiarów.
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwanym rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance threshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Zestawienie zmiany wartości średniej skuteczności dla różnych wartości współczynnika wiarygodności zostało umieszczone w Tabeli 7.13 dla wartości progowych odległości od 0,5 do 0,15. Pełne zestawienie znajduje się w załączniku B.

Tabela 7.13 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności dla wariantu z uśrednieniem lub indywidualną wartością współczynnika wiarygodności hipotez. Dane bez uwzględnienia podziału na rodzaj emocji [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	17,07%	24,18%	27,64%	26,31%	23,50%	26,12%	30,07%	29,43%	42,21%	26,52%
		(N=25200)	(N=823850)	(N=886868)	(N=282192)	(N=95334)	(N=408538)	(N=788008)	(N=487194)	(N=186576)	(N=3983760)
	2	23,50%	23,90%	27,77%	28,33%	24,07%	27,34%	31,37%	29,17%	34,16%	27,19%
		(N=28754)	(N=837338)	(N=891508)	(N=281532)	(N=136600)	(N=425066)	(N=939154)	(N=374896)	(N=68912)	(N=3983760)
0.10	1	24,84%	32,43%	34,84%	31,28%	36,92%	34,68%	37,73%	38,06%	45,82%	34,35%
		(N=14122)	(N=749774)	(N=948040)	(N=310282)	(N=103678)	(N=436862)	(N=816674)	(N=486264)	(N=118064)	(N=3983760)
	2	30,56%	33,01%	34,48%	35,94%	37,82%	36,29%	39,44%	36,62%	43,88%	35,70%
		(N=16130)	(N=751056)	(N=966218)	(N=312142)	(N=151322)	(N=441860)	(N=972952)	(N=336832)	(N=35248)	(N=3983760)
0.15	1	31,85%	38,20%	39,47%	37,96%	42,45%	39,72%	44,69%	44,32%	57,08%	40,43%
		(N=12290)	(N=699830)	(N=978794)	(N=333808)	(N=110168)	(N=455582)	(N=852096)	(N=457008)	(N=84184)	(N=3983760)
	2	38,41%	37,38%	39,61%	42,68%	43,30%	41,47%	45,96%	41,00%	48,06%	41,39%
		(N=14184)	(N=693672)	(N=1008464)	(N=334138)	(N=161700)	(N=453858)	(N=1004848)	(N=289656)	(N=23240)	(N=3983760)

Wariant współczynnika wiarygodności (zmiennej niezależnej) z uśrednieniem zwiększa nieznacznie skuteczność w ujęciu ogólnym (bez podziału na wymiary/emocje). Dla wariantu bez uśrednienia (1), zmiana średniej skuteczności jest lepiej przypisana do współczynnika wiarygodności uzgodnionego rozwiązania – co jest widoczne w dużej amplitudzie skuteczności wartości z przedziałów 0–20 oraz 95–100, zwłaszcza dla niższych wartości progowych odległości.

Następnie autor pracy przystąpił do analizy w podziale na poszczególne emocje. Dla emocji gniewu, wstrętu, radości oraz smutku skuteczność integracji dla przypadku z indywidualnym poziomem wiarygodności dla każdego wymiaru okazała się wyższa niż przypadku uśredniania. Przykładowo, dla emocji smutku, różnice te wynosiły około jednego punktu procentowego dla przypadku uśrednionego (General), a od 6 do 9 punktów procentowych dla przedziału wiarygodności 95–100, co jest widoczne w zestawieniu w Tabeli 7.14. Tabela ta zawiera zestawienie dla wartości progowych odległości od 0,5 do 0,15. Pełne zestawienie dla wszystkich wartości progowych oraz wszystkich badanych wymiarów znajduje się w załączniku B.

Jedynymi wyjątkami od wzrostu skuteczności dla wskazanych emocji była wartość progowa odległości D_{max} równa 0,3 dla emocji radości oraz smutku, wartość progowa odległości D_{max} równa 0,05 oraz 0,5 dla emocji wstrętu.

Tabela 7.14 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności dla wariantu z uśrednieniem lub indywidualną wartością współczynnika wiarygodności hipotez. Dane dla emocji smutku [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	25,03%	31,21%	17,27%	28,29%	20,41%	29,71%	28,22%	33,04%	30,40%	25,96%
		(N=954)	(N=66744)	(N=141704)	(N=52456)	(N=6108)	(N=33856)	(N=99810)	(N=80160)	(N=38528)	(N=520320)
	2	17,85%	29,90%	22,29%	26,06%	17,97%	30,38%	29,39%	30,90%	24,14%	24,94%
		(N=848)	(N=66384)	(N=141528)	(N=53552)	(N=9856)	(N=43508)	(N=120342)	(N=67718)	(N=16584)	(N=520320)
0.10	1	38,89%	41,00%	20,66%	34,30%	24,86%	40,36%	36,18%	41,91%	36,14%	33,41%
		(N=552)	(N=58728)	(N=146704)	(N=55968)	(N=6880)	(N=36640)	(N=102408)	(N=92888)	(N=19552)	(N=520320)
	2	28,61%	39,67%	32,29%	30,67%	23,80%	40,08%	36,89%	37,25%	27,22%	32,83%
		(N=432)	(N=56568)	(N=148048)	(N=57408)	(N=11026)	(N=48054)	(N=126136)	(N=64480)	(N=8168)	(N=520320)
0.15	1	46,11%	45,13%	23,75%	39,21%	25,83%	49,66%	43,01%	45,72%	56,29%	38,73%
		(N=552)	(N=53176)	(N=150040)	(N=58344)	(N=7624)	(N=39968)	(N=107336)	(N=90528)	(N=12752)	(N=520320)
	2	36,11%	43,73%	33,12%	37,46%	23,54%	49,15%	43,24%	46,26%	49,73%	37,84%
		(N=416)	(N=51048)	(N=151208)	(N=60064)	(N=12024)	(N=52240)	(N=129512)	(N=57920)	(N=5888)	(N=520320)

Z kolei dla emocji strachu oraz zaskoczenia wartości uśrednione okazały się lepsze we wszystkich przypadkach z wyjątkiem wartości progowej odległości równej 0,5 dla emocji strachu. Przypadek zaskoczenia spowodowany jest prawdopodobnie niskimi wartościami skuteczności dla rozpoznawania twarzy, dla którego zwiększenie współczynnika wiarygodności kanału audio (zwłaszcza dla mikrofonu na głowie) poprawiło uzgodnione rozwiązanie. Dla emocji strachu poziom wiarygodności indywidualny dla kanału audio był bliski zeru, także uśredniony współczynnik spowodował uwzględnienie kanału audio, co poprawiło uzgodnione rozwiązanie.

Podsumowując, poprawa dla wariantu 2 (uśrednienia współczynnika wiarygodności dla wszystkich wymiarów) jest widoczna podczas analizy danych w poszczególnych wymiarach, a przy analizie zagregowanej różnice się znoszą. Jedynie dla dwóch emocji uśredniony współczynnik wiarygodności okazał się lepszy. Jest to kolejny przykład wskazujący na potrzebę analizy każdego z wymiarów indywidualnie.

Analiza w rozbiciu na poszczególne emocje ukazała dodatkowo, że kanał audio wnosi poprawę do uzgodnionego rozwiązania, mimo niskiej skuteczności rozpoznania przez program automatycznego rozpoznawania emocji – przybliży zintegrowany wynik rozpoznawania do odpowiedzi referencyjnej.

7.7. Podsumowanie walidacji zaproponowanej metody kwantyfikacji informacji

Wnioski z analizy eksperymentu 2 dotyczące metody wyznaczania współczynnika wiarygodności pozwalają przyjąć poprawność zaproponowanej tezy pomocniczej nr 2, która mówi, że zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwia reprezentację niepewności oszacowania stanu emocjonalnego. Wraz ze wzrostem skuteczności, przypisana została wyższa wartość współczynnika wiarygodności.

W ramach przeprowadzonych eksperymentów zaobserwowano dużą różnicę w poziomie skuteczności rozwiązań jednodalnych dla poszczególnych wymiarów, co potwierdzają zarówno analiza danych, jak i wyniki eksperymentu nr 3, dotyczącego odpowiedniego doboru współczynnika wiarygodności w przypadku jego braku. Powodem różnic może być zarówno niedoskonałość metod rozpoznawania, ale także różna liczba i jakość próbek do uczenia. Rozbieżności te były widoczne zarówno dla kanału audio, jak i analizy mimiki twarzy np. emocja zaskoczenia lepiej była rozpoznawana przez program bazujących na prozodii mowy, a gniew lepiej był rozpoznawany przez program bazujący na mimice twarzy. Wnioskiem z tego jest potrzeba przypisania indywidualnego poziomu współczynnika wiarygodności dla każdego z wymiarów z osobna, zamiast stosować zagregowany współczynnik wiarygodności.

Dodatkowym osiągnięciem przeprowadzonych eksperymentów było potwierdzenie pozytywnego wpływu wykorzystania słabszego programu na wynik integracji. Uwzględnienie w analizie wskazania programu o niskim poziomie skuteczności, ale ze zbieżną hipotezą, pozwoliło na poprawę skuteczności uzgodnionego rozwiązania.

8. EWALUACJA METODY INTEGRACJI DANYCH Z WYKORZYSTANIEM NIEPEWNOŚCI

W celu udowodnienia trzeciej tezy pomocniczej autor pracy wykorzystał ogólnodostępny i uznany zbiór danych zaetykietowanych emocjonalnie SEMAINE [34] [150]. Opis zbioru danych został umieszczony w rozdziale 7.

Treść trzeciej tezy pomocniczej została sformułowana następująco:

Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwia poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

Przeprowadzono eksperyment 4. dotyczący niezawodności, zatytuowany: poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na zbiorze danych SEMAINE (rozdział 8.2), a także eksperymenty dotyczące porównania skuteczności i poprawności wskazań współczynnika wiarygodności dla zaproponowanej metody oraz rozwiązań referencyjnych:

- Eksperyment 5. Odpowiednia konfiguracja funkcji integracji – dobór odpowiednich parametrów (rozdział 8.3),
- Eksperyment 6. Porównanie skuteczności integracji zaproponowanej metody z rozwiązaniami jednomodalnymi oraz referencyjnymi na zbiorze danych SEMAINE (rozdział 8.4),
- Eksperyment 7. Porównanie poprawności wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania przez zaproponowaną metodę oraz rozwiązania referencyjne (rozdział 8.5).

Ewaluacja metody integracji CFIEE została przeprowadzona na zbiorze SEMAINE przy użyciu programu Noldus Face Reader oraz programu do analizy prozodii głosu opracowanej na Politechnice Gdańskiej [158].

Dane w każdym kanale charakteryzują się różną dostępnością, także przygotowano podział na 5 kategorii, dla których podano liczbę rekordów w 100 ms przedziałach czasowych:

- dane dostępne dla 5 kanałów – 35 360 rekordów,
- dane dostępne dla 4 kanałów – 669 456 rekordów,
- dane dostępne dla 3 kanałów – 39 033 rekordów
- dane dostępne dla 2 kanałów – 141 582 rekordów
- dane dostępne dla 1 kanału – 7 192 rekordów – dla tych okien czasowych nie przeprowadzono integracji.

W dalszej części pracy przeanalizowano głównie przypadki 5 dostępnych kanałów. Przyjęte konwencje notacyjne są tożsame z przyjętymi w rozdziale 7, dlatego też nie zostały tutaj powtórzone.

8.1. Opis rozwiązań referencyjnych

Do analizy zaproponowanego rozwiązania autor pracy zaproponował 4 rozwiązania referencyjne, do porównania skuteczności integracji oraz wyznaczenia wskaźnika wiarygodności. Zostały zaimplementowane na podstawie znanych w literaturze rozwiązań, opisanych w rozdziale 3.7, wykorzystując dwa kryteria oceny, jakimi są: zbieżność rozwiązań oraz poziom wiarygodności hipotez. Rozwiązaniami referencyjnymi są:

1. Średnia arytmetyczna (ang. MeanState) z wartości hipotez oraz współczynników wiarygodności, jako reprezentacja metod ciągłych [16] [110],
2. Średnia ważona poziomem wiarygodności (ang. NaiveWeightedByPlausibility), jako reprezentacja metod ciągłych [16] [110],
3. Program preferujący wiarygodność (ang. NaivePreferePlausibility) – program zwracający hipotezę o najwyższej wiarygodności, jako zmodyfikowane podejście specjalistów kaskadowych [16] [17], oparte na współczynniku wiarygodności,
4. Program preferujący zbieżność (ang. NaivePrefereConvergence) – jako rozwinięcie koncepcji głosowania większościowego [109] [111], program wyszukuje najliczniejszej grupy zbieżnych hipotez (oddalonych od siebie nie bardziej niż o wartość progową odległości D_{max}), a następnie wylicza z nich średnią arytmetyczną. W przypadku braku zbieżnych grup – działa jak rozwiązanie preferujące wiarygodność.

Zaproponowana w niniejszej pracy metoda integracji CFIEE traktuje każdy wymiar niezależnie, dlatego też przyjęto analogiczne założenia dla referencyjnych programów – aby dało się je w sposób rzetelny porównać. Ten sposób działania programu pozwala na wykorzystanie dowolnego ciągłego modelu reprezentacji emocji, o ile wszystkie wymiary mogą być traktowane niezależne tak jak w modelach emocji podstawowych czy wielowymiarowych.

8.2. Eksperyment 4. **Poprawa niezawodności rozpoznawana emocji poprzez zastosowanie obserwacji wielomodalnej na zbiorze danych SEMAINE**

W ramach eksperymentu 4 została przeanalizowana dostępność estymat stanu emocjonalnego modalności prozodii mowy, jak i mimiki twarzy dla danych SEMAINE.

Dla programu automatycznego rozpoznawania emocji z mimiki twarzy kluczowym elementem jest wykrycie twarzy. W badaniu zostały zastosowane 3 kamery – kamera z kolorowym obrazem ustawiona na wprost badanego (oznaczona jako FC), kamera monochromatyczna ustawiona na wprost badanego (oznaczona jako FBW) oraz kamera monochromatyczna nagrywająca profil badanego (oznaczona jako PBW). Do analizy przyjęty został 100 ms przedział czasowy, a wynik analizy przedstawia Tabela 8.1.

Tabela 8.1 Dostępność hipotez stanu emocjonalnego w zależności od kamery dla ramek czasowych 100 ms [opracowanie własne].

Kamera	N	Dostępność [%]
Wszystkie ramki	270 521	100,0%
Profilowa monochromatyczna (PBW)	9 959	3,7%
Frontalna Kolorowa (FC)	244 706	90,5%
FC+PBW	244 816	90,5%
Frontalna monochromatyczna (FBW)	256 272	94,7%
FBW+PBW	256 522	94,8%
FC+FBW	259 560	95,9%
FC+PBW+PBW	259 670	96,0%

Ze względu na niski poziom dostępności kamery profilowej (3,7%) można ją zaklasyfikować jako kamerę pomocniczą dla sytuacji, gdy użytkownik odwracał głowę w bok. Emocje z mimiki zostały określone najczęściej dla kamery monochromatycznej ustawionej na wprost (94,7%). Powiązanie danych z dwóch frontowych kamer pozwoliło osiągnąć wynik 95,9% (wzrost o 1,2 punktów procentowych).

Następnie przystąpiono do analizy dostępności kanału audio. Analizie poddano dane dotyczące tylko użytkownika. Wyodrębnienie głosu użytkownika zostało dokonane na podstawie zapisu transkrypcji. W udostępnionej bazie dostępne są dane dotyczące 141 sesji, z czego etykiety emocjonalne oraz dane audio/wideo zostały przypisane do 95 sesji. Z tego jedynie 77 posiada transkrypcję (z dokładnością do 100 ms). Transkrypcje czasami pomijały część wstępną (dostępną w nagraniu), w którym użytkownicy wybierali postać do rozmowy, w wyniku czego 5 min i 11s nagrań pozostaje bez transkrypcji, co stanowi 1,41% czasu nagrań. Pozostały czas stanowił w 66% wypowiedź użytkownika, w 35% wypowiedź operatora, a w 5% równoczesne wypowiedzi (uniemożliwiające jednoznaczne określenie właściciela emocji) oraz 4% ciszy. Czas użytkownika zajmowany przez reakcję niewerbalną był znikomy i stanowił 0,11% czasu użytkownika. Daje to dostępność estymat dla modalności prozodii mowy dla użytkownika na poziomie 55%, co sugeruje, aby wykorzystywać modalność audio raczej jako pomocniczą niż główną. Powyższe dane zostały zestawione w Tabeli 8.2

Tabela 8.2 Dostępność kanału audio dla użytkownika w bazie SEMAINE [opracowanie własne]

Kategoria danych audio	Dostępność [s]	Dostępność [%]
Całkowity czas nagrań	22 096,5	100%
Brak transkrypcji	310,9	1%
Wypowiedź rozmówcy	7 640,4	35%
Równoczesna mowa	1 120,1	5%
Cisza	878,1	4%
Komunikacja niewerbalna	16,1	0%
Wypowiedź użytkownika	12 130,9	55%

Dostępność została zweryfikowana także dla powiązania danych audio i wideo. Analiza kanału audio dotyczyła 2749 fragmentów wypowiedzi, dla których następnie sprawdzono dostępność modalności mimiki twarzy. Odpowiednio dla kamer przedniej monochromatycznej (FBW), przedniej kolorowej (FC) oraz bocznej monochromatycznej (PBW) otrzymano: 82 (2,98%), 233 (8,5%) oraz 2733 (99,45%) fragmentów, dla których nie wykryto emocji z modalności mimiki twarzy (brak wykrytej twarzy) – a wykryto z modalności prozodii mowy. Wynik ten wskazuje na możliwość zwiększenia dostępności danych poprzez dodanie dodatkowej modalności.

Modalność prozodii wykazuje zdecydowanie wyższą dostępność niż modalność mimiki twarzy z bocznej kamery, jednak dużo niższą dostępność niż pozostałe kamery. Dodanie kanału audio pozwoliło zwiększyć dostępność danych z modalności mimiki twarzy z 96,1% do 98,3%, co pokazuje potencjał analizy wielomodalnej w zwiększaniu dostępności. Najwyższą dostępność pojedynczego kanału osiągnęła przednia kamera monochromatyczna, która wyniosła 94,9%. Zestawienie dostępności emocji z mimiki twarzy na podstawie trzech kamer, wzbogacone o dane audio zostało zestawione w Tabeli 8.3.

Tabela 8.3 Zestawienie dostępności modalności związanej z analizą mimiki oraz prozodii mowy (wypowiedzi badanego) [opracowanie własne]

Dostępne kanały obserwacji	Liczba próbek	Dostępność [%]
Wszystkie ramki	270 074	100,0%
Brak dostępnych kanałów	4644	1,7%
Dane w co najmniej 4 kanałach obserwacji	7072	2,6%
Dane w co najmniej 3 kanałach obserwacji	9695	3,6%
Kamera PBW	9959	3,7%
Dane w co najmniej 2 kanałach obserwacji	85 353	31,6%
Kanał audio	190 843	70,7%
Kamera PBW + kanał audio	193 428	71,6%
Kamera FC	244 706	90,6%
Kamery FC+PBW	245 038	90,7%
Kamera FBW	256 272	94,9%
Kamery FBW+PBW	256 512	95,0%
Kamery FC+FBW	259 560	96,1%
Kamery FC+PBW+PBW	259 660	96,1%
Kamera FC + kanał audio	260 611	96,5%
Kamery FC+PBW + kanał audio	260 699	96,5%
Kamera FBW + kanał audio	262 846	97,3%
Kamery FBW+PBW + kanał audio	263 016	97,4%
Kamery FC+FBW + kanał audio	265 383	98,3%
Kamery FC+PBW+PBW + kanał audio	265 430	98,3%

W zestawieniu przyjęto dla kanału audio wszystkie fragmenty, w których występuje wypowiedź użytkownika (bez wyłączenia wspólnych fragmentów). Różnica w dostępności dla kanału audio z wcześniejszego zestawienia wynika z powodu przyjętej innej miary – tutaj analizowane są fragmenty 100 ms (analogicznie jak dla kanału wideo), w których występuje część wypowiedzi – we wcześniejszym zestawieniu prezentowano dostępność z dokładnością do 20 ms.

W Tabeli 8.3 przedstawiono także dane o dostępności danych we wszystkich 4 kanałach (które wyniosło 2,6%), 3 kanałach (które wyniosło 3,6%) oraz co najmniej dwóch kanałach (które wyniosło 31,6%). Przypadki te pokazują, że w 68,4% przypadków brak jest połowy kanałów, a brak co najmniej jednego kanału występuje w 96,4% przypadków. Jest to szczególnie ważne w kontekście wyboru metody fuzji, zwłaszcza dla fuzji wczesnej, dla której brak jednego z kanałów poważnie wpływa na skuteczność rozpoznawania (niepełny wektor cech). W przypadku późnej fuzji brak dostępności kanału nie stanowi problemu. Analiza wielomodalna pozwala na zwiększenie dostępności danych z 94,9% (dla kamery przedniej monochromatycznej) do 98,3% przy wykorzystaniu wszystkich dostępnych kanałów.

Wynik powyższej analizy potwierdza część tezy dotyczącą poprawy niezawodności rozwiązania dzięki zastosowaniu analizy wielomodalnej.

8.3. ***Eksperyment 5. Odpowiednia konfiguracja funkcji integracji – dobór odpowiednich parametrów***

Zaproponowana metoda integracji, opisana w rozdziale 5.4, pozwala na różną konfigurację wywołania w zależności od potrzeb.

Przyjęte zostały następujące parametry funkcji integracji:

- precyzja wyliczeń do 4 miejsc po przecinku,
- przyjęty rozkład prawdopodobieństwa: rozkład normalny,

a także cztery parametry, które zostały poddane analizie w poszczególnych podeksperymentach, w których są one traktowane jako zmienne zależne. Operacjonalizacja zmiennych całego eksperymentu została przedstawiona w Tabeli 8.4. W kolejnych sekcjach przedstawiono operacjonalizację zmiennych każdego z podanych parametrów.

Tabela 8.4 Operacjonalizacja zmiennych dla eksperymentu 5 [opracowanie własne].

Nazwa eksperymentu	Eksperyment 5 (od a do d)
Zmienne niezależne (w zależności od podeksperymentu) wraz z przyjmowanymi wartościami	<p>5a: Parametry rozkładu normalnego (c i d)</p> <ul style="list-style-type: none"> • 0,4; • 0,6; • Wyznaczona wzorem: $\sqrt{(2\pi)*D_{max}}$, na podstawie wartości progowej odległości – D_{max} (distance treshold). <p>5b: Funkcja zaufania (krok 3 zaproponowanej metody integracji CFIEE – opisanej w podrozdziale 5.4):</p> <ul style="list-style-type: none"> • Funkcja iloczynowa; • Funkcja bazująca na średniej arytmetycznej. <p>5c. Odrzucenie hipotez o niskiej wiarygodności, w dwóch wariantach:</p> <ul style="list-style-type: none"> • Brak odrzucenia. • Odrzucenie hipotezy przy różnicy współczynnika wiarygodności min 0,33 <p>5d: Wybór hipotezy (krok 8 zaproponowanej metody integracji CFIEE – opisanej w podrozdziale 5.4)</p> <ul style="list-style-type: none"> • Pierwsza hipoteza, • Hipoteza dalsza od wartości „0”, • Hipoteza bliższa wartości „0”, • Hipoteza uśredniona.
Zmienna zależna	<p>Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności</p> <p>Oczekiwany rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo)</p>
Pozostałe zmienne	<ul style="list-style-type: none"> • Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 • Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Eksperyment 5a. Parametry rozkładu normalnego

Pierwszym analizowanym parametrem jest dobór parametrów rozkładu normalnego. Analiza ich wpływu na skuteczność automatycznego rozpoznawania emocji, na przykładzie danych SEMAINE została nazwana jako Eksperyment 5a, którego operacjonalizacja zmiennych znajduje się w Tabeli 8.5.

Na początku przeprowadzono analizę przypadku uśrednionego bez podziału na poszczególne emocje, której wyniki dla wartości progowej odległości w przedziale od 0,05 do 0,15 zostały zestawione w Tabeli 8.6. Pełne zestawienie danych w rozbiciu na poszczególne emocje znajduje się w załączniku B.

Tabela 8.5 Operacjonalizacja zmiennych dla eksperymentu 5a – parametry rozkładu normalnego [opracowanie własne].

Nazwa eksperymentu	Eksperyment 5a
Zmienna niezależna	Parametry rozkładu normalnego (c i d): 1. c=d=0,4; 2. c=d=0,6; 3. Wyznaczona wzorem: $c=d=\sqrt{(2\pi)*D_{max}}$, na podstawie wartości progowej odległości – D_{max} (distance treshold).
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwany rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przeźrzyty współczynnika wiarygodności – przesunięte bardziej w prawo)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Najwyższą wartość skuteczności spośród badanych wariantów (zmiennych niezależnych) osiągnął wariant drugi (wartości parametrów rozkładu normalnego równe 0,6) dla wartości progowej odległości D_{max} (ang. distance treshold) między 0,1, a 0,2, co jest widoczne w kolumnie „General”, a wariant trzeci (zależny od współczynnika wartości progowej D_{max}) dla pozostałych wartości progowych odległości.

Tabela 8.6 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla parametrów rozkładu normalnego w trzech wariantach [opracowanie własne].

Distance treshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	20,57%	23,65%	27,38%	25,73%	25,51%	26,27%	29,70%	29,56%	40,97%	26,57%
		(N= 9922)	(N=478926)	(N=654136)	(N=213848)	(N=83908)	(N=274228)	(N=590456)	(N=301440)	(N= 48976)	(N=2655840)
	2	19,11%	21,57%	26,40%	27,82%	22,85%	23,23%	32,70%	31,57%	59,16%	26,12%
		(N=5912)	(N=420432)	(N=688824)	(N=247368)	(N=106764)	(N=367900)	(N=687544)	(N=127248)	(N=3848)	(N=2655840)
	3	21,28%	26,75%	29,34%	28,55%	22,80%	30,78%	29,75%	27,86%	33,98%	27,81%
		(N=38120)	(N=761830)	(N=435416)	(N=102508)	(N=41262)	(N=191476)	(N=449162)	(N=433402)	(N=202664)	(N=2655840)
0.10	1	28,08%	32,24%	35,01%	31,95%	34,44%	35,78%	38,12%	35,33%	45,30%	34,21%
		(N=9922)	(N=478926)	(N=654136)	(N=213848)	(N=83908)	(N=274228)	(N=590456)	(N=301440)	(N=48976)	(N=2655840)
	2	29,26%	30,96%	33,73%	36,45%	38,84%	32,82%	40,46%	38,74%	60,50%	35,53%
		(N=5912)	(N=420432)	(N=688824)	(N=247368)	(N=106764)	(N=367900)	(N=687544)	(N=127248)	(N=3848)	(N=2655840)
	3	25,99%	34,90%	35,21%	32,22%	39,05%	37,82%	37,17%	38,47%	40,54%	35,33%
		(N=14418)	(N=601472)	(N=571298)	(N=161208)	(N=64328)	(N=236594)	(N=511626)	(N=394408)	(N=100488)	(N=2655840)
0.15	1	35,48%	37,80%	39,97%	38,60%	43,66%	41,05%	44,73%	41,42%	51,91%	40,65%
		(N=9922)	(N=478926)	(N=654136)	(N=213848)	(N=83908)	(N=274228)	(N=590456)	(N=301440)	(N=48976)	(N=2655840)
	2	34,11%	37,54%	38,44%	43,09%	44,80%	39,32%	46,92%	48,70%	68,09%	41,77%
		(N=5912)	(N=420432)	(N=688824)	(N=247368)	(N=106764)	(N=367900)	(N=687544)	(N=127248)	(N=3848)	(N=2655840)
	3	35,94%	38,02%	40,20%	39,22%	40,11%	41,33%	44,33%	40,58%	50,98%	40,32%
		(N=10640)	(N=494144)	(N=644298)	(N=206730)	(N=81196)	(N=267312)	(N=578944)	(N=317976)	(N=54600)	(N=2655840)

W Tabeli 8.6 wartości skuteczności są bardziej adekwatne do przedziałów współczynnika wiarygodności dla wariantu drugiego. Jest to zauważalne między innymi w dużej amplitudzie skuteczności przedziałów zaufania 0–20 oraz 95–100, zwłaszcza dla emocji gniewu oraz wstrętu.

Dla drugiego, jak i trzeciego wariantu doboru parametrów rozkładu normalnego wartości współczynnika wiarygodności są średnio niższe niż dla wariantu pierwszego parametrów rozkładu normalnego (wartość równa 0,4). Jest to widoczne na przykładzie bez podziału na emocje – Tabela 8.6, gdzie wariant drugi dla wartości progowej odległości D_{max} od 0,05 do 0,15 ma niższą liczbę przypadków w wyższych progach współczynnika wiarygodności (80–100) niż w pozostałych, a dla wartości progowej odległości od 0,2 do 0,5 niższą liczbę przypadków ma wariant trzeci (dane w załączniku B). Jest to szczególnie widoczne w przedziałach wiarygodności 90–95 oraz 95–100. Dla przedziału 95–100 wariantu trzeciego nie odnotowano ani jednego przypadku.

Następnie przeanalizowano dane w podziale na poszczególne emocje. Najwyższe wartości skuteczności osiągnęły drugi oraz trzeci wariant doboru parametrów rozkładu. Pierwszy wariant (wartości równe 0,4) odnotował wyższą skuteczność jedynie dla pojedynczych przypadków (emocja radości dla wartości progowej odległości równej 0,15 oraz emocja smutku dla wartości progowej odległości równej 0,05).

Drugi wariant okazał się skuteczniejszy dla emocji gniewu dla wartości progowej odległości współczynnika D_{max} między 0,05 a 0,15 oraz między 0,3 a 0,5; wstrętu dla wartości progowej odległości współczynnika D_{max} między 0,15 a 0,2 oraz wartości 0,5; radości dla wartości progowej odległości współczynnika D_{max} między 0,05 a 0,1; smutku dla wartości progowej odległości współczynnika D_{max} równej 0,2 oraz 0,5; strachu dla wartości progowej odległości współczynnika D_{max} równej 0,15 0,3 oraz 0,5; a także zaskoczenia dla wartości progowej odległości współczynnika D_{max} równej 0,05, 0,15 oraz 0,2.

Dla pozostałych przypadków wariant trzeci okazał się skuteczniejszy, zwłaszcza dla emocji radości i zaskoczenia dla wartości progowej odległości między 0,2 a 0,5, gdzie odnotowano wyraźny wzrost w stosunku do pozostałych wariantów. Dla lepszej przejrzystości, w Tabeli 8.7 zestawiono powyższe dane – wartości progowe odległości w podziale na emocje, a w komórkach macierzy wskazano najlepszy wariant dla ich przecięcia. Najczęściej występującym wariantem doboru parametrów rozkładu normalnego był wariant 3.

Tabela 8.7 Macierz emocji oraz wartości progowych ze wskazaniem numeru wariantu, dla którego osiągnięto najwyższą skuteczność [opracowanie własne].

Emotion	0,05	0,1	0,15	0,2	0,3	0,5
General	3	2	3	3	3	3
Surprised	2	3	2	3	3	3
Happy	2	2	1	3	3	3
Sad	1	3	3	2	3	2
Angry	2	2	2	3	3	2
Disgusted	3	3	3	3	3	3
Scared	3	3	2	3	2	2

Dla emocji zaskoczenia żaden z wariantów doboru parametrów rozkładu normalnego nie wykazał adekwatnej zależności wzrostu skuteczności wraz ze wzrostem współczynnika wiarygodności, co jest prawdopodobnie spowodowane niskim poziomem wiarygodności danych wejściowych.

W związku z przedstawioną powyżej analizą autor pracy wskazał wariant drugi oraz trzeci za odpowiedni dla danych w bazie SEMAINE. Dlatego też w dalszej części pracy wykorzystane zostały oba warianty.

Eksperyment 5b. Warianty funkcji zaufania

Drugim analizowanym parametrem był dobór odpowiedniej funkcji zaufania. Analiza wpływu doboru funkcji na skuteczność automatycznego rozpoznawania emocji, na przykładzie zbioru SEMAINE została nazwana jako Eksperyment 5b, którego operacjonalizacja zmiennych znajduje się w Tabeli 8.8.

Tabela 8.8 Operacjonalizacja zmiennych dla eksperymentu 5b – funkcja zaufania [opracowanie własne].

Nazwa eksperymentu	Eksperyment 5b
Zmienna niezależna	Funkcja zaufania, wariant: 1. Funkcja iloczynowa 2. Funkcja bazująca na średniej arytmetycznej
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwanym rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Drugi wariant funkcji zaufania (oparty o średnią arytmetyczną) prezentuje wyższą wartość skuteczności – widoczną m.in. w kolumnie general w Tabeli 8.9, gdzie została zestawiona zmiana średniej skuteczności w podziale na przedziały współczynnika wiarygodności uzgodnionego rozwiązania oraz w podziale na warianty funkcji zaufania. W tabeli znajdują się dane dla wartości progowych odległości od 0,05 do 0,15. Pełne zestawienie danych znajduje się w załączniku B.

Jednocześnie widoczna jest wyraźna różnica w przedziałach wiarygodności. Funkcja integracji wykorzystująca średnią arytmetyczną (wariant 2) nigdy nie osiąga wyższej wartości współczynnika wiarygodności niż 80%. Wynika to z natury funkcji średniej, która nigdy nie zwróci wartości większej niż najwyższa wartość hipotezy – co ma miejsce przy funkcji iloczynowej w przypadkach zbieżności.



Funkcja iloczynowa we wszystkich analizowanych przypadkach lepiej dopasowuje współczynnik wiarygodności do skuteczności niż funkcja średniej arytmetycznej. Wartości dla przedziałów 60–100 są wyższe od średniej i przeważnie maleją wraz ze spadkiem współczynnika wiarygodności. Jednocześnie, przypadki z zerową skutecznością trafiły do najniższych przedziałów zaufania – co jest widoczne dla wartości progowej odległości (D_{max}) w przedziale od 0,1 do 0,2. Dla funkcji opartej o średnią arytmetyczną tendencja ta jest zaburzona i mniej widoczna.

Tabela 8.9 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych funkcji zaufania – iloczynowej (1) oraz opartej na średniej (2). Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].

Distance threshold (D_{max})	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	75,00%	0,00%	7,32%	14,48%	25,34%	27,15%	30,72%	29,31%	39,29%	24,46%
		(N=6)	(N=584)	(N=8824)	(N=75988)	(N=220438)	(N=833180)	(N=1727162)	(N=862090)	(N=255488)	(N=3983760)
	2	20,20%	28,05%	37,59%	36,46%	12,86%	7,13%				30,10%
		(N=53948)	(N=1660604)	(N=1769552)	(N=487736)	(N=11496)	(N=424)	(N=0)	(N=0)	(N=0)	(N=3983760)
0.10	1	0,00%	3,57%	15,74%	22,50%	39,19%	36,08%	38,58%	37,39%	45,20%	33,30%
		(N=2)	(N=512)	(N=10114)	(N=86160)	(N=242720)	(N=878218)	(N=1789626)	(N=823096)	(N=153312)	(N=3983760)
	2	27,80%	37,25%	43,90%	41,61%	24,94%	12,97%				37,33%
		(N=30250)	(N=1500318)	(N=1904144)	(N=536264)	(N=12280)	(N=504)	(N=0)	(N=0)	(N=0)	(N=3983760)
0.15	1		0,00%	18,72%	29,88%	42,72%	41,19%	45,33%	42,79%	54,50%	38,35%
		(N=0)	(N=568)	(N=11050)	(N=94778)	(N=257484)	(N=908848)	(N=1856944)	(N=746664)	(N=107424)	(N=3983760)
	2	35,19%	43,67%	49,60%	48,33%	43,62%	21,97%				44,27%
		(N=26474)	(N=1392934)	(N=1976208)	(N=573168)	(N=14384)	(N=592)	(N=0)	(N=0)	(N=0)	(N=3983760)

Oczekiwana tendencja wzrostu skuteczności wraz ze współczynnikiem zaufania jest szczególnie widoczna dla funkcji iloczynowej (wariant pierwszy) dla emocji smutku oraz strachu.

Dla emocji strachu funkcji iloczynowej udało się odseparować wyniki ze 100% skutecznością dla najwyższego progu współczynnika zaufania. Zakres współczynnika wiarygodności dla funkcji iloczynowej to 60–100%, a dla funkcji arytmetycznej 0–70%, przy średniej skuteczności na poziomie 51–92% (w zależności od przyjętego progu D_{max}).

Dane dla emocji smutku przedstawiono w Tabeli 8.10. Funkcja oparta na średniej arytmetycznej dla najwyższego wskazania (60–70%) wskazała przypadki z zerową skutecznością, a funkcja iloczynowa nie odnotowała podobnej anomalii, zachowując poprawną tendencję.

Tabela 8.10 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych funkcji zaufania – iloczynowej (1) oraz opartej na średniej (2). Tabela zawiera dane dla emocji smutku [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	100,00%		4,35%	18,06%	21,67%	30,05%	28,80%	32,12%	28,12%	23,61%
		(N=2)	(N=0)	(N=368)	(N=3592)	(N=15852)	(N=77364)	(N=220152)	(N=147878)	(N=55112)	(N=520320)
	2	20,81%	30,55%	31,52%	32,31%	0,00%					27,87%
		(N=1800)	(N=133128)	(N=282864)	(N=102416)	(N=112)	(N=0)	(N=0)	(N=0)	(N=0)	(N=520320)
0.10	1			12,50%	22,15%	27,53%	40,22%	36,53%	39,91%	33,17%	30,64%
		(N=0)	(N=0)	(N=376)	(N=3824)	(N=17794)	(N=84694)	(N=228544)	(N=157368)	(N=27720)	(N=520320)
	2	33,75%	40,34%	37,65%	38,56%	0,00%					36,36%
		(N=984)	(N=115296)	(N=294376)	(N=109552)	(N=112)	(N=0)	(N=0)	(N=0)	(N=0)	(N=520320)
0.15	1			8,70%	26,89%	28,68%	49,40%	43,12%	45,95%	54,42%	35,40%
		(N=0)	(N=0)	(N=368)	(N=4296)	(N=19512)	(N=92208)	(N=236848)	(N=148448)	(N=18640)	(N=520320)
	2	41,11%	44,43%	43,41%	45,90%	0,00%					41,97%
		(N=968)	(N=104224)	(N=300880)	(N=114112)	(N=136)	(N=0)	(N=0)	(N=0)	(N=0)	(N=520320)

Dla emocji zaskoczenia wariant iloczynowy funkcji zaufania osiągnął wyższą skuteczność niż wariant średniej arytmetycznej dla wartości progowej odległości równej 0,1 oraz od 0,3 do 0,5, jednakże mocno przeszacował współczynnik wiarygodności. Oba warianty funkcji zaufania nie poradziły sobie z przypadkiem emocji zaskoczenia – ze względu na niski poziom skuteczności rozwiązań jednokanałowych, co jest widocznym ograniczeniem zaproponowanej metody.

Wyniki dla pozostałych emocji są zbliżone do generalnej tendencji – dlatego też nie zostały przytoczone w osobnych tabelach. Pełne zestawienie danych znajduje się w załączniku B. Mimo wyższej skuteczności funkcji opartej o średnią arytmetyczną, w dalszej części pracy autor zdecydował się wykorzystać funkcję iloczynową, ze względu na lepsze wyznaczenie wartości funkcji wiarygodności.

Eksperyment 5c. Odrzucenie hipotez o niskiej wiarygodności

Kolejnym analizowanym parametrem było odrzucenie hipotez o niskiej wiarygodności. Analiza jego wpływu na skuteczność automatycznego rozpoznawania emocji, na przykładzie zbioru SEMAINE została nazwana Eksperymentem 5c, którego operacjonalizacja zmiennych znajduje się w Tabeli 8.11.

Tabela 8.11 Operacjonalizacja zmiennych dla eksperymentu 4c – odrzucenie hipotez o niskiej wiarygodności [opracowanie własne].

Nazwa eksperymentu	Eksperyment 4c
Zmienna niezależna	Odrzucenie hipotez o niskiej wiarygodności, w dwóch wariantach: 1. Brak odrzucenia. 2. Odrzucenie hipotezy przy różnicy współczynnika wiarygodności min 0,33
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwanym rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance threshold) (wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Wyniki analizy dla wartości progowych odległości od 0,05 do 0,15 znajdują się w Tabeli 8.12. Pełne zestawienie danych dla wszystkich wartości progowych odległości oraz w rozbięciu na poszczególne emocje znajduje się w załączniku B.

Wartości skuteczności, jak i przypisanych współczynników wiarygodności są zbliżone bez wyraźnej tendencji dla żadnej z metod – różnice są na poziomie 0,02 punktu procentowego. Może to być związane z wartościami współczynników wiarygodności hipotez, jak również z przyjętą dużą precyzją obliczeń. Z tego też powodu, w dalszej części pracy został przyjęty wariant bez odrzucania hipotez o niskiej wiarygodności (wariant 1).

Tabela 8.12 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla odrzucenia wartości o niskiej wiarygodności (2) oraz bez (1). Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	20,37%	24,04%	27,71%	27,33%	23,78%	26,69%	30,71%	29,31%	39,29%	26,85%
		(N=26977)	(N=830594)	(N=889188)	(N=281858)	(N=115939)	(N=416702)	(N=863641)	(N=431117)	(N=127744)	(N=3983760)
	2	20,37%	24,04%	27,71%	27,33%	23,78%	26,74%	30,73%	29,31%	39,29%	26,85%
		(N=26977)	(N=830594)	(N=889188)	(N=281866)	(N=115995)	(N=416902)	(N=863521)	(N=430973)	(N=127744)	(N=3983760)
0.10	1	27,76%	32,72%	34,65%	33,63%	37,36%	35,46%	38,61%	37,39%	45,20%	35,02%
		(N=15126)	(N=750415)	(N=957129)	(N=311212)	(N=127482)	(N=439367)	(N=894825)	(N=411548)	(N=76656)	(N=3983760)
	2	27,76%	32,72%	34,65%	33,63%	37,37%	35,46%	38,56%	37,38%	45,20%	35,01%
		(N=15126)	(N=750415)	(N=957129)	(N=311212)	(N=127518)	(N=439355)	(N=894801)	(N=411548)	(N=76656)	(N=3983760)
0.15	1	35,24%	37,79%	39,54%	40,37%	42,88%	40,63%	45,33%	42,79%	54,50%	40,91%
		(N=13237)	(N=696751)	(N=993629)	(N=333973)	(N=135928)	(N=454726)	(N=928472)	(N=373332)	(N=53712)	(N=3983760)
	2	35,15%	37,79%	39,54%	40,35%	42,87%	40,50%	45,32%	42,79%	54,50%	40,89%
		(N=13237)	(N=696751)	(N=993629)	(N=333973)	(N=135940)	(N=454714)	(N=928472)	(N=373332)	(N=53712)	(N=3983760)

Eksperyment 5d. Warianty wyboru hipotezy

Ostatnim analizowanym parametrem był wpływ wyboru hipotezy w przypadku równych wartości współczynnika wiarygodności po defuzyfikacji na skuteczność automatycznego rozpoznawania emocji. Eksperyment ten został nazwany 5d, a jego operacjonalizacja zmiennych znajduje się w Tabeli 8.13.

Tabela 8.13 Operacjonalizacja zmiennych dla eksperymentu 5d – sposób wyboru hipotezy [opracowanie własne].

Nazwa eksperymentu	Eksperyment 5d
Zmienna niezależna	Wybór hipotezy 1. Pierwsza hipoteza, 2. Hipoteza dalsza od wartości „0”, 3. Hipoteza bliższa wartości „0”, 4. Hipoteza uśredniona.
Zmienna zależna	Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%) Oczekiwany rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).
Pozostałe zmienne	<ul style="list-style-type: none">• Wartość progowa odległości – D_{\max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5• Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Wyniki analizy dla wartości progowych odległości od 0,05 do 0,15 znajdują się w Tabeli 8.14. Pełne zestawienie danych dla wszystkich wartości progowych odległości oraz w rozbiciu na poszczególne emocje znajduje się w załączniku B.

Podobnie jak w przypadku odrzucenia hipotez o niskiej wiarygodności skuteczność, jak i przypisane współczynniki wiarygodności są zbliżone bez wyraźnej tendencji dla żadnej z metod – różnice są na poziomie 0,03 punktu procentowego. Jest to spowodowane rzadkimi przypadkami, dla których występują dwie identyczne hipotezy. Wpływ na to ma również przyjęta dokładność – im większa precyzja, tym mniej takich przypadków wystąpi. Rozpatrywanie tego przypadku jest jednak konieczne – gdyż są to przypadki brzegowe konieczne do rozważenia. W dalszej części przyjęto drugi wariant (wartość dalsza od 0) – wskazany jako preferowany przez autora pracy w rozdziale 5.4.

Tabela 8.14 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych metod wyboru hipotezy. Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].

Distance threshold (Dmax)	Variant	0–20	20–40	40–50	50–60	60–70	70–80	80–90	90–95	95–100	General
0.05	1	20,20%	24,04%	27,74%	27,33%	23,78%	26,72%	30,72%	29,31%	39,29%	26,84%
		(N=13484)	(N=415300)	(N=444592)	(N=140930)	(N=57980)	(N=208402)	(N=431796)	(N=215524)	(N=63872)	(N=1991880)
	2	20,87%	24,04%	27,61%	27,31%	23,78%	26,71%	30,72%	29,31%	39,29%	26,87%
		(N=13502)	(N=415288)	(N=444600)	(N=140934)	(N=57994)	(N=208398)	(N=431774)	(N=215518)	(N=63872)	(N=1991880)
	3	20,20%	24,04%	27,74%	27,33%	23,77%	26,72%	30,72%	29,31%	39,29%	26,84%
		(N=13484)	(N=415300)	(N=444592)	(N=140930)	(N=57980)	(N=208402)	(N=431796)	(N=215524)	(N=63872)	(N=1991880)
	4	20,20%	24,04%	27,74%	27,33%	23,78%	26,72%	30,72%	29,31%	39,29%	26,84%
		(N=13484)	(N=415300)	(N=444592)	(N=140930)	(N=57980)	(N=208402)	(N=431796)	(N=215524)	(N=63872)	(N=1991880)
0.10	1	27,81%	32,72%	34,64%	33,63%	37,37%	35,46%	38,58%	37,39%	45,20%	35,02%
		(N=7562)	(N=375208)	(N=478564)	(N=155606)	(N=63747)	(N=219683)	(N=447408)	(N=205774)	(N=38328)	(N=1991880)
	2	27,62%	32,72%	34,70%	33,63%	37,36%	35,46%	38,58%	37,39%	45,20%	35,00%
		(N=7566)	(N=375206)	(N=478566)	(N=155606)	(N=63759)	(N=219673)	(N=447402)	(N=205774)	(N=38328)	(N=1991880)
	3	27,80%	32,72%	34,64%	33,64%	37,36%	35,46%	38,58%	37,39%	45,20%	35,02%
		(N=7562)	(N=375208)	(N=478564)	(N=155606)	(N=63747)	(N=219683)	(N=447408)	(N=205774)	(N=38328)	(N=1991880)
	4	27,81%	32,72%	34,64%	33,63%	37,37%	35,46%	38,58%	37,39%	45,20%	35,02%
		(N=7562)	(N=375208)	(N=478564)	(N=155606)	(N=63747)	(N=219683)	(N=447408)	(N=205774)	(N=38328)	(N=1991880)
0.15	1	35,20%	37,79%	39,53%	40,35%	42,87%	40,57%	45,33%	42,79%	54,50%	40,90%
		(N=6618)	(N=348376)	(N=496814)	(N=166986)	(N=67965)	(N=227363)	(N=464236)	(N=186666)	(N=26856)	(N=1991880)
	2	35,20%	37,79%	39,58%	40,37%	42,87%	40,57%	45,33%	42,79%	54,50%	40,91%
		(N=6620)	(N=348374)	(N=496816)	(N=166988)	(N=67973)	(N=227351)	(N=464236)	(N=186666)	(N=26856)	(N=1991880)
	3	35,18%	37,79%	39,53%	40,35%	42,87%	40,57%	45,33%	42,79%	54,50%	40,90%
		(N=6618)	(N=348376)	(N=496814)	(N=166986)	(N=67965)	(N=227363)	(N=464236)	(N=186666)	(N=26856)	(N=1991880)
	4	35,20%	37,79%	39,53%	40,35%	42,87%	40,57%	45,33%	42,79%	54,50%	40,90%
		(N=6618)	(N=348376)	(N=496814)	(N=166986)	(N=67965)	(N=227363)	(N=464236)	(N=186666)	(N=26856)	(N=1991880)

Podsumowanie eksperymentu 5

W ramach eksperymentu 5 zbadano wybór parametrów funkcji integracji na przykładzie zaetykietowanego emocjonalnie zbioru danych SEMAINE. Badanymi parametrami były:

- parametry rozkładu normalnego,
- wybór funkcji zaufania,
- decyzja o odrzuceniu hipotez o niskiej wiarygodności,
- wybór wariantu hipotezy.

Dwoma parametrami, które miały kluczowy wpływ na skuteczność rozpoznawania, jak i poprawność przypisania współczynnika wiarygodności uzgodnionej hipotezy były: parametry rozkładu normalnego oraz wybór funkcji zaufania.



Dla analizy parametrów rozkładu normalnego porównano ze sobą trzy wartości: (1) stałą wartość 0,4; (2) stałą wartość 0,6 oraz (3) wartość przypisaną w zależności od przyjętej wartości progowej odległości (określonej wzorem $\sqrt{(2\pi)*D_{max}}$). Wyższą skuteczność oraz dopasowanie współczynnika wiarygodności zaobserwowano dla drugiego i trzeciego wariantu. Z powodu ich różnorodności, oba te warianty zostały wykorzystane w kolejnych eksperymentach, jako dwie alternatywne metody.

Wybór funkcji zaufania wpływa na skuteczność integracji, ale jeszcze bardziej wpływa na przypisanie wartości współczynnika zaufania. Funkcja iloczynowa przypisuje wyższy współczynnik wiarygodności uzgodnionemu rozwiązaniu dla przypadków zbieżnych hipotez. Podobna zależność nie jest widoczna dla funkcji średniej arytmetycznej, dla której wzrost wartości wiarygodności nie jest nigdy możliwy, nawet w sytuacji, gdy wszystkie hipotezy są zgodne. Z tego powodu, mimo nieco wyższej skuteczności rozwiązań wykorzystujących średnią arytmetyczną, na funkcję zaufania została wybrana funkcja iloczynowa.

Zarówno dla przypadku z odrzuceniem wartości odstających, jak i bez, a także różnych wariantów wyboru hipotezy – wartości skuteczności oraz współczynnika wiarygodności nie różniły się znacząco dla danych ze zbioru SEMAINE. Z tego powodu do dalszej analizy wybrano domyślnie zalecane przez autora pracy warianty – braku odrzucenia wartości o niskiej skuteczności oraz wybór hipotezy dalszej od wartości 0.

8.4. Eksperyment 6. Porównanie skuteczności integracji zaproponowanej metody z rozwiązaniami jednododalnymi oraz referencyjnymi na zbiorze danych SEMAINE

Aby zweryfikować skuteczność rozpoznawania emocji, autor niniejszej pracy dokonał ewaluacji zaproponowanej przez siebie metody integracji (opisanej w rozdziale 5.4) po dostosowaniu parametrów funkcji integracji, których dobór opisano w Eksperymentie 5 (rozdział 8.3) oraz po doborze wartości współczynnika zaufania dla programu bez określonego współczynnika (dobór opisany w Eksperymentie 3, w rozdziale 7.6). Ewaluacji dokonano poprzez porównanie wyników zaproponowanej metody z wynikami poszczególnych programów jednododalnych (które zostały przedstawione w rozdziale 7.4) oraz z wynikami zaproponowanych programów referencyjnych (opisanych w rozdziale 8.1). Operacjonalizacja zmiennych eksperymentu znajduje się w Tabeli 8.15.

Tabela 8.15 Operacjonalizacja zmiennych dla eksperymentu 6 – metoda rozpoznawania emocji [opracowanie własne].

Nazwa eksperymentu	Ekspertyment 6
Zmienna niezależna	<p>Metoda rozpoznawania emocji:</p> <p>Dla rozwiązań jednomodalnych:</p> <ul style="list-style-type: none"> • FBW – program Face Reader, mimika twarzy, kamera monochromatyczna na wprost użytkownika, • FC – program Face Reader, mimika twarzy, kamera kolorowa na wprost użytkownika, • PBW – program Face Reader, mimika twarzy, kamera monochromatyczna profilowa, • Voice_VR – program opracowany na politechnice gdańskiej [158], prozodia mowy, mikrofon w pomieszczeniu, • Voice_VH – program opracowany na politechnice gdańskiej [158], prozodia mowy, mikrofon w słuchawkach. <p>Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> • MeanState – bazujące na średniej arytmetycznej, • NaivePreferability – preferujące wysoki współczynnik wiarygodności, • NaivePreferConvergence – preferujące wysoką zbieżność hipotez, • NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> • CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 • CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)*D_{max}}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.
Zmienna zależna	Skuteczność rozpoznawania emocji (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> • Wartość progowa odległości – D_{max} (distance treshold) (wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 • Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

Autor pracy porównał wyniki dla programów opartych o pojedyncze kanały, a także wyniki integracji dla rozwiązań referencyjnych oraz zaproponowanej metody integracji CFIEE w dwóch wariantach: z parametrami rozkładu normalnego równymi: (1) 0,6 oraz (2) zależnych od wartości progowej odległości D_{max} . Wyniki zostały podzielone na poszczególne emocje, a także na dane bez podziału na emocję (kolumna General). Ze względu na duży wpływ emocji zaskoczenia na uśrednione zestawienie (przy niskiej skuteczności) – zostało także dodane uśrednione zestawienie z wykluczeniem emocji zaskoczenia (General without Surprised). Dane dla wartości progowej odległości od 0,05 do 0,15 zostały umieszczone w Tabeli 8.16. Pełne zestawienie znajduje się w załączniku C.

Tabela 8.16 Skuteczność rozpoznawania emocji dla pojedynczych kanałów, programów referencyjnych oraz zaproponowanej metody CFIEE, dla wartości progowych odległości (D_{max}) [opracowanie własne].

	Distance threshold (D_{max})	Happy	Sad	Angry	Scared	Disgusted	Surprised	General	General without Surprised
n		3777	2168	2326	564	707	7057	16599	9542
FBW	0,05	10,27%	31,73%	45,83%	69,50%	33,10%	2,56%	17,77%	29,01%
FC		10,14%	31,04%	43,51%	69,50%	29,70%	3,30%	17,49%	27,98%
PBW		5,03%	29,75%	44,28%	38,65%	17,82%	0,72%	13,62%	23,15%
Voice_VR		2,52%	0,00%	1,85%	6,74%	0,00%	2,13%	1,96%	1,84%
Voice_VH		12,42%	11,85%	2,67%	1,95%	0,71%	2,59%	5,95%	8,43%
MeanState		18,85%	8,58%	7,14%	3,37%	49,22%	1,35%	9,19%	15,00%
NaivePreferePlausibility		11,15%	31,27%	43,59%	2,66%	60,40%	0,41%	15,57%	26,78%
NaivePrefereConvergence		9,72%	26,61%	38,01%	64,36%	29,84%	15,37%	21,01%	25,17%
NaiveWeightedByPlausibility		15,86%	9,92%	46,82%	2,66%	53,18%	0,14%	13,88%	24,04%
CFIEE 1		12,31%	28,60%	46,00%	62,59%	57,85%	0,31%	17,71%	30,57%
CFIEE 2		11,12%	31,69%	46,13%	64,36%	60,54%	0,35%	18,05%	31,14%
FBW	0,1	21,26%	38,56%	55,46%	73,23%	46,68%	4,97%	24,24%	38,48%
FC		20,57%	37,96%	54,60%	73,23%	44,55%	6,67%	24,51%	37,71%
PBW		11,33%	37,87%	55,42%	51,60%	30,69%	1,57%	19,02%	31,92%
Voice_VR		4,69%	0,28%	3,18%	9,93%	0,00%	5,38%	4,17%	3,28%
Voice_VH		24,09%	21,40%	4,69%	4,79%	1,70%	5,33%	11,43%	15,95%
MeanState		38,20%	16,65%	18,92%	9,04%	64,21%	3,37%	18,00%	28,81%
NaivePreferePlausibility		21,58%	38,10%	55,25%	7,80%	70,16%	0,68%	21,17%	36,32%
NaivePrefereConvergence		19,46%	35,29%	51,29%	69,15%	44,55%	24,49%	30,88%	35,61%
NaiveWeightedByPlausibility		31,85%	21,68%	58,34%	5,67%	65,06%	0,54%	21,45%	36,91%
CFIEE 1		24,44%	37,96%	54,90%	67,02%	70,58%	0,86%	23,86%	40,87%
CFIEE 2		22,80%	38,88%	55,12%	68,09%	70,44%	0,67%	23,59%	40,54%
FBW	0,15	31,37%	44,33%	64,45%	75,71%	56,72%	7,85%	30,28%	46,88%
FC		31,40%	44,19%	64,62%	75,71%	56,15%	10,43%	31,37%	46,86%
PBW		16,71%	43,22%	62,90%	57,62%	41,30%	2,18%	22,90%	38,23%
Voice_VR		9,80%	1,20%	4,30%	10,11%	0,28%	10,20%	7,68%	5,82%
Voice_VH		38,81%	31,60%	7,14%	8,87%	1,70%	9,22%	18,25%	24,93%
MeanState		56,50%	27,44%	28,89%	63,48%	71,57%	6,26%	28,36%	44,70%
NaivePreferePlausibility		31,98%	44,23%	64,32%	10,11%	76,38%	1,02%	26,10%	44,64%
NaivePrefereConvergence		30,00%	42,39%	60,71%	71,81%	56,72%	27,02%	37,21%	44,75%
NaiveWeightedByPlausibility		45,94%	41,56%	67,20%	8,87%	72,28%	0,78%	29,01%	49,88%
CFIEE 1		35,32%	45,99%	64,23%	71,63%	76,24%	1,26%	29,26%	49,97%
CFIEE 2		33,94%	44,97%	64,27%	71,45%	76,24%	1,06%	28,73%	49,19%

Dane dla wartości progowej odległości D_{max} równej 0,05.

Zaproponowana metoda integracji CFIEE w drugim wariacie (zależnym od wartości progowej odległości) w uśrednionym przypadku okazała się lepsza od wszystkich programów jednokanałowych, a także lepsza od referencyjnych rozwiązań, z wyjątkiem programu bazującego na zbieżności (NaivePrefereConvergence). Jednakże, pomijając emocję zaskoczenia, której skuteczność w jednokanałowych rozwiązaniach nie przekracza 4%, metoda CFIEE uzyskuje wyższy poziom skuteczności od wszystkich rozwiązań referencyjnych.

Dla emocji zniechęcenia oraz gniewu widoczny jest wzrost skuteczności metody CFIEE w stosunku do rozwiązań jednokanałowych, a wartości skuteczności dla emocji smutku i radości są zbliżone do jednokanałowych. Pogorszenie występuje jedynie dla emocji zaskoczenia oraz strachu. Dla emocji strachu – 64,86% wobec 69,5% dla najlepszego z rozwiązań jednokanałowych – wartym jednak uwagi jest, że jest to wynik równy wynikowi najlepszego z rozwiązań referencyjnych (NaivePrefereConvergence). Z kolei największa poprawa jest widoczna dla emocji zniechęcenia – z 33% (dla modalności FBW) do 60,54%. Metoda CFIEE w wariacie drugim uzyskała niższą skuteczność od rozwiązań referencyjnych jedynie dla emocji radości oraz zaskoczenia (MeanState oraz NaiveWeightedByConfidence).

Metoda CFIEE w wariacie pierwszym uzyskała analogiczne wyniki jak w wariacie drugim, z nieco wyższym poziomem skuteczności – uzyskując poprawę w stosunku do drugiego wariantu jedynie dla emocji radości.

Jednocześnie metoda CFIEE nie odnotowała spadku skuteczności dla emocji zniechęcenia oraz strachu jak rozwiązania referencyjne, dla których duży spadek widoczny jest w jednej lub w drugiej emocji.

Program bazujący na zbieżności hipotez (NaivePrefereConvergence) uzyskał najwyższą skuteczność wśród rozwiązań referencyjnych – 64,36% wobec 3,37% dla kolejnego rozwiązania referencyjnego (MeanState), ale niższą niż rozwiązanie jednokanałowe dla kamer przednich (69,50%). Jednocześnie program bazujący na zbieżności hipotez (NaivePrefereConvergence) odnotował przeciętny wynik dla emocji wstrętu (29,84%) w stosunku do pozostałych programów referencyjnych (od 49 do 60%).

Program preferujący wysoką wartość wiarygodności (NaivePreferePlausibility) otrzymał wyniki porównywalne z najwyższymi osiągniętymi przez rozwiązania jednokanałowe, z wyjątkiem emocji wstrętu, dla której odnotowano dużą poprawę (60% wobec 33%) oraz strachu – dla którego otrzymano znaczne pogorszenie (2,66% wobec 69,5% dla rozwiązań jednokanałowych).

Dane dla pozostałych wartości progowych odległości D_{max} .

Dalsze analizy zostały przeprowadzone dla kolejnych wartości progowych odległości D_{max} . Znajdują się one również w Tabeli 8.16.

Zaproponowana metoda integracji CFIEE okazała się skuteczniejsza od rozwiązań jednokanałowych dla emocji zniesmaczenia oraz smutku. Zbliżoną skuteczność otrzymała dla emocji gniewu, strachu oraz radości. Dla wartości progowej odległości równej 0,5 poziom skuteczności był wyższy niż dla rozwiązań jednokanałowych z wyjątkiem emocji strachu oraz zaskoczenia. Analogicznie jak dla wartości progowej D_{max} równej 0,05, nie zauważono wyraźnego spadku skuteczności w stosunku do rozwiązań jednokanałowych, co występuje dla rozwiązań referencyjnych.

Dla wysokich wartości progowej odległości D_{max} równej 0,3 oraz 0,5 wyraźnie wzrasta skuteczność rozwiązań opartych na średniej (MeanState oraz NaiveWeightedByPlausibility), co jest naturalne dla tego rodzaju integracji, gdyż dążą one do wartości środkowej.

Podsumowując, powyższe analizy zaproponowane rozwiązanie okazało się skuteczniejsze od rozwiązań referencyjnych dla wszystkich emocji z wyjątkiem zaskoczenia. Rozwiązania referencyjne również są skuteczne – na szczególną uwagę zasługuje program preferujący zbieżność (NaivePrefereConvergence), który uzyskał najwyższą wartość skuteczności przy bardzo niskim współczynniku zaufania emocji zaskoczenia (poniżej 25%), jednocześnie uzyskując bardzo niską skuteczność dla emocji zniesmaczenia.

Podsumowanie eksperymentu 6

Zaproponowana metoda poradziła sobie z rzeczywistym przypadkiem w sposób zadowalający. Uzyskana skuteczność była w większości przypadków wyższa niż dla rozwiązań jednomodalnych, a dla pozostałych była od nich nie gorsza.

W przypadkach wyłączenia z analizy wymiaru zaskoczenia (cechującego się niskimi wartościami wiarygodności) zaproponowana metoda (zwłaszcza wariant II) uzyskała skuteczność wyższą od rozwiązań referencyjnych. Uwzględniając emocję zaskoczenia, uzyskała wynik nie gorszy niż rozwiązania referencyjne.

Wartym zauważenia jest fakt, że zaproponowana metoda nie obniżyła wartości skuteczności uzgodnionego rozwiązania w stosunku do rozwiązań jednomodalnych dla żadnego wymiaru (emocji), co miało miejsce w przypadku rozwiązań referencyjnych – gdzie obserwowany był wyraźny spadek (rzędu 20–60 punktów procentowych) dla jednej z emocji – strachu lub zniesmaczenia (w zależności od rozwiązania referencyjnego).

Najwyższe wartości poprawy skuteczności zaproponowanej metody zauważono dla wymiarów ze zbieżnością hipotez oraz ze skutecznością rozwiązań jednomodalnych w okolicy 50–75%, co odpowiada obecnemu poziomowi skuteczności rozwiązań znanych z literatury [34] [40] [108]. Dla wyższych wartości (powyżej 75%) wzrost skuteczności jest mniejszy, a często wręcz utrzymany na poziomie hipotez rozwiązań jednomodalnych. Spadek skuteczności był zauważony głównie dla przypadków niskiego poziomu skuteczności rozwiązań jednomodalnych (na poziomie poniżej 10%), co można uznać za ograniczenie zaproponowanej metody.

Powyższe wyniki ukazują, że poprzez zastosowanie zaproponowanej metody integracji skuteczność uzgodnionego rozwiązania jest nie gorsza od rozwiązań jednokanałowych. Jednocześnie rozbieżność wyników dla każdego z wymiarów wskazuje na potrzebę analizy każdego z nich indywidualnie w celu optymalizacji skuteczności.

8.5. Eksperyment 7. Porównanie poprawności wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania przez zaproponowaną metodę oraz rozwiązania referencyjne

Celem eksperymentu nr 7 było porównanie poprawności wyznaczania współczynnika wiarygodności przez zaproponowaną metodę oraz rozwiązania referencyjne. Analiza została przeprowadzona na przykładzie zaetykietowanego emocjonalnie zbioru SEMAINE. Operacjonalizacja zmiennych eksperymentu znajduje się w Tabeli 8.17.

Tabela 8.17 Operacjonalizacja zmiennych dla eksperymentu 7 – sposób wyznaczania współczynnika wiarygodności przez metody rozpoznawania emocji [opracowanie własne].

Nazwa eksperymentu	Eksperyment 7
Zmienna niezależna	<p>Sposób wyznaczania współczynnika wiarygodności przez metody rozpoznawania emocji</p> <p>Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> • MeanState – bazujące na średniej arytmetycznej, • NaivePreferePlausibility – preferujące wysoki współczynnik wiarygodności, • NaivePrefereConvergence – preferujące wysoką zbieżność hipotez, • NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> • CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 • CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)*D_{max}}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.
Zmienna zależna	<p>Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności – wartości procentowe (0–100%)</p> <p>Oczekiwany rezultatem (pozytywnym) jest wzrost skuteczności (w komórkach) wraz ze wzrostem współczynnika wiarygodności (przedziały współczynnika wiarygodności – przesunięte bardziej w prawo).</p>
Pozostałe zmienne	<ul style="list-style-type: none"> • Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 • Rodzaj emocji (wymiar) – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie

W zaproponowanej metodzie integracji dwoma głównymi celami są: wskazanie jak najdokładniejszego wyniku (ang. accuracy) oraz wyznaczenie współczynnika wiarygodności (ang. plausibility) uzgodnionej hipotezy. W optymistycznym przypadku, gdy składowe hipotezy mają wysoką skuteczność, współczynnik wiarygodności nie jest zbyt istotny – nie jest też potrzebna złożona integracja (a tylko wybór jednej z dostępnych odpowiedzi). Jednakże w przypadku dużych niedokładności oraz rozbieżności wyników, które występują w dziedzinie automatycznego przetwarzania symptomów emocji, jest on istotny. Opisywane w literaturze rozwiązania najczęściej nie podają tego wskaźnika dla każdego wiersza wyniku, a dla całego programu. Może on zostać wyznaczony na testowym zbiorze danych – ten problem badawczy został omówiony w rozdziale 4.3. Preferowanym przypadkiem jest zwrócenie współczynnika wiarygodności, indywidualnie dla każdego z wymiarów (emocji).

Zaproponowana metoda integracji opiera się na dwóch kryteriach:

- zbieżność hipotez oraz
- współczynnik wiarygodności hipotez bazowych.

Zestawienie zmiany średniej skuteczności dla różnych wartości współczynnika wiarygodności, wybranej metody integracji i wartości progowych odległości (dla których hipotezy są zbieżne) od 0,05 do 0,15 został przedstawiony w Tabeli 8.18. Pełne zestawienie znajduje się w załączniku D. Przyjęte konwencje opisu zostały opisane w rozdziale 7.1. Dane w tabelach przedstawiają rozwiązania referencyjne oraz zaproponowaną metodę CFIEE (w dwóch wariantach).

Dane dla wartości progowej odległości D_{max} równej 0,05

Dla zaproponowanej metody CFIE w pierwszym wariantcie wzrost skuteczności jest proporcjonalny do wzrostu współczynnika wiarygodności uzgodnionego rozwiązania – co jest widoczne jako wzrost wartości skuteczności wraz ze zmianą progę wiarygodności (wartości w komórkach rosną od lewej do prawej). W przypadku rozwiązań referencyjnych nie widać podobnej zależności, przykładowo dla programów preferujących wysoki współczynnik wiarygodności (NaivePreferePlausibility) oraz zbieżność (NaivePrefereConvergence) najwyższa skuteczność (powyżej 50%) przypada na przedziały 40–50 oraz 50–60, a skuteczność w przedziale 95–100 nie przekracza 35%, co można zauważyć w zestawieniu w Tabeli 8.18.

Dane pozostałych wartości progowych odległości D_{max}

Dla kolejnych wartości progowych odległości D_{max} wskazania są analogiczne. Jedyną anomalią dla zaproponowanej metody są przypadki wzrostu skuteczności w przedziale wiarygodności uzgodnionej hipotezy równej 40–50 dla progę D_{max} równego 0,15 oraz 0,2, jednakże występuje on jedynie dla 23/16 (D_{max} równe 0,15 odpowiednio dla pierwszego i drugiego wariantu) oraz 13 (D_{max} równe 0,2 dla pierwszego wariantu) przypadków z 16599. Dla wartości progowej odległości równej 0,15 wyniki dla obu wariantów są zbliżone do siebie, a od wartości 0,2 drugi wariant zaczyna mieć wyższe wartości skuteczności od pozostałych rozwiązań z wyjątkiem wartości średniej (MeanState).

Z rozwiązań referencyjnych, jedynie program oparty o średnią (MeanState) wykazuje właściwe przypisanie współczynnika wiarygodności i to tylko dla wartości progowej odległości D_{max} równej 0,15 i 0,2, gdzie skuteczność dla przedziałów 70–80, 80–90, 90–95 jest wyraźnie wyższa od pozostałych przedziałów.

Dla rozwiązań opartych o zbieżność (NaivePrefereConvergence) oraz preferencję wysokiego współczynnika wiarygodności hipotez (NaivePreferePlausibility) widoczne są anomalie w postaci wzrostów skuteczności w przedziałach 40–50 oraz 50–60 dla wartości progowej odległości (D_{max}) od 0,05 do 0,15. Dla rozwiązań preferujących wysoki współczynnik wiarygodności hipotez bazowych (NaivePreferePlausibility) występują także przypadki klasyfikacji do przedziału wiarygodności 0–20% przypadków, dla których skuteczność jest równa 100% (zgodna ze wskazaniem ekspertów), co jest zachowaniem błędnym.

Tabela 8.18 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od wybranej metody integracji [opracowanie własne].

Distance threshold (Dmax)	Solution	0-20	20-40	40-50	50-60	60-70	70-80	80-90	90-95	95-100	General
0.05	MeanState	0,00% (N=8)	16,81% (N=673)	15,49% (N=6422)	7,01% (N=6077)	21,79% (N=2214)	22,24% (N=523)	14,65% (N=494)	16,17% (N=188)		14,72% (N=16599)
	Naive Prefere Plausibility	0,00% (N=1)	0,00% (N=4)	58,33% (N=13)	33,33% (N=7)	24,86% (N=24)	15,84% (N=85)	28,43% (N=1626)	29,98% (N=7464)	22,84% (N=7375)	27,60% (N=16599)
	Naive Prefere Convergence	12,19% (N=4345)	0 (N=7)	52,79% (N=648)	57,35% (N=22)	38,22% (N=397)	7,04% (N=219)	34,93% (N=1121)	35,64% (N=4851)	32,49% (N=4989)	31,63% (N=16599)
	Naive Weighted By Plausibility	25,00% (N=8)	20,18% (N=744)	29,16% (N=6875)	22,05% (N=6251)	8,24% (N=1890)	9,63% (N=360)	1,87% (N=375)	3,13% (N=96)		19,00% (N=16599)
	CFIEE 1		0,00% (N=0)	12,50% (N=20)	11,59% (N=450)	18,28% (N=690)	29,91% (N=4999)	28,68% (N=7078)	34,86% (N=3201)	76,85% (N=156)	26,42% (N=16599)
	CFIEE 2		0,00% (N=0)	12,50% (N=10)	22,92% (N=39)	11,17% (N=347)	29,52% (N=1396)	28,07% (N=5888)	15,46% (N=4544)	33,83% (N=4374)	21,95% (N=16599)
	MeanState	0,00% (N=8)	40,21% (N=399)	31,73% (N=6329)	20,29% (N=6436)	24,68% (N=2222)	35,03% (N=523)	26,02% (N=494)	27,31% (N=188)		27,46% (N=16599)
0.10	Naive Prefere Plausibility	100,00% (N=1)	0,00% (N=4)	70,83% (N=13)	33,33% (N=7)	24,86% (N=24)	31,61% (N=85)	35,90% (N=1626)	37,14% (N=7464)	30,18% (N=7375)	37,48% (N=16599)
	Naive Prefere Convergence	24,43% (N=3219)	0,00% (N=7)	54,58% (N=369)	64,00% (N=30)	38,60% (N=294)	14,47% (N=248)	39,05% (N=1305)	41,75% (N=5544)	39,19% (N=5583)	36,04% (N=16599)
	Naive Weighted By Plausibility	25,00% (N=8)	38,28% (N=470)	38,77% (N=6782)	31,33% (N=6610)	14,67% (N=1898)	17,22% (N=360)	3,73% (N=375)	9,38% (N=96)		27,92% (N=16599)
	CFIEE 1		0,00% (N=4)	30,26% (N=23)	21,96% (N=472)	38,08% (N=720)	38,28% (N=5201)	37,82% (N=8302)	42,12% (N=1819)	79,38% (N=58)	36,92% (N=16599)
	CFIEE 2		0,00% (N=0)	30,00% (N=16)	13,27% (N=184)	35,49% (N=553)	40,51% (N=1860)	34,46% (N=8219)	36,11% (N=4299)	45,89% (N=1467)	33,17% (N=16599)
	MeanState	0,00% (N=8)	46,04% (N=399)	39,48% (N=6329)	28,58% (N=6436)	38,23% (N=2222)	51,77% (N=523)	60,19% (N=494)	64,97% (N=188)		40,25% (N=16599)
	Naive Prefere Plausibility	100,00% (N=1)	0,00% (N=4)	70,83% (N=13)	33,33% (N=7)	24,86% (N=24)	40,34% (N=85)	41,12% (N=1626)	44,25% (N=7464)	35,99% (N=7375)	41,49% (N=16599)
0.15	Naive Prefere Convergence	33,45% (N=2346)	0,00% (N=7)	47,76% (N=240)	67,95% (N=44)	45,36% (N=242)	23,51% (N=283)	46,16% (N=1477)	48,51% (N=6007)	44,82% (N=5953)	41,36% (N=16599)
	Naive Weighted By Plausibility	25,00% (N=8)	46,19% (N=470)	50,66% (N=6782)	41,46% (N=6610)	24,23% (N=1898)	28,70% (N=360)	6,40% (N=375)	12,50% (N=96)		36,59% (N=16599)
	CFIEE 1		0,00% (N=4)	30,26% (N=23)	32,68% (N=472)	45,80% (N=720)	43,43% (N=5201)	46,21% (N=8302)	51,59% (N=1819)	86,67% (N=58)	44,21% (N=16599)
	CFIEE 2		0,00% (N=0)	53,85% (N=15)	30,80% (N=333)	42,87% (N=589)	44,83% (N=2629)	41,91% (N=9028)	37,56% (N=3089)	77,84% (N=915)	42,23% (N=16599)
	MeanState	0,00% (N=8)	46,04% (N=399)	39,48% (N=6329)	28,58% (N=6436)	38,23% (N=2222)	51,77% (N=523)	60,19% (N=494)	64,97% (N=188)		40,25% (N=16599)
	Naive Prefere Plausibility	100,00% (N=1)	0,00% (N=4)	70,83% (N=13)	33,33% (N=7)	24,86% (N=24)	40,34% (N=85)	41,12% (N=1626)	44,25% (N=7464)	35,99% (N=7375)	41,49% (N=16599)
	Naive Prefere Convergence	33,45% (N=2346)	0,00% (N=7)	47,76% (N=240)	67,95% (N=44)	45,36% (N=242)	23,51% (N=283)	46,16% (N=1477)	48,51% (N=6007)	44,82% (N=5953)	41,36% (N=16599)

Dane w podziale na poszczególne emocje

Następnie przeprowadzono szczegółową analizę dla poszczególnych emocji. Dla emocji gniewu zaproponowana metoda CFIEE zachowała się zgodnie z oczekiwaniami. Jedynym odstępstwem są wyniki dla przedziału współczynnika wiarygodności 60–70 (dla wariantu pierwszego) oraz 50–60 (dla wariantu drugiego) – występuje wyższa skuteczność niż w przedziale współczynnika wiarygodności 70–80 dla wartości progowej odległości D_{max} większej od 0,05. Dane dotyczące emocji gniewu w podziale na przedziały współczynnika wiarygodności zostały przedstawione w Tabeli 8.19. Pełne zestawienie dla wszystkich emocji z modelu reprezentacji oraz wartości progowych odległości znajduje się w załączniku C.

Dla przedziału współczynnika wiarygodności 70–80 zauważalny jest spory spadek skuteczności dla rozwiązań referencyjnych preferujących wysoką wiarygodność hipotez bazowych (NaivePreferePlausibility) oraz preferujących zbieżność (NaivePrefereConvergence), a następnie wzrost w przedziale 60–70. Rozwiązania referencyjne oparte na średniej (MeanState oraz NaiveWeightedByPlausibility) zwracają wartości współczynnika wiarygodności nie większe niż 60%.

Dla emocji wstrętu zaproponowana metoda CFIEE zachowała się zgodnie z oczekiwaniami. Jedynie dla przedziału współczynnika wiarygodności uzgodnionego rozwiązania 60–70 występuje wyższa skuteczność niż w przedziale 70–80 dla wartości progowej odległości D_{max} wyższej od 0,05. Dla rozwiązań referencyjnych jedynie dla rozwiązania preferujących zbieżność (NaivePrefereConvergence) oraz wysoką wartość współczynnika wiarygodności hipotez początkowych (NaivePreferePlausibility) widoczna jest właściwa tendencja wzrostu skuteczności wraz ze wzrostem współczynnika wiarygodności w przedziałach 60–100.

Dla emocji strachu trudno jest określić poprawność wyznaczenia współczynnika wiarygodności zintegrowanego wyniku dla zaproponowanej metody CFIEE, ze względu na wystąpienie wartości jedynie dla dwóch przedziałów współczynnika wiarygodności zintegrowanej hipotezy 70–80 oraz 80–90, dla których wartości skuteczności są zbliżone.

Dla emocji radości, smutku oraz zaskoczenia, dla żadnego z zaproponowanych rozwiązań, przypisanie współczynnika wiarygodności nie koresponduje z otrzymaną skutecznością (brak widocznego trendu).

Podsumowanie eksperymentu 7

Podsumowując rezultaty uzyskane w eksperymencie 7, dzięki zastosowaniu odpowiedniej metody integracji danych, a także odpowiedniej reprezentacji niepewności pomiarowej w rozpoznawaniu wielokanałowym, uzyskano poprawę skuteczności rozpoznawania emocji, a także adekwatne do skuteczności oszacowanie poziomu wiarygodności uzgodnionego rozwiązania.

Zaproponowana metoda integracji CFIEE uzyskała poziom skuteczności nie gorszy niż rozwiązania referencyjne. Jednakże rozwiązania referencyjne nie poradziły sobie z właściwym przypisaniem współczynnika wiarygodności, co z kolei jest niewątpliwie dużą przewagą zaproponowanej metody.

Tabela 8.19 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od wybranej metody integracji. Dane dla emocji gniewu [opracowanie własne].

Distance threshold (Dmax)	Solution	0-20	20-40	40-50	50-60	60-70	70-80	80-90	90-95	95-100	General
0.05	MeanState	0,00%	4,98%	8,78%	4,79%						4,64%
		(N=2)	(N=321)	(N=1356)	(N=647)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	Naive Prefere Plausibility			100,00%	100,00%	60,00%	14,29%	50,38%	47,35%	38,18%	58,60%
		(N=0)	(N=0)	(N=1)	(N=1)	(N=5)	(N=7)	(N=262)	(N=1018)	(N=1032)	(N=2326)
	Naive Prefere Convergence	4,93%		100,00%	100,00%	60,00%	7,69%	52,04%	51,23%	44,15%	52,51%
		(N=548)	(N=0)	(N=1)	(N=1)	(N=5)	(N=13)	(N=221)	(N=810)	(N=727)	(N=2326)
	Naive Weighted By Plausibility	50,00%	43,61%	42,40%	57,65%						48,42%
		(N=2)	(N=321)	(N=1356)	(N=647)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	CFIEE 1				7,41%	20,00%	19,27%	33,57%	54,93%	63,70%	33,14%
		(N=0)	(N=0)	(N=0)	(N=27)	(N=55)	(N=109)	(N=700)	(N=1289)	(N=146)	(N=2326)
	CFIEE 2				33,33%	4,55%	12,96%	20,69%	27,11%	52,49%	25,19%
		(N=0)	(N=0)	(N=0)	(N=3)	(N=22)	(N=54)	(N=145)	(N=273)	(N=1829)	(N=2326)
0.10	MeanState	0,00%	16,17%	20,44%	16,74%						13,34%
		(N=2)	(N=235)	(N=1414)	(N=675)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	Naive Prefere Plausibility			100,00%	100,00%	60,00%	42,86%	65,27%	58,84%	49,13%	68,01%
		(N=0)	(N=0)	(N=1)	(N=1)	(N=5)	(N=7)	(N=262)	(N=1018)	(N=1032)	(N=2326)
	Naive Prefere Convergence	11,50%		100,00%	100,00%	60,00%	23,08%	60,16%	59,91%	53,10%	58,47%
		(N=287)	(N=0)	(N=1)	(N=1)	(N=5)	(N=13)	(N=251)	(N=913)	(N=855)	(N=2326)
	Naive Weighted By Plausibility	50,00%	58,72%	55,73%	63,70%						57,04%
		(N=2)	(N=235)	(N=1414)	(N=675)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	CFIEE 1				17,86%	35,19%	34,19%	49,06%	62,71%	68,75%	44,63%
		(N=0)	(N=0)	(N=0)	(N=28)	(N=54)	(N=117)	(N=907)	(N=1172)	(N=48)	(N=2326)
	CFIEE 2				0,00%	26,67%	30,00%	37,89%	53,51%	61,81%	34,98%
		(N=0)	(N=0)	(N=0)	(N=8)	(N=45)	(N=50)	(N=161)	(N=970)	(N=1092)	(N=2326)
0.15	MeanState	0,00%	23,83%	31,82%	24,59%						20,06%
		(N=2)	(N=235)	(N=1414)	(N=675)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	Naive Prefere Plausibility			100,00%	100,00%	60,00%	57,14%	73,66%	68,96%	57,36%	73,88%
		(N=0)	(N=0)	(N=1)	(N=1)	(N=5)	(N=7)	(N=262)	(N=1018)	(N=1032)	(N=2326)
	Naive Prefere Convergence	15,20%		100,00%	100,00%	60,00%	30,77%	66,54%	69,39%	58,82%	62,59%
		(N=171)	(N=0)	(N=1)	(N=1)	(N=5)	(N=13)	(N=263)	(N=954)	(N=918)	(N=2326)
	Naive Weighted By Plausibility	50,00%	66,81%	64,64%	72,74%						63,55%
		(N=2)	(N=235)	(N=1414)	(N=675)	(N=0)	(N=0)	(N=0)	(N=0)	(N=0)	(N=2326)
	CFIEE 1				35,71%	42,59%	47,01%	58,77%	71,08%	83,33%	56,41%
		(N=0)	(N=0)	(N=0)	(N=28)	(N=54)	(N=117)	(N=907)	(N=1172)	(N=48)	(N=2326)
	CFIEE 2				25,00%	34,62%	48,08%	50,57%	63,22%	76,32%	49,63%
		(N=0)	(N=0)	(N=0)	(N=16)	(N=52)	(N=52)	(N=265)	(N=1278)	(N=663)	(N=2326)

8.6. Podsumowanie eksperymentów na zbiorze danych SEMAINE

W ramach ewaluacji metody integracji danych z wykorzystaniem niepewności na danych ze zbioru SEMAINE zostały przeprowadzone cztery eksperymenty.

Rezultaty eksperymentu nr 4 wykazały znacząco poprawę niezawodności rozwiązania poprzez zastosowanie analizy wielomodalnej. Dominującą modalnością była analiza mimiki twarzy, a głównym kanałem przednia kamera monochromatyczna. Już samo dodanie drugiego kanału w postaci kamery kolorowej spowodowało zwiększenie dostępności symptomów rozpoznawania emocji (wykryta twarz), a boczna kamera pomogła zwiększyć dostępność, gdy użytkownik nie patrzył na wprost. Dodatkowe kanały poprawiły dostępność z 94,9% do 96,1%. Dodanie drugiej modalności w postaci kanału audio spowodowało poprawę dostępności dla sytuacji, gdy symptomy emocji nie były dostępne dla modalności mimiki twarzy. Dodanie kanału audio spowodowało wzrost niezawodności rozwiązania, poprzez zwiększenie dostępności do poziomu 98,3% czasu badania.

Podczas eksperymentu 5 zostały zweryfikowane różnice w poziomie skuteczności integracji oraz poprawności przypisania współczynnika wiarygodności dla różnej konfiguracji parametrów funkcji integracji. Zbadane zostały cztery parametry, z czego różnice we wskazaniach zauważono dla dwóch z nich. Wyznaczono optymalne parametry rozkładu normalnego dla zbioru danych SEMAINE w dwóch wariantach – przypisania stałej wartości parametrów oraz w wersji adaptacyjnej, zależnej od wartości progowej odległości. Drugim wyznaczonym parametrem był wybór funkcji zaufania, dla której wybrano wariant iloczynowy ze względu na poprawne przypisanie wartości współczynnika wiarygodności uzgodnionego rozwiązania, powiązanej ze wzrostem skuteczności, a także zwiększającej wartość współczynnika wiarygodności dla przypadków zbieżnych hipotez, co nie zostało zaobserwowane dla funkcji opartej o średnią arytmetyczną.

W eksperymencie 6 skuteczność integracji zaproponowanej metody została porównana z rozwiązaniami jednomodalnymi oraz z rozwiązaniami referencyjnymi. Uzyskana skuteczność była w większości przypadków wyższa niż dla rozwiązań jednomodalnych, a dla pozostałych była nie gorsza od rozwiązań jednomodalnych. W przypadkach wyłączenia z analizy wymiaru zaskoczenia (cechującego się niskimi wartościami wiarygodności) zaproponowana metoda (zwłaszcza wariant II) uzyskiwała skuteczność wyższą od rozwiązań referencyjnych. Zaproponowana metoda, w przeciwieństwie do metod referencyjnych, nie obniżyła wartości skuteczności uzgodnionego rozwiązania w stosunku do rozwiązań jednomodalnych dla żadnego z wymiarów. Największą poprawę zauważono dla przypadków ze zbieżnymi hipotezami oraz wymiarów ze skutecznością rozpoznawania rzędu 50–75%. Zauważono także ograniczenie zaproponowanej metody: dla hipotez z niską wartością wiarygodności (poniżej 10%) zaobserwowano spadek skuteczności uzgodnionego rozwiązania.

W eksperymencie 7 porównano poprawność wyznaczania współczynnika wiarygodności w zaproponowanej metodzie oraz w rozwiązaniach referencyjnych. Zaproponowane rozwiązanie wyznaczyło wyższy współczynnik wiarygodności wraz ze wzrostem skuteczności rozpoznawania – w przeciwieństwie do rozwiązań referencyjnych, dla których podobna zależność nie została zauważona.

Podsumowując wyniki eksperymentów: analiza wielomodalna pozwala na poprawę niezawodności rozwiązania automatycznego rozpoznawania emocji. Zaproponowana metoda, po odpowiedniej kalibracji parametrów, pozwoliła na uzyskanie lepszej lub co najmniej nie gorszej skuteczności niż rozwiązania jednomodalne oraz referencyjne, a także pozwoliła na wyznaczenie współczynnika wiarygodności korespondującego z uzyskanym poziomem skuteczności.

Dalsze badania nad udowodnieniem tezy pomocniczej nr 3, dotyczącej poprawy wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji kontynuowano w kolejnych rozdziałach.

9. SYMULACJE SYTUACJI GRANICZNYCH

W uzupełnieniu do przeprowadzonych eksperymentów na zaetykietowanym emocjonalnie zbiorze SEMAINE autor pracy przygotował cztery symulacje, mające na celu zbadanie warunków brzegowych. Symulacje dotyczyły udowodnienia trzeciej tezy pomocniczej, brzmiącej:

Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwi poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

Weryfikacja niezawodności nie jest możliwa podczas symulacji, możliwa jest jednak weryfikacja poprawności wyznaczenia współczynnika wiarygodności. Przeprowadzonymi symulacjami były:

- Symulacja 2. Analiza zmiany wartości współczynnika wiarygodności dla rozwiązań zbieżnych i rozbieżnych (rozdział 9.1),
- Symulacja 3. Współczynnik wiarygodności uzgodnionego rozwiązania dla zmiennej wartości współczynnika wiarygodności hipotez (rozdział 9.2),
- Symulacja 4. Analiza wpływu wartości wiarygodności rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania (rozdział 9.3),
- Symulacja 5. Analiza wpływu odległości rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania (rozdział 9.4).

9.1. **Symulacja 2. Analiza zmiany wartości współczynnika wiarygodności dla rozwiązań zbieżnych i rozbieżnych**

Drugą symulacją była analiza zmiany wartości współczynnika wiarygodnością dla rozwiązań zbieżnych oraz rozbieżnych.

Zgodnie z założeniami rozwiązania wykorzystującego współczynnik wiarygodności, opisanymi w 5.4, gdy hipotezy są zbieżne – współczynnik wiarygodności powinien się zwiększać, a w przypadku rozbieżności – zmniejszać proporcjonalnie do odległości hipotez. W celu eliminacji zmiennych zaburzających, w symulacji założono identyczny współczynnik wiarygodności dla obu hipotez. Operacjonalizacja zmiennych eksperymentu znajduje się w Tabeli 9.1.

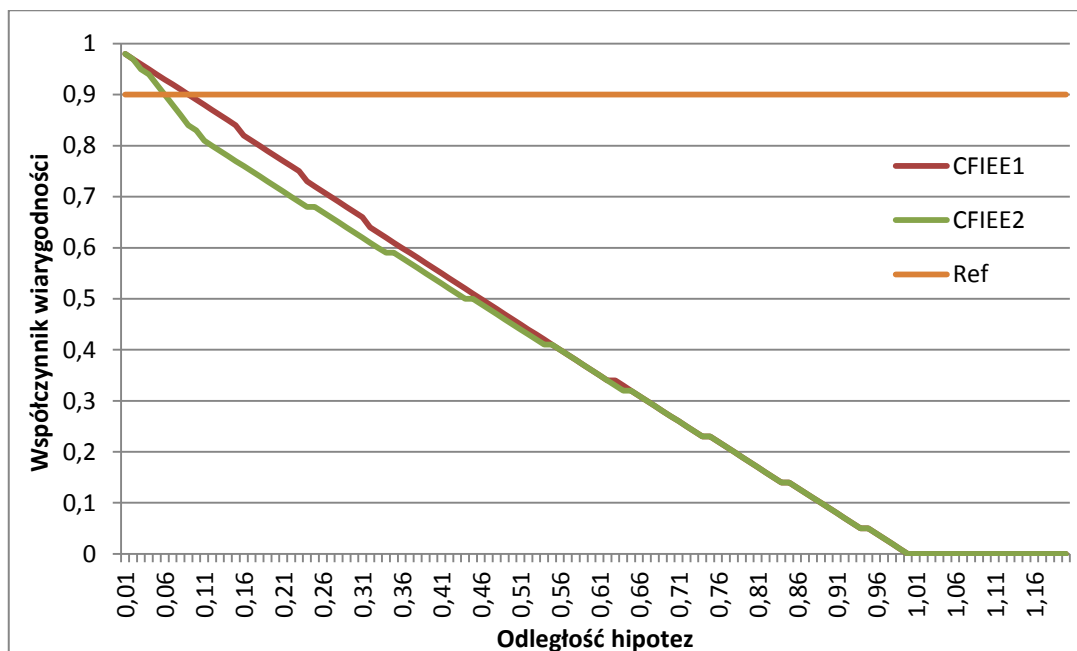
Symulacja dotyczyła przypadku dwóch hipotez o równych wartościach współczynnika wiarygodności, ale zmiennej odległości od siebie. Hipoteza badawcza eksperymentu brzmiała: dla zbieżnych hipotez o tym samym współczynniku wiarygodności, współczynnik wiarygodności uzgodnionego rozwiązania będzie wyższy niż współczynnik zbieżnych hipotez, a po przekroczeniu wartości progowej odległości (czyli dla hipotez rozbieżnych) będzie niższy.

Tabela 9.1 Operacjonalizacja zmiennych dla symulacji 2 [opracowanie własne].

Nazwa eksperymentu	Symulacja 2
Zmienna niezależna	Odległość hipotez – Wartości od 0,01 do 2, z krokiem równym 0,01
Zmienna zależna	Współczynnik wiarygodności ustalonego rozwiązania (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość współczynnika wiarygodności hipotez – tu wyjątkowo przyjęty tylko jedną wartość – 0,9. Wartość progowa odległości – D_{max} (distance treshold; wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 <p>Metoda rozpoznawania emocji: Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> MeanState – bazujące na średniej arytmetycznej, NaivePreferePlausibility – preferujące wysoki współczynnik wiarygodności, NaivePreferaConvergence – preferujące wysoką zbieżność hipotez, NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)} \cdot D_{max}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.

W analizie wykorzystano rozwiązania referencyjne opisane w rozdziale 8.1 oraz dwa warianty metody CFIEE – różniące się współczynnikami rozkładu normalnego. Wybór tych dwóch wariantów jest podyktowany wynikami eksperymentu 5 oraz faktem, że pozostałe parametry, takie jak eliminacja wartości czy różne warianty wyboru hipotezy nie mają wpływu na wynik w zaproponowanej symulacji.

Wykres zależności współczynnika wiarygodności od odległości hipotez, dla współczynnika wiarygodności równego 0,9 oraz wartości progowej odległości równej 0,05 został przedstawiony na Rys 9.1.



Rys 9.1 Zależność współczynnika wiarygodności od odległości hipotez, dla współczynnika wiarygodności równego 0,9 oraz wartości progowej odległości 0,05. Ref – rozwiązania referencyjne, CFIEE – zaproponowana metoda w dwóch wariantach [opracowanie własne]

Rozwiązania referencyjne zwracają wartość maksymalną lub średnią arytmetyczną współczynnika wiarygodności, w związku z czym dla przypadku identycznej wartości współczynnika w obu hipotezach – uzgodnione rozwiązanie dla każdego z rozwiązań referencyjnych zwraca zawsze tę samą wartość – wartość hipotezy początkowej. Jest to niezależne od odległości hipotez. Z tego powodu, zostały one przedstawione dalej zbiorczo jako „Rozwiązania referencyjne” lub w skrócie „Ref”, który równa się wartości współczynnika wiarygodności.

Wybrane fragmenty zależności współczynnika wiarygodności od odległości hipotez, dla współczynnika wiarygodności równego 0,9 został przedstawiony w Tabeli 9.2. Pełne zestawienie znajduje się w załączniku E.

Tabela 9.2 Zależność wartości współczynnika wiarygodności uzgodnionego rozwiązania od odległości hipotez, dla współczynnika wiarygodności hipotez równego 0,9. Wybrane wartości. Pogrubione zostały wartości graniczne (wartości 0,9 oraz 0) [opracowanie własne].

Distance threshold (Dmax)	Method	Distance											
		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1	0,99	1
0,05	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,95	0,94	0,92	0,9	0,88	0,86	0,84	0,83	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
0,1	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,96	0,95	0,94	0,92	0,91	0,9	0,88	0,87	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
0,15	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,89	0,88	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
0,2	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
0,3	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
0,5	CFIEE1	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	CFIEE2	0,98	0,97	0,96	0,95	0,94	0,93	0,92	0,91	0,9	0,89	0,01	0
	Ref	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9

Zaproponowane rozwiązanie dla wartości zbieżnych (odległość mniejsza od 0,05) zwraca wyższą wartość współczynnika wiarygodności niż wartość dla hipotez początkowych, zwłaszcza wariant CFIEE2, który dla wartości 0,06 prezentuje wyższą wartość współczynnika wiarygodności hipotez, a dla odległości 0,07 już niższą. Dla wariantu CFIEE1 (współczynniki rozkładu normalnego równe 0,6) punkt równy współczynnikowi hipotezy osiągnięty zostaje dla odległości 0,8.

Wraz ze wzrostem wartości progowej odległości, wartość graniczna (zrównania się z wartością hipotezy) przesuwa się bliżej punktu 0,1 – im wyższa wartość progowa odległości, tym wolniej. Punkt z zerową wiarygodnością zaś zawsze wypada w tej samej odległości, równej 1. Warty zauważanie jest również fakt, że referencyjne rozwiązania nie poradziły sobie w żadnym z przypadków powyższej symulacji.

Zwiększenie wartości progowej odległości zwiększa tolerancję dla metody CFIEE w wariancie 2 – wykorzystując wartość progową odległości do określenia parametrów rozkładu. Jest to szczególnie widoczne dla wartości współczynnika wiarygodności między 0,7 a 0,5, gdzie różnica między obydwoma wariantami jest dwukrotna (załącznik E).

W eksperymencie zbadano zachowanie współczynnika wiarygodności uzgodnionego rozwiązania wraz ze zwiększaniem odległości dwóch hipotez od siebie. Dzięki temu znaleziono wartości graniczne, do których osiągnięcia wartość ta była wyższa od hipotez początkowych, a po której zaczęła spadać. Dla niskich wartości progowej odległości (nieprzekraczających 0,1) – zachowanie zaproponowanej metody jest w pełni zgodne z oczekiwaniami – co jest widoczne w Tabeli 9.2.

Dla wyższych wartości progowych odległości współczynnik wiarygodności uzgodnionego rozwiązania jest mniej adekwatny w stosunku do oczekiwanej wartości, zwłaszcza dla wyższych wartości wiarygodności hipotez początkowych – jest to widoczne w danych w załączniku E. Dla współczynnika wiarygodności hipotez początkowych:

- Nie większego niż 0,7, wartości współczynnika wiarygodności uzgodnionego rozwiązania są adekwatne do wartości progowych odległości nie większych niż 0,2,
- Nie większego niż 0,5, wartości współczynnika wiarygodności uzgodnionego rozwiązania są adekwatne do wartości progowych odległości nie większych niż 0,3.
- Niższego niż 0,3, a zwłaszcza niższego niż 0,1, wyniki są bardzo rozmyte i mogą odbiegać od oczekiwanych.

Dla wartości progowej odległości powyżej 0,5 powyższa metoda nie daje optymalnych wyników.

Wnioskiem z powyższych obserwacji jest potrzeba zachowania ostrożności podczas dobierania wartości progowej odległości.

9.2. Symulacja 3. Współczynnik wiarygodności uzgodnionego rozwiązania dla zmiennej wartości współczynnika wiarygodności hipotez

Trzecią zaplanowaną symulacją była weryfikacja poprawności wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania dla sytuacji rozbieżnych hipotez o różnych wartościach wiarygodności. Symulacja ma na celu analizę wpływu poziomu wiarygodności hipotez na współczynnik wiarygodności uzgodnionego rozwiązania. Operacjonalizacja zmiennych znajduje się w Tabeli 9.4

Tabela 9.3 Operacjonalizacja zmiennych dla symulacji 3 [opracowanie własne].

Nazwa eksperymentu	Symulacja 3
Zmienna niezależna	Wiarygodność hipotez początkowych – wartości od 0,1 do 0,9, z krokiem równym 0,05
Zmienna zależna	Współczynnik wiarygodności ustalonego rozwiązania (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold; wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 Odległość hipotez – 0,5 oraz 1 <p>Metoda rozpoznawania emocji</p> <p>Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> MeanState – bazujące na średniej arytmetycznej, NaivePreferePlausibility – preferujące wysoki współczynnik wiarygodności, NaivePrefereConvergence – preferujące wysoką zbieżność hipotez, NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)} * D_{max}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.

Symulacja dotyczyła przypadku dwóch rozbieżnych hipotez o różnych wartościach współczynnika wiarygodności oraz stałej odległości od siebie – przebadano dwa warianty tej odległości (równej 0,5 oraz 1). Hipoteza badawcza eksperymentu brzmiała: dla rozbieżnych hipotez (zwłaszcza o tym samym współczynniku wiarygodności) współczynnik wiarygodności uzgodnionego rozwiązania będzie niższy niż współczynnik hipotez.

Do analizy wykorzystano rozwiązania referencyjne (opisane w rozdziale 8.1) oraz dwa warianty metody CFIEE – różniące się współczynnikami rozkładu normalnego – analogicznie jak dla Symulacji 2.

Zestawienie wyliczonego współczynnika wiarygodności uzgodnionego rozwiązania dla wartości progowej odległości równej 0,05 oraz odległości hipotez 0,5 zostały zestawione w Tabeli 9.5. Rozwiązania referencyjne preferujące zbieżność (NaivePrefereConvergence) oraz wiarygodność (NaivePreferePlausibility) dla wariantu dwóch hipotez działają w tożsamy sposób, także dalej będą prezentowane wspólnie pod nazwą programu preferującego wiarygodność. Również rozwiązania oparte na średnich zwracają tożsame wyniki, dlatego prezentowana będzie jedynie wartość średniej arytmetycznej.

Wszystkie rozwiązania referencyjne, dla równej wartości wiarygodności obu hipotez (przekątna tabeli) zwróciły wartość współczynnika wiarygodności hipotez początkowych. Metoda preferująca wiarygodność zwróciła najwyższą wartość z hipotez, co jest rozwiązaniem niezgodnym z oczekiwaniem. Nieco lepiej poradziła sobie metoda oparta o średnią, gdyż dla różnych wartości wiarygodności wynikowa hipoteza okazała się niższa. Niestety, rozwiązanie to nie bierze pod uwagę odległości (poziomu zbieżności hipotez), wobec czego dwie hipotezy o wysokim poziomie wiarygodności oraz rozbieżnych wynikach zostaną wskazane jako bardziej wiarygodne niż jedna hipoteza o wysokim poziomie i druga o niskim poziomie wiarygodności (np. 0,9 i 0,1), który mógł być spowodowany awarią sprzętu.

Tabela 9.4 Współczynnik wiarygodności uzgodnionego rozwiązania w zależności od współczynników wiarygodności rozbieżnych hipotez A i B. Dane w podziale na metodę. Wartość progowa odległości równa 0,05, odległość hipotez równa 0,5 [opracowanie własne].

Metoda	Wiarygodność hipotezy A	Wiarygodność hipotezy B				
		0,1	0,3	0,5	0,7	0,9
CFIIE1	0,1	0,1				
	0,3	0,21	0,25			
	0,5	0,37	0,37	0,36		
	0,7	0,57	0,54	0,48	0,4	
	0,9	0,79	0,72	0,64	0,55	0,46
CFIIE2	0,1	0,09				
	0,3	0,29	0,15			
	0,5	0,48	0,35	0,25		
	0,7	0,67	0,55	0,45	0,35	
	0,9	0,86	0,75	0,65	0,55	0,45
Mean State	0,1	0,1				
	0,3	0,2	0,3			
	0,5	0,3	0,4	0,5		
	0,7	0,4	0,5	0,6	0,7	
	0,9	0,5	0,6	0,7	0,8	0,9
Naive Prefere Plausibility	0,1	0,1				
	0,3	0,3	0,3			
	0,5	0,5	0,5	0,5		
	0,7	0,7	0,7	0,7	0,7	
	0,9	0,9	0,9	0,9	0,9	0,9

Zaproponowana metoda w obu wariantach zwróciła wyniki zgodne z oczekiwaniami. We wszystkich przypadkach otrzymano spadek poziomu wiarygodności dla uzgodnionej hipotezy wobec wyższej wartości współczynnika wiarygodności z dwóch hipotez. Im większa różnica w poziomie wiarygodności, tym spadek uzgodnionego rozwiązania jest mniejszy – jest to widoczne zwłaszcza dla wariantu 2. Z tego powodu wraz ze wzrostem wiarygodności hipotezy przeciwnej – spada wiarygodność uzgodnionego rozwiązania (wartości na wykresie maleją w prawo) – większa niewiara spowodowana drugą hipotezą rozbieżną.

Analiza dla innych wartości progowych odległości prezentuje analogiczną tendencję, dlatego też nie została tutaj przytoczona.

Zestawienie wyliczonego współczynnika wiarygodności uzgodnionego rozwiązania dla wartości progowej odległości równej 0,05, ale odległości równej 1 zaprezentowano w Tabeli 9.6. Nie zaprezentowano w niej zestawienia dla rozwiązań referencyjnych, gdyż odległość nie ma u nich wpływu na wartość wiarygodności, więc otrzymane wartości są zgodne z tymi w Tabeli 9.5.



Tabela 9.5 Współczynnik wiarygodności uzgodnionego rozwiązania w zależności od współczynników wiarygodności rozbieżnych hipotez A i B. Odległość hipotez równa 1 [opracowanie własne].

Metoda	Wartość progowa odległości (D_{max})	Wiarygodność hipotezy A	Wiarygodność hipotezy B				
			0,1	0,3	0,5	0,7	0,9
CFIEE1	Niezależnie od wartości progowej odległości	0,1	0				
		0,3	0,14	0,04			
		0,5	0,4	0,23	0		
		0,7	0,62	0,43	0,21	0	
		0,9	0,81	0,61	0,41	0,2	0
CFIEE2	0,05	0,1	0,01				
		0,3	0,21	0			
		0,5	0,41	0,2	0		
		0,7	0,6	0,4	0,2	0	
		0,9	0,8	0,6	0,4	0,2	0
CFIEE2	0,5	0,1	0				
		0,3	0,04	0,02			
		0,5	0,18	0,11	0,04		
		0,7	0,39	0,29	0,16	0,02	
		0,9	0,65	0,5	0,34	0,17	0

Przy zwiększonym dystansie hipotez wyraźnie widoczny jest wpływ różnicy wartości współczynnika wiarygodności dla zaproponowanej metody. Przy odległości hipotez równej 1 wiarygodność uzgodnionego rozwiązania w przybliżeniu równa jest ich różnicy, co w przypadku równej wartości hipotez zwraca wynik bliski 0.

Wraz ze wzrostem wartości progowej odległości dla wariantu 2, wartości współczynnika zaufania spadają (kara jest wyższa). Jest to spowodowane łagodniejszym rozmyciem wykresu przy fuzyfikacji (szerszy kształt rozkładu normalnego). Powoduje to większy wzrost wiarygodnością dla wartości zbieżnych, ale także większe kary przy mocno rozbieżnych wartościach (funkcja niewiary – opisana w 5.4). Wyniki dla wariantu 1 są tożsame dla wszystkich wartości progowych odległości, gdyż nie wpływa on na wartości parametrów rozkładu. Pełne zestawienie danych zostało umieszczone w załączniku F.

Wnioskami z tej symulacji jest to, że zaproponowana metoda zadziałała zgodnie z oczekiwaniami – w obu wariantach. Wraz ze wzrostem wartości współczynnika wiarygodności hipotezy przeciwnej – zaufanie do uzgodnionego rozwiązania spada – hipotezy się wzajemnie wykluczają. Rozwiązania referencyjne nie odnotowały podobnych zależności.

9.3. Symulacja 4. Analiza wpływu wartości wiarygodności rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania

Kolejną przeprowadzoną symulacją był pomiar wpływu na współczynnik wiarygodności uzgodnionego rozwiązania, współczynnika wiarygodności hipotezy rozbieżnej, do dwóch zbieżnych ze sobą hipotez o wysokiej wiarygodności, przy stałej odległości od siebie zbieżnych hipotez oraz stałej odległości hipotezy rozbieżnej od bliższej hipotezy zbieżnej. Hipoteza badawcza symulacji została sformułowana następująco: wpływ hipotezy rozbieżnej powinien być proporcjonalny do wartości jej współczynnika wiarygodności. Operacjonalizacja zmiennych znajduje się w Tabeli 9.7.

Tabela 9.6 Operacjonalizacja zmiennych dla symulacji 4 [opracowanie własne].

Nazwa symulacji	Symulacja 4
Zmienna niezależna	Wiarygodność hipotezy rozbieżnej – wartości od 0,2 do 0,8, z krokiem równym 0,2
Zmienna zależna	Współczynnik wiarygodności ustalonego rozwiązania (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Wartość progowa odległości – D_{max} (distance treshold) (wiersze): 0,1; 0,15; 0,2; 0,3; 0,5 Odległość hipotez – 0,5 oraz 1 <p>Metoda rozpoznawania emocji</p> <p>Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> MeanState – bazujące na średniej arytmetycznej, NaivePreferrePlausibility – preferujące wysoki współczynnik wiarygodności, NaivePreferreConvergence – preferujące wysoką zbieżność hipotez, NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)} \cdot D_{max}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.

Do analizy wykorzystano rozwiązania referencyjne (opisane w rozdziale 8.1) oraz dwa warianty metody CFIEE – różniące się współczynnikami rozkładu normalnego – analogicznie jak dla symulacji 2 i 3.

Zestawienie zależności wyliczonego współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej, dla wartości progowej odległości równej 0,05, odległości hipotez 0,5 i współczynnika wiarygodności zbieżnych hipotez równym 0,9, zostały zestawione w Tabeli 9.8. Pełne zestawienie znajduje się w załączniku G.

Tabela 9.7 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej dla odległości hipotezy rozbieżnej od bliższej zbieżnej hipotezy równej 0,5. Dla wartości progowej odległości równej 0,05 oraz współczynnika wiarygodności zbieżnych hipotez równych 0,9 [opracowanie własne].

Wiarygodność rozbieżnej hipotezy	Średnia	Metoda preferująca wysoką wiarygodność	CFIEE1	CFIEE2
0,2	0,67	0,9	0,84	0,82
0,4	0,73	0,9	0,81	0,77
0,6	0,8	0,9	0,77	0,72
0,8	0,87	0,9	0,73	0,67

Rozwiązania referencyjne preferujące zbieżność (NaivePreferConvergence) oraz wiarygodność (NaivePreferPlausibility) zwracają zawsze wartość równą hipotezom zbieżnym, ignorując rozbieżną hipotezę, co jest niezgodne z oczekiwanym.

Rozwiązania bazujące na średniej zwróciły niższą wiarygodność niż zbieżne hipotezy, jednak nie biorą one pod uwagę zbieżności hipotez (nie ma znaczenia, czy hipotezy są zbieżne, czy rozbieżne). Średnia zwraca wyższy poziom współczynnika wiarygodności wraz ze wzrostem wiarygodności przeciwstawnej hipotezy – co jest zachowaniem odwrotnym do oczekiwanego. W związku z powyższymi, żadne rozwiązanie referencyjne nie wskazuje współczynnika wiarygodności w sposób poprawny.

Oba warianty zaproponowanej metody zachowały się zgodnie z oczekiwaniem – wraz ze wzrostem wiarygodności przeciwstawnej hipotezy, wiarygodność rozwiązania spada w zbliżonym stopniu. Różnica między zaproponowanymi wariantami metody polegała na szybkości zmiany – ze względu na różne wartości parametrów rozkładu normalnego.

W Tabeli 9.9 zestawiono zależność wyliczonego współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej wraz ze wzrostem wartości progowej odległości. W zestawieniu ujęto jedynie drugi wariant zaproponowanej metody, gdyż pierwszy wariant, jak i rozwiązania referencyjne posiadają takie same wartości dla wszystkich wartości progowych odległości. Pełne zestawienie danych zostało umieszczone w załączniku G.

Tabela 9.8 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej dla odległości hipotezy rozbieżnej od bliższej zbieżnej hipotezy równej 0,5. Zestawienie w rozbiciu na różne wartości progowe odległości dla drugiego wariantu zaproponowanej metody [opracowanie własne].

Wiarygodność rozbieżnej hipotezy	Wartość progowa odległości (D_{max})					
	0,05	0,1	0,15	0,2	0,3	0,5
0,2	0,82	0,87	0,88	0,87	0,81	0,73
0,4	0,77	0,82	0,83	0,82	0,78	0,72
0,6	0,72	0,76	0,78	0,78	0,75	0,71
0,8	0,67	0,71	0,73	0,73	0,72	0,71

Dla zaproponowanej metody w drugim wariantcie, wraz ze wzrostem wartości progowej odległości do wartości 0,15 rośnie wartość współczynnika wiarygodności uzgodnionego rozwiązania – co jest poprawne i intuicyjne, ze względu na wypłaszczenie krzywej rozkładu normalnego. Jednocześnie, dla wyższych wartości progowych odległości niż 0,15 wartość współczynnika wiarygodności uzgodnionego rozwiązania spada. To mniej intuicyjne zachowanie jest związana ze znoszącymi się dwoma czynnikami – zwiększenia efektu zbieżności dwóch hipotez (wzrost wiarygodności) oraz zwiększeniu znaczenia rozbieżności trzeciej hipotezy (spadek wiarygodności). Działanie to nie jest do końca intuicyjne, jednakże jest poprawne i zgodne z założeniami metody.

Podsumowując wyniki symulacji 3, im wyższa jest wiarygodność rozbieżnej hipotezy, tym niższa wiarygodność uzgodnionego rozwiązania zwróconego przez zaproponowane rozwiązanie. Zachowanie to jest zgodne z oczekiwaniami.

9.4. Symulacja 5. Analiza wpływu odległości rozbieżnej hipotezy na wiarygodność uzgodnionego rozwiązania

Ostatnią przeprowadzoną symulacją był pomiar wpływu na współczynnik wiarygodności uzgodnionego rozwiązania odległości hipotezy rozbieżnej do dwóch zbieżnych ze sobą hipotez o wysokiej wiarygodności. Hipoteza badawcza symulacji została sformułowana następująco: wpływ hipotezy rozbieżnej powinien mieć proporcjonalny wpływ do odległości hipotezy rozbieżnej. Operacjonalizacja zmiennych znajduje się w Tabeli 9.10.

Tabela 9.9 Operacjonalizacja zmiennych dla symulacji 5 [opracowanie własne].

Nazwa symulacji	Symulacja 5
Zmienna niezależna	Odległość hipotezy rozbieżnej od bliższej ze zbieżnych hipotez – Wartości od 0,15 do 1,45, z krokiem równym 0,05
Zmienna zależna	Współczynnik wiarygodności ustalonego rozwiązania (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> Współczynnik wiarygodności hipotezy rozbieżnej, przyjmujący wartości: 0,1; 0,3; 0,5; 0,7; Wartość progowa odległości – D_{max} (distance treshold) (wiersze) – 0,05 <p>Metoda rozpoznawania emocji</p> <p>Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> MeanState – bazujące na średniej arytmetycznej, NaivePreferePlausibility – preferujące wysoki współczynnik wiarygodności, NaivePreferaConvergence – preferujące wysoką zbieżność hipotez, NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)*D_{max}}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości.

Do analizy wykorzystano rozwiązania referencyjne (opisane w rozdziale 8.1) oraz dwa warianty metody CFIEE – różniące się współczynnikami rozkładu normalnego – analogicznie jak dla symulacji 2, 3 oraz 4.

Zestawienie zależności wyliczonego współczynnika wiarygodności uzgodnionego rozwiązania od odległości rozbieżnej hipotezy od hipotez zbieżnych, dla wartości wiarygodności hipotezy rozbieżnej równej 0,4 oraz wartości progowej odległości równej 0,05 zostało zestawione w Tabeli 9.10. Pełne zestawienie danych zostało umieszczone w załączniku G.

Rozwiązania referencyjne preferujące zbieżność (NaivePreferaConvergence) i wiarygodność (NaivePreferePlausibility) zwracają najwyższą wartość współczynnika wiarygodności hipotez zbieżnych, co nie jest zgodne z oczekiwaniami. Rozwiązania oparte o średnie również zwracają wyniki niezależne od odległości hipotez, co w przypadku tej symulacji zwraca wartość stałą dla wszystkich wariantów, równa 0,73 (średnia ze współczynników wiarygodności wszystkich hipotez). Podobnie jak w poprzednich symulacjach – sposób przypisania współczynnika wiarygodności dla rozwiązań referencyjnych jest niezgodny z oczekiwaniami ze względu na brak uwzględnienia zbieżności oraz odległości hipotez.

Tabela 9.10 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od odległości rozbieżnej hipotezy, dla wartości wiarygodności hipotezy rozbieżnej równej 0,4 [opracowanie własne].

Dystans hipotezy rozbieżnej	Średnia	Metoda preferująca wysoką wiarygodność	CFIEE1	CFIEE2
0,15	0,73	0,9	0,87	0,86
0,2	0,73	0,9	0,85	0,84
0,25	0,73	0,9	0,84	0,83
0,3	0,73	0,9	0,82	0,81
0,35	0,73	0,9	0,82	0,8
0,4	0,73	0,9	0,81	0,79
0,45	0,73	0,9	0,81	0,78
0,5	0,73	0,9	0,81	0,77
0,55	0,73	0,9	0,8	0,76
0,6	0,73	0,9	0,8	0,75
0,65	0,73	0,9	0,8	0,74
0,7	0,73	0,9	0,79	0,73
0,75	0,73	0,9	0,78	0,72
0,8	0,73	0,9	0,77	0,71
0,85	0,73	0,9	0,77	0,7
0,9	0,73	0,9	0,76	0,69
0,95	0,73	0,9	0,75	0,68
1	0,73	0,9	0,74	0,67

Z kolei oba warianty zaproponowanej metody zachowały się zgodnie z oczekiwaniami, wraz ze wzrostem dystansu hipotezy rozbieżnej od zbieżnych hipotez, wiarygodność uzgodnionego rozwiązania spada. Wartości dla obu wariantów były zbliżone, pierwszy wariant wykazał wyższe wartości współczynnika wiarygodności.

9.5. Podsumowanie symulacji

Dla zaproponowanej metody integracji wszystkie badane przypadki brzegowe zostały obsłużone w sposób poprawny, z właściwymi tendencjami wzrostu i spadku wartości współczynnika wiarygodności. Wszystkie założone kryteria zostały spełnione.

Przeprowadzone symulacje ukazały duże ograniczenia metod referencyjnych w zakresie wyznaczenia wartości współczynnika wiarygodności uzgodnionego rozwiązania dla przypadków granicznych. Ignorują one w odległości hipotez od siebie, w tym ich zbieżność czy rozbieżność.

W symulacji 2 przeanalizowana została zmiana wartości współczynnika wiarygodności uzgodnionej hipotezy w zależności od odległości hipotez, przy założonym kryterium zbieżności w postaci wartości progowej odległości. Zaproponowana metoda w obu wariantach zadziałała zgodnie z oczekiwaniami dla wartości progowej odległości nie większej niż 0,1. Dla wartości progowej odległości nie większej niż 0,2 metoda działa poprawnie dla poziomu wiarygodności nie większego niż 70%, a dla wartości progowej odległości nie większej niż 0,3 dla poziomu wiarygodności nie większego niż 50%. Dla hipotez o wartości współczynnika wiarygodności poniżej 0,3, a zwłaszcza poniżej 0,1, wyniki są bardzo rozmyte i mogą odbiegać od oczekiwanych. Dla wartości progowej odległości powyżej 0,5 powyższa metoda nie daje optymalnych wyników.

W symulacji 3 przeanalizowano wpływ zmiany współczynnika wiarygodności jednej z hipotez na wskazania współczynnika wiarygodności uzgodnionej hipotezy, w przypadku dwóch rozbieżnych hipotez. Zaproponowana metoda została poprawnie zweryfikowana – wraz ze wzrostem wiarygodności hipotezy przeciwnej (wykluczającej) zaufanie do uzgodnionego rozwiązania spadało.

Czwarta symulacja dotyczyła przypadku trzech hipotez, gdzie dwie były zbieżne ze sobą, a trzecia rozbieżna z nimi. Analizowaną zmienną niezależną była wartość współczynnika wiarygodności hipotezy rozbieżnej. Zgodnie z oczekiwaniem wartość współczynnika wiarygodności uzgodnionego rozwiązania spadała wraz ze wzrostem współczynnika wiarygodności hipotezy rozbieżnej. Dodatkową obserwacją symulacji była analiza wpływu zmiany wartości progowej odległości, dla wariantu metody uwzględniającej wartość progową odległości przy wyznaczeniu parametru rozkładu normalnego. Wraz ze wzrostem wartości progowej odległości do wartości 0,15 – rosła wartość wiarygodności uzgodnionego rozwiązania, a powyżej tej wartości zaczynała maleć. Efekt ten jest spowodowany dwoma znoszącymi się czynnikami: zwiększenia efektu zbieżności dwóch hipotez (wzrost wiarygodności – funkcja zaufania) oraz zwiększeniu znaczenia rozbieżności trzeciej hipotezy (spadek wiarygodności – funkcja niewiary).

Ostatnia symulacja również dotyczyła przypadku trzech hipotez, gdzie dwie były zbieżne ze sobą, a trzecia rozbieżna z nimi. Analizowaną zmienną niezależną była odległość hipotezy rozbieżnej od bliższej ze zbieżnych hipotez. Obydwa warianty zaproponowanej metody zachowały się zgodnie z oczekiwaniami, wraz ze wzrostem dystansu hipotezy rozbieżnej od zbieżnych hipotez, wiarygodność uzgodnionego rozwiązania spada.

Podsumowując wyniki symulacji, sposób wyznaczania współczynnika wiarygodności przez zaproponowaną metodą jest poprawny i spełnił wszystkie sprawdzane kryteria. Zauważono ograniczenie związane z wartością progową odległości, które jednak nie występuje dla wartości progowej odległości mniejszej niż 0,1. Metody referencyjne nie zadziałały prawidłowo w ani jednej z przygotowanych symulacji.

Dalsze badania nad udowodnieniem tezy pomocniczej nr 3, dotyczącej poprawy wiarygodności i niezawodności w wielomodalnym rozpoznawaniu emocji zostały kontynuowane w kolejnym rozdziale.

10. EWALUACJA ZAPROPONOWANEGO ROZWIĄZANIA Z WYKORZYSTANIEM BADANIA Z CZTEREMA KAMERAMI

Dla udowodnienia trzeciej tezy pomocniczej, sformułowanej jako:

Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwi poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

Autor pracy wykorzystał trzecie źródło danych, jakim był eksperyment przeprowadzony na Politechnice Gdańskiej, w którym użytkownicy wykonywali zadania edukacyjnej na platformie e-nauczanie (moodle). Uczestnikami eksperymentu byli studenci, którzy mieli za zadanie przeczytanie wykładu, dodanie wpisu na forum, a także wypełnienie ankiety na temat stanu emocjonalnego oraz wiedzy o platformie e-nauczanie.

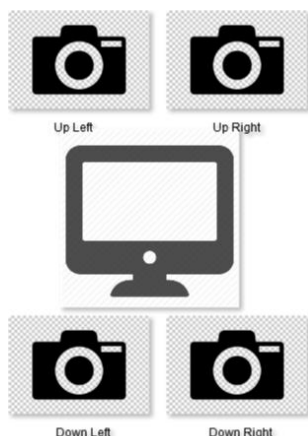
Ze względu na charakter badania, w którym nie jest znany stan emocjonalny wskazany przez eksperta – nie jest możliwa ocena skuteczności, a jedynie ocena poziomu wiarygodności na podstawie zbieżności hipotez. W tym celu została przeprowadzona dokładna analiza poziomu zbieżności danych z eksperymentu, która została opisana w rozdziale 10.3.

Przeprowadzone zostały dwa eksperymenty:

- Eksperyment 8. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na danych z eksperymentu z 4 kamerami (rozdział 10.2) oraz
- Eksperyment 9. Porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez (rozdział 10.4).

10.1. Opis przebiegu badania z czterema kamerami

Badanie zostało przeprowadzone w 2016 roku z użyciem stanowiska Monitora Emocji, którego opis został umieszczony w rozdziale 5.3. Uczestniczyło w nim 15 studentów w wieku od 20 do 21 lat. Czas badania wynosił od 35 do 49 minut, z wyjątkiem jednego badanego, który zakończył eksperyment po 15 minutach. Uczestnicy korzystali z klawiatury, myszy komputerowej oraz jednego monitora, na którego czterech rogach rozmieszczone zostały kamery: górna prawa (ang. up right), górna lewa (ang. up left), dolna prawa (ang. down right) oraz dolna lewa (ang. down left). Rozmieszczenie kamer zostało zobrazowane na Rys 10.1. Rozdzielczość nagrywania to 1280x720 oraz częstotliwość 30 klatek na sekundę (ang. fps). Parametry kamer były intencjonalnie standardowe, by przypominały jakością typowo wykorzystywane do użytku domowego. Jedynym niestandardowym ustawieniem było dodatkowe oświetlenie poprawiające jakość rozpoznawania emocji.



Rys 10.1 Schemat rozmieszczenia kamer w czasie eksperymentu [opracowanie własne]

Górna prawa kamera zwróciła mniejszą liczbę klatek (449 tys. wobec 565–568 tys. w pozostałych kamerach) – prawdopodobnie ze względu na zużycie procesora przez program synchronizujący kamery. Do redukcji różnic między kamerami wykorzystano synchronizację na poziomie 250 ms, w wyniku czego osiągnięto ok. 92 tys. klatek. Liczba klatek przed oraz po redukcji została zaprezentowana w Tabeli 10.1. Po redukcji liczba fragmentów we wszystkich kamerach jest zbliżona. Redukcja została dokonana z wykorzystaniem średniej arytmetycznej. Autor pracy jest świadom, że średnia arytmetyczna nie jest idealną metodą agregacji hipotez afektywnych, gdyż może doprowadzić do przesunięcia uzgodnionego rozwiązania bliżej stanu neutralnego (bliżej zera), jednak w krótkim czasie [40], gwałtowne zmiany emocji nie są duże, także efekt ten powinien być ograniczony.

Tabela 10.1 Liczba klatek z eksperymentu w rozbiciu na cztery kamery oraz uczestników badania, w ujęciu surowym (przed próbkowaniem) oraz po próbkowaniu do 250 ms [opracowanie własne].

Uczestnik	Przed próbkowaniem				Po próbkowaniu do 250ms			
	UL	DR	UR	DL	UL	DR	UR	DL
1	61 918	63 161	46 981	62 666	9 459	9 458	9 453	9 449
2	58 783	59 357	45 449	59 166	9 149	9 145	9 152	9 149
3	65 126	65 583	49 991	64 665	10 048	10 054	10 046	10 053
4	64 210	66 539	49 880	65 534	10 023	10 025	10 027	10 026
5	58 587	56 853	44 408	58 186	8 930	8 931	8 931	8 934
6	55 363	55 149	42 221	55 454	8 510	8 509	8 506	8 505
7	21 630	21 771	18 186	21 579	3 781	3 776	3 775	3 776
8	69 415	69 634	57 699	69 181	11 848	11 849	11 849	11 848
9	59 457	58 912	51 018	59 403	10 833	10 826	10 828	10 831
10	50 833	51 118	43 187	50 934	9 026	9 026	9 027	9 028
Suma	565 322	568 077	449 020	566 768	91 607	91 599	91 594	91 599

10.2. ***Eksperyment 8. Poprawa niezawodności rozpoznawania emocji poprzez zastosowanie obserwacji wielomodalnej na danych z eksperymentu z 4 kamerami***

W tym eksperymencie przeanalizowano dostępność estymat stanu emocjonalnego mimiki twarzy dla danych z eksperymentu na Politechnice Gdańskiej.

Dla programu automatycznego rozpoznawania emocji z mimiki twarzy kluczowym elementem jest wykrycie twarzy. W badaniu zostały zastosowane 4 kamery, ustawione w czterech rogach monitora – szczegóły konfiguracji zostały opisane w rozdziale 10.1. Do analizy przyjęty został 250 ms przedział czasowy, a wynik analizy przedstawia Tabela 10.2.

Tabela 10.2 Zestawienie dostępności modalności mimiki twarzy dla danych eksperymentu przeprowadzonego na Politechnice Gdańskiej [opracowanie własne].

Dostępne kanały obserwacji	Liczba próbek	Dostępność [%]
Brak dostępnych kanałów	3 346	3,65%
Dane w 4 kanałach obserwacji	69 675	76,04%
Kamera DR	78 074	85,21%
Dane w co najmniej 3 kanałach obserwacji	78 660	85,85%
Kamera DL	79 479	86,74%
Kamera UL	80 853	88,24%
Kamery DL+DR	82 738	90,30%
Kamera UR	82 882	90,46%
Dane w co najmniej 2 kanałach obserwacji	84 672	92,41%
Kamery DL+UL	85 050	92,82%
Kamery UL+UR	85 271	93,06%
Kamery UL+DR	85 410	93,21%
Kamery DL+DR+UL	86 307	94,19%
Kamery DL+UR	86 883	94,82%
Kamery DR+UR	87 050	95,00%
Kamery DL+UL+UR	87 432	95,42%
Kamery DR+UL+UR	87 882	95,91%
Kamery DL+DR+UR	87 894	95,93%
Kamery DL+DR+UL+UR	88 281	96,35%
Wszystkie ramki	91 627	100,00%

Dane z pojedynczych kamer osiągają dosyć wysoką dostępność – od 85,21% do 90,46%. Dla 3,65% przypadków w żadnej z kamer nie udało się wykryć twarzy, a połączenie danych ze wszystkich kamer pozwoliło uzyskać dostępność na poziomie 96,35% – wzrost o prawie 6 punktów procentowych wobec najkorzystniej ustawionej kamery.

Dane we wszystkich czterech kamerach uzyskano w 76% przypadków, w co najmniej trzech w 85% przypadków (co jest zbliżone dostępnością pojedynczej kamery), w co najmniej dwóch kamerach w 92% przypadków. Jest to szczególnie ważne w kontekście wyboru metody fuzji, zwłaszcza dla fuzji wczesnej, dla której brak jednego z kanałów poważnie wpływa na skuteczność integracji. W przypadku danych z tego badania jest to prawie 25% przypadków z niepełnym wektorem danych. Jest to jednak mniej znaczący problem niż w przypadku danych ze zbioru SEMAINE (opisane w rozdziale 7.4). W przypadku późnej fuzji brak dostępności kanału nie stanowi tak istotnego problemu.

Na podstawie obserwacji z tego eksperymentu oraz z eksperymentu 4 z analizy danych ze zbioru SEMAINE można potwierdzić część postawionej tezy pomocniczej, mówiącej o poprawie niezawodności poprzez analizę wielomodalną.

10.3. **Określenie poziomu zbieżności danych dla eksperymentu z Politechniki Gdańskiej dla czterech kamer**

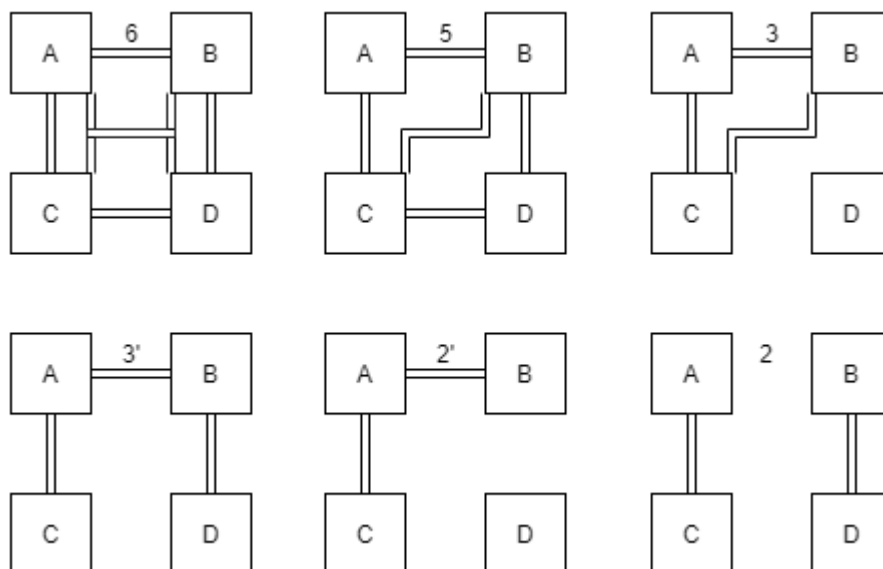
Aby sprawdzić poziom zbieżności danych z eksperymentu z czterema kamerami, wykorzystano próbkowanie danych z dokładnością do 250 ms, ze względu na wolniejsze działanie jednej z kamer. Szczegóły eksperymentu zostały opisane w rozdziale 10.1. Dane do eksperymentu zostały ograniczone do danych z dostępnymi wszystkimi czterema kamerami – czyli 76% wszystkich danych. Dane z kamer zostały przeanalizowane za pomocą programu Noldus FaceReader [50], który analizuje mimikę twarzy. Noldus FaceReader zwraca wynik w modelu 2D oraz modelu emocji podstawowych Ekmana.

Do opisu poziomu zbieżności danych z czterech kanałów autor pracy zaproponował skalę od 0 do 6, gdzie numer na skali informuje o liczbie „relacji zbieżności” między poszczególnymi hipotezami. Zbieżność jest określana na podstawie wartości progowej odległości. Opis skali został przedstawiony w Tabeli 10.3. Elementy skali oznaczone poprzez ‘ oznaczają typ „łańcuchowy” (nieprzechodni), gdzie zbieżność nie występuje w obrębie całej grupy – skrajne elementy nie są zbieżne ze sobą, mimo wspólnych zbieżnych hipotez np. 2’ oznacza sytuację, gdzie hipoteza B jest zbieżna z A i z C, ale A nie jest zbieżna z C.

Tabela 10.3 Opis zaproponowanej skali zbieżności dla czterech kanałów obserwacji [opracowanie własne].

Oznaczenie skali	Opis
6	Pełna zbieżność wszystkich 4 hipotez
5	Pełna zbieżność 3 hipotez i dwie hipotezy zbieżne z czwartą hipotezą – dwie skrajne hipotezy nie są zbieżne ze sobą.
4	Pełna zbieżność 3 hipotez i jedna relacja zbieżności z rozbieżną hipotezą
3	Pełna zbieżność 3 hipotez
3'	Typ „łańcuchowy” (nieprzechodni), gdzie jedna hipoteza jest w relacji zbieżności maksymalnie z dwoma hipotezami np. A jest zbieżna z B, B jest zbieżna z C, C jest zbieżna z D (A–B–C–D), ale C nie jest zbieżna z A, a D nie jest zbieżna ani z A ani z B.
2	Dwie grupy zbieżnych hipotez np. A zbieżna z B, C zbieżna z D
2'	Typ „łańcuchowy”, gdzie jedna hipoteza jest w relacji zbieżności maksymalnie z dwoma hipotezami np. B jest zbieżna z A i C, a hipoteza D nie jest zbieżna z żadną z hipotez.
1	Tylko dwie hipotezy są zbieżne
0	Brak zbieżności

Uproszczona wizualizacja kilku wybranych wartości skali została przez autora zaprezentowana na Rys 10.2. Kwadraty reprezentują poszczególne hipotezy, a linie reprezentują zbieżność pomiędzy hipotezami.



Rys 10.2 Wizualizacja kilku wybranych wartości zaproponowanej skali zbieżności dla czterech kanałów obserwacji [opracowanie własne]

Dla lepszej i lepszego zrozumienia i wizualizacji zaproponowanej skali, w Tabeli 10.4 zestawiono przykładowe wartości hipotez dla wartości progowej odległości równej 0,05.

Tabela 10.4 Przykładowe wartości czterech hipotez (oznaczonych jako A, B, C oraz D) dla poszczególnych klas zbieżności wraz z opisem [opracowanie własne].

Klasa zbieżności	A	B	C	D	Opis poziomu zbieżności (komentarz)
6	0,5	0,51	0,53	0,54	Wszystkie hipotezy są ze sobą zbieżne – najbardziej odległe hipotezy A i D są w odległości 0,04, mniejszej niż zadany próg
5	0,5	0,52	0,53	0,56	Wszystkie hipotezy są ze sobą zbieżne – z wyjątkiem hipotez A i D, których odległość wynosi 0,06 (więcej niż zadany próg)
4	0,5	0,52	0,54	0,58	Hipotezy A,B i C są ze sobą zbieżne oraz hipotezy C i D są zbieżne ze sobą.
3'	0,5	0,54	0,58	0,62	Hipoteza A i B są zbieżne, B i C są zbieżne ze sobą oraz C i D są zbieżne ze sobą.
3	0,5	0,52	0,54	0,7	Hipotezy A, B oraz C są zbieżne ze sobą, D jest rozbieżna z pozostałymi hipotezami.
2'	0,5	0,54	0,58	0,7	Hipotezy A i B są zbieżne ze sobą, B i C są zbieżne, a D jest rozbieżna z pozostałymi.
2	0,5	0,52	0,7	0,72	Hipotezy A i B są zbieżne ze sobą oraz C i D zbieżne ze sobą.
1	0,5	0,52	0,7	0,9	Hipotezy A i B są zbieżne ze sobą. Hipotezy C i D są rozbieżne z pozostałymi hipotezami.
0	0,5	0,56	0,65	0,7	Żadna z hipotez nie jest ze sobą zbieżna – odległość pomiędzy dowolną z hipotez jest większa niż wartość progowa odległości (0,05)

Procentowy udział przypadków z pełną zbieżnością był najwyższy dla emocji strachu oraz wstrętu i wyniósł odpowiednio 91,11% oraz 85,92% przypadków dla wartości progowej odległości równej 0,05. Wraz ze wzrostem wartości progowej odległości, liczba w pełni zbieżnych przypadków rosła, zaś przypadki braku jakiejkolwiek zbieżności zniknęły przy wartości progowej równej 0,2 – wartości pogrubione. Dane dla tych emocji zostały zebrane w Tabeli 10.5.

Tabela 10.5 Procentowany rozkład przypadków w rozbiciu na poziomy zbieżności dla emocji zniechęcenia i strachu w podziale na wartości progowe odległości (D_{max}) oraz zaproponowane klasy zbieżności [opracowanie własne].

Convergence	Disgusted						Scared					
	Distance treshold (D_{max}) [%]											
	0,05	0,10	0,15	0,20	0,30	0,50	0,05	0,10	0,15	0,20	0,30	0,50
6	85,92	89,87	91,99	93,42	95,32	97,40	91,11	94,29	95,84	96,88	98,00	99,08
5	0,29	0,33	0,26	0,22	0,15	0,16	0,10	0,07	0,07	0,04	0,02	0,02
4	0,98	0,80	0,80	0,75	0,72	0,67	0,84	0,65	0,57	0,42	0,30	0,20
3'	0,14	0,09	0,07	0,07	0,08	0,03	0,03	0,02	0,02	0,01	0,01	0,00
3	8,13	6,24	5,02	4,17	2,99	1,50	6,02	4,16	3,09	2,39	1,55	0,67
2'	0,40	0,19	0,14	0,12	0,04	0,00	0,11	0,04	0,02	0,01	0,01	0,00
2	0,73	0,68	0,65	0,62	0,51	0,24	0,51	0,30	0,17	0,14	0,10	0,03
1	3,00	1,69	1,05	0,63	0,20	0,00	1,24	0,45	0,21	0,11	0,02	0,00
0	0,42	0,11	0,02	0,00	0,00	0,00	0,05	0,02	0,00	0,00	0,00	0,00

Podobny odsetek przypadków zbieżności wykazała emocja radości oraz zaskoczenia. Emocja radości wykazała 78% przypadków pełnej zbieżności dla wartości progowej odległości równej 0,05, a emocja zaskoczenia 65% przypadków pełnej zbieżności oraz 20% przypadków pełnej zbieżności dla trzech hipotez.

Średni odsetek przypadków pełnej zbieżności osiągnęły emocje gniewu oraz smutku. Dla wartości progowej odległości równej 0,05 dla emocji gniewu tylko 29% przypadków osiągnęło pełną zbieżność danych z 4 kamer, 11% pełnej zbieżności dla trzech kamer, 14,5% przypadków z dwoma grupami zbieżnych hipotez oraz aż 37% przypadków, gdzie tylko dwie hipotezy były ze sobą zbieżne. Przy wartości progowej odległości równej 0,5, jedynie 68% danych było w pełni zbieżnych.

Dla wartości progowej odległości równej 0,05, 42% przypadków emocji smutku było zbieżnych dla 4 kamer, 16% dla trzech, 7% dla dwóch oraz 24% przypadków, gdzie tylko dwie hipotezy były ze sobą zbieżne. Przy wartości progowej odległości równej 0,5, jedynie 73% danych było w pełni zbieżnych. Dane dla emocji smutku oraz gniewu zestawiono w Tabeli 10.6

Tabela 10.6 Zbieżność danych z czterech kamer dla emocji gniewu oraz smutku w podziale na wartości progowe odległości (D_{max}) oraz zaproponowane klasy zbieżności [opracowanie własne].

Convergence	Angry						Sad					
	Distance treshold (D_{max}) [%]											
	0,05	0,1	0,15	0,2	0,3	0,5	0,05	0,1	0,15	0,2	0,3	0,5
6	29,12	37,28	44,07	49,23	58,24	68,96	42,81	51,6	56,75	60,21	65,48	73,88
5	0,43	0,78	0,67	0,92	0,87	0,63	0,23	0,28	0,34	0,46	0,74	1,31
4	3,01	4,93	5,74	6,49	5,65	6,93	2,94	2,76	2,79	3,01	3,82	5,74
3'	0,28	0,39	0,58	0,63	0,37	0,56	0,11	0,18	0,28	0,44	0,74	1,1
3	11,09	11,91	11,43	10,9	9,35	5,44	16,34	14,54	13,14	12,29	10,77	6,63
2'	0,98	1,02	0,89	0,69	0,43	0	0,79	0,9	1,05	1,1	0,94	0
2	14,59	19,17	20,95	21,06	21,17	17,49	7,01	9,97	12,24	13,58	14,33	11,33
1	36,86	23,28	15,26	9,93	3,91	0	23,69	17,31	12,43	8,57	3,17	0
0	3,65	1,23	0,42	0,14	0	0	6,08	2,46	0,98	0,35	0,01	0

Najniższy procentowy rozkład zbieżności zauważono dla wymiarów emocji neutralnej oraz wymiarów pobudzenia i wartościowości – zbieżność 4 kamer dla wartości progowej odległości równej 0,05 nie przekroczyła 8%! Zbieżność na poziomie 19% przypadków dla trzech kamer osiągnął tylko wymiar pobudzenia. Zbieżność dla dwóch kamer wyniosła odpowiednio: 49%, 27,5% oraz 43%. Przy wartości progowej odległości równej 0,5, dla pobudzenia zbieżność czterech kamer wyniosła 99,6%, co oznacza, że wartości były zbliżone do siebie, a dla emocji neutralnej i wartościowości wyniosły 60 i 56% – co wskazuje na duże różnice we wskazaniach. Pełne zestawienie danych znajduje się w załączniku G.

Analiza zbieżności dla poszczególnych wymiarów, wykazała duże różnice we wskazaniach dla poszczególnych wymiarów – jest to argument za indywidualną analizą integracji każdej z emocji, a zwłaszcza przypisaniem indywidualnego współczynnika wiarygodności. Dla czterech emocji zbieżność była wysoka, dla dwóch średnia, a dla 3 bardzo niska. Poziom współczynnika zaufania powinien być wskazany indywidualnie dla każdego wymiaru. Poziom rozbieżności wskazuje na potencjał zintegrowania danych.

10.4. ***Eksperyment 9. Porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez***

Celem ostatniego eksperymentu było porównanie wartości współczynnika wiarygodności uzgodnionej hipotezy z poziomem zbieżności hipotez. Teza badawcza eksperymentu brzmiała: im wyższy poziom zbieżności hipotez, tym wyższy powinien być współczynnik wiarygodności uzgodnionego rozwiązania. Eksperyment 9 został przeprowadzony na danych z badania przeprowadzonego na Politechnice Gdańskiej z wykorzystaniem czterech kamer (opis badania w rozdziale 10.1). Integracja została przeprowadzona poprzez wykorzystanie zaproponowanej metody CFIEE oraz zaproponowanych rozwiązań referencyjnych z rozdziału 0. Operacjonalizacja zmiennych dla eksperymentu znajduje się w Tabeli 10.7.

Tabela 10.7 Operacjonalizacja zmiennych dla eksperymentu 9 – porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez [opracowanie własne].

Nazwa eksperymentu	Eksperyment 9
Zmienna niezależna	<p>Poziom zbieżności hipotez, zgodnie ze skalą:</p> <ul style="list-style-type: none"> • 6 – Pełna zbieżność wszystkich 4 hipotez • 5 – Pełna zbieżność 3 hipotez i dwie hipotezy zbieżne z czwartą hipotezą – dwie skrajne hipotezy nie są zbieżne ze sobą. • 4 – Pełna zbieżność 3 hipotez i jedna relacja zbieżności z rozbieżną hipotezą • 3 – Pełna zbieżność 3 hipotez • 3' – Typ „łańcuchowy”, gdzie jedna hipoteza jest w relacji zbieżności maksymalnie z dwoma hipotezami np. A jest zbieżna z B, B jest zbieżna z C, C jest zbieżna z D (A–B–C–D), ale C nie jest zbieżna z A, a D nie jest zbieżna ani z A ani z B. • 2 – Dwie grupy zbieżnych hipotez np. A zbieżna z B, C zbieżna z D • 2' – Typ „łańcuchowy”, gdzie jedna hipoteza jest w relacji zbieżności maksymalnie z dwoma hipotezami np. B jest zbieżna z A i C, a hipoteza D nie jest zbieżna z żadną z hipotez. • 1 – Tylko dwie hipotezy są zbieżne • 0 – Brak zbieżności
Zmienna zależna	Współczynnik wiarygodności ustalonego rozwiązania (komórki) – wartości procentowe (0–100%)
Pozostałe zmienne	<ul style="list-style-type: none"> • Wartość progowa odległości – D_{max} (distance treshold) (wiersze) – 0,1; 0,15; 0,2; 0,3; 0,5 • Wymiar – Radość, Smutek, Gniew, Wstręt, Strach, Zaskoczenie, Wartościowość, Pobudzenie <p>Metoda rozpoznawania emocji: Dla rozwiązań referencyjnych, wielomodalnych/wielokanałowych:</p> <ul style="list-style-type: none"> • MeanState – bazujące na średniej arytmetycznej, • NaivePreferPlausibility – preferujące wysoki współczynnik wiarygodności, • NaivePreferConvergence – preferujące wysoką zbieżność hipotez, • NaiveWeightedByPlausibility – bazujące na średniej ważonej, gdzie wagą jest wartość współczynnika wiarygodności <p>Dla zaproponowanej metody</p> <ul style="list-style-type: none"> • CFIEE 1 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego równymi 0,6 • CFIEE 2 – zaproponowana metoda, z parametrami wyznaczonymi w eksperymencie 1, z parametrami rozkładu normalnego wyznaczanymi wzorem $\sqrt{(2\pi)} \cdot D_{max}$, gdzie D_{max} (distance treshold) oznacza wartość progową odległości,

Dane w przypadku ogólny bez podziału na emocje

Autor pracy porównał wyniki dla rozwiązań referencyjnych oraz zaproponowanej metody integracji CFIEE w dwóch wariantach: z parametrami rozkładu normalnego równymi: (1) 0,6 oraz (2) zależnych od wartości progowej odległości D_{max} . Zestawienie wartości współczynnika wiarygodności uzgodnionego rozwiązania w podziale na poziom zbieżności, metodę integracji oraz wartość progową odległości (dla wartości równych 0,05, 0,1 oraz 0,15) znajduje się w Tabeli 10.8. Pełne zestawienie danych wraz z rozbiem na poszczególne emocje znajdują się w załączniku I.

Oba warianty zaproponowanej metody zachowały się podobnie i zgodnie z oczekiwaniem – wraz ze spadkiem poziomu zbieżności (mniejsza liczba „relacji zbieżności”) wartość współczynnika wiarygodności również spadała. Od wartości 99,69% dla przypadków pełnej zbieżności wszystkich hipotez, do 69,3% dla przypadków braku zbieżności. Podczas zmiany wartości progowej odległości tendencja jest zachowana – jednak wartości współczynnika wiarygodności są niższe dla wyższych wartości progowych odległości – zarówno dla pełnej zbieżności, jak i braku zbieżności, aż do wartości 95% dla pełnej zbieżności oraz 22% dla braku zbieżności dla wartości progowej odległości równej 0,5.

Tabela 10.8 Wartość współczynnika wiarygodności uzgodnionego rozwiązania w podziale na poziom zbieżności, wartość progową odległości oraz metodę integracji. Bez podziału na emocję [opracowanie własne].

Distance threshold (Dmax)	Convergence scale	N	CFIEE 1	CFIEE 2	Mean State	Naive Prefere Plausibility	Naive Prefere Convergence	Naive Weighted By Plausibility
0.05	6	282 389	99,69%	99,69%	92,64%	94,86%	94,86%	92,64%
0.05	5	5 936	97,53%	97,55%	93,08%	95,06%	94,93%	93,08%
0.05	4	15 172	96,88%	96,92%	92,93%	94,93%	94,74%	92,93%
0.05	3'	4 920	95,45%	95,59%	93,05%	95,03%	94,76%	93,05%
0.05	3	71 720	92,52%	92,27%	91,55%	94,34%	93,93%	91,55%
0.05	2'	14 844	91,46%	91,49%	91,99%	94,50%	94,03%	91,99%
0.05	2	43 474	76,93%	77,16%	92,76%	94,91%	94,28%	92,76%
0.05	1	138 854	76,25%	75,91%	91,87%	94,46%	93,90%	91,87%
0.05	0	49 766	69,30%	72,11%	91,18%	94,10%	93,53%	91,18%
0.10	6	332 283	99,25%	99,26%	92,65%	94,86%	94,86%	92,65%
0.10	5	12 334	95,11%	95,12%	92,66%	94,81%	94,68%	92,66%
0.10	4	24 325	93,70%	93,72%	92,81%	94,89%	94,66%	92,81%
0.10	3'	7 711	91,23%	91,33%	92,16%	94,52%	94,20%	92,16%
0.10	3	70 123	88,93%	88,64%	91,26%	94,20%	93,78%	91,26%
0.10	2'	12 973	85,77%	85,67%	91,49%	94,22%	93,76%	91,49%
0.10	2	57 953	69,50%	69,61%	92,47%	94,79%	94,32%	92,47%
0.10	1	93 147	70,75%	70,30%	91,55%	94,32%	93,70%	91,55%
0.10	0	16 226	64,95%	66,74%	90,54%	93,77%	93,13%	90,54%
0.15	6	406 687	98,73%	98,73%	92,65%	94,85%	94,85%	92,65%
0.15	5	14 695	92,72%	92,74%	92,08%	94,49%	94,33%	92,08%
0.15	4	26 441	90,62%	90,63%	92,46%	94,68%	94,42%	92,46%
0.15	3'	7 669	86,96%	87,02%	91,87%	94,36%	94,03%	91,87%
0.15	3	56 338	85,63%	85,36%	91,05%	94,12%	93,69%	91,05%
0.15	2'	8 690	80,56%	80,43%	91,31%	94,17%	93,70%	91,31%
0.15	2	67 056	64,80%	64,84%	92,37%	94,76%	94,28%	92,37%
0.15	1	37 704	66,88%	66,48%	91,16%	94,12%	93,46%	91,16%
0.15	0	1 795	61,45%	62,46%	89,68%	93,27%	92,55%	89,68%

W rozwiązaniach referencyjnych brak jest tak wyraźnej różnicy w wartościach współczynnika wiarygodności, jak to ma miejsce w przypadku zaproponowanej metody. Wskazania współczynnika wiarygodności wahają się od 90 do 95%. Trend zgodny ze zmianą skali zbieżności jest najbardziej widoczny dla wymiaru pobudzenia oraz radości dla programu bazującego na średniej arytmetycznej. Dla pozostałych wymiarów tendencja zaczyna być widoczna od wartości progowej odległości 0,3. Pojawienie się tego trendu jest prawdopodobnie konsekwencją wskazań współczynnika wiarygodności hipotez, które mają wyższy współczynnik jakości dla zbieżnych przypadków (lepsze warunki). Różnice nie są wystarczająco znaczące, by twierdzić, że wpływ na to ma poziom zbieżności.

Poza przypadkami pełnej zbieżności, która są łatwiejsze do interpretacji, przeanalizowano także przypadki odstępstw od zaproponowanej skali. Dla zaproponowanej metody przy wartości progowej odległości od 0,1 do 0,2, wartości współczynnika wiarygodności są wyższe dla przypadku „1” (dwie zbieżne hipotezy) niż dla przypadku „2” (dwie grupy zbieżnych hipotez) – jest to przypadek trudny do rozstrzygnięcia – zależny od odległości zbieżnych grup hipotez od siebie. Może to być spowodowane niedużymi odległościami między rozbieżnymi hipotezami („prawie” zbieżnymi) oraz problemem z odpowiednim promowaniem zbieżności i karaniem rozbieżności dla wysokich wartości współczynnika wiarygodności (około 90%) dla wartości progowej odległości większej niż 0,1, który został opisany w symulacji 2 (rozdział 9.1). Dla wartości progowych odległości większych od 0,3 występuje także przypadek większego współczynnika wiarygodności dla „3” zamiast dla „3”.

Dane w podziale na poszczególne wymiary

Tendencje w podziale na poszczególne emocje dla zaproponowanej metody są zbliżone do uśrednionego przypadku – są zgodne z poziomem zbieżności. Jednakże występują tam również odstępstwa. Wraz ze wzrostem wartości progowej odległości, liczba odstępstw rośnie, czego przyczyny zostały opisane w przypadku ogólnym.

Najlepszy rezultat został osiągnięty dla wymiarów pobudzenia, prawdopodobnie ze względu na duże podobieństwo wyników dla wszystkich kamer. Dane dla wymiaru pobudzenia zestawiono w Tabeli 10.9. Przypadki braku zbieżności zniknęły przy wartości progowej odległości równej 0,15, przypadki pojedynczych zbieżności przy wartości równej 0,3 oraz przypadki dwóch zbieżności („2” oraz „2'”) zniknęły przy wartości progowej odległości 0,5.

Dla wartości progowej odległości nieprzekraczającej wartości 0,15 wszystkie wskazania są adekwatne do oczekiwanych – wraz z obniżeniem wartości na skali zbieżności, zwracana wartość współczynnika wiarygodności uzgodnionego rozwiązania jest coraz mniejsza.

Odstępstwa od prawidłowego działania pojawiają się dopiero dla wyższych wartości progowej odległości (powyżej 0,15), dla których wskazano ograniczenia w rozdziale 9. Dane te zostały zebrane w załączniku I.

Dla wymiaru pobudzenia odstępstwa pojawiają się dopiero od wartości równej 0,3 – przypadki „3” mają wyższy współczynnik wiarygodności od „3'” oraz „4”. Pierwsze odstępstwa dla emocji radości, zaskoczenia oraz wartościowości pojawiły się dla wartości progowej odległości większej od 0,15. Współczynnik wiarygodności dla przypadku „1” był wyższy niż dla przypadku „2” (dwie zbieżne grupy hipotez), co mogło być spowodowane dużą odległością od siebie tych dwóch zbieżnych grup. Dla wartości progowej odległości równej 0,3 przypadki „3” (pełna zbieżność trzech hipotez) mają wyższy współczynnik wiarygodności od „3'”, a dla wartości progowej równej 0,5 również wyższy współczynnik wiarygodności od „4” (pełna zbieżność 3 hipotez i dodatkowa relacja zbieżności z czwartą hipotezą).

Tabela 10.9 Wartość współczynnika wiarygodności uzgodnionego rozwiązania w podziale na poziom zbieżności, wartość progową odległości oraz metodę integracji. Wartości dla wymiaru pobudzenia [opracowanie własne].

Distance threshold (D_{max})	Convergence scale	N	CFIEE 1	CFIEE 2	Mean State	Naive Prefere Plausibility	Naive Prefere Convergence	Naive Weighted By Plausibility
0.05	6	5 043	98,40%	98,40%	93,79%	95,60%	95,60%	93,79%
0.05	5	3 261	97,48%	97,50%	93,33%	95,33%	95,15%	93,33%
0.05	4	4 702	96,70%	96,74%	93,00%	95,11%	94,87%	93,00%
0.05	3'	3 042	95,41%	95,56%	92,81%	94,96%	94,64%	92,81%
0.05	3	13 887	95,10%	95,00%	92,77%	95,01%	94,62%	92,77%
0.05	2'	7 000	93,31%	93,42%	91,97%	94,54%	94,04%	91,97%
0.05	2	9 727	91,42%	91,91%	92,48%	94,73%	94,25%	92,48%
0.05	1	19 224	89,50%	89,96%	91,36%	94,12%	93,45%	91,36%
0.05	0	3 789	83,96%	86,63%	90,15%	93,41%	92,64%	90,15%
0.10	6	20 778	97,02%	97,02%	93,30%	95,29%	95,29%	93,30%
0.10	5	7 924	95,11%	95,13%	92,51%	94,79%	94,66%	92,51%
0.10	4	10 296	93,62%	93,65%	92,38%	94,73%	94,51%	92,38%
0.10	3'	4 323	91,28%	91,40%	91,60%	94,28%	93,88%	91,60%
0.10	3	12 075	92,37%	92,24%	91,83%	94,48%	94,11%	91,83%
0.10	2'	3 082	88,82%	88,83%	90,62%	93,69%	93,03%	90,62%
0.10	2	7 179	87,12%	87,40%	91,46%	94,08%	93,66%	91,46%
0.10	1	3 868	84,79%	84,89%	90,42%	93,60%	92,77%	90,42%
0.10	0	150	79,55%	80,69%	88,40%	92,63%	90,91%	88,40%
0.15	6	38 292	95,85%	95,85%	92,99%	95,10%	95,10%	92,99%
0.15	5	9 222	92,76%	92,78%	91,74%	94,37%	94,19%	91,74%
0.15	4	8 832	90,67%	90,68%	91,58%	94,25%	93,89%	91,58%
0.15	3'	2 564	87,34%	87,41%	90,70%	93,71%	93,25%	90,70%
0.15	3	5 909	89,77%	89,63%	91,12%	94,08%	93,62%	91,12%
0.15	2'	794	84,41%	84,31%	89,73%	93,32%	92,11%	89,73%
0.15	2	3 521	83,20%	83,34%	90,95%	93,75%	93,17%	90,95%
0.15	1	541	80,64%	80,49%	89,03%	92,59%	91,28%	89,03%

Sytuacje, w których przypadki pełnej zbieżności dla trzech hipotez mają wyższy współczynnik wiarygodności od przypadków „3”, występują dla emocji wstrętu (od wartości progowej odległości 0,2), neutralnej (od wartości progowej 0,3), strachu (od wartości progowej odległości równej 0,15) oraz smutku (od wartości progowej odległości równej 0,5). Dla większości emocji z wartością progową równą 0,5 wskazania dla przypadków „3” są również wyższe niż dla „4”.

Dla emocji gniewu, smutku, neutralnej oraz strachu odstępstwa występują także od wartości progowej odległości równej 0,05 – w obrębie przypadków „0”, „1”, „2” oraz „2’”. Pełna zbieżność we wszystkich przypadkach osiąga zawsze najwyższy poziom współczynnika wiarygodności.



Podsumowując wyniki eksperymentu – zaproponowana metoda bardzo wyraźnie różnicuje wartość współczynnika wiarygodności względem poziomów zbieżności, czego nie można powiedzieć o rozwiązaniach referencyjnych. Jest to szczególnie istotne przy integracji w rozwiązaniach czasu rzeczywistego oraz takich, gdzie nie mamy informacji na temat prawdziwego stanu emocjonalnego, a możemy bazować jedynie na zbieżności i poziomie współczynnika wiarygodności.

10.5. **Podsumowanie eksperymentów**

Przeprowadzone eksperymenty na danych z badania na czterech kamerach pozwoliły na ukończenie ewaluacji zaproponowanej metody integracji danych. Uzyskane wyniki potwierdzają możliwość zwiększenia niezawodności poprzez analizę wielomodalną oraz poprawność wyznaczania współczynnika wiarygodności uzgodnionego rozwiązania.

Wyniki eksperymentu nr 8 wykazały poprawę niezawodności rozwiązania poprzez zastosowanie analizy wielokanałowej. Dostępność kanału z pojedynczej kamery wahała się od 85 do 90%, a wykorzystując cztery kamery – uzyskano dostępność na poziomie 96,35%. Wykorzystanie dwóch najlepiej ustawionych kamer (w tym wypadku prawych) pozwoliło już na poprawę dostępności do 95%, co pokazuje, że dodanie nawet jednego dodatkowego kanału pozwala na poprawę dostępności.

Wyniki eksperymentu 9 dotyczyły weryfikacji powiązania współczynnika wiarygodności uzgodnionego rozwiązania, wyznaczonego przez zaproponowaną metodę ze spójnością wskazań poszczególnych kanałów. Wyznaczony współczynnik wiarygodności był tym większy, im większa spójność rozwiązań, a rozwiązania o niskiej spójności wyników otrzymywały wyraźnie niższy współczynnik wiarygodności. Wynik ten pozwala na wykorzystanie metody w przypadkach rzeczywistych, w których nie znamy referencyjnego rozwiązania, a możemy korzystać tylko ze wskazań wielu rozwiązań automatycznego rozpoznawania emocji.

Wraz z wynikami poprzednich badań pozwala to przyjąć ostatnią tezę pomocniczą mówiącą, że wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnych umożliwia poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji. Udowodnienie wszystkich tez pomocniczych, pozwala również na przyjęcie tezy głównej rozprawy doktorskiej.

11. PODSUMOWANIE

Głównym celem rozprawy doktorskiej było zaproponowanie metody integracji danych o stanie emocjonalnym użytkowników komputerów, która uwzględnia różne programy, kanały obserwacji oraz modele reprezentacji emocji, a także kontekstową niepewność pomiarów symptomów emocji. Cel nadrzędny został zdekomponowany na trzy cele:

- zaproponowanie metody wyznaczania niepewności pomiarowej związanej z kanałem obserwacji na podstawie danych dostępnych w czasie trwania pomiaru;
- zaproponowanie metody integracji bazującej na współczynniku niepewności wyznaczanym dynamicznie w czasie działania programu;
- przeprowadzenie analizy zaproponowanej metody integracji oraz wyników z poszczególnych programów automatycznego rozpoznawania emocji.

Realizacja powyższych celów posłużyła do potwierdzenia tez sformułowanych w rozdziale 1.3. Wszystkie tezy zostały potwierdzone. Aby tego dokonać, przeprowadzono przegląd literatury dotyczący istniejących modeli reprezentacji emocji, sposobów ich odwzorowań, a także opis istniejących metod rozpoznawania w podziale na najczęściej wykorzystywaną modalności. Następnie omówiono obszary zastosowań automatycznego rozpoznawania emocji wraz z ich aktualnymi wyzwaniami badawczymi.

Ze względu na rosnące zainteresowanie i częstotliwość stosowania rozpoznawania wielomodalnego oraz wielokanałowego w aktualnych rozwiązaniach, przeprowadzono przegląd literatury dotyczący obecnie stosowanych metod integracji. Omówiono scenariusz wczesnej, późnej oraz hybrydowej fuzji, ze szczególnym naciskiem na istniejące metody integracji w scenariuszu późnej fuzji. Na podstawie istniejących rozwiązań autor pracy wyodrębnił dwa kryteria funkcji integracji: zbieżności oraz preferencję najbardziej wiarygodnych odpowiedzi programów automatycznego rozpoznawania emocji. Na podstawie istniejących rozwiązań zostały przygotowane cztery rozwiązania referencyjne, a także opracowana nowa metoda integracji.

Przegląd literatury został zakończony poprzez analizę aktualnych problemów badawczych w dziedzinie rozpoznawania wielomodalnego oraz zaadresowaniem w zaproponowanej metodzie takich problemów badawczych jak niekompletność danych, różny zakres stosowalności modalności, określenie niepewności związanej z pomiarem symptomów emocji, rozbieżność wskazań programów automatycznego rozpoznawania emocji czy problem synchronizacji czasowej.

W kolejnym etapie pracy opracowano zaproponowane rozwiązanie, na które składają się trzy elementy:

- architektura rozwiązania oparta o wzorzec architektoniczny Scoreboard,
- propozycja metody kwantyfikacji informacji o niepewności pomiarów w danym kanale obserwacji oraz
- metoda integracji informacji o stanie emocjonalnym uwzględniająca niepewność i niedostępność kanałów obserwacji.

Zaproponowane rozwiązanie zostało opracowane na bazie istniejących rozwiązań. W szczególności bazuje ono na wykorzystaniu ciągłego modelu reprezentacji emocji (takiego jak modele wielowymiarowe czy modele emocji podstawowych) oraz zastosowaniu metody fuzji hybrydowej – ze szczególnym naciskiem na elementy późnej fuzji. Wykorzystano również koncepcję metod rozmytych, miar: niepewności, zaufania i wiarygodności z teorii ewidencji Dempstera–Shaffera, a także założenie z filtru Kalmana o niepewności każdego pomiaru oraz zmniejszenie zaufania do rozwiązań, które mają zbyt mało danych (niższą jakość).

Dla każdego z elementów rozwiązania została przygotowana teza pomocnicza. W odpowiedzi na postawione tezy zostało przygotowanych 9 eksperymentów, 5 symulacji oraz studium przypadku.

11.1. **Teza pomocnicza 1**

Na początku przeprowadzono badania nad udowodnieniem pierwszej tezy pomocniczej, brzmiącej następująco:

Zastosowanie wzorca architektonicznego Scoreboard pozwala na poprawę modularności, łatwości integracji, adaptowalności i odporności w wielokanałowym rozpoznawaniu emocji.

W tym celu przeprowadzono eksperyment integracji danych wielokanałowych w oparciu o analizę ładunku emocjonalnego tekstu, symulację oraz studium przypadku Monitora emocji wraz z jakościową analizą ATAM. Na podstawie przeprowadzonych badań można sformułować następujące wnioski:

1. Rozwiązania bazujące na wzorcu Scoreboard są modularne w sposób porównywalny z wzorcem Blackboard, który został użyty jako wzorzec referencyjny. Możliwe jest wykorzystanie wielu różnych programów automatycznego rozpoznawania emocji.
2. Wykorzystanie wzorca Scoreboard poprawia skuteczność rozwiązania. Uzgodnione rozwiązanie w większości przypadków wykazuje nieco wyższą skuteczność niż rozwiązania jednokanałowe oraz oparte na wzorcu Blackboard przy założeniu, że liczba programów „silnych” oraz „słabych” jest odpowiednio zbalansowana, a rozwiązania „silne” są zbieżne. Wzorzec Scoreboard wspiera zarówno metodę wczesnej, późnej, jak i hybrydowej fuzji.



3. Rozwiązanie oparte o wzorzec Scoreboard cechuje się adaptowalnością nie gorszą niż w przypadku wzorca Blackboard. Programy automatycznego rozpoznawania emocji są łatwo podłączane, odłączane oraz modyfikowane bez wpływu na pozostałe moduły. Złożoność adaptacji nowych programów zależy głównie od wybranej technologii. Dodatkowo modyfikacja funkcji integracji oraz oceny nie wymaga modyfikacji pozostałych modułów – zmiana dotyczy tylko jednej klasy.
4. Wiele programów może być ocenianych dynamicznie w czasie działania (ang. run-time). Możliwość oceny oraz adaptowalność do zmieniających się okoliczności (takich jak dostępność kanału, rozpoznawanie emocji w zależności od kontekstu) są wbudowane we wzorzec Scoreboard, który wspiera ciągłą ocenę programów. Rozwiązanie to pomaga budować zaufanie do wyniku rozwiązania i wykorzystywanych komponentów.
5. Wzorzec Scoreboard cechuje się odpornością na czasową niedostępność kanałów obserwacji oraz programów automatycznego rozpoznawania emocji. Możliwe jest automatyczne wykrycie programów o niskiej skuteczności i dostosowanie częstości ich wywołania w czasie działania aplikacji (dynamicznie). Odporność ta nie występuje w referencyjnym wzorcu Blackboard.

Dane ilościowe oraz jakościowe obserwacje pozwoliły na zaakceptowanie pierwszej tezy pomocniczej mówiącej, że rozwiązanie oparte o wzorzec Scoreboard zapewnia modularyzację, łatwość integracji, odporność oraz adaptowalność dla rozwiązań automatycznego rozpoznawania emocji użytkownika komputera.

11.2. *Teza pomocnicza 2*

Następnie przeprowadzono badania nad udowodnieniem drugiej tezy pomocniczej, brzmiącej następująco:

Zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji umożliwia reprezentację niepewności oszacowania stanu emocjonalnego

W celu jej udowodnienia przeprowadzono dwa eksperymenty na zaetykietowanym emocjonalnie zbiorze danych SEMAINE z wykorzystaniem zaproponowanych metod wyznaczania współczynnika niepewności dla modalności mimiki twarzy. **Przeprowadzone badania pozwalają sformułować wniosek, że wraz ze wzrostem skuteczności rozpoznawania emocji, zaproponowana metoda kwantyfikacji informacji kontekstowych o danym kanale obserwacji w automatycznym rozpoznawaniu emocji przypisuje wyższą wartość współczynnika wiarygodności. Pozwala to na przyjęcie drugiej tezy pomocniczej.**

W ramach przeprowadzonych eksperymentów zaobserwowano duże różnice w poziomie skuteczności rozwiązań jednomodalnych dla poszczególnych wymiarów, co potwierdza zarówno analiza danych, jak i wyniki eksperymentu nr 3, dotyczącego odpowiedniego doboru współczynnika wiarygodności w przypadku jego braku.

Dotyczy to przede wszystkim przypisania indywidualnego współczynnika niepewności. Jest to umotywowane różnorodnym poziomem skuteczności dla każdej emocji wskazywanym przez rozwiązania jednomodalne. Poziom zbieżności danych różnił się znacząco w zależności od poszczególnych wymiarów dla tych samych programów i kanałów, co zostało szczególnie uwidocznione w eksperymencie z czterema kamerami. Dla czterech emocji zbieżność była wysoka, dla dwóch średnia, a dla 3 bardzo niska.

Rozbieżności te były widoczne zarówno dla kanału audio, jak i analizy mimiki twarzy np. emocja zaskoczenia lepiej była rozpoznawana przez program bazujących na prozodii mowy, a gniew lepiej był rozpoznawany przez program bazujący na mimice twarzy. Wnioskiem z tego jest potrzeba przypisania indywidualnego poziomu współczynnika wiarygodności dla każdego z wymiarów z osobna, zamiast stosować zagregowany współczynnik wiarygodności.

Dodatkowym osiągnięciem przeprowadzonych eksperymentów było potwierdzenie pozytywnego wpływu wykorzystania słabszego programu na wynik integracji. Uwzględnienie w analizie wskazania programu o niskim poziomie skuteczności, ale ze zbieżną hipotezą, pozwoliło na poprawę skuteczności uzgodnionego rozwiązania.

11.3. **Teza pomocnicza 3**

Następnie przeprowadzono badania nad udowodnieniem trzeciej tezy pomocniczej, brzmiącej następująco:

Wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwi poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji.

W celu udowodnienia części dotyczącej niezawodności przeprowadzono dwa eksperymenty (nr 4 oraz nr 8) analizujące poprawę niezawodności podczas wykorzystania wielokanałowej i wielomodalnej analizy na przykładzie dwóch różnych źródeł danych.

Rezultaty eksperymentu nr 4, przeprowadzone na zaetykietowanym emocjonalnie zbiorze danych SEMAINE wykazały znacząco poprawę niezawodności rozwiązania poprzez zastosowanie analizy wielomodalnej. Dominującą modalnością była analiza mimiki twarzy, a głównym kanałem obserwacji przednia kamera monochromatyczna. Dodatkowe kanały (tu: kamery) poprawiły dostępność z 94,9% do 96,1%. Dodanie modalności audio do modalności mimiki twarzy spowodowało wzrost niezawodności rozwiązania, poprzez zwiększenie dostępności do poziomu 98,3% czasu badania.

Wyniki eksperymentu nr 8, przeprowadzonego na danych z badania przeprowadzonego na Politechnice Gdańskiej z wykorzystaniem czterech kamer, wykazały poprawę niezawodności rozwiązania poprzez zastosowanie analizy wielokanałowej. Dostępność kanału z pojedynczej kamery wahała się od 85 do 90%, a wykorzystując cztery kamery, uzyskano dostępność na poziomie 96,35%. Wykorzystanie dwóch najlepiej ustawionych kamer (w tym wypadku prawych) pozwoliło już na poprawę dostępności do 95%, co pokazuje, że dodanie nawet jednego dodatkowego kanału pozwala na poprawę dostępności.

Wyniki tych eksperymentów pozwalają sformułować wniosek, że metody analiza wielomodalna umożliwiają poprawę niezawodności rozpoznawania emocji.

Następnie poprzez zaplanowane eksperymenty i symulacje przeanalizowano kwestię poprawy wiarygodności przy zastosowaniu zaproponowanej metody. Na zaetykietowanym emocjonalnie zbiorze danych SEMAINE zostały przeprowadzone cztery eksperymenty, przeprowadzono cztery symulacje przypadków brzegowych oraz jeden eksperyment na danych z badania na Politechnice Gdańskiej z czterema kamerami.

Podczas eksperymentu nr 5 zostały zweryfikowane różnice w poziomie skuteczności integracji oraz poprawności przypisania współczynnika wiarygodności uzgodnionego rozwiązania dla różnej konfiguracji parametrów funkcji integracji. Zbadane zostały cztery parametry, z czego różnice we wskazaniach zauważono dla dwóch z nich – doboru parametrów rozkładu normalnego oraz funkcji zaufania. Określono optymalne parametry zaproponowanej metody dla zaetykietowanego emocjonalnie zbioru SEMAINE.

W eksperymencie 6 skuteczność integracji zaproponowanej metody została porównana z rozwiązaniami jednomodalnymi oraz z rozwiązaniami referencyjnymi. Uzyskana skuteczność była w większości przypadków wyższa niż dla rozwiązań jednomodalnych, a dla pozostałych była nie gorsza od rozwiązań jednomodalnych. Zaproponowana metoda w przeciwieństwie do metod referencyjnych nie obniżyła wartości skuteczności uzgodnionego rozwiązania w stosunku do rozwiązań jednomodalnych dla żadnego z wymiarów.

W eksperymencie 7 porównano poprawność wyznaczania współczynnika wiarygodności w zaproponowanej metodzie oraz w rozwiązaniach referencyjnych. Zaproponowane rozwiązanie wyznaczyło wyższy współczynnik wiarygodności wraz ze wzrostem skuteczności rozpoznawania – w przeciwieństwie do rozwiązań referencyjnych, dla których podobna zależność nie została zauważona.

W symulacji 2 przeanalizowana została zmiana wartości współczynnika wiarygodności uzgodnionego rozwiązania w zależności od odległości dwóch hipotez, przy założonym kryterium zbieżności w postaci wartości progowej odległości. Zaproponowana metoda zadziałała zgodnie z oczekiwaniami dla wartości progowej odległości nie większej niż 0,1. Dla wartości progowych 0,2 metoda działa poprawnie dla poziomu wiarygodności nie większego niż 70%, a dla wartości progowej odległości równej 0,3 dla poziomu wiarygodności nie większego niż 50%. Dla hipotez o współczynniku o wartości poniżej 0,3, a zwłaszcza poniżej 0,1, wyniki są bardzo rozmyte i mogą odbiegać od oczekiwanych. Dla wartości progowej odległości powyżej 0,5 powyższa metoda nie daje optymalnych wyników.

W symulacji 3 przeanalizowano wpływ zmiany współczynnika wiarygodności jednej z hipotez na wyznaczenie współczynnika wiarygodności uzgodnionego rozwiązania, w przypadku dwóch rozbieżnych hipotez. Zaproponowana metoda została poprawnie zweryfikowana – wraz ze wzrostem wiarygodności hipotezy przeciwnej (wykluczającej) zaufanie do uzgodnionego rozwiązania spadało.



Czwarta symulacja dotyczyła przypadku trzech hipotez, gdzie dwie były zbieżne ze sobą, a trzecia rozbieżna z nimi. Analizowaną zmienną niezależną była wartość współczynnika wiarygodności hipotezy rozbieżnej. Zgodnie z oczekiwaniem wartość współczynnika wiarygodności uzgodnionego rozwiązania spada wraz ze wzrostem współczynnika wiarygodności hipotezy rozbieżnej.

Ostatnia symulacja również dotyczyła przypadku trzech hipotez, gdzie dwie były zbieżne ze sobą, a trzecia rozbieżna z nimi. Analizowaną zmienną niezależną była odległość hipotezy rozbieżnej od bliższej ze zbieżnych hipotez. Obydwa warianty zaproponowanej metody zachowały się zgodnie z oczekiwaniami, wraz ze wzrostem dystansu hipotezy rozbieżnej od zbieżnych hipotez, wiarygodność uzgodnionego rozwiązania spada. Wartości dla obu wariantów były zbliżone.

Wyniki ostatniego eksperymentu dotyczyły weryfikacji powiązania wartości metodę współczynnika wiarygodności uzgodnionego rozwiązania wyznaczanego przez zaproponowaną ze spójnością poszczególnych kanałów obserwacji. Wyznaczony współczynnik wiarygodności był tym wyższy, im wyższa była spójność rozwiązań, a rozwiązania o niskiej spójności wyników otrzymywały wyraźnie niższy współczynnik wiarygodności. Wynik ten pozwala na wykorzystanie metody w przypadkach rzeczywistych, w których nie znamy referencyjnego rozwiązania, a możemy korzystać tylko ze wskazań wielu rozwiązań automatycznego rozpoznawania emocji.

Przeprowadzone badania pozwalają sformułować wniosek, że wykorzystanie reprezentacji niepewności w integracji danych o stanie emocjonalnym umożliwia poprawę wiarygodności oraz niezawodności w wielomodalnym rozpoznawaniu emocji. Pozwala to na przyjęcie trzeciej tezy pomocniczej.

11.4. **Teza główna oraz osiągnięcia pracy**

Podsumowując wyniki symulacji i eksperymentów – sposób wyznaczenia współczynnika wiarygodności przez zaproponowaną metodą jest poprawny i spełnił wszystkie sprawdzane kryteria.

W świetle uzyskanych wyników symulacji oraz eksperymentów i wymienionych wniosków można stwierdzić, że dowiedziono prawdziwości tezy głównej oraz tez pomocniczych.

Do głównych oryginalnych osiągnięć niniejszej rozprawy należą:

- Zaproponowanie nowej metody integracji,
- Zaproponowanie metody kwantyfikacji informacji o niepewności dla modalności mimiki twarzy,
- Potwierdzenie poprawności działania metody wyznaczenia współczynnika wiarygodności uzgodnionego rozwiązania.

Pozostałymi osiągnięciami pracy są:

- Potwierdzenie, że analiza wielomodalna i wielokanałowa poprawia niezawodność rozwiązania,
- Potwierdzenie podobnego poziomu dostępności kanałów obserwacji w obrębie jednej modalności oraz różnego poziomu dostępności w przypadku kanałów z różnych modalności.
- Potwierdzenie pozytywnego wpływu wykorzystania słabszego programu (o niższej skuteczności) na wynik integracji,
- Potwierdzenie, że słusznym podejściem jest odrębna dla każdego wymiaru (emocji) analiza poziomu skuteczności, a zatem również i współczynnik wiarygodności hipotez, jak i współczynnik wiarygodności uzgodnionego rozwiązania. Jest to spowodowane wieloma czynnikami, takimi jak
 - różny poziom skuteczności dla tego samego programu, tego samego badania i tej samej modalności i kanału, ale różnego analizowanego wymiaru,
 - inny poziom zbieżności danych z wymiaru z poszczególnych kanałów w przypadku różnicy jedynie położenia czujnika (kąąt kamery czy umiejscowienie mikrofonu),
 - ogólnodostępne zaetykietowane emocjonalnie zbiory danych mają różną dostępność etykiet dla poszczególnych wymiarów, dlatego też poziom nauczania programów może się różnić między poszczególnymi wymiarami.

Podczas przeprowadzonych symulacji oraz eksperymentów zostały zidentyfikowane ograniczenia metody integracji CFIEE związane z przyjmowanymi wartościami progowymi odległości. Dla wartości progowej odległości nie większej niż 0,1 zaproponowana metoda działa zgodnie z oczekiwaniami. Dla wartości progowej odległości nie większej niż 0,2 metoda działa poprawnie dla poziomu wiarygodności nie większego niż 70%, a dla wartości progowej odległości nie większej niż 0,3 dla poziomu wiarygodności nie większego niż 50%. Dla hipotez o wartości współczynnika wiarygodności poniżej 0,3, a zwłaszcza poniżej 0,1, wyniki są bardzo rozmyte i mogą odbiegać od oczekiwanych. Dla wartości progowej odległości powyżej 0,5 powyższa metoda nie daje optymalnych wyników.

Autor pracy zebrał kilka kluczowych obserwacji, jako zbiór wniosków dla praktyków, implementujących rozwiązania wielomodalnego rozpoznawania emocji:

- W przypadku większej różnorodności kanałów (w szczególności różne modalności, różne czasy otrzymywania informacji) rekomendacją autora pracy jest wykorzystanie fuzji późnej, a do analizy danych w obrębie jednej modalności lub modalności ze zbliżoną częstotliwością zwracania informacji o symptomach emocji – wykorzystanie dowolnej fuzji – wczesnej, późnej lub hybrydowej fuzji,
- Wykorzystanie wielowymiarowych modeli reprezentacji emocji jest bardziej uniwersalne i pozwala na większą elastyczność rozwiązania – są one też coraz częściej wybierane przez badaczy,
- Projektując program automatycznego rozpoznawania emocji oraz rozwiązanie integrujące – warto przemyśleć indywidualną analizę poziomu skuteczności, jakości danych oraz określenie poziomu niepewności dla każdego z analizowanych wymiarów (emocji) – ze względu na duże rozbieżności pomiędzy nimi,
- Wykorzystanie większej liczby modalności lub choćby dodatkowego kanału (nawet, poprzez jego zdublowanie w postaci dodatkowej kamery) – zwiększa niezawodność rozwiązania. Dodatkowa modalność zwiększa niezawodność w większym stopniu.

Ponadto, rezultatem badań prowadzonych w ramach tej rozprawy są następujące publikacje, których współautorem jest autor rozprawy:

1. Landowska, A., Brodny, G., Scoreboard Architectural Pattern and Integration of Emotion Recognition Results, 2019, IEEE Access, DOI: 10.1109/access.2018.2889557 [4]
2. Brodny G., Kołakowska A., Landowska A., Szwoch M., Szwoch W., Wróbel M.: Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions, W: 2016 9th International Conference on Human System Interactions (HSI), DOI: 10.1109/HSI.2016.7529664, 2016, IEEE [159]
3. Brodny G., Landowska A., "Integration in multichannel emotion recognition," in *2018 11th International Conference on Human System Interaction (HSI)*, 2018, pp. 35–41, doi: 10.1109/HSI.2018.8431343 [3]
4. Landowska A., Brodny G., Wróbel M., "Limitations of Emotion Recognition from Facial Expressions in e-Learning Context, Proceedings of the 9th International Conference on Computer Supported Education", Scitepress, 2017, s.383–389 [2]
5. Landowska, A., Brodny, G. Investigation of educational processes with affective computing methods. e-mentor, (3 (70)), 2017, s.15–24.[1]
6. Landowska A., Brodny G., „Postrzeganie inwazyjności automatycznego rozpoznawania emocji w kontekście edukacyjnym”, EduAkcja. Magazyn edukacji elektronicznej, Nr 1 (13)/2017 [64]
7. Brodny G., Witkowicz, M. "AFFITS Serwer algorytmów rozpoznających emocje – Architektura systemu," Gdańsk, 2016, dokument techniczny [5].

11.5. **Kierunki dalszych prac.**

W trakcie swoich badań autor pracy znalazł wiele kierunków rozwoju zaproponowanej metody integracji oraz przypadków wartych dalszej analizy, które wykraczały poza ramy tej pracy. Kierunki dalszych prac zostały zebrane w tym rozdziale.

Niewątpliwie wartym zastanowienia i dalszych badań jest zbadanie alternatywnych funkcji zaufania oraz niewiary tak, aby poziom zbieżności był zgodny z systemem promowania za zbieżność i brak zbieżności, także dla wyższych wartości progowych odległości.

Metoda wyznaczania współczynnika wiarygodności może zostać udoskonalona dla modalności mimiki twarzy. Jednocześnie, warto rozszerzyć badania na inne modalności – takie jak prozodia mowy oraz dane fizjologiczne.

Przykładem wartym przebadania, wykraczających poza tę pracę, jest analiza skuteczności rozpoznawania danych audio w zależności od umieszczenia mikrofonu – zwłaszcza porównanie wartości dla emocji zaskoczenia z pozostałymi emocjami, ze względu na inne zachowanie niż pozostałe emocje. Wymagana jest analiza, czy różnice występują także w innych źródłach danych.

Wszelkie optymalizacje czasowe takie jak możliwości zrównoleglenia pracy wywoływanych ekspertów będą przedmiotem dalszych prac w trakcie wykorzystania zaproponowanej metody przy kolejnych badaniach oraz projektach.

Jednym z obiecujących kierunków jest wykorzystania śledzenia historii zmian emocji do poprawy jej predykcji. Właściwy sposób jego zastosowania jest tematem szerszych badań. Przykładowo, można wykorzystać w tym celu filtr Kalmana lub wykorzystać uzgodnione rozwiązanie z poprzedniego kroku jako dodatkowego eksperta, ale z obniżonym współczynnikiem zaufania – poziom obniżenia wskazania powinien zostać zbadany eksperymentalnie. Element ten, choć wart jest dalszych badań, nie został wykorzystany w tej pracy ze względu na skupienie się na poprawie samej metody integracji dla pojedynczych sekwencji, które są bazą dla analizy i korekt związanych z predykcją na podstawie historii. Innym powodem niewykorzystania filtra Kalmana, była obawa przed brakiem przejrzystości, czy poprawa została osiągnięta dzięki zaproponowanemu rozwiązaniu, czy samemu wykorzystaniu filtra.

Innym kierunkiem badań jest analiza specyfiki każdej z modalności tak, by dostosować rozmiar ramki danych do specyfiki modalności, takich jak zduplikowanie wartości z kanału audio na sąsiednie ramki, w przypadku braku nowszych informacji [116].

Dokładniejsze zbadanie wartości błędu dla odwzorowań modeli – porównanie rozbieżności wyników przy wykorzystaniu bazowych wartości oraz rozbieżności po wykorzystaniu jednego lub dwóch odwzorowań. Badanie to wymagałoby wykorzystanie dwóch programów automatycznego rozpoznawania emocji wykorzystujących różne modele, a także wykorzystanie odpowiedniej referencyjnej bazy, w której dane wejściowe miałyby zaetykietowane dane w obu modelach. Porównanie rozbieżności w obu przypadkach pozwoliłoby oszacować poziom błędu.

Obecnie większość programów jest uczona na uśrednionych reakcjach ludzkich, które jednakże różnią się ze względu na indywidualne doświadczenia jednostek, kultury czy języka. Warty rozważenia jest próba kalibracji programu do użytkownika, by na podstawie przypisania badanego do profilu osobowości, można było zastosować lepsze dopasowanie parametrów programu np. wskazanie modalności dominującej (niosącej najwięcej informacji) np. 50% gesty, 30% mimika, 20% pozostałe modalności. W przypadku braku kluczowej modalności zmniejszyć końcowy współczynnik zaufania. Idealnym rozwiązaniem byłoby wyjście od generalnego modelu, który z czasem byłby dopasowywany do użytkownika [16]. Odpowiednia kalibracja programu rozpoznawania emocji jest jednym z czterech głównych czynników regulującym ryzyko związane z odmową świadczenia usług dla grup osób, które są nie dość reprezentowane w danych uczących programy automatycznego rozpoznawania emocji [9] [160], co mogłoby zapobiec pogłębianiu stereotypów.

LITERATURA

- [1] A. Landowska and G. Brodny, "Investigation of educational processes with affective computing methods," *e-mentor*, vol. 2017, no. 70(3), pp. 15–24, Jun. 2017, doi: 10.15219/em70.1304.
- [2] A. Landowska, G. Brodny, and M. R. Wrobel, "Limitations of emotion recognition from facial expressions in e-learning context," *CSEDU 2017 - Proc. 9th Int. Conf. Comput. Support. Educ.*, vol. 2, 2017.
- [3] G. Brodny and A. Landowska, "Integration in multichannel emotion recognition," in *2018 11th International Conference on Human System Interaction (HSI)*, 2018, pp. 35–41, doi: 10.1109/HSI.2018.8431343.
- [4] A. Landowska and G. Brodny, "Scoreboard architectural pattern and integration of emotion recognition results," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2018.2889557.
- [5] G. Brodny and M. Witkowicz, "AFFITS Serwer algorytmów rozpoznających emocje - Architektura systemu," Gdańsk, 2016.
- [6] A. Landowska, "Przetwarzanie emocjonalne i scenariusze jego zastosowania w edukacji i e-edukacji," in *Postępy e-edukacji*, 2013.
- [7] A. Landowska, "Technical Report no 2/2020 AFFECTIVE COMPUTING -Glossary of terms in English," Gdańsk, 2020.
- [8] J. N. Bailenson *et al.*, "Real-time classification of evoked emotions using facial feature tracking and physiological responses," *Int. J. Hum. Comput. Stud.*, 2008, doi: 10.1016/j.ijhcs.2007.10.011.
- [9] J. Hernandez *et al.*, *Guidelines for Assessing and Minimizing Risks of Emotion Recognition Applications*. Institute of Electrical and Electronics Engineers (IEEE), 2021, pp. 1–8.
- [10] "SJP - prozodia." [Online]. Available: <https://sjp.pl/prozodia>.
- [11] F. Karray, M. Alemzadeh, J. Abou Saleh, and M. Nours Arab, "Human-Computer Interaction: Overview on State of the Art," *Int. J. Smart Sens. Intell. Syst.*, 2008, doi: 10.21307/ijssis-2017-283.
- [12] E. D. Ferguson, K. R. Scherer, and P. Ekman, "Approaches to Emotion," *Am. J. Psychol.*, 1985, doi: 10.2307/1422518.
- [13] A. Mehrabian, "Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament," *Curr. Psychol.*, 1996, doi: 10.1007/bf02686918.
- [14] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, P. C. Ellsworth, and P. C. E. Johnny R.J. Fontaine, Klaus R. Scherer, Etienne B. Roesch, "The world of emotions is not two-dimensional," *Psychol. Sci.*, vol. 18, no. 12, pp. 1050–1057, Dec. 2007, doi: 10.1111/j.1467-9280.2007.02024.x.
- [15] V. A. Petrushin, "Emotion recognition in speech signal: Experimental study, development, and application," in *6th International Conference on Spoken Language Processing, ICSLP 2000*, 2000.
- [16] J. Wagner, F. Lingenfelser, E. André, and J. Kim, "Exploring fusion methods for multimodal emotion recognition with missing data," *IEEE Trans. Affect. Comput.*, 2011, doi: 10.1109/T-AFFC.2011.12.
- [17] J. Kim and F. Lingenfelser, "Ensemble approaches to parametric decision fusion for bimodal emotion recognition," in *BIOSIGNALS 2010 - Proceedings of the 3rd International Conference on Bio-inspired Systems and Signal Processing, Proceedings*, 2010.
- [18] D. Ruta and B. Gabrys, "An Overview of Classifier Fusion Methods," *Comput. Inf. Syst.*, 2000.
- [19] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*. 2006, doi: 10.1109/MCAS.2006.1688199.
- [20] A. Landowska, "Uncertainty in emotion recognition," *Journal of Information, Communication and Ethics in Society*, vol. 17, no. 3. 2019, doi: 10.1108/JICES-03-2019-0034.
- [21] A. Shaik, B. Manda, C. Prakashini, and K. Deepthi, "Metrics for object oriented design software systems: a survey," *J. Emerg.*, 2010.
- [22] R. Willink, "Guide to the Expression of Uncertainty in Measurement," in *Measurement*

Uncertainty and Probability, 2013.

- [23] A. Kołakowska, A. Landowska, and M. Szwoch, *Modeling emotions for affect-aware applications*. 2015.
- [24] Z. Kowalczyk and M. Czubenko, "Computational approaches to modeling artificial emotion - an overview of the proposed solutions," *Frontiers Robotics AI*. 2016, doi: 10.3389/frobt.2016.00021.
- [25] E. L. Rosenberg, "Levels of analysis and the organization of affect," *Rev. Gen. Psychol.*, 1998, doi: 10.1037/1089-2680.2.3.247.
- [26] W. Szwoch, "Using physiological signals for emotion recognition," in *2013 6th International Conference on Human System Interactions, HSI 2013*, 2013, pp. 556–561, doi: 10.1109/HSI.2013.6577880.
- [27] W. D. Tenhouten, "Anger and Contested Place in the Social World," *Sociol. Mind*, vol. 8, pp. 226–248, 2018, doi: 10.4236/sm.2018.83018.
- [28] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, doi: 10.1007/978-3-642-34584-5_11.
- [29] R. Plutchik, "The circumplex as a general model of the structure of emotions and personality.," in *Circumplex models of personality and emotions.*, 2004.
- [30] R. Plutchik, "Emotion: A psychoevolutionary synthesis," *Harpercollins Coll. Div.*, 1980.
- [31] Z. Shi, J. Wei, Z. Wang, J. Tu, and Q. Zhang, "Affective transfer computing model based on attenuation emotion mechanism," *J. Multimodal User Interfaces*, vol. 5, no. 1–2, pp. 3–18, Mar. 2012, doi: 10.1007/s12193-011-0071-7.
- [32] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Trans. Affect. Comput.*, 2012, doi: 10.1109/T-AFFC.2011.30.
- [33] X. Hu, "Music and mood: Where theory and reality meet," *Proc. iConference*, 2010.
- [34] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, 2017, doi: 10.1016/j.inffus.2017.02.003.
- [35] O. Bălan, G. Moise, L. Petrescu, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "Emotion Classification Based on Biophysical Signals and Machine Learning Techniques," *Symmetry (Basel)*, vol. 12, no. 1, p. 21, Dec. 2019, doi: 10.3390/sym12010021.
- [36] A. Ortony, G. Clore, and A. Collins, "The cognitive structure of emotions," 1990.
- [37] A. Mehrabian, "Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model," *Aust. J. Psychol.*, vol. 48, no. 2, pp. 86–92, Aug. 1996, doi: 10.1080/00049539608259510.
- [38] G. Yang, Z. Wang, G. Wang, and F. Chen, "Affective computing model based on emotional psychology," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, doi: 10.1007/11881070_37.
- [39] P. Gebhard and Patrick, "ALMA," in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems - AAMAS '05*, 2005, p. 29, doi: 10.1145/1082473.1082478.
- [40] I. Hupont, S. Ballano, S. Baldassarri, and E. Cerezo, "Scalable multimodal fusion for continuous affect sensing," in *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*, 2011, doi: 10.1109/WACI.2011.5953150.
- [41] A. Landowska, "Towards New Mappings between Emotion Representation Models," *Appl. Sci.*, vol. 8, no. 2, p. 274, Feb. 2018, doi: 10.3390/app8020274.
- [42] H. Gunes and M. Piccardi, "Affect recognition from face and body: early fusion vs. late fusion," *Syst. Man Cybern. 2005 IEEE Int. Conf.*, vol. 4, pp. 3437–3443, 2005, doi: 10.1109/ICSMC.2005.1571679.
- [43] Z.-Q. Wang and Ivan Tashev, "Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks," *Int. Conf. Acoust. speech signal Process.*, 2017.
- [44] A. J. Gill, R. M. French, D. Gergle, and J. Oberlander, "Identifying Emotional Characteristics from Short Blog Texts," *30th Annu. Conf. Cogn. Sci. Soc.*, 2008.
- [45] A. Kołakowska, A. Landowska, and K. Karpianko, "Gyroscope-based game revealing progress of children with autism," in *ACM International Conference Proceeding Series*,

2017, doi: 10.1145/3036290.3036324.

- [46] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Soc. Psychol.*, 1971, doi: 10.1037/h0030377.
- [47] P. Ekman and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *J. Pers. Soc. Psychol.*, 1971, doi: 10.1037/h0030377.
- [48] L. Liu, D. Preoțiu-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice," in *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 2016.
- [49] K. I. Bența and M. F. Vaida, "Towards real-life facial expression recognition systems," *Adv. Electr. Comput. Eng.*, 2015, doi: 10.4316/AECE.2015.02012.
- [50] "FaceReader solution description." [Online]. Available: <http://www.noldus.com>. [Accessed: 30-Mar-2016].
- [51] A. Dinculescu, C. Vizitiu, A. Nistorescu, M. Marin, and A. Vizitiu, "Novel approach to face expression analysis in determining emotional valence and intensity with benefit for human space flight studies," in *2015 E-Health and Bioengineering Conference, EHB 2015*, 2016, doi: 10.1109/EHB.2015.7391378.
- [52] S. J. Burton, A. A. Samadani, R. Gorbet, and D. Kulić, "Laban movement analysis and affective movement generation for robots and other near-living creatures," in *Springer Tracts in Advanced Robotics*, 2016.
- [53] K. Kaza *et al.*, "Body motion analysis for emotion recognition in serious games," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, doi: 10.1007/978-3-319-40244-4_4.
- [54] S. Piana, A. Stagliaño, F. Odone, and A. Camurri, "Adaptive Body Gesture Representation for Automatic Emotion Recognition," *ACM Trans. Interact. Intell. Syst.*, 2016, doi: 10.1145/2818740.
- [55] S. Senecal, L. Cuel, A. Aristidou, and N. Magnenat-Thalmann, "Continuous body emotion recognition system during theater performances," in *Computer Animation and Virtual Worlds*, 2016, doi: 10.1002/cav.1714.
- [56] A. Aristidou *et al.*, "Emotion control of unstructured dance movements," in *Proceedings - SCA 2017: ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2017, doi: 10.1145/3099564.3099566.
- [57] T. Balomenos, A. Raouzaïou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Emotion analysis in man-machine interaction systems," in *Lecture Notes in Computer Science*, 2005, doi: 10.1007/978-3-540-30568-2_27.
- [58] E. Van Dyck, P. Vansteenkiste, M. Lenoir, M. Lesaffre, and M. Leman, "Recognizing induced emotions of happiness and sadness from dance movement," *PLoS One*, 2014, doi: 10.1371/journal.pone.0089773.
- [59] E. Hudlicka, "To feel or not to feel: The role of affect in human-computer interaction," *Int. J. Hum. Comput. Stud.*, 2003, doi: 10.1016/S1071-5819(03)00047-8.
- [60] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Trans. Affect. Comput.*, vol. 2, no. 2, pp. 106–118, Apr. 2011, doi: 10.1109/T-AFFC.2011.7.
- [61] M. Valstar *et al.*, "AVEC 2016 - Depression, mood, and emotion recognition workshop and challenge," in *AVEC 2016 - Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, co-located with ACM Multimedia 2016*, 2016, doi: 10.1145/2988257.2988258.
- [62] J. Han, Z. Zhang, F. Ringeval, and B. Schuller, "Reconstruction-error-based learning for continuous emotion recognition in speech," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017, pp. 2367–2371, doi: 10.1109/ICASSP.2017.7952580.
- [63] A. Landowska and J. Miler, "Limitations of Emotion Recognition in Software User Experience Evaluation Context," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, 2016, pp. 1631–1640, doi: 10.15439/2016F535.
- [64] A. Landowska and G. Brodny, "Postrzeganie inwazyjności automatycznego rozpoznawania emocji w kontekście edukacyjnym," pp. 26–41, 2017.
- [65] K. Y. Huang, C. H. Wu, Q. B. Hong, M. H. Su, and Y. H. Chen, "Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech

- Sounds,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2019, doi: 10.1109/ICASSP.2019.8682283.
- [66] W. Lim, D. Jang, and T. Lee, “Speech emotion recognition using convolutional and Recurrent Neural Networks,” in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2016*, 2017, doi: 10.1109/APSIPA.2016.7820699.
- [67] G. Trigeorgis *et al.*, “Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2016, doi: 10.1109/ICASSP.2016.7472669.
- [68] J. A. Russell and A. Mehrabian, “Evidence for a three-factor theory of emotions,” *J. Res. Pers.*, 1977, doi: 10.1016/0092-6566(77)90037-X.
- [69] M. Bradley and P. Lang, “Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings,” *Psychology*, 1999, doi: 10.1109/MIC.2008.114.
- [70] U. Krcadinac, P. Pasquier, J. Jovanovic, and V. Devedzic, “Synesketch: An Open Source Library for Sentence-Based Emotion Recognition,” *IEEE Trans. Affect. Comput.*, 2013, doi: 10.1109/T-AFFC.2013.18.
- [71] M. Wierzba *et al.*, “Basic emotions in the nencki affective word list (NAWL be): New method of classifying emotional stimuli,” *PLoS One*, 2015, doi: 10.1371/journal.pone.0132305.
- [72] M. Riegel *et al.*, “Nencki Affective Word List (NAWL): the cultural adaptation of the Berlin Affective Word List–Reloaded (BAWL-R) for Polish,” *Behav. Res. Methods*, 2015, doi: 10.3758/s13428-014-0552-1.
- [73] C. H. Wu and W. Bin Liang, “Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels,” *IEEE Trans. Affect. Comput.*, 2011, doi: 10.1109/T-AFFC.2010.16.
- [74] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, “Facial expression analysis with AFFDEX and FACET: A validation study,” *Behav. Res. Methods*, 2018, doi: 10.3758/s13428-017-0996-1.
- [75] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, doi: 10.1109/34.954607.
- [76] M. Wöllmer, M. Kaiser, F. Eyben, B. Schuller, and G. Rigoll, “LSTM-modeling of continuous emotions in an audiovisual affect recognition framework,” *Image Vis. Comput.*, 2013, doi: 10.1016/j.imavis.2012.03.001.
- [77] S. K. D’Mello and J. Kory, “A review and meta-analysis of multimodal affect detection systems,” *ACM Computing Surveys*. 2015, doi: 10.1145/2682899.
- [78] Picard R.W, *Affective Computing*. Massachusetts: MIT press. Cambridge, 1997.
- [79] Wydział Projektowy Akademii Sztuk Pięknych w Katowicach, “Badania użyteczności a pomiar zachowań użytkowników,” *Zakład Badań Wizualnych i Interakcji*. [Online]. Available: <https://aspzbwii.wordpress.com/2014/07/17/badania-uzytecznosci-a-pomiar-zachowan-uzytownikow/>. [Accessed: 04-Feb-2020].
- [80] “7 Sentiment Analysis Tools to Improve Employee Engagement in 2020 | HR Technologist.” [Online]. Available: <https://www.hrtechnologist.com/articles/employee-engagement/sentiment-analytics-tools-features-price/>. [Accessed: 15-Dec-2021].
- [81] H. Deakin and K. Wakefield, “Skype interviewing: reflections of two PhD researchers,” *Qual. Res.*, 2014, doi: 10.1177/1468794113488126.
- [82] M. Zetlin, “AI Is Now Analyzing Candidates’ Facial Expressions During Video Job Interviews,” *Inc.com*, 2018. .
- [83] M. Butcher, “The robot-recruiter is coming - VCV’s AI will read your face in a job interview,” *techcrunch*, 2019. .
- [84] A. T. Rupasinghe, N. L. Gunawardena, S. Shujan, and D. A. S. Atukorale, “Scaling personality traits of interviewees in an online job interview by vocal spectrum and facial cue analysis,” in *16th International Conference on Advances in ICT for Emerging Regions, ICTer 2016 - Conference Proceedings*, 2017, doi: 10.1109/ICTER.2016.7829933.
- [85] K. Anderson *et al.*, “The TARDIS framework: Intelligent virtual agents for social coaching in job interviews,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, doi: 10.1007/978-3-319-03161-3_35.

- [86] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, 2012, doi: 10.1109/T-AFFC.2011.15.
- [87] "Wzruszające filmy | Oficjalna witryna Netflix." [Online]. Available: <https://www.netflix.com/pl/browse/genre/4136>. [Accessed: 15-Dec-2021].
- [88] "Mood and Genre filters for 'Liked Songs' - The Spotify Community." [Online]. Available: <https://community.spotify.com/t5/Community-Blog/Mood-and-Genre-filters-for-Liked-Songs/ba-p/5160106>. [Accessed: 15-Dec-2021].
- [89] S. D'Mello and A. Graesser, "The half-life of cognitive-affective states during complex learning," *Cogn. Emot.*, 2011, doi: 10.1080/02699931.2011.613668.
- [90] M. Csikszentmihalyi, "The flow experience and its significance for human psychology," in *Optimal Experience*, 2012.
- [91] R. Corive *et al.*, "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, 2001, doi: 10.1109/79.911197.
- [92] S. Petrovica, A. Anohina-Naumeca, and H. K. Ekenel, "Emotion Recognition in Affective Tutoring Systems: Collection of Ground-truth Data," in *Procedia Computer Science*, 2016, doi: 10.1016/j.procs.2017.01.157.
- [93] S. Swauger, "Software that monitors students during tests perpetuates inequality and violates their privacy," *MIT Technology Review*, 2020.
- [94] "Stress Management - Stress Watch & Monitoring | Fitbit." [Online]. Available: <https://www.fitbit.com/global/us/technology/stress>. [Accessed: 15-Dec-2021].
- [95] "Facebook Increasingly Reliant on A.I. To Predict Suicide Risk : NPR." [Online]. Available: <https://www.npr.org/2018/11/17/668408122/facebook-increasingly-reliant-on-a-i-to-predict-suicide-risk?t=1639600302665>. [Accessed: 15-Dec-2021].
- [96] D. DeVault *et al.*, "SimSensei kiosk: A virtual human interviewer for healthcare decision support," in *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, 2014.
- [97] K. Chang, D. Fisher, and J. Canny, "AMMON: A Speech Analysis Library for Analyzing Affect, Stress, and Mental Health on Mobile Phones," in *Proceedings of the 2011 PhoneSense conference*, 2011, doi: 10.1.1.232.365.
- [98] "Upgraded Google Glass Helps Autistic Kids 'See' Emotions - IEEE Spectrum." [Online]. Available: <https://spectrum.ieee.org/upgraded-google-glass-helps-autistic-kids-see-emotions>. [Accessed: 15-Dec-2021].
- [99] W. PG, "E-technologie dla dzieci z autyzmem." [Online]. Available: <http://autyzm.eti.pg.gda.pl/>. [Accessed: 04-Feb-2020].
- [100] M. Jedrzejewska-Szczerska, K. Karpieńko, and A. Landowska, "System supporting behavioral therapy for children with autism," in *Journal of Innovative Optical Health Sciences*, 2015, doi: 10.1142/S1793545815410084.
- [101] N. Roy, J. Pineau, and S. Thrun, "Spoken dialogue management using probabilistic reasoning," 2000, doi: 10.3115/1075218.1075231.
- [102] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans*, 2011, doi: 10.1109/TSMCA.2011.2116000.
- [103] J. Miler and A. Landowska, "Designing effective educational games-a case study of a project management game," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, FedCSIS 2016*, 2016, pp. 1657–1661, doi: 10.15439/2016F434.
- [104] S. Shimojo and L. Shams, "Sensory modalities are not separate modalities: Plasticity and interactions," *Current Opinion in Neurobiology*. 2001, doi: 10.1016/S0959-4388(00)00241-5.
- [105] A. Mehrabian and Susan R. Ferris, "Inference of attitudes from nonverbal communication in two channels," *J. Consult. Psychol.*, 1967, doi: 10.1037/h0024648.
- [106] G. Subramanian, N. Cholendiran, K. Prathyusha, N. Balasubramanain, and J. Aravinth, "Multimodal Emotion Recognition Using Different Fusion Techniques," in *Proceedings of 2021 IEEE 7th International Conference on Bio Signals, Images and Instrumentation, ICBSII 2021*, 2021, doi: 10.1109/ICBSII51839.2021.9445146.
- [107] F. Ringeval *et al.*, "AV+EC 2015 - The first affect recognition challenge bridging across audio, video, and physiological data," in *AVEC 2015 - Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge, co-Located with MM 2015*, 2015, doi: 10.1145/2808196.2811642.
- [108] G. Castellano, L. Kessous, and G. Caridakis, "Emotion recognition through multiple

- modalities: Face, body gesture, speech,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, doi: 10.1007/978-3-540-85099-1_8.
- [109] M. Lanze *et al.*, “Multimodal emotion recognition using a spontaneous Filipino emotion database,” in *2010 3rd International Conference on Human-Centric Computing, HumanCom 2010*, 2010, doi: 10.1109/HUMANCOM.2010.5563314.
- [110] C. Busso *et al.*, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in *ICMI'04 - Sixth International Conference on Multimodal Interfaces*, 2004, doi: 10.1145/1027933.1027968.
- [111] G. Caridakis *et al.*, “Multimodal emotion recognition from expressive faces, body gestures and speech,” in *IFIP International Federation for Information Processing*, 2007, doi: 10.1007/978-0-387-74161-1_41.
- [112] I. Kanluan, M. Grimm, and K. Kroschel, “Audio-visual emotion recognition using an emotion space concept,” in *European Signal Processing Conference*, 2008.
- [113] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. 2014.
- [114] F. Schwenker, C. R. Dietrich, C. Thiel, and G. Palm, “Learning of Decision Fusion Mappings for Pattern Recognition,” 2006.
- [115] J. C. Lin, C. H. Wu, and W. L. Wei, “Error weighted semi-coupled hidden markov model for audio-visual emotion recognition,” *IEEE Trans. Multimed.*, 2012, doi: 10.1109/TMM.2011.2171334.
- [116] G. Caridakis, L. Malatesta, L. Kessous, N. Amir, A. Raouzaiou, and K. Karpouzis, “Modeling naturalistic affective states via facial and vocal expressions recognition,” in *ICMI'06: 8th International Conference on Multimodal Interfaces, Conference Proceeding*, 2006, doi: 10.1145/1180995.1181029.
- [117] L. Todorovski and S. Džeroski, “Combining classifiers with meta decision trees,” *Mach. Learn.*, 2003, doi: 10.1023/A:1021709817809.
- [118] M. Glodek, S. Scherer, and F. Schwenker, “Conditioned hidden Markov model fusion for multimodal classification,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2011.
- [119] M. Glodek, M. Schels, G. Palm, and F. Schwenker, “Multi-modal Fusion based on classifiers using reject options and Markov Fusion Networks,” in *Proceedings - International Conference on Pattern Recognition*, 2012.
- [120] G. Krell *et al.*, “Fusion of fragmentary classifier decisions for affective state recognition,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, doi: 10.1007/978-3-642-37081-6_13.
- [121] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Fluids Eng. Trans. ASME*, 1960, doi: 10.1115/1.3662552.
- [122] R. Blackman, S. S. and Popoli, “Design and analysis of modern tracking systems,” *Artech House*. 1999.
- [123] M. Glodek, S. Reuter, M. Schels, K. Dietmayer, and F. Schwenker, “Kalman filter based classifier fusion for affective state recognition,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, doi: 10.1007/978-3-642-38067-9_8.
- [124] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: A survey,” *Multimed. Syst.*, 2010, doi: 10.1007/s00530-010-0182-0.
- [125] H. Gunes and B. B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” *Image Vis. Comput. Affect Anal. Contin. Input*, vol. 31, no. 2, pp. 120–136, Feb. 2013.
- [126] S. Poria, E. Cambria, N. Howard, G. Bin Huang, and A. Hussain, “Fusing audio, visual and textual clues for sentiment analysis from multimodal content,” *Neurocomputing*, 2016, doi: 10.1016/j.neucom.2015.01.095.
- [127] M. Pantic and L. J. M. Rothkrantz, “Toward an affect-sensitive multimodal human-computer interaction,” in *Proceedings of the IEEE*, 2003, doi: 10.1109/JPROC.2003.817122.
- [128] L. S. Chen, T. S. Huang, T. Miyasato, and R. Nakatsu, “Multimodal human emotion/expression recognition,” in *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, 1998, doi: 10.1109/AFGR.1998.670976.

- [129] A. Landowska, "Towards emotion acquisition in IT usability evaluation context," in *ACM International Conference Proceeding Series*, 2015, doi: 10.1145/2814464.2814470.
- [130] G. Wickstrom and T. Bendix, "The "Hawthorne effect" - What did the original Hawthorne studies actually show?," *Scandinavian Journal of Work, Environment and Health*. 2000, doi: 10.5271/sjweh.555.
- [131] J. Kiefer, "Conditional confidence statements and confidence estimators," *J. Am. Stat. Assoc.*, vol. 72, no. 360, 1977, doi: 10.1080/01621459.1977.10479956.
- [132] M. Schmitt, F. Ringeval, and B. Schuller, "At the border of acoustics and linguistics: Bag-of-audio-words for the recognition of emotions in speech," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016, doi: 10.21437/Interspeech.2016-1124.
- [133] A. Landowska, "Emotion monitor - Concept, construction and lessons learned," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, 2015, no. October 2015, pp. 75–80, doi: 10.15439/2015F264.
- [134] G. Brodny, "Monitor emocji w interakcji człowiek-komputer," Politechnika Gdańska, 2015.
- [135] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of the 13th ACM International Conference on Multimedia, MM 2005*, 2005, doi: 10.1145/1101149.1101300.
- [136] L. Cyra and J. Górski, "Expert assessment of arguments: A method and its experimental evaluation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, doi: 10.1007/978-3-540-87698-4_25.
- [137] A. Landowska, "Raport techniczny 34/2012 - Metoda zapewniania wiarygodności procesu rozpoznawania stanów emocjonalnych w interakcji człowiek-komputer," Gdańsk, 2012.
- [138] Statsoft, "Internetowy Podręcznik Statystyki." [Online]. Available: https://www.statsoft.pl/textbook/stathome_stat.html. [Accessed: 31-Jan-2020].
- [139] A. F. M. Smith and G. Shafer, "A Mathematical Theory of Evidence.," *Biometrics*, 1976, doi: 10.2307/2529769.
- [140] K. S. Małgorzata Lewicka, "System biometryczny identyfikujący osoby na podstawie cech osobniczych twarzy," AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE, 2008.
- [141] R. van Solingen, V. Basili, G. Caldiera, and H. D. Rombach, "Goal Question Metric (GQM) Approach," in *Encyclopedia of Software Engineering*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2002.
- [142] M. Babar, L. Zhu, and R. Jeffery, "A framework for classifying and comparing software architecture evaluation methods," *Softw. Eng. Conf.*, 2004.
- [143] R. Kazman, M. Klein, and P. Clements, "ATAM: Method for architecture evaluation," 2000.
- [144] K. A. M. Ferreira, M. A. S. Bigonha, R. S. Bigonha, L. F. O. Mendes, and H. C. Almeida, "Identifying thresholds for object-oriented software metrics," in *Journal of Systems and Software*, 2012, doi: 10.1016/j.jss.2011.05.044.
- [145] T. L. Alves, C. Ypma, and J. Visser, "Deriving metric thresholds from benchmark data," in *2010 IEEE International Conference on Software Maintenance*, 2010, pp. 1–10, doi: 10.1109/ICSM.2010.5609747.
- [146] P. Oliveira, M. T. Valente, and F. P. Lima, "Extracting relative thresholds for source code metrics," in *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*, 2014, pp. 254–263, doi: 10.1109/CSMR-WCRE.2014.6747177.
- [147] S. Herbold, J. Grabowski, and S. Waack, "Calculation and optimization of thresholds for sets of software metrics," *Empir. Softw. Eng.*, vol. 16, no. 6, pp. 812–841, Dec. 2011, doi: 10.1007/s10664-011-9162-z.
- [148] A. Landowska, "Web questionnaire as construction method of affect-annotated lexicon - Risks reduction strategy," in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, 2015, doi: 10.1109/ACII.2015.7344605.
- [149] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, "Evaluation criteria for affect-annotated databases," in *Communications in Computer and Information Science*, 2015, doi: 10.1007/978-3-319-18422-7_52.
- [150] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, "AVEC 2011 - The first international audio/visual emotion challenge," in *Lecture Notes in Computer*

- Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, doi: 10.1007/978-3-642-24571-8_53.
- [151] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, 2012, doi: 10.1109/T-AFFC.2011.25.
- [152] E. Douglas-cowie, R. Cowie, C. Cox, N. Amir, and D. Heylen, "The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation," in *LREC Workshop on Corpora for Research on Emotion and Affect*, 2008.
- [153] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder, "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Trans. Affect. Comput.*, 2012, doi: 10.1109/T-AFFC.2011.20.
- [154] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'Feeltrace': An instrument for recording perceived emotion in real time.," *ISCA Work. Speech {&} Emot.*, 2000.
- [155] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, doi: 10.1007/978-3-642-15892-6_39.
- [156] S. Mariooryad and C. Busso, "Analysis and compensation of the reaction lag of evaluators in continuous emotional annotations," in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 2013, doi: 10.1109/ACII.2013.21.
- [157] G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic, "The semaine corpus of emotionally coloured character interactions," in *2010 IEEE International Conference on Multimedia and Expo, ICME 2010*, 2010, doi: 10.1109/ICME.2010.5583006.
- [158] A. Różyński, "Rozpoznawanie emocji na podstawie analizy głosu," Prace dyplomowe WETI Politechnika Gdańska, 2020.
- [159] G. Brodny, A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, "Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions," in *Proceedings - 2016 9th International Conference on Human System Interactions, HSI 2016*, 2016, doi: 10.1109/HSI.2016.7529664.
- [160] J. Buolamwini, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *," *Proc. Mach. Learn. Res.*, vol. 81, 2018.

WYKAZ RYSUNKÓW

Rys 2.1 Struktura emocji w procesie podejmowania decyzji [31].....	17
Rys 2.2 Kontekst rozpoznawania emocji [opracowanie własne].....	19
Rys 2.3 Model przestrzeni wartość–pobudzenie [33].....	22
Rys 2.4 Emocje podstawowe z modelu reprezentacji emocji podstawowych Ekmana w przestrzeni modelu reprezentacji emocji PAD [35].....	23
Rys 3.1 Klasyfikacja metod integracji późnej fuzji [opracowanie własne].....	41
Rys 3.2 Schemat działania filtra Kalmana [34].....	48
Rys 5.1 Koncepcyjny model integracji w wielomodalnym rozpoznawaniu emocji [3].....	59
Rys 5.2 Diagram klas wzorca projektowego Scoreboard [4].....	61
Rys 5.3 Diagram komponentowy monitora emocji [3].....	67
Rys 5.4 Architektura Monitora Emocji zaimplementowanego z pomocą Web Serwisów [5].....	68
Rys 5.5 Schemat blokowy algorytmu CFIEE [Opracowanie własne].....	70
Rys 5.6 Wykres rozmycia hipotezy w punkcie 0, przy współczynniku wiarygodności równym: a) 1 b) 0,75 c) 0,5 d) 0,25 dla danych z przykładu kroku 2 [opracowanie własne].....	76
Rys 5.7 Wartości funkcji oceny w zależności od współczynnika pewności hipotezy oraz wartości błędu bezwzględnego [opracowanie własne].....	80
Rys 5.8 Osie obrotów twarzy w przestrzeni 3D [139].....	81
Rys 7.1 Dekompozycja wiarygodności w rozpoznawaniu emocji [20].....	104
Rys 9.1 Zależność współczynnika wiarygodności od odległości hipotez, dla współczynnika wiarygodności równego 0,9 oraz wartości progowej odległości 0,05. Ref – rozwiązania referencyjne, CFIEE – zaproponowana metoda w dwóch wariantach [opracowanie własne].....	149
Rys 10.1 Schemat rozmieszczenia kamer w czasie eksperymentu [opracowanie własne].....	161
Rys 10.2 Wizualizacja kilku wybranych wartości zaproponowanej skali zbieżności dla czterech kanałów obserwacji [opracowanie własne].....	164

WYKAZ TABEL

Tabela 2.1 Przykłady obserwacji jednomodalnych, wielomodalnych oraz jednokanałowych i wielokanałowych [opracowanie własne].	19
Tabela 2.2 Wybrane modele reprezentacji emocji, bazujące na emocjach podstawowych [opracowanie własne].	21
Tabela 2.3 Zestawienie par: modalności i modelu reprezentacji emocji stosowane przez różnych autorów [opracowanie własne].	29
Tabela 3.1 Statystyki dotyczące metod fuzji [77]	34
Tabela 4.1 Dostępność modalności w zależności od kontekstu e-nauczania [1].	53
Tabela 4.2 Porównanie modalności w kontekście edukacyjnym. Wartość Wysoka – oznacza lepsze dostosowanie, Niska – oznacza gorsze dopasowanie [1].	53
Tabela 5.1 Przykładowe wartości dla normalizacji danych z różnych modeli reprezentacji emocji [opracowanie własne].	73
Tabela 5.2 Przykładowe wartości błędów i wyliczone przy nich oceny wyznaczone przez proponowaną funkcję oceny [opracowanie własne].	79
Tabela 5.3 Wartości kątów odchylenia od twarzy na wprost kamery oraz wartości proponowanych miar niepewności [opracowanie własne].	82
Tabela 5.4 Proponowane komponenty rozwiązania i ich konsekwencje [opracowanie własne].	84
Tabela 6.1 Cel i pytania do metody GQM, na podstawie [140].	88
Tabela 6.2 Pokrycie miar jakości przez eksperymenty i symulacje [opracowanie własne].	90
Tabela 6.3 Skuteczność rozpoznawania emocji w podziale na zbiory danych i wykorzystane rozwiązanie integracji dla wyników eksperymentu. Na podstawie [4].	91
Tabela 6.4 Miary adaptawalności (A1–A6) dla rozwiązania z eksperymentu z programami do analizy ładunku emocjonalnego z tekstu. Na podstawie [4].	92
Tabela 6.5 Miary adaptawalności (narzuty – A7 i A8) dla eksperymentu z programami do analizy ładunku emocjonalnego z tekstu. Na podstawie [4].	93
Tabela 6.6 Wariant I analizy odporności – trzy „dobre” i jeden „słaby” program. Skuteczność rozpoznawania emocji w podziale na zbiory danych z eksperymentu. Na podstawie [4].	94
Tabela 6.7 Scenariusz II analizy odporności – trzy „dobre” i jeden „słaby” program. Skuteczność rozpoznawania emocji w podziale na zbiory danych z eksperymentu. Na podstawie [4].	94
Tabela 6.8 Miary odporności dla symulacji ze atrapami (ang. mock-up) ekspertów. Na podstawie [4].	95
Tabela 6.9 Miary modularyzacji dla studium przypadku. Na podstawie [4].	97
Tabela 6.10 Miary adaptawalności (narzuty – A7 i A8) dla studium przypadku Monitora emocji (zróżnicowane technologie). Na podstawie [4].	97
Tabela 6.11 Scenariusze ATAM dla studium przypadku Monitora Emocji. Na podstawie [4].	99
Tabela 6.12 Podejścia architektoniczne ATAM dla studium przypadku Monitora Emocji. Na podstawie [4].	100
Tabela 6.13 Powiązanie scenariuszy ATAM z podejściami ATAM zaprojektowanymi dla studium przypadku Monitora Emocji. Na podstawie [4].	101
Tabela 7.1 Operacjonalizacja zmiennych dla określenia skuteczności rozpoznania kanału audio w bazie SEMAINE [opracowanie własne].	108
Tabela 7.2 Skuteczność (ang. accuracy) rozpoznania emocji poprzez program automatycznego rozpoznawania emocji bazujący na kanałach audio z bazy SEMAINE na 3983 próbkach w podziale na poszczególne emocje [opracowanie własne].	108
Tabela 7.3 Operacjonalizacja zmiennych dla określenia skuteczności rozpoznania kanału wideo w bazie SEMAINE [opracowanie własne].	109
Tabela 7.4 Skuteczność rozpoznawania emocji programu Noldus FaceReader na danych z trzech kamer z bazy SEMAINE. Analiza w podziale na lokalizację kamery, emocję oraz wartość progową odległości [opracowanie własne].	110
Tabela 7.5 Wartości zaproponowanych współczynników wiarygodności oraz wskaźnika jakości programu Noldus FaceReader w podziale na kanały dostępu [opracowanie własne].	111
Tabela 7.6 Wartości zaproponowanych współczynników wiarygodności w podziale na kanały, poszczególne emocje oraz przypadki rozpoznane poprawnie (oznaczone jako S) i nieprawidłowo (oznaczone jako F) dla wartości progowej odległości równej 0,05	

[opracowanie własne].....	113
Tabela 7.7 Operacjonalizacja zmiennych dla eksperymentu 3a – dobór współczynnika zaufania [opracowanie własne].....	114
Tabela 7.8 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od doboru współczynnika wiarygodności rozwiązań jednomodalnych [opracowanie własne].....	115
Tabela 7.9 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od doboru współczynnika wiarygodności rozwiązań jednomodalnych. Tabela zawiera dane dla emocji wtrętu (ang. disgust) [opracowanie własne].	117
Tabela 7.10 Zestawienie najlepszych wariantów współczynnika wiarygodności dla kanału audio w podziale na emocje oraz wartości progowych odległości (D_{max}) [opracowanie własne].	117
Tabela 7.11 Wartości współczynników wiarygodności dla poszczególnych emocji, dla kanału audio w podziale na wartości progowe odległości (D_{max}) [opracowanie własne].	118
Tabela 7.12 Operacjonalizacja zmiennych dla eksperymentu 3b – uśrednianie wartości współczynnika wiarygodności [opracowanie własne].	119
Tabela 7.13 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności dla wariantu z uśrednieniem lub indywidualną wartością współczynnik wiarygodności hipotez. Dane bez uwzględnienia podziału na rodzaj emocji [opracowanie własne].....	119
Tabela 7.14 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności dla wariantu z uśrednieniem lub indywidualną wartością współczynnik wiarygodności hipotez. Dane dla emocji smutku [opracowanie własne].	120
Tabela 8.1 Dostępność hipotez stanu emocjonalnego w zależności od kamery dla ramek czasowych 100 ms [opracowanie własne].	124
Tabela 8.2 Dostępność kanału audio dla użytkownika w bazie SEMAINE [opracowanie własne]	124
Tabela 8.3 Zestawienie dostępności modalności związanej z analizą mimiki oraz prozodii mowy (wypowiedzi badanego) [opracowanie własne]	125
Tabela 8.4 Operacjonalizacja zmiennych dla eksperymentu 5 [opracowanie własne].	127
Tabela 8.5 Operacjonalizacja zmiennych dla eksperymentu 5a – parametry rozkładu normalnego [opracowanie własne].	128
Tabela 8.6 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla parametrów rozkładu normalnego w trzech wariantach [opracowanie własne].....	128
Tabela 8.7 Macierz emocji oraz wartości progowych ze wskazaniem numeru wariantu, dla którego osiągnięto najwyższą skuteczność [opracowanie własne].	129
Tabela 8.8 Operacjonalizacja zmiennych dla eksperymentu 5b – funkcja zaufania [opracowanie własne].	130
Tabela 8.9 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych funkcji zaufania – iloczynowej (1) oraz opartej na średniej (2). Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].	131
Tabela 8.10 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych funkcji zaufania – iloczynowej (1) oraz opartej na średniej (2). Tabela zawiera dane dla emocji smutku [opracowanie własne].	132
Tabela 8.11 Operacjonalizacja zmiennych dla eksperymentu 4c – odrzucenie hipotez o niskiej wiarygodności [opracowanie własne].....	133
Tabela 8.12 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla odrzucenia wartości o niskiej wiarygodności (2) oraz bez (1). Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].	133
Tabela 8.13 Operacjonalizacja zmiennych dla eksperymentu 5d – sposób wyboru hipotezy [opracowanie własne].....	134
Tabela 8.14 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności dla różnych metod wyboru hipotezy. Tabela zawiera dane bez rozróżnienia emocji [opracowanie własne].....	135
Tabela 8.15 Operacjonalizacja zmiennych dla eksperymentu 6 – metoda rozpoznawania emocji [opracowanie własne].....	137
Tabela 8.16 Skuteczność rozpoznawania emocji dla pojedynczych kanałów, programów referencyjnych oraz zaproponowanej metody CFIEE, dla wartości progowych odległości (D_{max}) [opracowanie własne].	138
Tabela 8.17 Operacjonalizacja zmiennych dla eksperymentu 7 – sposób wyznaczania	

współczynnika wiarygodności przez metody rozpoznawania emocji [opracowanie własne].	141
Tabela 8.18 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od wybranej metody integracji [opracowanie własne].	143
Tabela 8.19 Zmiana średniej skuteczności dla różnych wartości współczynnika wiarygodności w zależności od wybranej metody integracji. Dane dla emocji gniewu [opracowanie własne].	145
Tabela 9.1 Operacjonalizacja zmiennych dla symulacji 2 [opracowanie własne].	149
Tabela 9.2 Zależność wartości współczynnika wiarygodności uzgodnionego rozwiązania od odległości hipotez, dla współczynnika wiarygodności hipotez równego 0,9. Wybrane wartości. Pogrubione zostały wartości graniczne (wartości 0,9 oraz 0) [opracowanie własne].	150
Tabela 9.3 Operacjonalizacja zmiennych dla symulacji 3 [opracowanie własne].	152
Tabela 9.4 Współczynnik wiarygodności uzgodnionego rozwiązania w zależności od współczynników wiarygodności rozbieżnych hipotez A i B. Dane w podziale na metodę. Wartość progowa odległości równa 0,05, odległość hipotez równa 0,5 [opracowanie własne].	153
Tabela 9.5 Współczynnik wiarygodności uzgodnionego rozwiązania w zależności od współczynników wiarygodności rozbieżnych hipotez A i B. Odległość hipotez równa 1 [opracowanie własne].	154
Tabela 9.6 Operacjonalizacja zmiennych dla symulacji 4 [opracowanie własne].	155
Tabela 9.7 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej dla odległości hipotezy rozbieżnej od bliższej zbieżnej hipotezy równej 0,5. Dla wartości progowej odległości równej 0,05 oraz współczynnika wiarygodności zbieżnych hipotez równych 0.9 [opracowanie własne].	155
Tabela 9.8 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od wartości hipotezy rozbieżnej dla odległości hipotezy rozbieżnej od bliższej zbieżnej hipotezy równej 0,5. Zestawienie w rozbiu na różne wartości progowe odległości dla drugiego wariantu zaproponowanej metody [opracowanie własne].	156
Tabela 9.9 Operacjonalizacja zmiennych dla symulacji 5 [opracowanie własne].	157
Tabela 9.10 Zależność współczynnika wiarygodności uzgodnionego rozwiązania od odległości rozbieżnej hipotezy, dla wartości wiarygodności hipotezy rozbieżnej równej 0,4 [opracowanie własne].	158
Tabela 10.1 Liczba klatek z eksperymentu w rozbiu na cztery kamery oraz uczestników badania, w ujęciu surowym (przed próbowaniem) oraz po próbkowaniu do 250 ms [opracowanie własne].	161
Tabela 10.2 Zestawienie dostępności modalności mimiki twarzy dla danych eksperymentu przeprowadzonego na Politechnice Gdańskiej [opracowanie własne].	162
Tabela 10.3 Opis zaproponowanej skali zbieżności dla czterech kanałów obserwacji [opracowanie własne].	163
Tabela 10.4 Przykładowe wartości czterech hipotez (oznaczonych jako A, B, C oraz D) dla poszczególnych klas zbieżności wraz z opisem [opracowanie własne].	164
Tabela 10.5 Procentowany rozkład przypadków w rozbiu na poziomy zbieżności dla emocji zniesmaczenia i strachu w podziale na wartości progowe odległości (D_{max}) oraz zaproponowane klasy zbieżności [opracowanie własne].	165
Tabela 10.6 Zbieżność danych z czterech kamer dla emocji gniewu oraz smutku w podziale na wartości progowe odległości (D_{max}) oraz zaproponowane klasy zbieżności [opracowanie własne].	166
Tabela 10.7 Operacjonalizacja zmiennych dla eksperymentu 9 – porównanie wartości współczynnika wiarygodności z poziomem zbieżności hipotez [opracowanie własne].	167
Tabela 10.8 Wartość współczynnika wiarygodności uzgodnionego rozwiązania w podziale na poziom zbieżności, wartość progową odległości oraz metodę integracji. Bez podziału na emocję [opracowanie własne].	168
Tabela 10.9 Wartość współczynnika wiarygodności uzgodnionego rozwiązania w podziale na poziom zbieżności, wartość progową odległości oraz metodę integracji. Wartości dla wymiaru pobudzenia [opracowanie własne].	170

WYKAZ ZAŁĄCZNIKÓW

- A Wartości zaproponowanych współczynników wiarygodności w podziale na kanały oraz poszczególne emocje
- B Parametry konfiguracyjne metody CFIEE oraz dobór współczynnika wiarygodności dla kanału audio – porównanie poprawności wyznaczenia współczynnika wiarygodności
- C Porównanie skuteczności rozpoznawania emocji dla pojedynczych kanałów, programów referencyjnych oraz zaproponowanej metody CFIEE
- D Skuteczność metod integracji w podziale na przedziały współczynnika wiarygodności hipotezy końcowej
- E Wartość współczynnika wiarygodności dla uzgodnionego rozwiązania dla dwóch hipotez o równej wartości współczynnika wiarygodności w zależności od wartości progowej odległości oraz współczynnika wiarygodności hipotez oraz odległości między hipotezami
- F Wartość współczynnika wiarygodności dla uzgodnionego rozwiązania dla dwóch hipotez o stałej odległości oraz różnej wartości współczynnika wiarygodności w zależności od wartości progowej odległości między hipotezami
- G Wartość współczynnika wiarygodności dla uzgodnionego rozwiązania dla dwóch hipotez zbieżnych o stałej odległości oraz trzeciej hipotezie rozbieżnej o różnej wartości współczynnika wiarygodności oraz odległości od bliższej ze zbieżnych hipotez.
- H Procentowy rozkład przypadków w podziale na poziomy zbieżności.
- I Wartość współczynnika wiarygodności uzgodnionego rozwiązania w zależności od poziomu zbieżności.