



**AKADEMIA GÓRNICZO-HUTNICZA**

**im. Stanisława Staszica w Krakowie**

**WYDZIAŁ INŻYNIERII  
MECHANICZNEJ I ROBOTYKI**

---

# **Praca dyplomowa inżynierska**

**Magdalena Karpierz**

*Imię i nazwisko*

**Inżynieria Akustyczna**

*Kierunek studiów*

**Ocena emocji w korpusie mowy  
emocjonalnej**

*Temat pracy dyplomowej*

**Dr inż. Bartosz Ziółko**

*Promotor pracy*

.....

*Ocena*

Kraków, rok  
2015/2016

Kraków, dn 15.01.2016

Imię i nazwisko: Magdalena Karpierz

Nr albumu: 259219

Kierunek studiów: Inżynieria Akustyczna

Specjalność: -

### OŚWIADCZENIE

Świadoma odpowiedzialności karnej za poświadczanie nieprawdy oświadczam, że niniejszą inżynierską pracę dyplomową wykonałam osobiście i samodzielnie oraz nie korzystałam ze źródeł innych niż wymienione w pracy.

Jednocześnie oświadczam, że dokumentacja praca nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (Dz. U. z 2006 r. Nr 90 poz. 631 z późniejszymi zmianami) oraz dóbr osobistych chronionych prawem cywilnym. Nie zawiera ona również danych i informacji, które uzyskałam w sposób niedozwolony. Wersja dokumentacji dołączona przeze mnie na nośniku elektronicznym jest w pełni zgodna z wydrukiem przedstawionym do recenzji.

Zaświadczam także, że niniejsza inżynierska praca dyplomowa nie była wcześniej podstawą żadnej innej urzędowej procedury związanej z nadawaniem dyplomów wyższej uczelni lub tytułów zawodowych.

.....

*podpis dyplomanta*

Kraków, dn 15.01.2016

Imię i nazwisko: Magdalena Karpierz

Nr albumu: 259219

Kierunek studiów: Inżynieria Akustyczna

Specjalność: -

## OŚWIADCZENIE

Niniejszym oświadczam, że zachowując moje prawa autorskie, udzielam Akademii Górniczo-Hutniczej im. S. Staszica w Krakowie nieograniczonej w czasie nieodpłatnej licencji niewyłącznej do korzystania z przedstawionej dokumentacji inżynierskiej pracy dyplomowej, w zakresie publicznego udostępniania i rozpowszechniania w wersji drukowanej i elektronicznej.

Kraków, .....

.....

*data*                      *podpis dyplomanta*

---

<sup>i</sup> Na podstawie Ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (Dz.U. 2005 nr 164 poz. 1365) Art. 239. oraz Ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz.U. z 2000 r. Nr 80, poz. 904, z późn. zm.) Art. 15a. "Uczelni w rozumieniu przepisów o szkolnictwie wyższym przysługuje pierwszeństwo w opublikowaniu pracy dyplomowej studenta. Jeżeli uczelnia nie opublikowała pracy dyplomowej w ciągu 6 miesięcy od jej obrony, student, który ją przygotował, może ją opublikować, chyba że praca dyplomowa jest częścią utworu zbiorowego."

Kraków, dnia 15.01.2015

**AKADEMIA GÓRNICZO-HUTNICZA  
WYDZIAŁ INŻYNIERII MECHANICZNEJ I ROBOTYKI**

**TEMATYKA PRACY DYPLOMOWEJ  
INŻYNIERSKIEJ**  
dla studenta IV roku studiów stacjonarnych

**Magdalena Karpierz**

*Imię i nazwisko studenta*

TEMAT PRACY DYPLOMOWEJ INŻYNIERSKIEJ:

**Ocena emocji w emocjonalnym korpusie mowy**

*Promotor pracy:* Dr inż. Bartosz Ziółko

*Recenzent pracy:* [Tytuł, imię i nazwisko Recenzenta]

.....  
*Podpis dziekana:*

**PLAN PRACY DYPLOMOWEJ**

1. Omówienie tematu pracy i sposobu realizacji z promotorem.
2. Zebranie i opracowanie literatury dotyczącej tematu pracy.
3. Zebranie i opracowanie wyników badań.
4. Analiza wyników badań, ich omówienie i zatwierdzenie przez promotora.
5. Opracowanie redakcyjne.

Kraków, 15.01.2015 .....

*data*

*podpis dyplomanta*

**TERMIN ODDANIA DO DZIEKANATU:**

..... 20.....

.....  
*podpis promotora*

Akademia Górniczo-Hutnicza im. Stanisława Staszica  
**Wydział Inżynierii Mechanicznej i Robotyki**

Kraków, .....

Kierunek:

Profil dyplomowania:

[imię i nazwisko autora]

**Praca dyplomowa inżynierska**

[temat]

Opiekun: [stopień naukowy, imię i nazwisko promotora]

### STRESZCZENIE

[Treść streszczenia, maksymalnie do końca strony, Times New Roman 12 pkt]

AGH University of Science and Technology  
**Faculty of Mechanical Engineering and Robotics**

Kraków, the.....

Field of Study:

Specialisations:

**[First name and family name of the Author]**

**Engineer Diploma Thesis**

**[Title of the project in English]**

Supervisor: [degree, first name and family name of the Supervisor]

#### SUMMARY

[The summary content, must fit within the page limit Times New Roman 12 pkt]

# Spis treści

1. Wstęp
  - 1.1 Wprowadzenie
  - 1.2 Cele pracy
2. Część teoretyczna
  - 2.1 Definicja emocji i ich właściwości
  - 2.2 Emocje, a inne pojęcia afektywne
  - 2.3 Neurofizjologia emocji
  - 2.4 Uniwersalność emocji
  - 2.5 Funkcje emocji
3. Emocje w głosie
  - 3.1 Wokalna ekspresja emocji
    - 3.1.1 Czynniki kształtujące ekspresję
    - 3.1.2 Parametry akustyczne mowy emocjonalnej
  - 3.2 Percepcja emocji w głosie
    - 3.2.1 Percepcja, a parametry akustyczne
    - 3.2.2 Rozpoznanie emocji
    - 3.2.3 Płeć, a emocje
4. Testowanie baz nagrań mowy emocjonalnej
  - 4.1 Stosowane modele emocji
  - 4.2 Przegląd wybranych korpusów mowy
5. Test odsłuchowy
  - 5.1 Korpus mowy emocjonalnej AGH
  - 5.2 Grupa ankietowanych
  - 5.3 Metoda badania
6. Analiza wyników
  - 4.1 Średnie rozpoznanie
  - 4.2 Rozpoznanie poszczególnych emocji
  - 4.3 Nagrania z referencją
  - 4.4 Samoocena, a wynik
  - 4.5 Nastrój ankietowanego, a wynik
  - 4.6 Wykształcenie ankietowanego, a wynik
  - 4.7 Ocena testu przez ankietowanych

## 7. Wnioski i dalszy rozwój pracy

# 1. Wstęp

## 1.1 Wprowadzenie

Mowa zawiera i odzwierciedla wiele informacji na temat mówcy. Zarówno jej lingwistyczne jak i paralingwistyczne (np. ton głosu, tempo mowy, intonacja) elementy są ważną częścią międzyludzkiej komunikacji. Rozpoznawanie emocji innych osób z wyrazu twarzy, głosu czy postawy ciała to kluczowe umiejętności w społecznych interakcjach. Badania nad rozpoznawaniem emocjonalnych stanów z niewerbalnych komunikatów mają więc długą historię w psychologii.

Rozpoznanie emocji również w komunikacji ludzi z komputerami ma ogromne znaczenie. W książce „The media equation” wykazano, że ludzie kontaktują się z komputerami w taki sam sposób, jak ze sobą nawzajem [1]. Oczekują grzeczności i odpowiedniego nastawienia do swoich postaw i stanów. Rozpoznanie i synteza emocji odgrywa coraz ważniejszą rolę i służy różnorodnym celom: pomaga ulepszyć aplikacje tworząc środowiska przyjazne użytkownikom, umożliwia tworzenie bardziej naturalnych wirtualnych charakterów, a także wspomaga rozwój prac psychologicznych. Rozwój systemów rozpoznawania emocji w mowie wymaga tworzenia baz nagrań mowy emocjonalnej, na których można trenować systemy. Najczęstszym sposobem realizacji takich nagrań jest rejestracja odegranych przez (profesjonalnych lub nie) aktorów emocji. Emocje są doświadczeniem subiektywnym tzn. wyrażanie i odbiór emocji jest odmienne u różnych osób, dlatego ważnym elementem korpusu mowy emocjonalnej jest informacja o tym, jak postrzega go reprezentatywna grupa ludzi.

Wybór tematu pracy jest umotywowany zaciekawieniem autorki percepcją dźwięku oraz emocjami. Pogłębianie tego tematu w czasie realizacji tej pracy, pozwoliło na pozyskanie naukowego spojrzenia na tę dziedzinę oraz zdobycia pierwszego doświadczenia w przeprowadzeniu testów odsłuchowych.



Drugi rozdział praca skupia się na ogólnym zjawisku emocji podając ich definicje i cechy. Kolejny opisuje zjawiska związane z emocjami w głosie. Czwarty rozdział jest przybliżeniem stosowanych rozwiązań podczas tworzenia baz nagrań mowy emocjonalnej oraz testowania ich na grupie słuchaczy. W dalszej części pracy opisane jest przeprowadzenie oceny korpusu mowy emocjonalnej przez autorkę pracy, wyniki testu odsłuchowego, ich analiza oraz wnioski.

## **1.2 Cele pracy**

Przedmiotem badania jest opracowana w Akademii Górniczo-Hutniczej baza danych nagrań mowy emocjonalnej zgromadzona w celu badań nad zawartością afektywną sygnału mowy [2]. Za cele niniejszej pracy obrano przygotowanie i przeprowadzenie testu odsłuchowego, który pozwoli na ocenienie tych nagrań przez reprezentatywną grupę słuchaczy. Kolejnym etapem badania jest analiza wyników przeprowadzonego testu w celu wskazania rozpoznania poszczególnych emocji, ewentualnych mocniejszych i słabszych stron korpusu. Ocena nagrań pozwoli na wzbogacenie bazy danych o opinie słuchaczy. W przyszłości może również stanowić podstawę do porównania rozpoznania emocji w nagraniach przez ludzi z rozpoznaniem emocji przez program. Sprawdzone zostały również hipotezy czy wykształcenie słuchaczy, ich samoocena umiejętności rozpoznawania emocji i aktualny nastrój w czasie testu odsłuchowego mają związek z ich wynikami.

## **2. Część teoretyczna**

### **2.1 Definicja emocji i ich właściwości**

Emocje na ogół są definiowane jako złożony zespół zmian cielesnych i psychicznych, zawierających pobudzenie fizjologiczne, procesy poznawcze, reakcje behawioralne i uczucia [3]. Pobudzeniem fizjologicznym są zmiany neuronalne, hormonalne, trzewiowe oraz mięśniowe. Na procesy poznawcze składają się interpretacje, wspomnienia i oczekiwania. Emocje mogą mieć charakter ekspresyjny np. śmiech lub płacz, mogą być również ukierunkowane na działanie. Emocje są reakcją na

zdarzenia postrzegane przez daną osobę jako ważne ( sytuację może być postrzegana jako ważna świadomie lub nieświadomie), skupiają one naszą uwagę na danych sytuacjach, oznaczając je jako szczególne pod pewnym względem i pobudzając nas do działania [3].

To jaką emocję odczuwamy, zależy od naszej interpretacji sytuacji oraz od znaczenia jakie jej nadajemy. Emocje są formą interakcji i wyrażają w jakich sposób jej doświadczamy. Określają co jest dla nas korzystne, co zagrażające, co obojętne, a co ważne. Oznacza to, że procesy emocjonalne obok procesów poznawczych odgrywają podstawową rolę w całości kształcie czynności regulacyjnych [4].

## **2.2 Emocje, a inne pojęcia afektywne**

Pojęcia takie jak emocje, nastrój lub uczucia często są używane zamiennie i nieprecyzyjnie. Co je rozróżnia? Dla niektórych badaczy podstawową różnicą jest kryterium długości trwania [5]. Emocje uważane są za krótkotrwałe, nastrój zaś za długotrwały. Ponadto emocje mają zazwyczaj jednorodny charakter, co oznacza, że w danej chwili może występować tylko jedna emocja. Są one zawsze reakcją na coś i na ogół są wzbudzone w sytuacjach wymagających działań przystosowawczych. Emocje są wrodzone, a pod względem neurofizjologicznym związane z procesami zachodzącymi w pniu mózgu [5]. W emocjach impulsywność przeważa nad refleksyjnością. Wiążą się z procesami fizjologicznymi człowieka. Zwłaszcza te najprostsze, występują zarówno u ludzi, jak i u zwierząt [4].

Psychologowie poznawczy dowodzą, że stany emocjonalne mogą wpływać na uczenie się, pamięć, oceny społeczne, procesy kategoryzacji i twórczości. Uważa się, że kiedy człowiek doświadcza danej emocji w określonej sytuacji, to emocja ta jest przechowywana jest w pamięci razem z zachodzącymi aktualnie zdarzeniami. Emocje modulują zachowanie, natomiast nastrój moduluje przetwarzanie informacji [5]. Obserwuje się zależność przypominania sobie doświadczonego w przeszłości zdarzenia emocjonalnego wtedy, gdy dana osoba jest w tym samym nastroju w jakim była podczas tego zdarzenia np. Ludzie pamiętają więcej smutnych zdarzeń, wtedy gdy czują się smutni. Przetwarzanie zgodnie z nastrojem polega na tym, że ludzie są wybiórczo uwrażliwieni na przyswajanie informacji, która jest zgodna z ich aktualnym nastrojem. Jest bardziej prawdopodobne, że na materiał zgodny z dominującym nastrojem danej

osoby zwróci ona uwagę i przetworzy dokładniej zestawiając z bardziej wypracowanymi skojarzeniami.

Uczucia dotyczą specyficznie ludzi. Są trwałe, cechuje je brak bezpośredniej zależności od aktualnej sytuacji człowieka. Uczucia są niejednorodne pod względem treściowym, co oznacza, że różne uczucia mogą występować równocześnie i wzajemnie się przenikać. Mają one charakter społeczny i są zawsze formą interakcji, a także wyrażają sposób doświadczania tej interakcji. Od strony neurofizjologicznej wiąże się to z procesami zachodzącymi w korze mózgowej i stąd w uczuciach występuje przewaga refleksyjności nad impulsywnością [4]. Uczucia nie mają charakteru wrodzonego i człowiek nabywa je w trakcie doświadczeń związanych z kontaktami z innymi ludźmi.

### **2.3 Neurofizjologia emocji**

Układ autonomiczny dzieli się na układ współczulny (pobudzający) oraz przywspółczulny (hamujący) [3]. To właśnie on, dzięki działaniu obu tych części przygotowuje organizm do reakcji emocjonalnych. Silne emocje jak strach czy gniew przygotowują organizm na potencjalne niebezpieczeństwo. Współczulny układ nerwowy wydzielają hormony, które pobudzają narządy wewnętrzne do wydzielania cukru do krwi, zwiększają wydzielanie potu i śliny, podnoszą ciśnienie krwi. Gdy zagrożenie minie, hamująca część układu nerwowego ograniczy wydzielanie aktywujących hormonów [3].

Przedmiotem współczesnych badań jest ciało migdałowate [3]. To właśnie ono nadaje znaczenie informacjom otrzymany zmysłów i działa jako filtr dla pamięci. Ludzie, których ciało migdałowate zostało uszkodzone w wyniku wypadku nie wykazują żadnych reakcji w sytuacjach normalnie wywołujących silne reakcje emocjonalne [3]. Kora mózgowa dostarcza wspomnień i skojarzeń. Odgrywa znaczącą rolę we wszystkich złożonych emocjach. Badanie wskazują na istnienie odrębnych ośrodków przetwarzania emocji pozytywnych i negatywnych. Dzięki analizie EEG wykryto związek lewej półkuli mózgowej z emocjami pozytywnymi, a prawej z negatywnymi [3].

### **2.4 Uniwersalność emocji**

Paul Ekman, wybitny badacz ekspresji mimicznej twierdzi, że wszyscy ludzie stosują i rozumieją ten sam 'język twarzy' [3]. Udowodnił, że ten sam zbiór wyrazów twarzy jest uniwersalny dla całego gatunku ludzkiego, przypuszczalnie dlatego, że są one naszym dziedzictwem ewolucyjnym. Jego zdaniem do emocji podstawowych zaliczają się strach, złość, smutek, radość, wstręt i zaskoczenie, gdyż są one doświadczane i rozpoznawane przez wszystkich ludzi. Pozostałe emocje są jego zdaniem mieszanką emocji podstawowych, a ich wyrażanie zależy od czynników kulturowych.

## **2.5 Funkcje emocji**

Trudno określić funkcje emocji, ponieważ są one skomplikowanym zjawiskiem o różnorodnych konsekwencjach. Szukając funkcji emocji należy analizować je oddzielnie, ponieważ nie można stwierdzić, że wszystkie spełniają te same funkcje [3]. Emocje są czynnikiem motywującym do zachowań ukierunkowanych na radzenie sobie z wydarzeniami, które je wywołują np. strach ma na celu ochronę przed czynnikiem go wywołującym. Decydują również o zasobach energetycznych organizmu wykorzystywanych w danej interakcji z otoczeniem. Kolejną ich funkcją jest ukierunkowanie i podtrzymanie procesów poznawczych, dzięki czemu zbliżamy się do pozytywnych bodźców i unikamy negatywnych. Im większa intensywność stymulacji, tym bardziej sterują one naszą uwagą. Emocje odgrywają ogromną rolę w komunikacji regulując nasze interakcje społeczne. Potrafią demaskować to co czujemy i zamierzamy. Znaczna część komunikacji odbywa się w języku niewerbalnych komunikatów [3]. Pozytywne emocje wiążą nas z niektórymi ludźmi, negatywne odstraszaają i sprawiają, że zachowujemy dystans wobec innych.

Wokalne sygnały mogą wywoływać i często wywołują emocjonalną odpowiedź. W takim ujęciu podstawową funkcją wokalnego przekazywania emocji nie jest ich wyrażenie, ale wpłynięcie na słuchacza i jego zachowanie [6]. Słuchacz bazując na swoim doświadczeniu z danym dźwiękiem i kontekście w jakim on się pojawia, wnioskuje w jakim stanie emocjonalnym jest mówca. Emocja nadaje wypowiedzi kontekst, dlatego ludzie są wrażliwi na stan emocjonalny mówcy i jego ukryte intencje. Umiejętności odczytywania znaczenia wypowiedzi z emocji zawartych w głosie oraz

niewerbalnych wskazówek uczymy się jako dzieci. Wykazały to badania, w których prezentowano dzieciom wypowiedzi zawierające niespójne informacje językowe i paralingwistyczne (wyraz twarzy, intonację głosu). Oczekiwana jest zgodność tych informacji, jednak w przypadku ich konfliktu czterolatki zazwyczaj uznają znaczenie słów (bardziej niż wyraz twarzy) za decydujące o charakterze wypowiedzi [7]. Większość dziesięciolatków i dorośli interpretują te same zdania korzystając głównie z komunikatów niewerbalnych [7].

### **3. Emocje w głosie**

#### **3.1 Wokalna ekspresja emocji**

##### **3.1.1 Czynniki kształtujące ekspresję**

Jeśli jedną z funkcji emocji jest motywowanie do reakcji w sposób przystosowujący do wymagań życia, to dwie zdolności są niezbędne do koordynowania zachowania społecznego: zdolność skutecznego komunikowania innym naszych odczuć oraz umiejętność odczytywania odczuć innych. Emocja jako mechanizm adaptacyjny ma na celu ułatwienie organizmowi radzenie sobie z ważnymi zdarzeniami [8]. Scherer uznał, że zmiany odpowiednich elementów są synchronizowane w obrębie epizodu emocjonalnego, ponieważ organizm mobilizuje swoje zasoby i łączy systemy, aby zwiększyć swoje szanse w obliczu ważnego wydarzenia [8]. Składniki pozostają w interakcji, więc zmiana jednego z nich skutkuje zmianą innego. Na przykład wzrost częstości oddechu, zapewniający większą ilość tlenu w momencie pobudzenia wpływa na wyraz twarzy i ekspresję wokalną (zmiana ciśnienia powietrza w pogłosie), a także na liczne fizjologiczne parametry obwodowe. Nasza ekspresja rozszerza reakcje adaptacyjną poza granicę jednostki, stanowiąc komunikat dla środowiska społecznego o reakcji na wydarzenie i chęci działania w określony sposób [8].

Najbardziej podstawowymi czynnikami determinującymi afekt wokalny są zmiany fizjologiczne towarzyszące emocjom, które powodują zmiany w funkcjonowaniu systemów wytwarzania głosu. Z emocjami łączą się rozmaite reakcje adaptacyjne w układzie nerwowym, autonomicznym i obwodowym [8]. Te reakcje prowadzą do zmian w funkcjonowaniu części systemu wytwarzania mowy: w oddechu,

drgania fałdów głosowych i w artykulacji np. wraz z silnie pobudzającą emocją taką jak wściekłość, wzrasta napięcie mięśni krtani, a to w połączeniu ze wzrostem ciśnienia pogłośni powoduje zmianę dźwięku powstającego w głośni, a w związku z tym zmianę jakości głosu. Oczekiwanie cierpienia wpływa na płynność mowy, pojawienie się nowej nieoczekiwanej informacji może objawić się błędami mowy oraz jękaniem się. Lęk niekorzystnie wpływa na planowanie wypowiedzi i jej realizację. Z kolei zadowolenie polepsza płynność mowy.

Drugim rodzajem czynników kształtujących ekspresję wokalną są normy społeczne i oczekiwania słuchaczy. Zgodnie z teorią społecznego konstruktywizmu, emocje są wytworem kultury i swoje znaczenie, a także konsekwencje zawdzięczają wyuczonym zasadom społecznym. Inteligencja społeczna i społeczne poznanie obejmują nasze zdolności do wychwycenia, zapamiętania i zinterpretowania naszych i innych zachowań w kontekście umysłowych stanów (myśli, intencji, wierzeń) oraz do oddziaływania w złożonym społecznie świecie [9].

### **3.1.2 Parametry akustyczne mowy emocjonalnej**

Zmienne akustyczne są silnie związane z przekazywaniem emocji [10]. Poziom, zakres i obwiednia częstotliwości podstawowej (która jest powiązana z częstotliwością drgań fałdów głosowych) są postrzegane jako wysokość dźwięku. Ludzie intonując wypowiedź wykorzystują modulacje częstotliwości podstawowej w kontrolowany sposób (by coś przekazać) [11]. Energia (lub amplituda) jest identyfikowana z intensywnością głosu. Rozkład energii w zakresie częstotliwości wpływa na postrzeganą barwę głosu.

Emocjonalnie nacechowana wypowiedź jest związana ze zmianą parametrów akustycznych [10]. Oznacza to, że specyficzne wzory czy konfiguracje parametrów są powiązane z danym emocjonalnym stanem. Informację dotyczące stanu emocjonalnego są zakodowane w parametrach akustycznych, a następnie dekodowane przez słuchacza w celu przypisania osobie mówiącej stanu emocjonalnego [8]. Związki pomiędzy percepcyjnymi aspektami mowy, a akustycznymi pomiarami ludzkiej mowy i dźwięków wokalnych zostały dość dobrze poznane. Przykładowo złość charakteryzuje się wzrostem średniej częstotliwości podstawowej i średniej energii, a radość ma związek ze wzrostem zakresu, zmiennością i średnią częstotliwości podstawowej [9].

## **3.2 Percepcja emocji w głosie**

### **3.2.1 Percepcja, a parametry akustyczne**

W badaniach wpływu manipulacji parametrami akustycznymi w próbie na rozpoznanie emocji najbardziej wpływową zmienną okazał się być zakres częstotliwości podstawowej [6]. Wąski zakres częstotliwości podstawowej był oceniany jak sygnalizowanie smutku, a szeroki jako silne pobudzenie negatywną emocją jak złość czy irytacja. Mowa o wysokim natężeniu również była postrzegana jako negatywna emocja lub agresja. Szybka mowa jest interpretowana jako wyraz radości, wolna jako oznaka smutku. Dużą rolę odgrywa również głośność nagrania [6].

Banse i Scherer [10, 8] dokonali również porównania wyników ocen dokonanych przez ludzi z algorytmami statystycznymi takimi jak analiza dyskryminacyjna i analiza interpolacyjna. Stwierdzili podobieństwo współczynników dyskryminacji oraz wysoką odpowiedniość między układami pomyłek, co świadczy o tym, że ludzie rozpoznając emocje w wokalu korzystają ze wskazówek zakodowanych w parametrach akustycznych oraz, że ludzie modyfikują mowę- aby wyrazić swoje emocje w sposób spójny i specyficzny [8].

### **3.2.2 Rozpoznanie emocji**

Duża liczba badań wykazała, że słuchacze potrafią dokładnie rozpoznać stan afektywny mówcy [6]. W najpowszechniej przyjmowanej procedurze badawczej mówcy (często profesjonalni aktorzy) proszeni są o wokalne przedstawienie różnych stanów emocjonalnych za pomocą standardowych wypowiedzi. Próbkę są nagrywane, a następnie prezentowane grupie słuchaczy, którą prosi się o identyfikację lub kategoryzację emocji. Scherer dokonał przeglądu trzydziestu tego rodzaju badań i wyliczył, że średnie rozpoznanie w takich badaniach wynosi około 60%, co znacznie przekracza losowy wskaźnik 12% - gdyby oceniający zgadywali [8]. Późniejsze badania potwierdziły jego ogólną ocenę trafności- ich rozpoznanie w zależności od badanych emocji i różnych grup słuchaczy wynosiło około 60%. Jest to wynik zbliżony do

rozpoznawalności emocji na podstawie wyrazu twarzy. Choć całkowita trafność rozpoznania emocji wokalnych jest nieco niższa niż stwierdzona w ekwiwalentnych badaniach nad dekodowaniem emocjonalnej ekspresji twarzy. Rozpoznanie jest zależne od różnych czynników badawczych: wyboru emocji, ilości kategorii emocji, jakości odegrania emocji [8]. Wpływ ma również dobór grupy słuchaczy.

Tab.1 Porównanie rozpoznania (w procentach) dla poszczególnych emocji w 2 doświadczalnych badaniach [10]

Badanie	Strach	Wstręt	Radość	Smutek	Złość
Van Bezooijen (1984)	58	49	72	67	74
Scherer et al. (1991)	52	28	59	72	68

Cechy przekazu głosowego sprawiają, że przypuszczalnie jedne emocje sygnalizuje on lepiej, inne zaś gorzej [8]. Ta hipoteza zgodna jest z faktem, że różne emocje łatwiej jest wyrażać i spostrzegać kiedy przekazywane są za pomocą głosu. Smutek, złość i strach należą do emocji, których ekspresja wokalna jest najwyraźniejsza [8]. Możliwość ostrzeżenia grupy o niebezpieczeństwie poprzez komunikowanie stanu afektywnego z odległości jest oczywistą korzyścią adaptacyjną. Z kolei wstręt- jak stwierdzono- trudno komunikować za pomocą głosu prawdopodobnie dlatego, że z punktu widzenia ewolucji ta emocja powinna być przekazywana z bliskiej odległości, a wtedy można ją bardzo dobrze wyrazić wizualnie- emocja ta ma bardzo wysokie rozpoznanie na podstawie wyrazu twarzy [8]. Wyniki dla pozytywnych emocji są zróżnicowane, Są wysokie kiedy słuchacz ma daną tylko jedną pozytywną opcję do wyboru jak np. szczęście. Jednak kiedy ma ich do wyboru więcej, przykładowo zadowolenie czy zainteresowanie, ilość prawidłowych odpowiedzi znacząco spada.

Wyniki rozpoznawalności emocji mówią o konieczności analizy każdej emocji z osobna. Warto również zwracać uwagę na kontekst, w jakim dochodzi do pomyłki, ponieważ błędy nie mają charakteru przypadkowego. Wiele błędów polega na myleniu emocji należących do tej samej rodziny, więc powinno się je rozważać nie jako błędy,



ale jako wskaźniki podobieństwa wokalnego pomiędzy dwoma kategoriami emocji [6]. Czas trwania próbki może odgrywać znaczącą rolę. Na przykład pojedyncze słowa złości są mniej naładowane emocjonalnie (lub nawet są neutralne), niż w sytuacjach, gdy są częścią dłuższej wypowiedzi [6].

### **3.2.3 Płeć, a emocje**

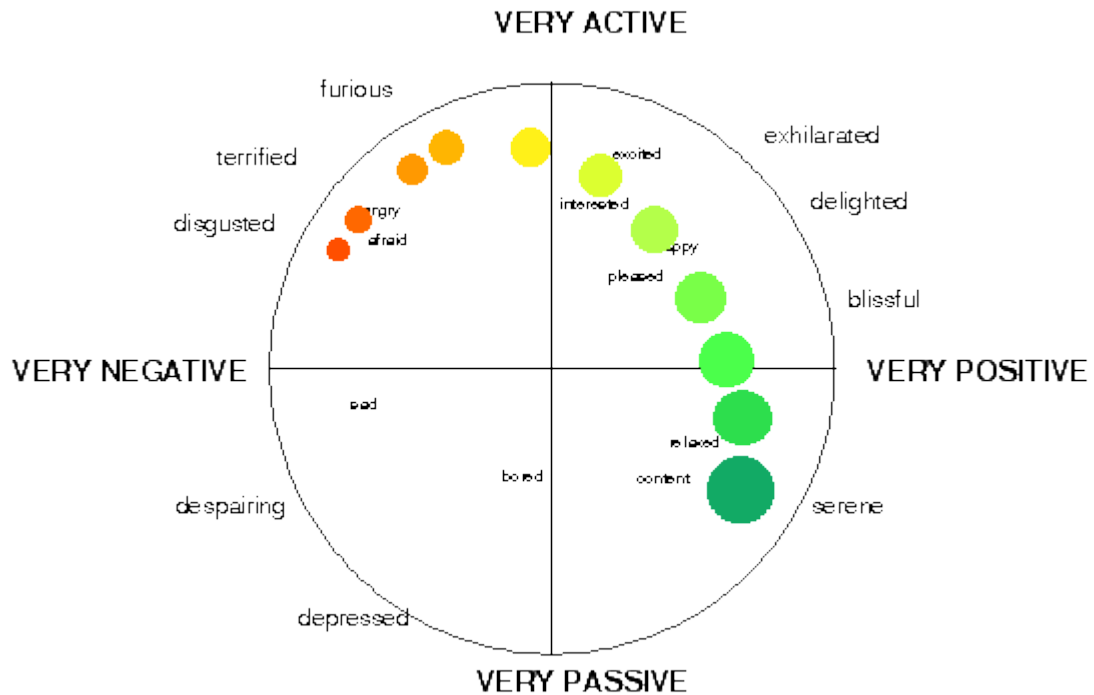
Powszechnym stereotypem jest, że kobiety są bardziej emocjonalne, częściej wyrażają emocję poprzez płacz, śmiech itp. [6]. Uważa się, że są również bardziej uzdolnione w odczytywaniu niewerbalnych informacji związanych z emocjami takich jak zmiany w wyrazie twarzy, głosu, w wykonywanych ruchach ciała czy poprzez uśmiech. Stereotypy dotyczą jednak poszczególnych emocji. Kobiety częściej niż mężczyźni okazują szczęście, zaskoczenie, smutek, strach, wstyd i winę. Natomiast mężczyźni częściej wyrażają dumę, pogardę i gniew [6]. Stereotypy te odnoszą się bardziej do ekspresji tych emocji niż ich doświadczania. Uprzedzenia kształtują rzeczywistość różnic w wyrażaniu emocji. Generują oczekiwania w wyrażaniu emocji oraz mają wpływ na zasady jak, kiedy i gdzie emocje mogą być wyrażane przez mężczyzn i kobiety w poszczególnych kulturach [6]. W badaniach nad rozpoznaniem emocji kobiety mają wyższe rozpoznanie emocji od mężczyzn, są to jednak niewielkie różnice rzędu 3-5%.

## **4. Testowanie baz nagrań mowy emocjonalnej**

### **4.1 Stosowane modele emocji**

Istnieją różne podejścia do konceptualizacji emocji i ich rozróżniania. Częstym rozwiązaniem jest przyjęcie ograniczonej ilości podstawowych emocji, z których wywodzi się reszta, jako ich mieszanka. Przykładem tego podejścia jest model emocji Roberta Plutchika, który zaproponował model ośmiu podstawowych emocji. Emocje te są wrodzone i bezpośrednio odnoszą się do zachowań adaptacyjnych, które mają na celu





Kolejnym podejściem jest kompromis pomiędzy dwoma wymienionymi koncepcjami. Koncepcja, która zakłada, że język i wiedza kształtują sposób kategoryzowania informacji przez ludzi [11]. Ten model emocji ma cechy zarówno indywidualnych emocji jak i hierarchicznej struktury między nimi.

## 4.2 Metody indukowania emocji

Istnieją różne metody indukcji emocji stosowane w badaniach kodowania afektu wokalnego. Należą do nich emocjonalne opowieści, obrazy, muzyka, filmy i wyobrażenia, gry komputerowe oraz zadania [8]. Gry komputerowe można zaprogramować do manipulowania sytuacjami wywołującymi emocje, przy jednoczesnej kontroli czynników ubocznych. Problemem są tu jednak ograniczenia etyczne.

Zrealizowanie nagrań naturalnych emocji nieindukowanych jest bardzo trudne z powodu zazwyczaj złych warunków akustycznych. Dobrą okolicznością do badania mogą być wydarzenia sportowe, podczas których zawodnicy i kibice doznają emocji o dużym nasileniu oraz programy telewizyjne na żywo. Często stosowanym rozwiązaniem pozwalającym na kontrolowanie wyniku nagrania jest rejestracja emocji odegranych przez aktorów. Wadą tej metody może być sztuczność i przerysowanie

emocji, jednak- ponieważ funkcją mowy jest również wpływanie na innych ludzi część badaczy uważa, że odgrywane emocje są bardzo zbliżone do tych występujących w normalnych sytuacjach [8].

### **4.3 Przegląd rozpoznania emocji w wybranych korpusów mowy**

W tabeli 2 autorka pracy dokonała zestawienia informacji na temat wybranych korpusów mowy emocjonalnej podając ich przeciętne rozpoznanie emocji, dostępne informacje na temat grupy słuchaczy i realizacji testu. Większość z nich zawiera nagrania tylko podstawowych emocji odegranych przez profesjonalnych aktorów. Przeprowadzanie testów różni się ilością odsłuchiwanym przykładów i możliwych odpowiedzi. Przykładowo we włoskiej bazie nagrań mowy słuchacze wybierali jedną z dwóch emocji po odsłuchaniu dwóch zdań wypowiedzianych przez tego samego mówcę. Natomiast w międzynarodowym badaniu przeprowadzonym przez badaczy z Göteborg University słuchacze opisywali rozpoznany stan emocjonalny dowolnym słowem, które następnie było tłumaczone na język szwedzki. W niektórych testach słuchacze mieli możliwość odsłuchania nagrania tylko jeden raz, w innych nieograniczoną ilość razy.

<b>BAZA NAGRAŃ</b>	<b>BADANE EMOCJE</b>	<b>NAGRANIA</b>	<b>JAK PRZPROWADZONO BADANIA</b>	<b>GRUPA SŁUCHACZY</b>	<b>SKUTECZNOŚĆ OCENY</b>
Berlin Database of Emotional Speech	Złość Znużenie Wstręt Strach Szczęście Smutek	10 wyselekcjonowanych (naturalność, rozpoznanie emocji) przez sędziów aktorów (5 kobiet, 5 mężczyzn)	Zaprezentowano wypowiedzi w losowej kolejności przed komputerem. Słuchacze mieli możliwość odsłuchania przykładu tylko raz przed podjęciem decyzji.	20 osób	Złość 96,9% Neutralne 88,2% Strach 87,3 % Znużenie 86,2 % Szczęście 83,7 % Smutek 80,7 % Wstręt 79,6%
EMOVO Corpus-Italian Emotional Speech Database	Strach Złość Radość Zaskoczenie Wstręt Smutek Stan neutralny	6 profesjonalnych aktorów (3 kobiety, 3 mężczyzn) 84 nagrania Zdanie w nagraniach są neutralne emocjonalnie lub absurdalne w kontekście sensu	Zaprezentowano 2 zdania tego samego aktora następnie słuchacz wybierał jedną z pośród 2 emocji.	24 osoby	Neutralny 93% Wstręt 67% Radość 65% Strach 74 % Złość 92% Zaskoczenie 81% Smutek 92%
Danish emotional speech database	Stan neutralny Radość Zaskoczenie Smutek Złość	4 mówców aktorów radiowych (2 kobiety, 2 mężczyzn)	Przeprowadzono 4 osobne test dla nagrań każdego z aktorów. Słuchacze nie otrzymywali informacji czy ich odpowiedź była poprawna, mogli słuchać nagrań kilka razy, ale nie mogli wracać do poprzednich przykładów i porównywać. Po danym teście określali jego trudność (75% określiło	20 osób (10 kobiet, 10 mężczyzn)  Wiek słuchaczy: < 20 lat 2 os. 21-30 5 os. 31-40 5 os. 41-50 4 os. 51-60 4 os.	Neutralny 60,8% Zaskoczenie 59,1% Radość 56,4% Smutek 85,2% Złość 75,1%  Kobiety 69% Mężczyźni 66%

			jako trudne)		
The Magdeburger Prosodie-Korpus	Stan neutralny Radość Smutek Złość Strach Wstręt	2 mówców profesjonalnych (mężczyzna i kobieta) -200 pseudo-słów wybranych jako reprezentatywne	Zaprezentowano 180 pseudo-słów w losowej kolejności	73 słuchaczy w wieku 24-56 lat	Stan neutralny 83,65% Radość 82,05% Smutek 57,67% Złość 93,49% Strach 78,06% Wstręt 74,4%
T. S. Polzin and A. H. Waibel at the Carnegie Mellon University database	Radość Smutek Złość Strach Stan neutralny	5 studentów aktorstwa 50 zdań w każdej z emocji	Słuchał nagrań jednego aktora w losowej kolejności i wybierał jedną z 4 opcji (radość, smutek, złość, strach)	Nie podano	Ok. 70%
Asa Abelin, Göteborg University	Radość Zaskoczenie Smutek Złość Strach Nieśmiałość Dominacja Wstręt	1 mówca (mężczyzna) wypowiedział 1 zdanie	Słuchacze opisywali stan emocjonalny mówcy dowolnym słowem, które następnie było tłumaczone na język szwedzki	35 szwedów 12 anglików 23 finów 23 hiszpan	Średnio: Radość 76% Zaskoczenie 72% Smutek 83% Złość 85% Strach 64% Nieśmiałość 37% Dominacja 76% Wstręt 22%

Tab. 2 Rozpoznanie emocji w wybranych korpusach mowy emocjonalnej [14,15,16,17,18,19]

## **4. Test odsłuchowy**

### **4.1 Korpus mowy emocjonalnej**

Korpus mowy emocjonalnej AGH zawiera nagrania wyrażające pięć spośród emocji podstawowych (radość, smutek, złość, strach, zdziwienie), ironię oraz stan neutralny jako sygnał referencyjny. Ironia nie jest emocją w rozumieniu teorii emocji podstawowych. W nagraniach wzięło udział 12 mówców (6 kobiet, 6 mężczyzn) w wieku 20-30 lat. Część z nich to profesjonalni aktorzy lub amatorzy, a część - studenci wolontariusze [2].

Po uprzednim zapoznaniu się z tekstami, mówcy zostali poproszeni o przeczytanie ich kolejno w sposób wyrażający poszczególne emocje. Dla każdego mówcy zarejestrowano nagrania o tej samej treści. Treść nagrań stanowią pojedyncze słowa, zdania dialogowe oraz jeden tekst ciągły. Treść dobrano tak, żeby była możliwie neutralna i nie indukowała konkretnej emocji [2]. Treść zaprojektowano pod kątem użycia w interfejsach człowiek-komputer oraz różnorodności typów wypowiedzi. Łącznie dla każdego mówcy zarejestrowano 282 słowa, każde w 6 stanach emocjonalnych. Łączny czas wszystkich nagrań wynosi ponad 3.5 godziny. Baza nagrań może być zastosowana w identyfikowaniu wzorców reakcji emocjonalnych w sygnale mowy oraz w zweryfikowaniu możliwej jakości automatycznego wnioskowania na temat nastawienia emocjonalnego mówcy (cechy wysokopoziomowe mowy) przez algorytmy przetwarzania sygnałów i automatycznej klasyfikacji.

### **4.2 Grupa ankietowanych**

W teście odsłuchowym wzięły udział 33 osoby (18 kobiet i 15 mężczyzn) w wieku 19- 55 lat (średnia 24,1). 23 osoby (69,7 % ankietowanych) określiły swoje wykształcenie jako techniczne, 3 (9% ankietowanych) - matematyczno-przyrodnicze, 3- artystyczne, 2 (6% ankietowanych)- humanistyczne, 1 osoba- społeczne i 1 osoba jako 'inne' niż wymienione kategorie.

### 4.3 Metoda badania

W celu przeprowadzenia testu powstała strona internetowa zawierająca instrukcje, wylosowane nagrania oraz następujące pytania do słuchacza:

-W skali od 1-5 jak dobrze Twoim zdaniem rozpoznajesz emocje u innych osób? (5-bardzo dobrze, 4-dobrze, 3-przeciętnie, 2-słabo, 1-źle)

-Wyształcenie (Studia: techniczne, ekonomiczne, matematyczne/przyrodnicze, społeczne, humanistyczne, językowe, artystyczne, inne)

-W skali od 1-5 oceń swój nastrój (5-bardzo dobry, 4-dobry, 3-neutralny, 2-kiepski, 1-zły)

-Określ poziom trudności testu: bardzo trudny/trudny/średni/łatwy

W celu zbadania percepcji nagrań został wykonany test odsłuchowy składający się z dwóch części. W obu z nich słuchacz miał za zadanie przypisać nagraniu rozpoznany przez siebie stan emocjonalny. Do wyboru miał jedną z 8 odpowiedzi (radość, smutek, złość, strach, zdziwienie, stan neutralny oraz odpowiedź 'nie wiem'). Pierwsza część testu zawierała 91 nagrań: 8 zdań, 2 polecenia, 2 cyfry oraz tekst ciągły w każdym z 7 stanów emocjonalnych. Nagrania wykorzystane w teście zostały wylosowane i nie zostały poddane wcześniejszej ocenie i selekcji przez autorkę testu. Rozkład nagrań użytych w teście w zależności od mówcy znajduje się w tabeli 1. Instrukcja do pierwszej części testu wyglądała następująco: Poniżej znajdują się nagrania mowy. Odsłuchaj je i zaznacz jakie nastawienie emocjonalne w nich rozpoznajesz. Możesz odsłuchać daną próbkę wielokrotnie. Po wybraniu odpowiedzi nie wracaj do poprzednich przykładów.

Tab.1 Rozkład nagrań użytych w teście odsłuchowym w zależności od mówców. Kolorem żółtym oznaczeni są aktorzy, kolorem niebieskim oznaczeni są amatorzy.

Emocja	AKA	AKL	HKR	MMA	BTO	JMI	MCH	MGR	MPO	PJU	PKE
Ironia	2	1	1	0	2	1	2	0	3	0	1
Radość	2	0	2	1	0	1	3	0	2	2	0
Smutek	0	0	3	3	0	1	1	1	1	3	0



Stan neutralny	3	2	0	1	2	1	0	2	2	0	0
Strach	2	1	1	2	0	3	1	2	0	0	1
Zdziwienie	2	2	0	0	1	1	1	2	0	1	3
Złość	0	3	1	1	0	1	1	0	2	4	0
Suma	11	9	8	8	5	9	9	7	10	10	5

Druga część testu składała się z 18 nagrań wylosowanych z pośród nagrań z pierwszej części, po 3 nagrania w 6 stanach emocjonalnych, wykluczony został stan neutralny. Przykłady składały się z nagrania wraz z sygnałem referencyjnym (tym samym zdaniem o neutralnym zabarwieniu emocjonalnym, wypowiedzianym przez tego samego mówcę) przed i po zdaniu, którego stan emocjonalny miał za zadanie rozpoznać słuchacz. Instrukcja do drugiej części testu: Poniżej znajdują się nagrania wraz z nagraniem referencyjnym. Usłyszysz to samo zdanie 3 razy. Pierwsza i ostatnia wypowiedź ma (intencyjnie) neutralne zabarwienie emocjonalne. Zaznacz jakie nastawienie emocjonalne rozpoznajesz w drugiej wypowiedzi.

Kolejność nagrań była losowa dla każdego otworzenia testu. Ankietowany miał możliwość odsłuchania każdego nagrania dowolną ilość razy. Test został dobrany, tak aby ocenić możliwie dużo nagrań, jednocześnie unikając przemęczenia słuchacza co skutkowałoby mniejszą koncentracją i miałyby negatywny wpływ na wyniki testu. Zrobienie testu zajmowało około 20 minut.

Rys. 1 Fragment internetowego arkusza odpowiedzi [20]

Pytanie 1. Jakie nastawienie emocjonalne rozpoznajesz?

00:00 / 00:00

- ironia
- stan neutralny
- radość
- smutek
- strach
- zdziwienie
- złość
- nie wiem

## 6. Analiza wyników

### 6.1 Średnie rozpoznanie

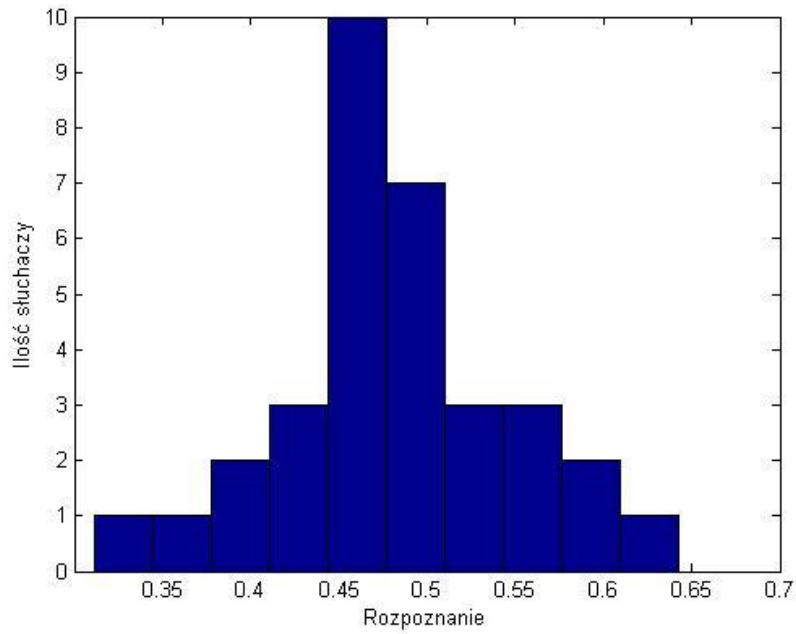
Tab. 1 Rozpoznanie w zależności od płci ankietowanych i mówców

Rozpoznanie [%]						
	cz. I Nagrania bez referencji		cz. II Nagrania z referencją		Wszystkie przykłady	
Całkowite	46,45		57,24		48,23	
Emocji podstawowych	48,56		59,39		50,35	
Płeć ankietowanego	Kobiety	Mężczyźni	Kobiety	Mężczyźni	Kobiety	Mężczyźni
	47,80	44,84	61,42	52,22	50,05	46,05
Płeć mówcy	Kobiety	Mężczyźni	Kobiety	Mężczyźni	Kobiety	Mężczyźni
	46,71	46,24	54,54	58,95	48	48,34

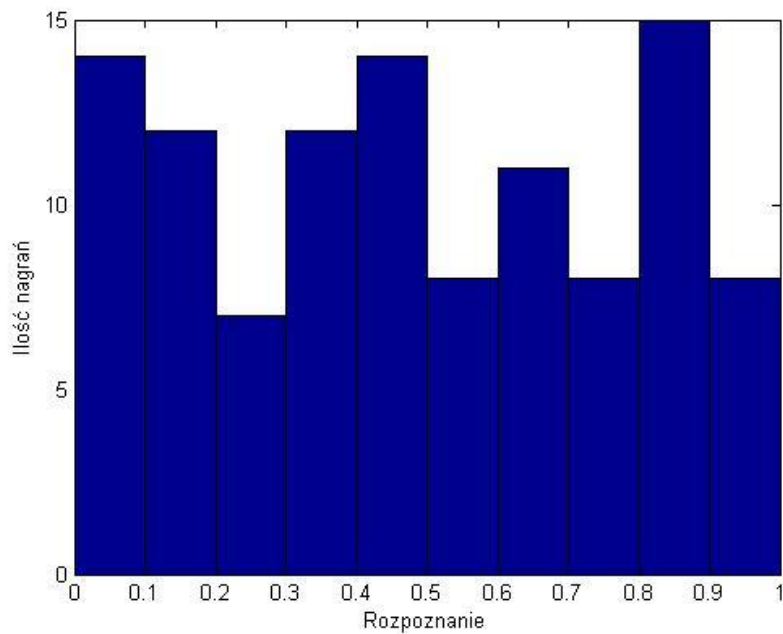
Średnie rozpoznanie wszystkich nagrań w pierwszej części testu wynosi 46,5%-znacznie przekracza wynik losowy (12,4% nie uwzględniając korzystania z odpowiedzi 'nie wiem'). Średnie rozpoznanie emocji podstawowych (wyłączając ironię, która nie jest emocją podstawową) wynosi 48,6% (+/- 9,63%) . Różnice w rozpoznaniu emocji w nagraniach przez kobiety (47,8%), a przez mężczyzn (44,84%) są niewielkie i nieistotne statystycznie ( $p=0,2179$ ). Również w przypadku różnicy płci mówców wyniki różnią się bardzo nieznacznie (46,71% i 46,24%,  $p=0,07741$ ).

Średnie rozpoznanie emocji w nagraniach z referencją wynosi 57,24%. W tej części testu różnica między ilością poprawnych odpowiedzi kobiet i mężczyzn jest większa (61,42% i 52,22%), ale również nieistotna statystycznie ( $p=0,3784$ ).

Rys.2 Rozkład rozpoznania wszystkich nagrań dla wszystkich słuchaczy



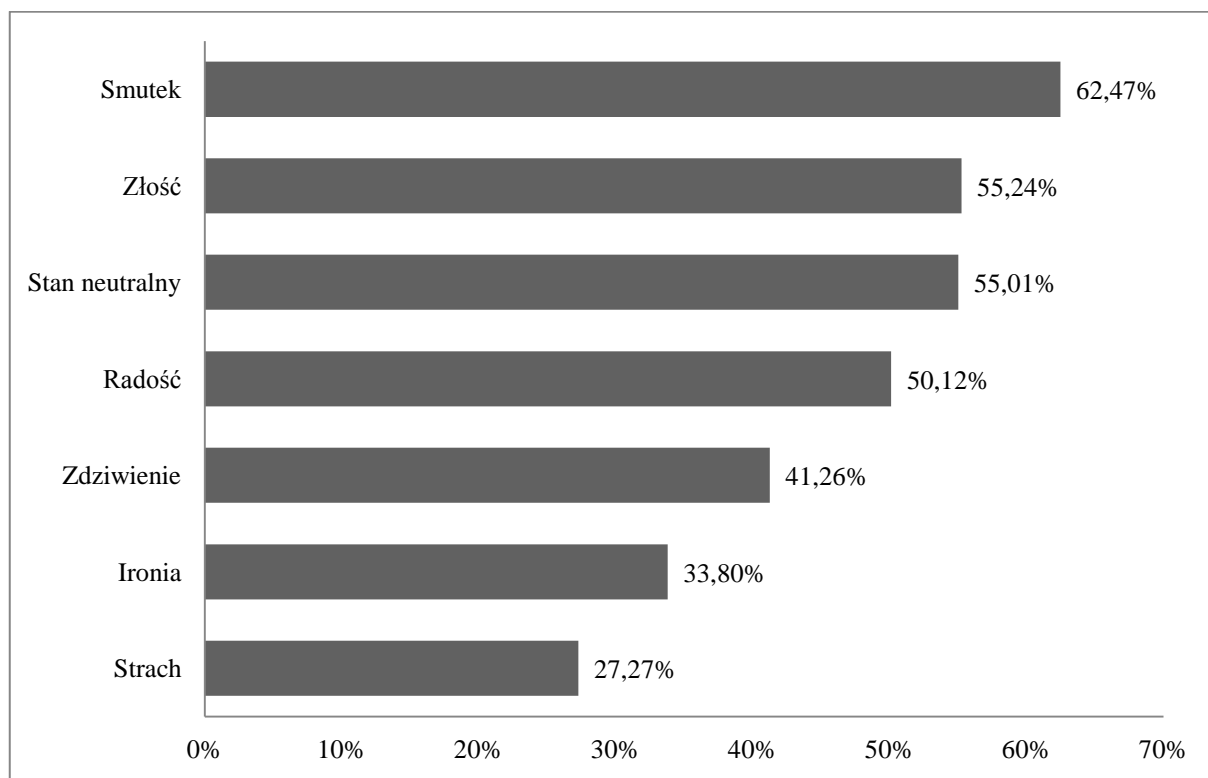
Rys.3 Rozkład rozpoznania wszystkich nagrań



Rozkład rozpoznania wszystkich nagrań w zależności od ilości słuchaczy przedstawiony jest na rysunku 1 i jest rozkładem normalnym zgodnie z testem zgodności Chi-kwadrat. Rozkład rozpoznania nagrań w zależności od ich ilości nie jest rozkładem normalnym.

## 6.2 Rozpoznanie poszczególnych emocji

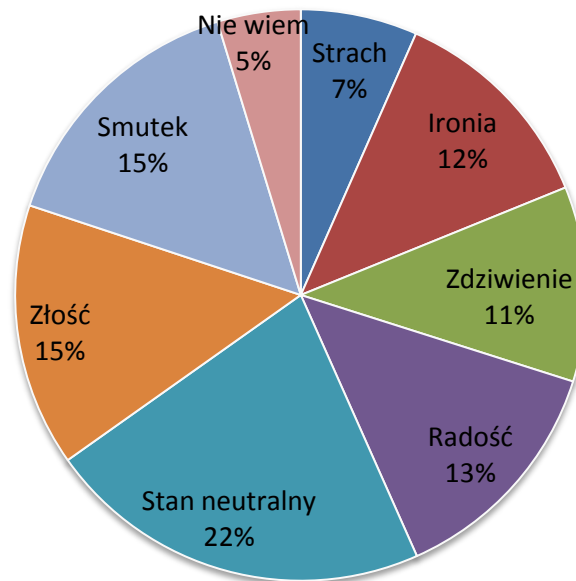
Rys. 4 Średnie rozpoznanie nagrań w zależności od emocji



Tab. 2 Odchylenie standardowe rozpoznania emocji

Emocja	Odchylenie standardowe rozpoznania [%]
Strach	17,98
Ironia	21,34
Zdziwienie	23,65
Radość	23,72
Stan neutralny	25,24
Złość	17,35
Smutek	19,26

Rys. 5 Rozkład odpowiedzi



Tab. 3 Macierz błędu (St= Strach, Ir= Ironia, Zd= Zdziwienie, Ra= Radość, Sn= Stan neutralny, Zł= Złość, Sm= Smutek, Nw= Nie wiem)

	St	Ir	Zd	Ra	Sn	Zł	Sm	Nw
Strach	0,27	0,06	0,16	0,02	0,13	0,14	0,15	0,07
Ironia	0,02	0,34	0,07	0,13	0,27	0,08	0,05	0,05
Zdziwienie	0,05	0,15	0,41	0,07	0,12	0,06	0,08	0,05
Radość	0,01	0,13	0,09	0,50	0,19	0,03	0,00	0,05
Stan neutralny	0,05	0,06	0,03	0,00	0,55	0,12	0,14	0,05
Złość	0,03	0,07	0,01	0,21	0,07	0,55	0,02	0,04
Smutek	0,03	0,05	0,00	0,01	0,20	0,06	0,62	0,02

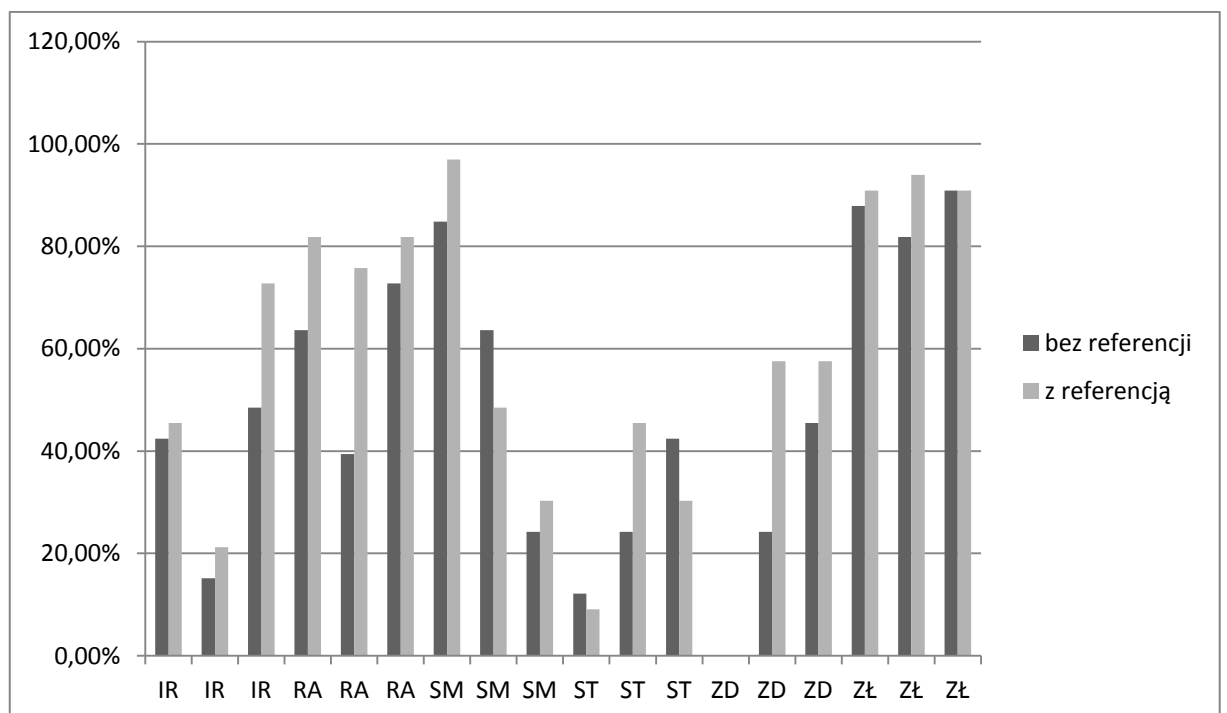
Najwyższe rozpoznanie mają smutek (62%) i złość (55%), co zgodne jest z faktem, że te emocje te mają najlepszą wokalną ekspresję (ref. Psychologia emocji). Zaskakujący okazał się być duży procent pomyłek złości z radością (21%). Jest to

spowodowane stylem mówcy (który nie jest profesjonalnym aktorem), którego próby przedstawienia złości aż 66% słuchaczy oceniło jako radość. Zgodnie z literaturą emocją o trzecim najwyższym rozpoznaniu jest strach, jednak bywa też mylony z emocjami pokrewnymi jeśli takie są do wyboru [10]. W teście miał on średnie rozpoznanie tylko 27% i często mylony był ze zdziwieniem (16%), smutkiem (15%) i złością (14%) . Jest też emocją przy której ankietowani najczęściej użyli odpowiedzi „nie wiem”. Strach i zdziwienie są emocjami podobnymi- znajdują się obok siebie na kole emocji Plutchika. Najczęściej wybieraną odpowiedzią był stan neutralny (22%). Można z tego wnioskować, że brak emocji był często wybierany w przypadku kiedy słuchacz nie wiedział co wybrać.

Ironia, która z definicji nie jest emocją i nie była uwzględniana w innych znanych autorce pracy korpusach mowy ma średnie rozpoznanie 34% i bardzo często była uznawana za stan neutralny (27%).

### 6.3 Nagrania z referencją

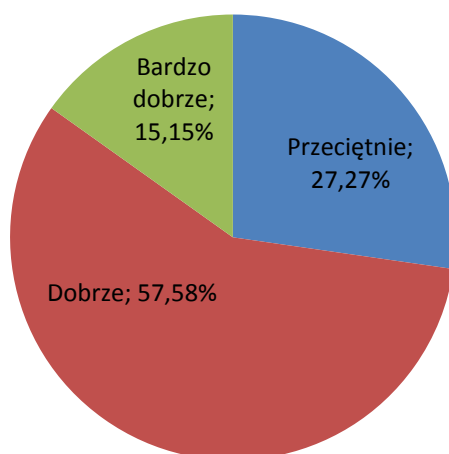
Rys. 6 Porównanie rozpoznania 18 nagrań ocenionych bez i z referencją (ST= Strach, IR= Ironia, ZD= Zdziwienie, RA= Radość, ZŁ= Złość, SM= Smutek)



18 nagrań, ocenionych przez słuchaczy w pierwszej części testu, zostało ocenione również w drugiej części testu, gdzie występowały razem z sygnałem referencyjnym. Średni wzrost rozpoznania po odsłuchaniu przykładu z referencją wynosi 9,26 punktów procentowych (poziom istotności  $p=0,011$ ). W podobnym teście porównującym rozpoznanie emocji w węgierskiej mowie średni wzrost w przypadku słuchania nagrania po referencji wyniósł 8,96% [21]. Największy wzrost rozpoznania odnotowano w przypadku radości (średnio o 21,21%) i zdziwienia (średnio o 15,15%). 3 z 18 nagrań miało niższe rozpoznanie po odsłuchaniu nagrania referencyjnego (strach i smutek). Nagrania, w którym mówca chciał wyrazić smutek przez 30% osób zostało uznane za neutralne emocjonalnie. W przypadku strachu aż 67% słuchaczy rozpoznało w jednym z nagrań zdziwienie, a 42% w drugim złość. Jeden przykład w całym teście ma zerowe rozpoznanie, zarówno w przypadku referencji jak i jej braku, większość słuchaczy (51,5%) rozpoznała w nim smutek zamiast zdziwienia.

## 6.4 Samoocena, a wynik

Rys. 7: Rozkład samooceny zdolności rozpoznawania emocji u innych osób przez uczestników badania



Tab.4 Porównanie samooceny ankietowanych ze średnim rozpoznaniem

Samooocena	Przeciętnie	Dobrze	Bardzo dobrze
Średnie rozpoznanie [%]	47,09	48,44	49

Ankietowani zostali poproszeni o ocenienie swoich zdolności rozpoznawania emocji u innych osób w skali 1-5 (1-źle, 2-słabo, 3-przeciętnie, 4-dobrze, 5-bardzo dobrze). Średnio im lepsza samooocena umiejętności w rozpoznawaniu emocji, tym lepszy wynik. Jednak różnice w średnim rozpoznaniu dla poszczególnych grup są niewielkie.

## 6.5 Nastrój ankietowanego, a wynik

Nastrój modeluje przetwarzanie informacji co oznacza, że pamiętamy więcej zdarzeń związanych z nastrojem, w którym aktualnie się znajdujemy. Zgodnie z tym faktem zweryfikowana została hipoteza, że nastrój ankietowanego ma wpływ na rozpoznawanie przez niego poszczególnych emocji. Ankietowani ocenili swój nastrój w skali 1-5 (1-zły, 2-kiepski, 3-neutralny, 4-dobry, 5- bardzo dobry). Porównane zostały podany nastrój z wynikiem rozpoznawania poszczególnych emocji. Nie udało się zaobserwować istotnych statystycznie różnic ( $p=0,7435$ ).

## 5.7 Wykształcenie ankietowanego, a wynik

Tab. 5 Wykształcenie ankietowanych, a średnie rozpoznanie

Wykształcenie	Średnie rozpoznanie [%]	Ilość osób
artystyczne	49,33	3
techniczne	42,83	23
społeczne	40,00	1
humanistyczne	40,00	2
inne	39,56	1
matematyczne/przyrodnicze	35,33	3



Najwyższe rozpoznanie (49,33%) ma grupa o wykształceniu artystycznym, najniższe grupa o wykształceniu matematycznym/przyrodniczym. Najlicniejsza grupa (69,7%) o wykształceniu technicznym osiągnęła drugi z kolei wynik (42,83%). Niestety z powodu niewielkiej liczności grup nie można mówić o zależnościach istotnych statystycznie ( $p=0,0835$ ).

## **6.6 Ocena testu przez ankietowanych**

54,55% ankietowanych określiło test jako trudny, 15,15% jako bardzo trudny, a 30,3% jako średni. Jego trudność polegała częściowo na krótkim czasie trwania większości przykładów- część z nich zawierała tylko jedno słowo. Zdaniem słuchaczy w niektórych przykładach w teście brakowało innych niż dostępne odpowiedzi, które ich zdaniem lepiej opisywały nastawienie emocjonalne mówcy.

## **7. Wnioski i dalszy rozwój pracy**

Średnie rozpoznanie emocji podstawowych 48,56% jest wynikiem niższym niż podawane w innych badaniach 60%. Przyczyną tej różnicy może być dobór bardzo krótkich nagrań do testu utrudniających wychwycenie w nich odegranej emocji. Wpływ ma również dostępność odpowiedzi 'nie wiem', która stanowi 5% wszystkich odpowiedzi w teście. Kolejnym czynnikiem może być też nieudane odegranie konkretnych emocji przez niektórych aktorów, jak w przypadku mówcy, którego przedstawienie złości 66% słuchaczy zinterpretowało jako radość.

Podobieństwem do wyników innych badań jest zgodne z cechami przekazu głosowego najwyższe rozpoznanie złości i smutku. Również częste wybieranie stanu neutralnego przez słuchaczy jest zgodne z obserwacjami innych testów. Ironia, która nie jest zawarta w innych znanych autorce pracy bazach nagrań, ma niski stopień rozpoznania (34%).

Nie zaobserwowano istotnych różnic w rozpoznaniu w zależności od płci słuchacza i mówcy. W przypadku nagrań słuchanych razem z referencją odnotowano

średni wzrost o 9,26 punktów procentowych. Były też jednak przypadki spadku rozpoznania w porównaniu z odsłuchem bez referencji.

W porównaniu średniego rozpoznania emocji z samooceną ankietowanych, powstała zależność wyższego wyniku dla grup o lepszej samoocenie. Różnice pomiędzy poszczególnymi grupami są jednak niewielkie. Zestawienie nastroju ankietowanych z ich rozpoznaniem poszczególnych emocji wskazuje na brak istotnego związku. Nierównomierność liczebności grup w podziale ze względu na wykształcenie osób ankietowanych, nie pozwala na wyciągnięcie jednoznacznych wniosków co do wpływu wykształcenia na rozpoznawanie emocji.

Ograniczeniem zastosowanej metody był brak możliwości opisanie emocji usłyszanych w nagraniach słownictwem innym niż dostępne w odpowiedziach.

Możliwe jest dalsze przeprowadzanie testu odsłuchowego i gromadzenie większej ilości danych o percepcji nagrań korpusu mowy emocjonalnej. Przeprowadzenie testu wśród osób z wykształceniem innym niż techniczne pozwoliłoby na zweryfikowanie hipotezy, że wykształcenie ma wpływ na rozpoznawanie emocji. Ocena nagrań przez ludzi może zostać porównana z oceną programu analizującego zmiany parametrów akustycznych.

## **Bibliografia**

- [1] Reeves B. and Nass C., *The media equation*, Cambridge University press, 1996.
- [2] M. Igras, B. Ziółko, *Baza danych nagrań mowy emocjonalnej*, Studia Informatica, vol. 34, 2013, s.67-77.
- [3] Zimbardo P.: *Psychologia i Życie*, Wydawnictwo Naukowe PWN, Warszawa 1999
- [4] Artykuł: Emocje, uczucia, pragnienia – jak je rozpoznawać i nad nimi panować  
źródło: [http://rodzina.ipjp2.pl/index.php?option=com\\_content&view=article&id=774:emocje-uczucia-pragnienia-jak-je-rozpoznawac-i-nad-nimi-panowac&catid=176&Itemid=329](http://rodzina.ipjp2.pl/index.php?option=com_content&view=article&id=774:emocje-uczucia-pragnienia-jak-je-rozpoznawac-i-nad-nimi-panowac&catid=176&Itemid=329)
- [5] Ekman P, Davidson R.J: *Natura emocji- podstawowe zagadnienia*, Gdańskie wydawnictwo psychologiczne, Gdańsk 2002

- [6] Lewis M., Jeannette M. Haviland-Jones, Lisa Feldman Barrett: *Handbook of Emotions* Third edition
- [7] Quam, C. and Swingley, D. (2012), *Development in Children's Interpretation of Pitch Cues to Emotions*. *Child Development*, 83: 236–250. doi: 10.1111/j.1467-8624.2011.01700.x
- [8] Lewis M., Haviland-Jones J. M.: *Psychologia emocji*, Gdańskie wydawnictwo psychologiczne, Gdańsk 2005
- [9] Brehm S. S., Kassin S. M., *Social Psychology*, Houghton Mifflin Company, Geneva, Illinois, 1996.
- [10] Banse R., Scherer K. R.: Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology*, 70, 614–636.
- [11] Sobol-Shikler T.: *Analysis of affective expression in speech*, University of Cambridge, 2009
- [12] Obraz: Koło emocji Plutchika, Źródło:  
<http://www.historiasztuki.com.pl/ilustracje/film/Plutchik-wheel.jpg>
- [13] *A Database of German Emotional Speech*, źródło: [https://www.kw.tu-berlin.de/fileadmin/a01311100/A\\_Database\\_of\\_German\\_Emotional\\_Speech\\_-\\_Burkhardt\\_01.pdf](https://www.kw.tu-berlin.de/fileadmin/a01311100/A_Database_of_German_Emotional_Speech_-_Burkhardt_01.pdf), dostęp (12.01.2016)
- [13] Obraz: 'Feeltrace' źródło: Źródło: <http://emotion-research.net/toolbox/feeltrace.gif>
- [15] *EMOVO Corpus: an Italian Emotional Speech Database*, źródło: [http://www.lrec-conf.org/proceedings/lrec2014/pdf/591\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/591_Paper.pdf), dostęp (12.01.2016)
- [16] *Documentation of the Danish Emotional Speech Database*, źródło: <http://kom.aau.dk/~tb/speech/Emotions/des.pdf>, dostęp (12.01.2016)
- [17] The "Magdeburger Prosodie-Korpus", źródło: [http://www.isca-speech.org/archive\\_open/sp2002/sp02\\_699.pdf](http://www.isca-speech.org/archive_open/sp2002/sp02_699.pdf), dostęp (12.01.2016)
- [18] Polzin T. S., Waibel A. H.: *Detecting Emotions in Speech*, źródło: [https://www.ri.cmu.edu/pub\\_files/pub1/polzin\\_thomas\\_1998\\_1/polzin\\_thomas\\_1998\\_1.pdf](https://www.ri.cmu.edu/pub_files/pub1/polzin_thomas_1998_1/polzin_thomas_1998_1.pdf), dostęp (12.01.2016)

[19] A. Abelin, J. Allwood, *Cross Linguistic Interpretation of Emotional Prosody*, Proceedings ISCA Workshop (ITRW) on Speech and Emotion: a Conceptual Framework for Research, Belfast, 2000

[20] Strona internetowa do testu odsłuchowego, Źródło: <https://lit-brook-1783.herokuapp.com/> (dostęp: 8.01.2016 r.)

[21] Tóth S. L., Sztahó D., Visci K.: *Speech Emotion Perception by Human and Machine*, in Proceedings of COST Action, Patras, Greece, 2012, s. 213–224.