ZASTOSOWANIE MULTIMODALNEJ KLASYFIKACJI W ROZPOZNAWANIU STANÓW EMOCJONALNYCH NA PODSTAWIE MOWY SPONTANICZNEJ

Dorota Kamińska, Adam Pelikant
Politechnika Łódzka, Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki

Streszczenie. *Artykuł prezentuje zagadnienie związane z rozpoznawaniem stanów emocjonalnych na podstawie analizy sygnału mowy. Na potrzeby badań stworzona została polska baza mowy spontanicznej, zawierająca wypowiedzi kilkudziesięciu osób, w różnym wieku i różnej płci. Na podstawie analizy sygnału mowy stworzono przestrzeń cech. Klasyfikację stanowi multimodalny mechanizm rozpoznawania, oparty na algorytmie kNN. Średnia poprawność rozpoznawania wynosi 83%.*

Słowa kluczowe: rozpoznawanie emocji, sygnał mowy, algorytm kNN

SPONTANEOUS EMOTION RECOGNITION FROM SPEECH SIGNAL USING MULTIMODAL CLASSIFICATION

Abstract. *The article presents the issue of emotion recognition from a speech signal. For this study, a Polish spontaneous database, containing speech from people of different age and gender, was created. Features were determined from the speech signal. The process of recognition was based on multimodal classification, related to kNN algorithm. The average of accuracy performance was up to 83%*

Keywords: emotion recognition, speech signal, kNN algorithm

Wstęp

W komunikacji międzyludzkiej sygnał mowy, poza przekazem semantycznym, niesie ze sobą informacje dotyczące stanu emocjonalnego mówcy. W celu polepszenia komunikacji człowiek -komputer/człowiek-robot (HCI/HRI) powstają systemy rozpoznawania emocji, dzięki czemu stałaby się ona bardziej naturalna i wiarygodna.

Dotychczasowe badania opierają się głównie na próbkach mowy odegranej, w której zdefiniowane jest konkretne zabarwienie emocjonalne głosu. Uzyskiwane są w ten sposób bardzo dobre wyniki rozpoznawania. Jednak spontaniczna mowa może stanowić zbiór różnych emocji bądź ich mieszaninę [9]. Zdarza się, że etykietowanie mowy przez ludzi stanowi problem, a emocje są przez nich różnie identyfikowane [5]. Dlatego tworząc system, który miałby działać w warunkach naturalnych, należy wziąć pod uwagę złożoność emocji zawartych w mowie spontanicznej.

Przedmiotem niniejszych badań jest opracowanie systemu realizującego identyfikację stanu emocjonalnego mówcy. Podczas eksperymentów dokonano porównania cech reprezentujących zarówno mowę spontaniczną jak i odegraną przez profesjonalistów oraz ich wpływ na identyfikację emocji naturalnych. Biorąc pod uwagę złożoność emocji w mowie spontanicznej oraz ich zmienność w trakcie wypowiedzi, zaproponowano multimodalny proces klasyfikacji.

Pozostała część niniejszej pracy została podzielona na cztery rozdziały. Pierwszy rozdział prezentuje krótką analizę omawianego zagadnienia. Bazy mowy wykorzystane w niniejszych badaniach opisane są w rozdziale 2. Następny rozdział prezentuje przegląd metod i algorytmów badawczych wykorzystanych w badaniach. W rozdziale 4. opisane zostało autorskie podejście do klasyfikacji emocji w mowie naturalnej. Rozdział 5. ukazuje osiągnięte rezultaty i wyniki klasyfikacji, zaś rozdział ostatni stanowi krótkie podsumowanie wykonanych badań oraz przyszłe kierunki rozwoju.

1. Analiza zagadnienia

Prace nad systemem rozpoznającym emocje rozpoczynają się od zgromadzenia odpowiedniej bazy plików dźwiękowych. Większość naukowców korzysta z gotowej, ogólnodostępnej bazy próbek nacechowanych emocjami tzw. Berlin Database [21]. Są to nagrania dziesięciu profesjonalnych aktorów (kobiet i mężczyzn), wypowiadających dziesięć zdań w siedmiu różnych stanach emocjonalnych (złość, strach, zadowolenie, smutek, znudzenie, mowa neutralna) [10]. Inni nagrywają dźwięki z audycji radiowych, filmów czy programów telewizyjnych [6].

Kolejną fazą automatycznego rozpoznawania jest dobór odpowiednich cech. Zasadniczo zbiór deskryptorów powszechnie stosowanych do analizy mowy spełnia się również przy rozpoznawaniu emocji. Większość naukowców opiera swoje badania o częstotliwość podstawową, formanty, energię sygnału i prozodia [18]. Czasami sięgają jednak do bardziej złożonych cech, jak współczynniki MFCC [13] czy LPC [11], które są standardem w rozpoznawaniu mowy [8].

Na podstawie zgromadzonych cech tworzone są wektory cech używane w następnym kroku – klasyfikacji. Metody te, to narzędzia standardowe, ale ich dobranie jest również ważnym elementem. Spośród prostych statystycznych metod najczęściej używany jest algorytm kNN, który daje bardzo dobre wyniki [15]. Z bardziej zaawansowanych metod największą popularnością cieszą się ukryte modele Markowa oraz coraz częściej wykorzystywane sztuczne sieci neuronowe [14]. Najczęściej jednak dokonywane jest porównanie skuteczności kilku metod [4].

2. Materiał badawczy

Na potrzeby niniejszych badań stworzona została polska baza mowy spontanicznej nacechowanej emocjami. Głównym źródłem nagrań są programy telewizyjne i audycje radiowe. Zebrano ponad 500 nagrań o czasie trwania kilka-kilkanaście sekund pochodzących od kilkudziesięciu osób w różnym wieku i różnej płci. Nagrania zapisano w formacie PCM WAVE 44,1 kHz.

Na podstawie skompletowanych nagrań, ośmiu decydentów dokonało ich klasyfikacji w sześć podstawowych grup (klas) emocji: radość (H), smutek (S), złość (A), strach (F), znudzenie (B) oraz mowa neutralna (N). W ten sposób dokonano selekcji nagrań niejednoznacznie określanych. Ostatecznie wybrano tylko te nagrania, które oceniono jednoznacznie. Rozkład próbek na dane grupy przedstawiono w tabeli numer 1.

Tabela 1. Liczba nagrań przypisanych do poszczególnych emocji

Nazwa emocji	Ilość nagrań w grupie
Radość	95
Smutek	78
Złość	65
Strach	40
Znudzenie	40
Mowa naturalna	34

Dodatkowo dokonano porównania jakości klasyfikacji emocji spontanicznych oraz odegranych. W tym celu zastosowano również polską bazę emocji udostępnianą przez Zakład Elektroniki Medycznej Politechniki Łódzkiej. Stanowi ona zbiór 240 nagrań

Artykuł recenzowany *otrzymano: 17.07.2012r., przyjęto do druku: 28.07.2012r.*

pięciu różnych zdań wypowiadanych przez ośmiu aktorów (4 kobiety, 4 mężczyzn). Zbiór podzielono na te same grupy emocji, co baza opisana powyżej [22].

3. Metody i algorytmy badawcze

Z wyselekcjonowanych próbek mowy wyznaczono parametry sygnału szeroko stosowane w rozpoznawaniu mowy ludzkiej. Wśród nich znalazły się niżej opisane parametry.

3.1. Częstotliwość podstawowa *F0*

Podczas mówienia struny głosowe rozchylają się energicznie na całej długości (głoski dźwięczne), lub są całkowicie otwarte (głoski bezdźwięczne). W przypadku głosek dźwięcznych pobudzenie traktu głosowego jest sygnałem okresowym, składającym się z szeregu delt Diraca. Odległość między impulsami stanowi okres, zaś jego odwrotność częstotliwość tonu podstawowego [19].

Częstotliwość tonu podstawowego jest cechą osobniczą, wynika z rozmiaru krtani oraz napięcia i rozmiaru strun głosowych, jest zależna od płci oraz wieku. Przykładowo dla mężczyzn zawiera się w przedziale od 80 do 480 Hz, natomiast dla kobiet w przedziale od 160 do 960 Hz. Jest zatem odpowiedzialna za skalę głosu [20]. Podczas rozmowy zakres jej zmian związany jest głównie z intonacją, która odgrywa ogromną rolę w ekspresji emocji [7].

W opisywanym systemie detekcja częstotliwości podstawowej została wykonana przy pomocy algorytmu analizującego funkcję autokorelacji. Na podstawie przebiegu *F0* wyznaczono 23 cechy statystyczne (takie jak średnia, mediana, odchylenie standardowe itp.), które stanowią jeden z wektorów cech.

3.2. Formanty

Trakt głosowy składa się z szeregu struktur posiadających zdolność do drgań własnych. Przechodzący przez niego ton krtaniowy podlega modyfikacjom wskutek drgań własnych gardła, jamy nosowej czy też jamy ustnej. Dzięki temu określone składowe pierwotnego tonu krtaniowego ulegają wzmocnieniu, inne natomiast osłabieniu. Maksima wzmocnionych częstotliwości nazywane są formantami [12]. Ich wartości zależą od cech osobniczych (długość przewodu głosowego), ale również od sposobu artykulacji (stopień zaokrąglenia ust).

W opisywanych badaniach wyznaczone zostały cztery pierwsze formanty. Na ich podstawie, podobnie jak w przypadku *F0*, wyznaczono podstawowe parametry statystyczne, otrzymując w ten sposób piętnasto-elementowy wektor cech.

3.3. Współczynniki *MFCC*

Współczynniki *MFCC*, które uwzględniają proces percepcji ucha ludzkiego, obecnie są standardem w rozpoznawaniu mowy. W metodzie tej, sygnał na początku zostaje poddany preemfazie, następnie wymnażany jest z oknem Hamminga opisanego wzorem:

$$\mathrm{w(n)} = 0{,}53836 - 0{,}46164\cos\left(\frac{2\pi \mathrm{n}}{\mathrm{N}-1}\right) \tag{1}$$

gdzie N – długość okna.

Kolejno oblicza się *FFT*, a następnie wartości prążków widma są podnoszone do kwadratu (wyznaczając w ten sposób estymatę funkcji gęstości widmowej mocy sygnału) i uśredniane za pomocą nakładających się funkcji wagowych o kształcie trójkątnym, których szerokość rośnie wraz z częstotliwością. Przy projektowaniu funkcji trójkątnych uwzględnia się psychoakustyczną skalę melową, opisaną poniższym wzorem, gdzie *m, f* jest tą samą częstotliwością w melach i hercach:

$$m = 1127{,}01048 \cdot \ln(1 + f/700); \quad f = 700(e^{m/1127{,}01048} - 1) \tag{2}$$

Następnie estymata jest logarytmowana i obliczana jest transformata kosinusowa. W ten sposób wyznaczane są współczynniki *MFCC*, według wzoru:

$$c_k = \sqrt{\frac{2}{L}} \sum_{l=0}^{L-1} \ln\left(\tilde{S}(l)\right) \cos\left(\frac{\pi k}{L}(L + 1/2)\right); \quad k = 0,1,2,\ldots,q-1 \tag{3}$$

gdzie: L – liczba zastosowanych filtrów, q – liczba wyznaczanych współczynników mel-cepstralnych [19].

Na potrzeby badań wyznaczono 12 współczynników *MFCC*, następnie na ich podstawie wyznaczono podstawowe parametry statystyczne, takiej jak średnia, odchylenie standardowe, mediana, minimum i maksimum.W ten sposób uzyskano 60 parametrów, które stanowią dane wejściowe do jednego z wektorów cech, zastosowanych w klasyfikacji.

3.4. Współczynniki *LPC*

Liniowe kodowanie predykcyjne to metoda szeroko stosowana w zadaniach rozpoznawania mowy. Ideą tej metody jest przewidywanie każdej nowej próbki na podstawie ważonej liniowej kombinacji n próbek ją poprzedzających. Wagi należy dobierać tak, aby zminimalizować błąd średniokwadratowy predykcji. Aproksymację bieżącej próbki p_n można zapisać w następujący sposób:

$$p_n = -a_1 p_{n-1} - a_2 p_{n-2} - \ldots - a_r p_{n-r} + e_n \tag{4}$$

gdzie r – rząd predyktora, e_n – pozostałość predykcj (residuum). Wagi a_1, a_2, … a_r nazywamy współczynnikami predykcji. Można je wyznaczyć np. poprzez minimalizację residułów, uśrednioną dla N próbek:

$$E : \sum_{n=1}^{N} e_n^2 = \sum_{n=1}^{N} \left(\sum_{k=0}^{r} a_k p_{n-k} \right)^2 ; a_0 = 1 \tag{5}$$

W celu minimalizacji E ze względu na residua należy przyrównać odpowiednie pochodne cząstkowe do zera:

$$\frac{\partial E}{\partial p_m} = \sum_{n=1}^{N} 2 p_{n-m} \sum_{k=0}^{r} a_k p_{n-k} = 0; \quad m \in \langle 1, r \rangle \tag{6}$$

Odwracając porządek sumowania:

$$\sum_{k=0}^{r} s_{mk} a_k = 0 \tag{7}$$

gdzie:

$$s_{mk} = \sum_{n=1}^{N} p_{n-m} p_{n-k} \tag{8}$$

Otrzymane w ten sposób współczynniki korelacji s_{mk} mogą posłużyć do obliczenia współczynników predykcji. Można zapisać:

$$\sum_{k=1}^{r} s_{mk} a_k = -s_{m0} \tag{9}$$

lub w notacji macierzowej:

$$S \cdot a = -s_0 \tag{10}$$

gdzie: S - macierz kwadratowa o wymiarze r z elementami s_{mk}, a i s_0 - wektory

Odwracając macierz S otrzymujemy ostatecznie:

$$a = -S^{-1} s_0 \tag{11}$$

Eksperymenty pokazują, że optymalną liczbą współczynników *LPC* jest 12 [2], dlatego też w niniejszych rozważaniach również przyjęto taką wartość. Na ich podstawie stworzono wektor cech statystycznych o liczności 60.

3.5. Współczynniki *PLP*

LPC przedstawia trakt głosowy jako model zbliżony do przebiegu sygnału mowy, w równym stopniu dla wszystkich częstotliwości pasma analizy. Jednakże taka reprezentacja nie jest zgodna z faktycznym procesem percepcji ucha ludzkiego, w której to czułość zmienia się wraz z częstotliwością. Dlatego analiza *PLP* została opracowana jako próba wyeliminowania tej niezgodności. Jest ona zbliżona do analizy mel, aczkolwiek zamiast skali mel wykorzystywana jest skala bark i przy ostatecznym obliczaniu współczynników zamiast *DCT* wykorzystywany jest model autoregresywny z wszystkimi biegunami [3]. Na podstawie współczynników *PLP* wyznaczono wektor parametrów analogiczny jak w przypadku współczynników *MFCC* oraz *LPC*.

3.6. Wektory cech

Na podstawie wyżej opisanych grup parametrów oraz cech związanych z energią sygnału, stworzone zostały oddzielne wektory cech. Parametry znormalizowano do przedziału (0,1). Ostateczne wyliczono 224 parametry reprezentujące sygnał, a rozkład ich ilości w obrębie danej grupy przedstawiony jest w tabeli numer 2. Wstępne badania przy użyciu klasyfikatora kNN wykazały dokładność poszczególnych parametrów, co wykorzystano w późniejszym wprowadzaniu wag podczas opisanego poniżej głosowania. Dokładność parametrów oraz przypisane wagi również przedstawiono w tabeli numer 2.

Tabela 2. Wyniki klasyfikacji emocji dla każdej grupy cech

Badane parametry	Ilość parametrów	Dokładność rozpoznawania (waga)
Parametry statystyczne *F0*	23	57.1%, (3)
Energia sygnału	6	52.6%, (2)
Parametry statystyczne *F1-F4*	15	49.7%, (1)
Parametry statystyczne dla *MFCC*	**60**	**75.8%, (6)**
Parametry statystyczne dla *PLP*	60	65%, (4)
Parametry statystyczne dla *LPC*	60	75%, (5)

Do klasyfikacji multimodalnej, szerzej opisanej w następnym rozdziale, wykorzystano algorytm *k* najbliższych sąsiadów. Klasyfikacja obiektu dokonywana jest poprzez liczenie odległość między reprezentującym go wektorem cech, a wszystkimi wektorami zbioru treningowego. Do obliczenia odległości w niniejszej pracy użyta została metryka Manhattan, obliczana na podstawie wzoru:

$$d(x_j, x_i) = \sum_{i=1}^{n} \left| x_j^i - x_k^i \right| \tag{12}$$

Nowy obiekt zaliczany jest do tej klasy, która jest najczęściej reprezentowana wśród *k* najbliższych obiektów zbioru treningowego. Algorytm ten, mimo prostoty, jest efektywny obliczeniowo oraz pomocny w rozwiązaniu złożonych problemów [1].

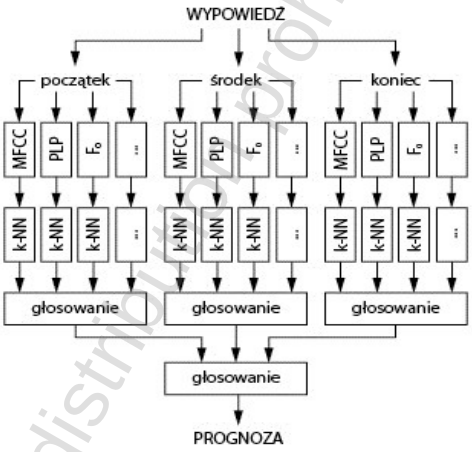
4. Proponowany algorytm klasyfikacji emocji

Typowy algorytm przetwarzania sygnału na potrzeby rozpoznawania emocji składa się z trzech podstawowych elementów. Pierwszym jest wstępna obróbka sygnału, którą w niniejszej pracy stanowi jedynie odszumianie nagrań. Następnie tworzona jest przestrzeń cech, na podstawie której przeprowadzane jest rozpoznawanie. Ostatni etap stanowi klasyfikacja, czyli określenie do jakiej klasy należy badany obiekt.

Biorąc pod uwagę możliwość zmiany zabarwienia emocjonalnego w czasie, zaproponowano algorytm klasyfikacji, przedstawiony na rysunku nr 1.

Algorytm składa się z czterech etapów. Pierwszy z nich stanowi podział wypowiedzi na trzy równe części: początek, środek i koniec. Każda z nich została poddana indywidualnej klasyfikacji przy użyciu algorytmu kNN. Należy podkreślić, iż proces ten również został podzielony na oddzielne elementy, a ich liczność odpowiada liczności różnych grup parametrów szerzej opisanych w poprzednim rozdziale.

W ten sposób rezultat pierwszego etapu, to sześć klas dla każdej części wypowiedzi. Do pierwszego głosowania dodatkowo zastosowano wagi, które dobrano na podstawie jakości grup parametrów. Tak więc współczynniki *MFCC* otrzymują najwyższą wagę - 6, natomiast wektor parametrów związanych z formantami, którego jakość klasyfikacji była najniższa, wagę 1. Na tej podstawie dokonano głosowania, zaś jego rezultat stanowi dane wejściowe kolejnego, jakim jest głosowanie w obrębie całej wypowiedzi. Na podstawie liczności klas wskazanych przez klasyfikatory w danej części wypowiedzi, wybierana jest najliczniejsza. W wyniku tego głosowania uzyskiwana jest prognoza.



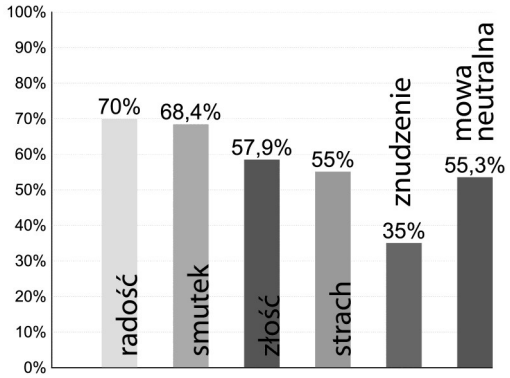
Rys. 1. Algorytm klasyfikacji

5. Eksperymenty i otrzymane rezultaty

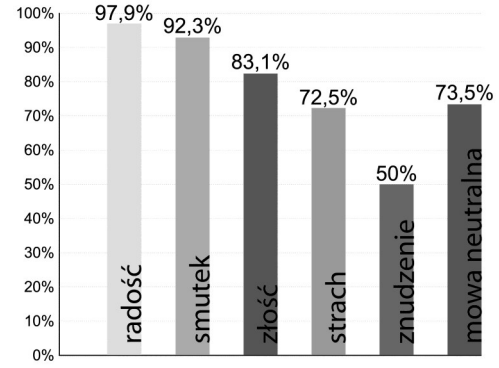
Dla porównania, badania przeprowadzono w dwóch grupach:

- baza emocji odegranych stanowi zbiór uczący i testowy (grupa I);
- baza emocji spontanicznych stanowi zbiór uczący i testowy (grupa II).

Rezultaty dla obu grup przedstawione są na poniższych wykresach.



Rys. 2. Wyniki klasyfikacji grupy pierszej



Rys. 3. Wyniki klasyfikacji grupy drugiej

Macierze pomyłek klasyfikacji wymienionych grup przedstawiają tabele nr 3 oraz 4. Wyniki na przekątnej oznaczają liczby poprawnie rozpoznanych przez system plików audio dla danej emocji.

Tabela 3. Macierz pomyłek klasyfikacji emocji dla bazy grupy I

Nazwa emocji	A	B	F	H	N	S
Złość (A)	**22**	2	5	7	1	1
Znudzenie (B)	2	**14**	5	0	14	5
Strach (F)	5	4	**22**	3	1	5
Radość (H)	8	0	3	**28**	1	0
Mowa naturalna (N)	1	8	1	0	**21**	7
Smutek (S)	0	5	4	0	3	**26**

Tabela 4. Macierz pomyłek klasyfikacji emocji dla bazy grupy II

Nazwa emocji	A	B	F	H	N	S
Złość (A)	**54**	0	0	1	0	10
Znudzenie (B)	0	**20**	6	0	14	0
Strach (F)	0	9	**29**	0	2	0
Radość (H)	1	0	0	**94**	0	0
Mowa naturalna (N)	0	8	1	0	**25**	0
Smutek (S)	6	0	0	0	0	**72**

Macierze pomyłek klasyfikacji wymienionych grup przedstawiają tabele nr 3 oraz 4. Wyniki na przekątnej oznaczają liczby poprawnie rozpoznanych przez system plików audio dla danej emocji.

6. Wnioski

Jak pokazują badania, rozpoznawanie emocji w głosie jest zadaniem trudnym, a póki co osiągane rezultaty dalekie są od ideału. Ocena stanu emocjonalnego na podstawie mowy stanowi problem nawet dla człowieka. Szczególnie trudnym, choć bardzo ważnym, zagadnieniem jest rozpoznawanie emocji w mowie spontanicznej.

Dzięki podziałowi na oddzielne wektory można zaobserwować jaki wkład mają one w poprawne rozpoznawanie. Najbardziej wydajne okazały się cechy statystyczne wyznaczone ze współczynników *MFFC* oraz *PLP*. Najgorsze rezultaty otrzymano dla parametrów statystycznych wyznaczonych z formantów. Nie dokonano selekcji cech, aczkolwiek wprowadzono wagi, co wyraźnie wskazywało na to, które z nich są istotne w klasyfikacji.

Wbrew oczekiwaniom klasyfikacja grupy I, zawierającej zarówno w zbiorze testowym jak i treningowym staranne nagrania aktorskie, okazała się wyraźnie słabsza niż ta, zawierającej nagrania mowy spontanicznej. Średnie rozpoznawanie wynosiło 56.8%. Najlepiej rozpoznawalną emocją w tej grupie była radość (70%) oraz smutek (68%). Stosunkowo często zdarzały się błędy w rozpoznaniu pomiędzy znudzeniem, a stanem neutralnym, którego rozpoznawalność była najniższa (35%).

Osiągnięte wyniki dla grupy II pokazują, że wykorzystany algorytm okazał się przydatny do rozpoznania stanów emocjonalnych w mowie spontanicznej. Średnie rozpoznawanie wynosiło 85.5%. Najlepiej rozpoznawalną emocją również w tej grupie była radość (98.9%) oraz smutek (92%). W tym przypadku wyniki mogą wiązać się ze znacznie większą liczbą badanych próbek właśnie dla tych emocji. Najniższą rozpoznawalność zanotowano również dla znudzenia (65.5%), którego to liczba próbek w zbiorze, w porównaniu z radością i smutkiem również była niska.

Przeprowadzone badania pokazują jak ważnym elementem jest baza mowy. Dlatego stworzenie odpowiednio efektywnej bazy będzie kolejnym krokiem badań. Dodatkowo trzeba napomnieć, że poprawność rozpoznania stanu emocjonalnego nieznanego mówcy przez człowieka wynosi zaledwie 60% [16]. Dodatkowo słuchacz ocenia również treść wypowiedzi.

Połączenie systemu analizującego sygnał mowy z analizą semantyczną wypowiedzi, a także z systemem wizyjnym powinno zwiększyć skuteczność rozpoznawania [17].

Naturalnym kierunkiem rozwoju prowadzonych badań jest przede wszystkim sprawdzenie możliwości innych algorytmów klasyfikacji jak również analiza innych parametrów sygnału mowy.

Literatura

[1] Basztura C.: *Komputerowe systemy diagnostyki akustycznej*, Wydawnictwo Naukowe PWN, Warszawa 1996.

[2] Ciota Z.: *Metody przetwarzania sygnałów akustycznych w komputerowej analizie mowy,* Akademicka Oficyna Wydawnicza EXIT, Warszawa 2010.

[3] Chiangkai E.: *Speech recognition by clustering wavelet and PLP coefficients,*Massachusetts Institute of Technology, Massachusetts, 1997.

[4] Gaurav M.: *Performance analysis of spectral and prosodic features and their fusion for emotion recognition in speech.* Proc. Spoken Language Technology Workshop 2008, Goa, India, p. 313 - 316.

[5] Izdebski K.: *Emotions in the Human Voice*, Volume I Foundations, San Diego, 2007.

[6] Kamaruddin N., Wahab A.: *Driver behavior analysis through speech emotion understanding,* Intelligent Vehicles Symposium (IV), 2010 IEEE, p. 238 - 243.

[7] Narayanan S., Busso C., Lee S.: *Analysis of emotionally salient aspects of fundamental frequency for emotion detection.* IEEE Transactions on audio, speech, and language processing, Volume: 17 , Issue: 4, 2009, p. 582 - 596.

[8] Niewiadomy D., Pelikant A.: *Implementation of isolated words boundaries recognition*, Proc. XII International Conference System Modeling and Control SMC'2007, Zakopane 2006.

[9] Plutchik R.: *The nature of emotion*, American Scientist, Volume 89, July-August 2001, p. 344-350.

[10] Polzehl T., Schmitt A., Metze F.: *Approaching multi-lingual emotion recognition-from speech - on language dependency of acoustic/prosodic features for anger recognition.* Proc. of Speech Prosody, Chicago 2010.

[11] Razak A., Komiya R., Abidin M.: *Comparison between fuzzy and nn method for speech emotion recognition.* Proc. Information Technology and Applications, ICITA 2005, p. 297 - 302.

[12] Scherer K., Goudbeek M., Goldman J.P.: *Emotion dimensions and formant position*, Proc. Interspeech 2009, Brighton UK, 2009.

[13] Shaukat A., Chen K.: *Emotional state recognition from speech via soft-competition on different acoustic representations.* Proc. Neural Networks (IJCNN), 2011, p. 1910 - 1917.

[14] Soltani K., Ainon R.: *Speech emotion detection based on neural networks*, Proc. Signal Processing and Its Applications, 2007, p. 1 – 3.

[15] Ślot K.: *Rozpoznawanie biometryczne*, WKiŁ, Warszawa, 2010.

[16] Turkot M., Janicki A.: *Rozpoznawanie stanu emocjonalnego mówcy z wykorzystaniem maszyny wektorów wspierających (SVM)*, Materiały konferencyjne: KSTiT, Bydgoszcz, 2008.

[17] Wang Y., Guan L.: *Recognizing human emotional state from audiovisual signals*, Proc. IEEE Transactions on multimedia, vol. 10, 2008, p. 659 - 668.

[18] Yeqing Y., Tao T.: *An new speech recognition method based on prosodic analysis and SVM in Zhuang language*, Proc. 2011 International Conference on Mechatronic Science, Electric Engineering and Computer, 2011, p. 1209 - 1212.

[19] Zieliński T.: *Cyfrowe przetwarzanie sygnałów. Od teorii do zastosowań.* WKiŁ, Warszawa 2007.

[20] https://www.msu.edu/course/

[21] http://pascal.kgw.tu-berlin.de

[22] http://www.eletel.p.lodz.pl/med/

Mgr inż. Dorota Kamiska

e-mail: dorota.kaminska@p.lodz.pl

Dorota Kamińska uzyskała tytuł mgr inż. w 2009 roku na wydziale Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej. Obecnie jest doktorantką w Instytucie Mechatroniki i Systemów Informatycznych Politechniki Łódzkiej. Główne zainteresowania badawcze obejmują przetwarzanie sygnałów, metody klasyfikacji oraz bazy danych.

Dr hab. inż. Adam Pelikant

e-mail: adam.pelikant@p.lodz.pl

Profesor nadzwyczajny w Instytucie Mechatroniki i Systemów Informatycznych Wydziału Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej. Główne zainteresowania badawcze obejmują bazy danych, hurtownie danych i eksplorację danych.