STUDIA INFORMATICA 2010
Volume 31 Number 2A (89)

Tomasz POTEMPA
Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Informatyki
Robert WIELGAT
Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Elektroniki i Telekomunikacji
Daniel KRÓL
Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Informatyki
Paweł KOZIOŁ
Regionalna Dyrekcja Ochrony Środowiska w Krakowie
Agnieszka LISOWSKA-LIS
Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Elektrotechniki

WSPOMAGANIE AUTOMATYCZNEGO ROZPOZNAWANIA SYGNAŁÓW DŹWIĘKOWYCH Z WYKORZYSTANIEM MULTIMEDIALNEJ BAZY DANYCH*

Streszczenie. W artykule opisano sposób usprawnienia rozpoznawania sygnałów dźwiękowych z wykorzystaniem multimedialnej bazy danych. Założono rozpoznawanie sygnałów poprzez porównywanie ich ze wzorcami. Duża liczba wzorców wydłuża czas i może obniżać skuteczność rozpoznawania. Rozwiązaniem problemu może być wstępna preselekcja wzorców na podstawie wybranych parametrów sygnału. Dokonano porównania jakości preselekcji dla poszczególnych parametrów.

Słowa kluczowe: multimedialna baza danych, rozpoznawanie sygnałów

SUPPORTING AUTOMATIC AUDIO SIGNAL RECOGNITION USING MULTIMEDIA DATABSE

Summary. Method for improving audio signals recognition using multimedia database is presented in the paper. Recognition using signals patterns matching was assumed. Large number of signal patterns prolongs recognition time and may decrease recognition accuracy. Initial preselection of signal patterns based on chosen

* Badania zostały wykonane i sfinansowane w ramach grantu MNiSW nr N N519 402934, „Opracowanie automatycznego systemu akustycznego monitoringu ptaków dla Ciężkowicko-Rożnowskiego Parku Krajobrazowego”.

signal parameters can alleviate problem. Comparison of effectiveness of preselection method for parameters was examined.

Keywords: multimedia database, signal recognition

1. Wprowadzenie

Istotnym elementem ochrony przyrody jest przeciwdziałanie wymieraniu rzadkich gatunków zwierząt. Dużą grupę wymierających lub zagrożonych gatunków zwierząt stanowią ptaki. Aby je chronić należy najpierw zadać sobie pytanie, które gatunki są zagrożone. Odpowiedzi na to pytanie dostarczają różnego rodzaju badania terenowe, zmierzające do wyznaczenia populacji oraz siedlisk wielu gatunków ptaków [1]. Badania terenowe polegające głównie na obserwowaniu i zliczaniu osobników poszczególnych gatunków są bardzo kosztowne i czasochłonne oraz angażują wiele osób. Sytuację pogarsza fakt, że wiele instytucji nie posiada odpowiednich baz danych, w których można by było zapisywać wyniki badań terenowych, co znacząco utrudnia późniejsze opracowanie wyników. Jednak nawet dobrze zaprojektowana baza z dostępem przez sieć internetową nie rozwiązuje wszystkich problemów.

Głównym problemem w przypadku baz danych przechowujących dane na temat rozmieszczenia i liczebności populacji poszczególnych gatunków ptaków jest pozyskiwanie i wprowadzanie informacji do bazy danych. W chwili obecnej dominującym sposobem wprowadzania informacji jest ich ręczne wpisywanie, co jest czasami dość czasochłonne. Ponadto wprowadzane dane mogą być obarczone błędami ze względu na złe oznaczanie gatunków ptaków, zwłaszcza przez osoby z mniejszym doświadczeniem. Pewnym problemem jest również nieuwzględnianie w obserwacjach terenowych ptaków, których nie widać, ale słychać. Może być to spowodowane brakiem doświadczenia w rozpoznawaniu głosów ptaków, jak również innymi czynnikami natury obiektywnej.

Rozwiązaniem, które mogłoby w pewnym stopniu rozwiązać zarysowane powyżej problemy jest nagrywanie głosów ptaków w trakcie badań terenowych i ich późniejsze automatyczne lub półautomatyczne rozpoznawanie oraz wprowadzanie do bazy danych. Nie bez znaczenia jest fakt, że baza danych, w której są gromadzone duże ilości informacji w połączeniu z systemem eksperckim stwarzają nieporównanie większe możliwości rozpoznawania sygnałów niż programy rozpoznające sygnały bez wsparcia bazodanowego.

Wśród metod automatycznego rozpoznawania głosów ptaków, którego celem jest rozpoznawanie gatunku, często spotyka się rozwiązania bazujące na metodach rozpoznawania mowy ludzkiej. Wśród tych metod można wymienić metody rozpoznawania mówców [2], jak również metody rozpoznawania wypowiadanych treści: niejawne modele

Markowa (HMM) [3], nieliniowa transformacja czasowa (DTW) [4]. Ta druga grupa metod oprócz rozpoznawania gatunków ptaków pozwala również na rozpoznanie sygnałów komunikacyjnych ptaków. Zatem najbardziej uniwersalnym sposobem automatycznego rozpoznawania głosów ptaków wydaje się być zastosowanie wybranej metody rozpoznawania treści wypowiedzi (w rozważanym przypadku sygnałów komunikacyjnych ptaków).

Obecnie najczęściej używaną metodą do rozpoznawania treści wypowiedzi jest metoda HMM. Niemniej jednak w kontekście opisywanego zagadnienia posiada ona pewne niedogodności, które utrudniają lub wręcz uniemożliwiają jej efektywne zastosowanie do rozpoznawania głosów ptaków. Pierwszą niedogodnością jest wciąż niewystarczająca wiedza na temat podstawowych jednostek leksykalnych tworzących sygnały komunikacyjne ptaków, która jest niezbędna w celu stworzenia niejawnych modeli Markowa dla wspomnianych jednostek leksykalnych. Istnieją co prawda w literaturze światowej opracowania dotyczące sygnałów komunikacyjnych ptaków oraz ich morfologii [3], jednak są one niekompletne. Ponadto w przypadku dużej liczby rozpoznawanych gatunków liczba jednostek leksykalnych znacząco wzrasta, co stwarza ryzyko pojawienia się bardzo podobnych modeli Markowa dla różnych jednostek leksykalnych i w efekcie pogorszenie jakości rozpoznawania. Kolejną niedogodnością w przypadku niejawnych modeli Markowa jest stosunkowo czasochłonny proces trenowania modeli Markowa, co stwarza problemy w przypadku ich uaktualniania na podstawie nowych sygnałów dźwiękowych. Niemniej jednak, w przypadku gdyby modelowaną jednostką był cały sygnał komunikacyjny, zastosowanie metody HMM wydaje się uzasadnione.

Pozbawiona wymienionych niedogodności jest metoda DTW. Nie wymaga ona stosowania zaawansowanej analizy lingwistycznej rozpoznawanej wypowiedzi. Do rozpoznawania potrzebna jest oznaczona baza wzorców, gdzie etykietą wzorca jest nazwa gatunku ptaka oraz opcjonalnie rodzaj sygnału komunikacyjnego. Sposób aktualizacji bazy wzorców jest stosunkowo prosty, ponieważ wymaga jedynie dodania nowego odpowiednio oznaczonego wzorca do bazy. W przypadku przepełnienia bazy wzorców z bazy są usuwane wzorce najrzadziej używane lub te, które są najczęstszym źródłem pomyłek. Ponadto metoda nieliniowej transformacji czasowej jest prostsza w implementacji.

Pomimo przedstawionych zalet metody DTW posiada ona jedną zasadniczą wadę – konieczność przechowywania w bazie dużej liczby wzorców proporcjonalnej do liczby rozpoznawanych sygnałów komunikacyjnych ptaków. Duża liczba wzorców powoduje zwiększenie czasu rozpoznawania oraz może prowadzić do zwiększenia liczby błędnie rozpoznawanych sygnałów. Poza tym duża liczba wzorców powoduje również dużą zajętość pamięci, jednak przy obecnym szybkim wzroście pojemności nośników pamięci ta niedogodność traci na znaczeniu.

Rozwiązaniem problemu może być szybka i efektywna metoda ograniczania liczby wzorców, w najbardziej korzystnym przypadku wyłącznie do wzorców rozpoznawanego gatunku (sygnału komunikacyjnego) ptaka. W kolejnych podpunktach rozdziału zostaną przedstawione badania zmierzające do znalezienia jak najbardziej efektywnej metody ograniczania liczby wzorców – inaczej mówiąc preselekcji wzorców – na podstawie wybranych prostych parametrów sygnału.

2. Preselekcja wzorców na podstawie parametrów sygnału

Skuteczną pod względem obliczeniowym metodą preselekcji wzorców wydaje się być wyznaczenie prostych parametrów rozpoznawanego sygnału oraz ograniczanie grupy wzorców do wzorców o wartościach parametru zbliżonych do wartości parametru rozpoznawanego sygnału. Przykładowo, wyznaczając długość sygnału w próbkach można ograniczyć liczbę wzorców do tych, które mieszczą się w zakresie od 0,5 do 1,5 długości rozpoznawanego sygnału.

2.1. Parametry sygnału

Dla dowolnego sygnału można wyznaczyć bardzo wiele charakteryzujących go parametrów. W opisywanych badaniach przetestowano efektywność preselekcji dla następujących parametrów:

1) długość sygnału w próbkach,
2) odcięta środka ciężkości kwadratu sygnału w próbkach,
3) odcięta środka ciężkości kwadratu modułu średniego FFT,
4) SNR sygnału.

Długość sygnału w próbkach jest parametrem, który można odczytywać bezpośrednio z nagłówka plików w formacie WAV. Odciętą środka ciężkości kwadratu sygnału można obliczyć wg wzoru [5]:

$$OSKS = \frac{\sum_{n=1}^{N} n \cdot s^2(n)}{\sum_{n=1}^{N} s^2(n)} \tag{1}$$

Odciętą środka ciężkości kwadratu modułu średniego FFT obliczano na podstawie średniego wektora FFT, k-ty element średniego wektora modułu FFT obliczano jako średnia arytmetyczna k-tych elementów wektorów modułu FFT wyznaczanych w ramkach 1024 próbek przesuniętych co 512 próbek. Przed wyznaczaniem FFT sygnał w ramce wymnażano

przez funkcję okna Hamminga. Odciętą środka ciężkości kwadratu modułu FFT wyznaczano zgodnie ze wzorem (1), gdzie s(k) jest k-tym elementem średniego wektora modułu FFT. W średnim wektorze modułu FFT nie uwzględniano pierwszego elementu transformaty FFT reprezentującego składową stałą sygnału, zatem N = 511.
SNR wyznaczano zgodnie ze wzorem:

$$SNR = 20\log_{10}\frac{P_{S\max}}{P_{N\min}}, \tag{2}$$

gdzie: P_{Smax} – maksymalna moc sygnału, P_{Nmin} – minimalna moc szumów.

Wartość PSmax była obliczana jako maksymalna wartość mocy ze wszystkich ramek sygnału. Założono, że sygnał jest tą częścią nagrania, która zaczyna się 100 ms za początkiem nagrania, a kończy 100 ms przed końcem nagrania. Do obliczania SNR przyjęto długość ramki równą 20 ms, a przesunięcie między ramkami równe 10 ms. Wartość PNmin była obliczana jako minimalna wartość mocy ze wszystkich ramek szumu. Założono, że fragmenty szumu zajmują pierwsze 100 ms nagrania oraz ostatnie 100 ms nagrania. Zarówno moc sygnału, jak i moc szumu obliczano na podstawie próbek z pojedynczej ramki nagrania zgodnie ze wzorem:

$$P = \sum_{k=1}^{N} n^2(k)/N, \tag{3}$$

gdzie: n(k) – k-ta próbka nagrania (sygnału lub szumu) w ramce, N – liczba próbek w ramce równa 960 dla częstotliwości próbkowania 48 kHz.

Liczba próbek w ramce dla sygnału równego lub dłuższego niż 1 ramka wynosiła 960. Jeżeli sygnał był krótszy niż jedna ramka, wtedy liczba próbek w ramce była równa liczbie próbek sygnału.

2.2. Metody preselekcji

W prezentowanej pracy zbadano efekty działania metod preselekcji, które nazwano:

- metoda przedziałowa,
- twarda metoda rozkładu normalnego,
- miękka metoda rozkładu normalnego.

2.2.1. Metoda przedziałowa

W metodzie przedziałowej na początku dla każdej klasy i każdego parametru preselekcji są wyznaczane 2 współczynniki – dolny i górny, przez które przemnaża się następnie odpowiedni parametr rozpoznawanego sygnału (np. jego długość) w celu zawężenia liczby wzorców. Następnie na etapie preselekcji parametr sygnału przemnaża się przez dolny i górny współczynnik otrzymując odpowiednio dolną i górną granicę przedziału preselekcji dla

każdej klasy. Jeżeli wartość parametru danego wzorca zawiera się w przedziale, to wzorzec zostaje wybrany, jeżeli nie znajduje się w przedziale, to zostaje odrzucony. Na podstawie wybranych wzorców można rozpoznawać sygnał, podczas gdy wzorce odrzucone nie są w procesie rozpoznawania wykorzystywane.

Obliczanie dolnego i górnego współczynnika preselekcji odbywa się wg poniższego algorytmu:

1. $n = 0$.
2. Odczytaj z bazy danych wartości parametru dla wzorców n-tej klasy sygnałów komunikacyjnych:

$$p_n(0), p_n(1),\ldots, p_n(k),\ldots p_n(N_n - 1), \tag{4}$$

gdzie: $p_n(k)$ – wartość parametru dla k-tego wzorca n-tej klasy; N_n – liczba wzorców dla n-tej klasy.

3. Oblicz znormalizowane różnice między uszeregowanymi w kolejności rosnącej sąsiednimi wartościami parametru.

$$d_n(k) = \frac{p_n(k+1) - p_n(k)}{(p_n(k+1) + p_n(k))/2} \tag{5}$$

4. Oblicz dolny i górny współczynnik preselekcji dla n-tej klasy na podstawie wartości parametrów dających maksymalną znormalizowaną różnicę.

$$wd(n) = \frac{2 \cdot p_n(k_{\max})}{p_n(k_{\max}+1) + p_n(k_{\max})}, \qquad wg(n) = \frac{2 \cdot p_n(k_{\max}+1)}{p_n(k_{\max}+1) + p_n(k_{\max})}, \tag{6}$$

gdzie: $k_{\max}$= arg max $(d_n(k))$, k=0,1, …, N_n–2.

5. n=n+1, jeżeli n jest równe liczbie klas, to zakończ, w przeciwnym przypadku idź do punktu 2.

Za pomocą opisanego powyżej algorytmu wyznacza się dolny i górny współczynnik preselekcji dla każdego parametru preselekcji z osobna. Metoda przedziałowa opisana tym algorytmem pozwala na proste wyznaczenie dolnego i górnego współczynnika preselekcji i równie prostą preselekcję w oparciu o obliczone współczynniki. Wyznaczanie współczynników dla wartości parametrów dających największą różnicę (największy przedział) pozwala na uniknięcie sytuacji, w której żaden wzorzec nie jest wybrany, pomimo tego, że wartość parametru sygnału rozpoznawanego zawiera się między minimalną a maksymalną wartością parametru wyznaczoną na podstawie wzorców.

W przedziałowej metodzie preselekcji po wyznaczeniu współczynników wd i wg oblicza się dla rozpoznawanego słowa górną i dolną granicę przedziału wartości danego parametru. Następnie wybiera się wzorce, dla których wartość parametru należy do przedziału. Proces wybierania wzorców zaczyna się biorąc na początek parametr, który najskuteczniej wybiera wzorce, a następnie kolejne parametry, które wybierają wzorce coraz mniej skutecznie.

2.2.2. Twarda metoda rozkładu normalnego

W metodzie tej kryterium preselekcji stanowi przedział ufności, którego granice są obliczane na podstawie funkcji gęstości rozkładu normalnego ze średnią μ i odchyleniem standardowym σ dla każdej klasy rozpoznawanych gatunków:

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right) \tag{7}$$

Do wyznaczenia rozkładu normalnego wykorzystano średnią arytmetyczną wartości próby μ oraz estymator odchylenia standardowego dla próby w postaci:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{8}$$

gdzie: x_i – wartości parametru w próbie, $\bar{x}$ średnia arytmetyczna wartości próby, n liczba elementów próby.

Granice przedziału ufności stanowią zakres:

$$\langle \mu - w \cdot \sigma; \mu - w \cdot \sigma \rangle, \tag{9}$$

gdzie: w – waga, $w>0$.

W metodzie dla każdego rozpoznawanego sygnału komunikacyjnego wyznacza się listę wzorców, które zawierają się w przedziale, do którego należy przedmiotowy sygnał.

2.2.3. Miękka metoda rozkładu normalnego

W tej metodzie preselekcji na podstawie wartości parametru dla wzorców każdej klasy są obliczane estymaty parametrów rozkładu normalnego, czyli średnia wartość parametru w grupie wzorców należących do tej samej klasy oraz odchylenie standardowe dla wzorców tej samej klasy. Parametry rozkładu normalnego oblicza się dla wzorców każdej klasy.

Następnie jest wyznaczana gęstość prawdopodobieństwa dla parametru rozpoznawanego sygnału na podstawie rozkładu normalnego dla danej klasy i rozpatrywanego parametru. W dalszej kolejności wyznacza się gęstości prawdopodobieństwa dla kolejnych parametrów w tej samej klasie. Otrzymane wartości gęstości prawdopodobieństwa mnoży się przez siebie otrzymując globalną gęstość prawdopodobieństwa w danej klasie. Jeżeli globalna gęstość prawdopodobieństwa będzie mniejsza od progu kwalifikacji dla danej klasy, wówczas podejmuje się decyzję, że rozpoznawany sygnał komunikacyjny nie należy do klasy, jeżeli jest większa lub równa, to podejmuje się decyzję, że rozpoznawany sygnał może należeć do tej klasy. Opisany proces powtarza się dla wszystkich klas otrzymując zbiór klas, do których może należeć sygnał komunikacyjny. Następnie do etapu rozpoznawania wybiera się wszystkie wzorce z zakwalifikowanych klas. Próg kwalifikacji dla każdej klasy został

wyznaczony na podstawie krzywych ROC (*Receiver Operating Characteristic*) przy założeniu TPR ≥ 0,99, gdyż w przypadku rozważanej preselekcji negatywne skutki odrzucenia poprawnego wzorca (FN) są zdecydowanie większe od negatywnych skutków przyjęcia wzorca niepoprawnego (FP).

3. Opis doświadczeń

W opisywanych w niniejszym rozdziale doświadczeniach dokonywano preselekcji sygnałów ze zbioru nagrań głosów ptaków. Zbiór nagrań podzielono na zbiór uczący, czyli wzorce oraz zbiór testowy. W tabeli 1 Zestawiono podstawowe informacje na temat zbioru nagrań.

Tabela 1

Struktura zbioru nagrań głosów ptaków

Lp.	Gatunek ptaka	Liczba wzorców	Liczba nagrań w zbiorze testowym
1.	Bogatka (Parus major)	10	31
2.	Dzięcioł duży (Dendrocopos major)	10	40
3.	Gawron (Corvus frugilegus)	10	40
4.	Kos (Turdus merula)	10	89
5.	Sójka (Garrulus glandarius)	10	51
6.	Trzcinniczek (Acrocephalus scirpaceus)	10	41
7.	Ziemba (Fringilla coelebs)	10	67

Zbiór testowy został utworzony z reprezentatywnej próby wzorców dobranej w taki sposób, aby każdy rodzaj sygnału komunikacyjnego był przynajmniej raz reprezentowany w zbiorze uczącym. Nagrania były dokonywane za pomocą mikrofonu kardioidalnego lub hiperkardioidalnego z częstotliwością próbkowania 48 kHz i rozdzielczością bitową 16 bitów/próbkę.

Do oceny jakości preselekcji wykorzystano wybrane wskaźniki analizy ROC (Receiver Operating Characteristic) [6] przedstawione poniżej:

- Dokładność (precision)

$$Pc = \frac{TP}{TP + FP} \tag{10}$$

- Kompletność/czułość (sensitivity)

$$Cz = \frac{TP}{P} \tag{11}$$

- Specyficzność (specificity):

$$Sp = \frac{TP}{FP + TN} \tag{12}$$

- Trafność (accuracy):

$$Tr = \frac{TP + TN}{P + N} \tag{13}$$

- Miara F (F-measure):

$$Fm = \frac{1}{\frac{1}{Tr} + \frac{1}{Cz}}, \tag{14}$$

gdzie:

TP – liczba prawidłowo wybranych wzorców;

TN – liczba prawidłowo odrzuconych wzorców;

FP – liczba nieprawidłowo wybranych wzorców;

P – liczebność wzorców prawidłowych (należących do danej klasy);

N – liczebność wzorców nieprawidłowych (należących do pozostałych klas).

Na potrzeby opisywanych badań wprowadzono również nowy wskaźnik wynikający z analizy ROC:

- Ograniczanie (ang. Limitation)

$$Og = \frac{TP + FP}{P + N} \tag{15}$$

Ponadto w celu dokładniejszego scharakteryzowania jakości preselekcji wprowadzono dodatkowe wskaźniki, nie wynikające bezpośrednio z analizy ROC. Tymi wskaźnikami są: skuteczność preselekcji (SP) oraz wskaźnik prawidłowo wybranych klas (WPK). Skuteczność preselekcji (SP) została obliczona jako stosunek liczby sygnałów danej klasy ze zbioru testowego, dla których został wybrany przynajmniej jeden prawidłowy wzorzec (PZ) do wszystkich sygnałów tej klasy ze zbioru testowego (WS):

$$SP = \frac{PZ}{WS} \cdot 100\% \tag{16}$$

Wzorzec uważa się za prawidłowy, jeżeli należy do tej samej klasy, co sygnał ze zbioru testowego, dla którego jest dokonywana preselekcja .

Kolejnym wskaźnikiem jest wskaźnik prawidłowo wybranych klas (WPK), będący stosunkiem liczby prawidłowo wybranych klas (LPK) do liczby wszystkich klas (LWK).

$$WPK = \frac{LPK}{LWK} \cdot 100\% \tag{17}$$

Klasę uważa się za wybraną, jeżeli przynajmniej 1 wzorzec tej klasy został wybrany w procesie preselekcji.

W tabelach 2, 3, 4 przedstawiono wartości poszczególnych wskaźników preselekcji dla trzech parametrów preselekcji dokonywanej za pomocą metody przedziałowej w zależności od gatunku ptaka.

Tabela 2

Współczynniki jakości preselekcji na podstawie długości sygnału dla różnych gatunków ptaków dla metody przedziałowej

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	216	687	100	0,69	17,71	0,31	0,7	0,75	0,74	0,43
2.	Dzięcioł duży	352	418	100	0,88	45,45	0,84	0,88	0,97	0,96	0,86
3.	Gawron	242	1013	100	0,6	16,81	0,24	0,61	0,68	0,67	0,34
4.	Kos	349	1827	100	0,39	18,09	0,19	0,39	0,72	0,68	0,26
5.	Sójka	213	1199	100	0,41	17,41	0,18	0,42	0,68	0,64	0,25
6.	Trzcinniczek	259	776	100	0,63	18,14	0,33	0,63	0,79	0,77	0,44
7.	Ziemba	298	1613	100	0,44	17,18	0,18	0,44	0,67	0,64	0,26
8.	ŚREDNIE WARTOŚCI	-	-	100	0,53	21,54	0,33	0,58	0,75	0,73	0,41

Tabela 3

Współczynniki jakości preselekcji na podstawie odciętej środka ciężkości kwadratu sygnału dla różnych gatunków ptaków

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	71	739	97	0,34	15,31	0,1	0,23	0,64	0,58	0,14
2.	Dzięcioł duży	136	835	100	0,29	16,39	0,16	0,34	0,71	0,66	0,22
3.	Gawron	52	886	78	0,31	13,30	0,06	0,13	0,65	0,58	0,08
4.	Kos	274	2590	99	0,41	14,36	0,11	0,31	0,57	0,53	0,16
5.	Sójka	451	1657	100	0,46	14,33	0,28	0,88	0,62	0,66	0,43
6.	Trzcinniczek	79	863	61	0,3	10,59	0,09	0,19	0,68	0,61	0,12
7.	Ziemba	339	1530	100	0,32	15,99	0,22	0,51	0,7	0,68	0,31
8.	ŚREDNIE WARTOŚCI	-	-	90,71	0,35	14,32	0,15	0,37	0,65	0,61	0,21

W tabelach 5, 6, 7 przedstawiono wartości poszczególnych wskaźników preselekcji dla trzech parametrów preselekcji dokonywanej za pomocą twardej metody rozkładu normalnego.

Tabela 4

Współczynniki jakości preselekcji na podstawie odciętej środka ciężkości kwadratu średniego FFT sygnału dla różnych gatunków ptaków

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	93	359	81	0,16	17,36	0,26	0,3	0,86	0,78	0,28
2.	Dzięcioł duży	180	496	98	0,17	20,86	0,36	0,45	0,87	0,81	0,4
3.	Gawron	68	632	90	0,22	14,81	0,11	0,17	0,77	0,68	0,13
4.	Kos	285	1320	100	0,21	18,98	0,22	0,32	0,81	0,74	0,26
5.	Sójka	133	498	86	0,13	22,45	0,27	0,26	0,88	0,79	0,26
6.	Trzcinniczek	70	633	95	0,22	15,85	0,11	0,17	0,77	0,69	0,13
7.	Ziemba	196	1092	97	0,23	15,26	0,18	0,29	0,78	0,71	0,22
8.	ŚREDNIE WARTOŚCI	-	-	92,43	0,19	17,94	0,21	0,28	0,82	0,74	0,24

Tabela 5

Współczynniki jakości preselekcji na podstawie długości sygnału dla różnych gatunków ptaków (dla wagi $w = 3{,}0$)

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	310	1887	100	0,86	14,28	0,16	1	0,15	0,27	0,28
2.	Dzięcioł duży	400	2751	100	0,98	14,28	0,15	1	0,02	0,16	0,25
3.	Gawron	400	2400	100	0,85	16,66	0,17	1	0,17	0,29	0,29
4.	Kos	890	5083	100	0,81	14,28	0,18	1	0,21	0,33	0,3
5.	Sójka	510	2810	100	0,78	16,66	0,18	1	0,25	0,36	0,31
6.	Trzcinniczek	410	2514	100	0,87	14,28	0,16	1	0,14	0,27	0,28
7.	Ziemba	670	3840	100	0,81	16,66	0,17	1	0,21	0,32	0,3
8.	ŚREDNIE WARTOŚCI	-	-	100	0,85	15,3	0,17	1	0,17	0,28	0,29

Krzywe ROC w metodzie miękkiej rozkładu normalnego zostały wyznaczone metodą One-Versus-Rest, w której problem wieloklasowy rozwiązywany jest poprzez wyznaczenie wartości TPR, FPR oraz wykreślenie krzywej ROC dla każdej klasy oddzielnie. W rozpatrywanym przypadku dla każdej z klas wykonano obliczenia przyjmując wzorce rozpatrywanej klasy jako prawidłowe, natomiast wzorce wszystkich pozostałych klas jako błędne. Ponadto wyznaczono wartość parametru określającego wydajność klasyfikacji AUC (*Area Under Curve*) na podstawie [7]:

$$AUC_t = \sum_{c_i \in C} AUC(c_i) \cdot p_i(c_i), \tag{18}$$

gdzie: $AUC(c_i)$ – powierzchnia pod krzywą ROC dla klasy c_i; p_i – waga określająca dominancję (liczbę) wzorców klasy c_i w zbiorze C.

Wartość AUC dla rozpatrywanego przypadku wynosi 0,8169.

Tabela 6

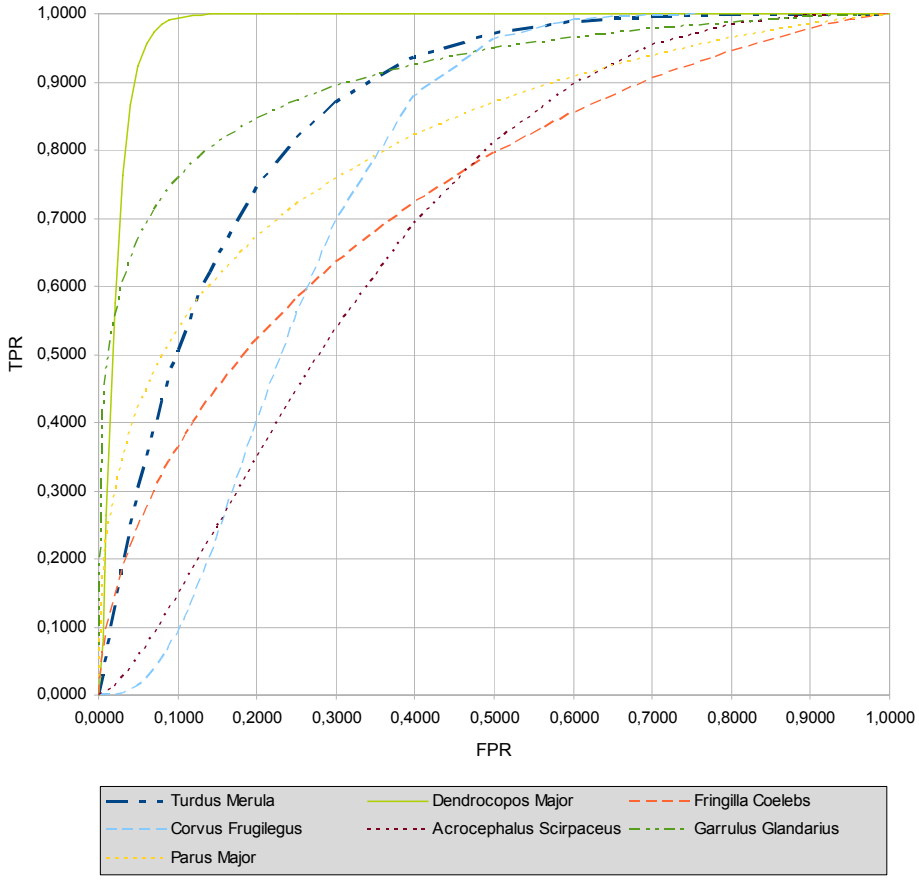
Współczynniki jakości preselekcji na podstawie odciętej środka ciężkości kwadratu sygnału dla różnych gatunków ptaków (dla wagi w = 4,0)

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	310	2170	100	1	14,28	0,14	1	0	0,14	0,25
2.	Dzięcioł duży	400	2750	100	0,98	14,28	0,15	1	0,02	0,16	0,25
3.	Gawron	400	2760	100	0,98	14,28	0,14	1	0,02	0,16	0,25
4.	Kos	890	6230	100	1	14,28	0,14	1	0	0,14	0,25
5.	Sójka	510	3570	100	1	14,28	0,14	1	0	0,14	0,25
6.	Trzcinniczek	410	2840	100	0,98	14,28	0,14	1	0,01	0,15	0,25
7.	Ziemba	670	4670	100	0,99	14,28	0,14	1	0	0,15	0,25
8.	ŚREDNIE WARTOŚCI	-	-	100	0,99	14,28	0,14	1	0,01	0,15	0,25

Tabela 7

Współczynniki jakości preselekcji na podstawie odciętej środka ciężkości kwadratu średniego FFT sygnału dla różnych gatunków ptaków (dla wagi w = 11,0)

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	310	1820	100	0,83	14,28	0,17	1	0,19	0,3	0,29
2.	Dzięcioł duży	400	2380	100	0,85	14,28	0,17	1	0,18	0,29	0,29
3.	Gawron	400	2550	100	0,91	14,28	0,16	1	0,1	0,23	0,27
4.	Kos	890	5420	100	0,86	14,28	0,16	1	0,15	0,27	0,28
5.	Sójka	510	3570	100	1	14,28	0,14	1	0	0,14	0,25
6.	Trzcinniczek	410	2590	100	0,9	14,28	0,16	1	0,11	0,24	0,27
7.	Ziemba	670	4380	100	0,93	14,28	0,15	1	0,08	0,21	0,27
8.	ŚREDNIE WARTOŚCI	-	-	100	0,9	14,28	0,16	1	0,12	0,24	0,27



Rys. 1. Krzywe ROC dla iloczynu gęstości prawdopodobieństw parametrów sygnału ze zbioru wzorców dla różnych gatunków ptaków

Fig. 1. ROC curves traced on product of signal parameters probability density for patterns set of different bird species

W tabeli 8 przedstawiono wartości poszczególnych wskaźników preselekcji dokonywanej za pomocą miękkiej metody rozkładu normalnego.

Tabela 8

Współczynniki jakości preselekcji na podstawie długości sygnału, odciętej środka ciężkości kwadratu sygnału, odciętej środka ciężkości kwadratu średniego FFT, SNR dla różnych gatunków ptaków

Lp.	Gatunek ptaka	Długość sygnału									
		PW	*WW*	SP [%]	*Og*	*WPK* %]	*Pc*	*Cz*	*Sp*	*Tr*	*Fm*
1.	Bogatka	270	1220	100	0,64	14,28	0,22	0,87	0,49	0,54	0,35
2.	Dzięcioł duży	390	1990	97,5	0,71	14,28	0,2	0,98	0,33	0,43	0,33
3.	Gawron	380	1920	95	0,68	14,28	0,2	0,95	0,36	0,44	0,33
4.	Kos	860	4550	96,62	0,73	14,28	0,19	0,97	0,31	0,4	0,32
5.	Sójka	500	2410	98,03	0,67	16,66	0,21	0,98	0,38	0,46	0,34
6.	Trzcinniczek	380	2170	92,68	0,75	14,28	0,18	0,93	0,27	0,37	0,29
7.	Ziemba	630	3560	94,02	0,75	14,28	0,18	0,94	0,27	0,37	0,3
8.	ŚREDNIE WARTOŚCI	-	-	96,26	0,7	14,62	0,2	0,95	0,34	0,43	0,32

4. Dyskusja wyników i wnioski

Zanim zostanie przedstawiona dyskusja wyników oraz będą zaprezentowane wnioski, należy zastanowić się nad interpretacją poszczególnych wskaźników jakości preselekcji. Wskaźnik *SP* – skuteczność preselekcji mówi o tym, jak dużo sygnałów jest prawidłowo zaklasyfikowanych. Skuteczność preselekcji powinna wynosić 100%. Jeżeli jest inna, oznacza to, że pewne sygnały ze zbioru testowego zostały błędnie zaklasyfikowane i nie ma możliwości ich poprawnego rozpoznania na etapie ostatecznej klasyfikacji.

Precyzja (*Pc*) również powinna wynosić w idealnym przypadku 100%. Wówczas oznacza to, że wybrane wzorce należą do jednej klasy i w dodatku jest to prawidłowa klasa. Jest to bardzo korzystna sytuacja, ponieważ nie jest wówczas konieczne dalsze rozpoznawanie, a rozpoznaną klasą jest ta, do której należą wybrane wzorce. Niemniej jednak obniżenie wartości wskaźnika do wartości mniejszej niż 100% nie powoduje tak poważnych konsekwencji jak w przypadku obniżenia się skuteczności preselekcji.

Niewielka wartość wskaźnika czułość (*Cz*) sygnalizuje, że zostało wybranych niewiele wzorców prawidłowej klasy, w związku z czym rozpoznawanie sygnału może być trudniejsze. Z kolei wybór dużej liczby wzorców pociąga za sobą zwiększenie czasu obliczeń. Jednak w kontekście opisywanego zastosowania czas obliczeń nie jest krytycznym parametrem w systemie, dlatego pożądane jest, aby czułość była możliwie bliska 100%.

Ograniczanie (*Og*) powinno mieć jak najmniejszą wartość większą od 0. Małe wartości wskaźnika *Og* oznaczają mniejszą liczbę wybranych wzorców, przez co czas rozpoznawania następującego po etapie preselekcji jest również krótki.

Opisane powyżej wskaźniki mają znaczenie w przypadku stosowania metody DTW. W przypadku stosowania metody HMM z modelowaniem całych sygnałów znaczenie ma ostatni wskaźnik WPK . Im bliższy 100%, tym lepiej.

W praktyce wskaźniki preselekcji mniej lub bardziej odbiegają od wzorcowych wartości. W przeprowadzonych badaniach najlepsze wskaźniki uzyskano dla dzięcioła dużego. Analizując nagrania głosu dzięcioła dużego należy stwierdzić, że wykonane nagrania sygnałów komunikacyjnych dzięcioła dużego niewiele różniły się między sobą, a dość znacząco od sygnałów komunikacyjnych innych gatunków. Sygnały komunikacyjne pozostałych gatunków cechowały się dużą różnorodnością w obrębie gatunku i były dość podobne w kategoriach obliczanych parametrów do sygnałów innych gatunków, co mogło powodować wybieranie nieprawidłowych wzorców. Płynie stąd wniosek, że w kolejnych badaniach należałoby podzielić zbiór sygnałów dla gatunku na mniejsze grupy, w których sygnały nie będą tak bardzo różniły się od siebie pod kątem określonego parametru. Jednym ze sposobów dzielenia sygnałów na grupy jest klasteryzacja np. za pomocą algorytmu *k*-średnich.

Z przebadanych parametrów najlepszym okazała się długość sygnału, co należy uznać za dobry rezultat ze względu na to, że długość sygnału jest bardzo prostym do obliczenia parametrem sygnału. Pozostałe parametry, jak odcięta środka ciężkości kwadratu sygnału oraz odcięta środka ciężkości kwadratu modułu średniego FFT wykazywały gorsze wskaźniki jakości. Przyczyną mogły być stosunkowo duże szumy otoczenia – średni SNR był na poziomie około 11,33 dB.

Z wyników otrzymanych w twardej metodzie rozkładu normalnego oraz analizując histogram badanych parametrów dla poszczególnych gatunków można wyciągnąć wniosek, że głównym czynnikiem powodującym w niektórych przypadkach nieskuteczność preselekcji wzorców są fluktuacje przebiegu histogramu. Przykładowo, dla gatunku kos w przypadku parametru długość obserwuje się dość znaczną liczbę skupionych sygnałów komunikacyjnych o długości większej od $\mu + 2\sigma$. W związku z tym polepszenie efektywności metody można by osiągnąć poprzez klasteryzację histogramu, w wyniku czego możliwe byłoby uzyskanie wielu rozkładów normalnych w poszczególnych elementarnych klastrach.

W metodzie miękkiej rozkładu normalnego satysfakcjonujące wyniki parametrów analizy ROC nie przełożyły się w oczekiwanym stopniu na końcową jakość preselekcji. Głównym problemem, który wpływa na niską efektywność eliminacji nieprawidłowych wzorców (wskaźnik *Og*), jest konieczność zapewnienia bardzo wysokiej skuteczności (*SP*), wymaganej dla prawidłowego i skutecznego przeprowadzenia drugiego etapu, czyli rozpoznawania. W metodzie miękkiej rozkładu normalnego polepszenie klasyfikacji może zostać wykonane poprzez dodanie nowych parametrów sygnału, gdyż, jak zaobserwowano podczas wykony-wania badań, wprowadzanie każdego kolejnego parametru sygnału do wyznaczenia globalnej gęstości prawdopodobieństwa poprawiało wyniki. Wartość AUC zwiększyła się z wartości 0.6314 dla parametru długość sygnału do wartości 0,8169 dla wszystkich parametrów sygnału. Ponadto w dalszych badaniach należy poddać ocenie bardziej złożoną obliczeniowo metodę One-Versus-One wyznaczania krzywych ROC.

Z trzech prezentowanych metod preselekcji najlepsza wydaje się być metoda przedziałowa. Nie dała ona co prawda 100% skuteczności preselekcji w każdym przypadku, ale skutecznie ograniczała grupę wzorców. Ponadto w tej metodzie skuteczność preselekcji można łatwo poprawić poprzez uczenie się systemu, polegające na dodawaniu do zbioru wzorców sygnałów niepoprawnie odrzuconych na etapie preselekcji oraz odpowiednie przeliczenie progów wd i wg. W twardej i miękkiej metodzie rozkładu normalnego poprawa skuteczności preselekcji prowadzi jednocześnie do zwiększania się parametru ograniczanie (Og), co nie jest z punktu widzenia dalszego rozpoznawania korzystne.

BIBLIOGRAFIA

1. Bibby C. J., Burgess N. D., Hill D. A., Mustoe S. H.: Bird census techniques. Academic Press, London, 2000.
2. Revathi A., Ganapathy R., Venkataramani Y.: Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach. International Journal of Computer science & Information Technology (IJCSIT), Vol. 1, No. 2, November 2009.
3. Chou Ch. H., Lee Ch. H., Ni H. W.: Bird Species Recognition by Comparing the HMMs of the Syllables, Second International Conference on Innovative Computing, Information and Control (ICICIC 2007), 2007, s.143.
4. Wielgat R., Zieliński T. P., Potempa T., Lisowska-Lis A., Król D.: HFCC Based Recognition of Bird Species. Miesięcznik naukowo-techniczny Elektronika - konstrukcje, technologie, zastosowania, zeszyt XLIX, nr 4/2008.
5. Zieliński T.P.: Cyfrowe przetwarzanie sygnałów – Od teorii do zastosowań, WKŁ, Warszawa, 2005.
6. Fawcett T.: An introduction to ROC analysis. Pattern Recognition Letters 27, Elsevier, 2006, s. 861÷874.
7. Provost F., Domingos P.: Well-trained PETs: Improving probability estimation trees. CeDER Working Paper #IS-00-04, Stern School of Business, New York University, NY, 2001.

Recenzenci: Dr inż. Adam Duszeńko
Dr inż. Michał Kawulok

Wpłynęło do Redakcji 31 stycznia 2010 r.

Abstract

Method for improving audio signals recognition using multimedia database is presented in the paper. Recognition using signal pattern matching by dynamic time warping (DTW) or whole-word hidden Markov model (HMM) was assumed. Large number of signal patterns prolongs recognition time and may decrease recognition accuracy. Initial preselection of signal patterns based on chosen signal parameters, being the process of decreasing the number of patterns can alleviate problem.

The signals analyzed in the paper were bird voices. There were four signal parameters examined: length of signal, x-coordinate of the signal square center of mass (1), x coordinate of the signal FFT square center of mass, signal to noise ratio (2). Three preselection methods were tested: interval methods (section 2.2.1), hard normal distribution method (section 2.2.2) and soft normal distribution method (2.2.3). In order to compare the effectiveness of the preselection methods for the parameters, several coefficients resulting from ROC analysis were calculated. These parameters are: precision (10), sensitivity (11), specificity (12), accuracy (13), F-measure (14). Moreover there was proposed new coefficients: limitation (15), preselection accuracy (16) and correctly chosen class coefficient (17). The best preselection method seems to be the interval method. The best preselection signal parameter was length of signal. Further research involving signal clusterization is required.

Adresy

Tomasz POTEMPA: Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Informatyki, ul. Mickiewicza 8, 33-100 Tarnów, Polska, tpotempa@pwsztar.edu.pl .

Robert WIELGAT: Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Elektroniki i Telekomunikacji, ul. Mickiewicza 8, 33-100 Tarnów, Polska, rwielgat@poczta.onet.pl .

Daniel KRÓL: Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Informatyki, ul. Mickiewicza 8, 33-100 Tarnów, Polska, danielkrol@poczta.onet.pl .

Paweł KOZIOŁ: Regionalna Dyrekcja Ochrony Środowiska w Krakowie, Wydział Spraw Terenowych w Tarnowie, al. Solidarności 5-9, 33-100 Tarnów, Polska, pawel.koziol@rdos.krakow.pl .

Agnieszka LISOWSKA-LIS: Państwowa Wyższa Szkoła Zawodowa w Tarnowie, Zakład Elektrotechniki, ul. Mickiewicza 8, 33-100 Tarnów, Polska, lisowskalis@pwsztar.edu.pl .