Instytut Podstawowych Problemów Techniki
Polskiej Akademii Nauk

# Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato

mgr Krzysztof Tyburek

Rozprawa doktorska napisana pod kierunkiem
prof. dr. hab. Witolda Kosińskiego

Warszawa, Listopad 2006

## Podziękowania

Pragnę złożyć serdeczne podziękowania wszystkim osobom, które swoimi cennymi uwagami przyczyniły się do realizacji niniejszej rozprawy. Szczególnie pragnę podziękować promotorowi **prof. Witoldowi Kosińskiemu** za opiekę naukową. Serdecznie dziękuję **dr Alicji Wieczorkowskiej** oraz **dr Waldemarowi Cudnemu** za poświęcony czas i bardzo cenne uwagi, które w znacznym stopniu pomogły mi zrealizować badania zawarte w niniejszej rozprawie.

*Pracę tę dedykuję wszystkim moim najbliższym, a w szczególności moim Rodzicom.*

*Krzysztof Tyburek*

# Spis treści

# Wstęp

Większość dotychczasowych rozwiązań związanych z wydobywaniem wiedzy bazuje na technice etykietowania przechowywanych informacji — do pewnego czasu technika ta dotyczyła również danych opisujących dźwięk. Cała procedura etykietowania jest dość pracochłonna i czasochłonna. Poza tym takie rozwiązanie nie zawsze daje rzetelny wynik — to znaczy wysyłane zapytanie nie zawsze jest zgodne z oczekiwaniami osoby (czy systemu) pytającej. Istnieje wielkie prawdopodobieństwo, że dwie zupełnie różne (binarnie) informacje dźwiękowe mogą się okazać tą samą sekwencją utworu muzycznego zagranego z różną dynamiką lub w pomieszczeniu o różnej akustyce. Kolejny problem, który występuje w procesie rozpoznawania sygnałów dźwiękowych, jest właściwa interpretacja źródła dźwięku. Rozpoznanie dźwięku pochodzącego na przykład z drgającej struny gitary może być bardzo trudne. Trudność ta najczęściej wynika z doskonałych procesorów muzycznych, za pomocą których z łatwością można "podrobić" oryginalny instrument. W obliczu pojawiających się problemów związanych z możliwością wyszukiwania informacji audio najistotniejszym zagadnieniem jest opracowanie stosownych algorytmów wyszukujących właściwy system kodowania informacji multimedialnych oraz klasyfikujący typ źródła dźwięku (na przykład instrumenty strunowe, dęte drewniane i tym podobne).

Drogą do rozwiązania problemu klasyfikacji i agregacji danych multimedialnych jest nowo powstały (posiadający certyfikat ISO) standard MPEG-7, który dostarcza szereg podstawowych deskryptorów opisujących dźwięk. Na bazie standardu MPEG 7 stworzono nowe deskryptory rozpoznające pewne instrumenty muzyczne. W części audio tego standardu jest 17 deskryptorów (widmowych i czasowych). MPEG-7 zawiera warstwę nośną niskiego poziomu narzędzi, które stosowane są ogólnie we wszelkich dźwiękach. Mechanizmy te ustalają podstawowy poziom kompatybilności wśród opisów audio i pozwalają tworzyć nowe aplikacje. Warstwa ta również integruje MPEG-7 z innymi częściami standardu. Na warstwie nośnej zbudowana jest też seria narzędzi wysokiego poziomu, które są dostosowywane do poszczególnych grup aplikacji.

Zadaniem tych grup jest przeszukiwanie i wyszukiwanie cech sygnałów. Usługa audio realizowana jest z wykorzystaniem bazy danych skompresowanego MPEG-4 znajdującego się na jednym (lub więcej) serwerze medialnym oraz skojarzoną z nim asocjacyjną bazą metadanych MPEG-7 na serwerze kwerendowym. Dla każdego indeksowanego skompresowanego sygnału audio MPEG-4, baza metadanych MPEG-7 przechowuje pełną reprezentację melodii oraz mechanizm łączący metadane ze skojarzonym nośnikiem, na przykład adres URL. Baza danych może również przechowywać inne deskryptory opisujące, np. styl i gatunek muzyki oraz bazę dźwięków instrumentów w charakterystycznym pasażu. Wykorzystując opis sygnału za pośrednictwem deskryptorów istnieje możliwość wysłania kwerendy w formie fonicznej. Kształt fali, sygnału próbnego jest przenoszony na serwer zapytań (kwerend), gdzie

przeprowadzany jest proces, w którym uzyskiwane są jego metadane, które są celem kwerendy. Serwer MPEG-7 wyszukuje zatem odpowiedników melodii z baz danych. Kilka najlepszych odpowiedników przenoszonych jest z powrotem do urządzenia [intr1].

MPEG 7 dostarcza 17 podstawowych deskryptorów sklasyfikowanych w poszczególnych klasach:

- Basic,
- Basic Spectral,
- Signal Parameters,
- Timbral Temporal,
- Timbral Spectral,
- Spectral Basis,
- Silence Descriptor.

Najogólniej deskryptory można podzielić na dwa podtypy:

1. deskryptory funkcji widma,
2. deskryptory funkcji czasu.

W swoich pracach, związanych z automatyczną klasyfikacją instrumentów muzycznych, autorzy (na przykład Xavier Serra [6] [5], Pedro Cano [4], José M. Martínez [5]) często ograniczają się między innymi do instrumentów dętych (zarówno blaszanych jak i drewnianych), niektórych perkusyjnych oraz elementów wokalizy.

Celem, jaki przyświecał autorowi niniejszej rozprawy, było dążenie do utworzenia mechanizmów pozwalających na rozpoznanie źródła dźwięku ze szczególnym uwzględnienie instrumentów strunowych, które nie są wyczerpująco opisane w pracach naukowych. Postacie czasowe przebiegów instrumentów strunowych (chordofonów) charakteryzują się brakiem stanu quasi-ustalonego oraz bardzo krótkim transjentem początkowym. Oznacza to, że proces parametryzacji może odbywać się tylko z wykorzystaniem transjentu końcowego, co jest znacznym utrudnieniem podczas procesu ekstrakcji cech (zagadnienie to szerzej opisano w rozdziale 7). Rozwiązanie problemu klasyfikacji źródła dźwięku pozwoli na skuteczne przeszukiwanie multimedialnych baz danych, w których kwerendy będą przyjmowały postać foniczną, a źródłem dźwięku jest instrument strunowy z artykulacją pizzicato. Bardzo istotnym powodem, dla których wszczęto poszukiwania nowych deskryptorów treści multimedialnych jest aspekt praktyczny. Algorytmy przeszukiwania zasobów multimediów powinny zostać zaimplementowane miedzy innymi w aplikacjach wspomagających pracę w studiach nagraniowych, studiach radiowych i telewizyjnych oraz w studiach produkcyjnych dźwiękowe materiały reklamowe. Założono, że podczas prowadzonych badań uda się odszukać taki wektor cech, który w swojej treści będzie zawierał tylko deskryptory wynikające z analizy postaci widmowej badanego przebiegu (swoje obawy związane z zasadnością wykorzystania deskryptorów postaci czasowej autor przedstawił w 7.3.1).

Pozwoliło to sformułować tezy pracy:

**TEZA 1** : Istnieje taki wektor cech, który pozwoli na skuteczne rozpoznanie przebiegów dźwiękowych instrumentów strunowych z artykulacją pizzicato.

**TEZA 2** : Nowo zaproponowany wektor cech zawiera deskryptory wynikające tylko z analizy przestrzeni widmowej badanych próbek dźwięków.

Dla osiągnięcia tez wykonano niżej wymienione czynności.

W rozdziale 1 wprowadzono do zagadnień związanych z ogólnym pojęciem fal, pojęciem fal harmonicznych, podłużnych i poprzecznych. Zaakcentowano takie zjawiska jak nakładanie się i interferencja fal, odbicie i załamanie fali. Wyjaśniono zagadnienia dotyczące pola dźwiękowego z uwzględnieniem amplitudy, częstotliwości, wysokości dźwięku, widma dźwięku, długości fal dźwiękowych, natężenia fali dźwiękowej, dudnienia oraz zniekształceń dźwięku. Ponadto zdefiniowano i opisano sygnał cyfrowy z uwzględnieniem definicji próbkowania, rozdzielczości kodowania, kwantyzacji oraz zaszumienia sygnału.

W rozdziale 2 opisano podstawowe mechanizmy wykorzystywane w trakcie procesu ekstrakcji cech przebiegów cyfrowych (FFT, DFT oraz analizę falkową). Ponadto wprowadzono do zagadnienia przecieku widma, które jest obecne podczas analizy badanych próbek dźwięków. Poza tym przedstawiono metodę okien czasowych, powodujących redukcję przecieku częstotliwości przez zminimalizowanie listków bocznych bez konieczności poszerzania okna.

W rozdziale 3 przedstawiono ogólnie przyjętą klasyfikację instrumentów muzycznych, w której podstawowym kryterium jest określenie źródła dźwięku. Wyeksponowano przyjęty przez Carla Sachsa 5-klasowy podział instrumentów.

W rozdziale 4 przeprowadzono przegląd metod analizy przebiegów dźwiękowych. Opisywane w tym rozdziale metody są ogólnie przyjęte przez autorów prac związanych z parametryzacją dźwięków. Przedstawiono metody analizy postaci czasowej oraz postaci widmowej przebiegu. Deskryptory uzyskane przy pomocy tych metod, w niniejszej rozprawie zostały potraktowane jako klasyczne deskryptory opisu przebiegów muzycznych

Rozdział 5 jest poświęcony wprowadzeniu w tematykę baz danych, uwzględniając zarówno ogólne pojęcia związane z realizacją transakcji, jak i pojęcia związane z definicją systemu zarządzania bazą danych, modelem relacyjnym, obiektowym oraz relacyjno-obiektowym. Temat związany z bazami danych jest nierozerwalny z realizacją niniejszej rozprawy. Przyjęto, że wynikiem realizacji badań jest zbiór deskryptorów, tworzący taki wektor cech, który pozwoli realizować zapytania foniczne wytypowane do multimedialnych baz danych. Efekt pracy związanej z niniejszą rozprawą powinien przyczynić się do optymalizacji zapytań fonicznych, a tym samym do poprawy funkcjonowania multimedialnych systemów baz danych.

W rozdziale 6 przedstawiono mechanizm działania podstawowych algorytmów klasyfikujących wykorzystywanych w procesie automatycznej klasyfikacji instrumentów muzycznych. Wprowadzono do zagadnień związanych z metodą reprezentacji obiektów, uwzględniając metodę wektorową i strukturalną. Poza tym opisano metody eksperymentalne takie jak metoda holdout oraz metoda k-krotnej walidacji krzyżowej. Omówiono interpretację macierzy przekłamań, jako podstawowego elementu interpretacji wyniku klasyfikacji obiektów. Następnie uwzględniono zagadnienia związane z procesem selekcji cech oraz normalizacji atrybutów.

W rozdziale 7 opisano procedurę przygotowania danych eksperymentalnych oraz przyjętą metodologię badań. W rozdziale tym uwzględniono fizyczne cechy badanych

próbek dźwięków oraz fragmenty widma przeznaczone do analizy. Przedstawiono przykładowe wyniki klasyfikacji z uwzględnieniem tradycyjnych deskryptorów. Ponadto porównano klasyfikację 8 klas instrumentów z wykorzystaniem analizy pełnej reprezentacji widma oraz widma składowych harmonicznych. Poza tym, opisano zakres częstotliwości próbek przeznaczonych do eksperymentów (wraz z przynależnością do określonych oktaw) oraz klasy instrumentów przeznaczone do badań.

W rozdziale 8 przedstawiono propozycję parametryzacji widma z uwzględnieniem określonego obszaru. Zaproponowano podział fragmentu widma ze względu na rozkład częstotliwościowy oraz energetyczny. Opisano wpływ ilości i szerokości warstw na ogólny wynik klasyfikacji. Ponadto wprowadzono metodę siatki, jako metodę analizy wybranych przestrzeni fragmentu widma. Przedstawiono wyniki klasyfikacji z uwzględnieniem przykładowych algorytmów klasyfikujących z uwzględnieniem metod opisanych w rozdziale 6.

W rozdziale 9 opisano rozwiązanie problemu optymalizacji szerokości warstw. Poza tym, zaakcentowano wpływ doboru ilości kolumn (rozkładu częstotliwościowego) na ogólny wynik klasyfikacji. Zaproponowano podział 4- i 7-warstwowy, oraz podział 10- i 8-kolumnowy do celów poprawy skuteczności klasyfikacji. Ostatecznie zaproponowano 2 wektory cech (28- i 12-elementowy) uwzględniające deskryptory pozyskane w wyniku tylko analizy przestrzeni widmowej badanych próbek dźwięku. Wskazano na skuteczną klasyfikację 8 klas instrumentów strunowych, 12-elementowego wektora cech.

# Spis oznaczeń

**Rozdział 1**

- $\nu$ — prędkość fali
- $A$ — amplituda
- $T$ — okres fali
- $\lambda$ — długość fali
- $\omega$ — pulsacja
- $p_0$ — ciśnienie powietrza w obecności fali akustycznej
- $\varphi$ — gęstość powietrza
- $f$ — częstotliwość drgań [Hz]
- $c$ — prędkość rozchodzenia się fali [m/s]
- $I$ — natężenie fali dźwiękowej [W/m$^2$]
- $t$ — czas [s]
- $S$ — pole powierzchni, na którą pada energia dźwiękowa [$m^2$]
- $P$ — moc fali dźwiękowej
- $W$ — energia niesiona przez fale
- $B$ — poziom natężenia dźwięku [dB]
- $f_s$ — częstotliwość próbkowania
- $\tau$ — przesunięcie czasowe

**Rozdział 2**

- $x(t)$ — sygnał ciągły w dziedzinie czasu
- $T$ — okres próbkowania
- $f_m$ — częstotliwość dla prążka o numerze $m$
- $N$ — liczebność próby (liczba próbek poddanych DFT lub FFT)
- $f_r$ — rozdzielczość widma
- $s(n)$ — sygnał wejściowy
- $v(n)$ — sygnał otrzymany w wyniku okienkowania
- $w(n)$ — funkcja okna
- $g_{b,a}(t)$ — funkcja analizująca
- $a$ — współczynnik rozszerzenia
- $b$ — parametr przesunięcia czasowego

**Rozdział 3**

$d$ — długość struny

$T$ — siła naciągu struny

$P$ — gęstość materiału

**Rozdział 4**

$l_{tn}$ — logarytm czasu narastania dźwięku

$t_{max}$ — czas osiągnięcia maksymalnej amplitudy dźwięku

$t_{pp}$ — czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie początkowym

$l_{tk}$ — logarytm czasu wybrzmiewania dźwięku

$t_{pk}$ — czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie końcowym

$ZC$ — gęstość przejść przez zero sygnału (zero crossings)

$Br$ — środek ciężkości widma

$A(i)$ — amplituda $i$-tej składowej

$m_k$ — moment widmowy k-tego rzędu

$L_n$ — średnia arytmetyczna poziomu amplitudy $n$-tej składowej dla M ramek

$\Delta\%$ — suma procentowych zmian częstotliwości n-tej harmonicznej w stosunku do częstotliwości podstawowej

$Ev$ — zawartości składowych parzystych widma

$Od$ — zawartości składowych nieparzystych widma

$Tr1$, $Tr2$, $Tr3$ — parametry tristimulus

**Rozdział 5**

$PK$ — klucz główny

$FK$ — klucz obcy

$A$ oraz $B$ — wejściowe zbiory lub wielozbiorów

$R$ — oznaczenie wyjściowego zbioru lub wielozbioru

$a$ — element wejściowy zbioru lub wielozbioru $A$

$f$, $g$ oraz $h$ — reprezentacja funkcji

$id$ — funkcja tożsamościowa

$p$ — predykat

$T$ — typ wynikowy operatora

$\langle\ \rangle$ — oznaczenie krotek

$L$ — nazwa pola krotki

$a/L$ — wartość krotki a minus pole oznaczone L

**Rozdział 6**

$Ir$ — nieregularność widma

$c$ — ilość klas

$U_i$ — zbiór wektorów zbioru uczącego $?U$

$N_i$ — liczebność zbioru $U_i$

$g_i$ — funkcja dyskryminacyjna klasyfikatora minimalno-odległościowego

$\psi_{mo}$ — minimalno-odległościowy algorytm klasyfikacji

$\mathbf{x}$ — wektor cech

$\hat{P}(j)$, $\hat{f}(x|j)$ — estymatory wielkości

$N_i$ — ilość wektorów z klasy $i$ w zbiorze uczącym

$R$ — obszar zainteresowań klasyfikatora

$v$ — objętość obszaru $R$

$k$ — $k$ najbliższych sąsiadów wektora $\mathbf{x}$ (zbioru uczącego)

$k_i$ — liczba wektorów pośród $k$ najbliższych sąsiadów wektora $\mathbf{x}$

$TD$ — tablica decyzyjna

$C$ — atrybuty warunkowe tablicy decyzyjnej

$D$ — atrybuty decyzyjne tablicy decyzyjnej

$f$ — funkcja decyzyjna tablicy decyzyjnej

$A$ — niepusty skończony zbiór atrybutów tablicy decyzyjnej

$U$ — niepusty zbiór (uniwersum) tablicy decyzyjnej

$SI$ — system informacyjny

$\overline{B}X$ — B-górne przybliżenie

$\underline{B}X$ — B-dolne przybliżenie

$POS_B(X)$ — B-pozytywny obszar zbioru $X$

$BN_B(X)$ — B-brzeg zbioru $X$

$NEG_B(X)$ — B-negatywny obszar zbioru $X$

$n_{\text{blad}}$ — liczba błędnie sklasyfikowanych przykładów testowych

$n_{\text{ltest}}$ — liczba przykładów testowych

$n_{\text{ppr}}$ — liczba poprawnie sklasyfikowanych przykładów testowych

$x_{ij}$ — wartość $j$-tej współrzędnej $i$-tego wektora cech po normalizacji

$x^*_{ij}$ — przed normalizacją

$śr_i$ — średnia wartość $i$-tej cechy

ROZDZIAŁ 1

# Fale i ruch falowy

## 1.1. Ogólne pojęcie fali

Fala to zaburzenie stanu ośrodka (pola) rozchodzące się w przestrzeni ze skończoną, charakterystyczną dla danego skupienia prędkością i niosące ze sobą pewną energię. W zależności od rodzaju fal zaburzenie to będzie miało swoja charakterystykę. Na przykład dla fali akustycznej zaburzeniem będą drgania powietrza i związane z nimi lokalne zmiany jego ciśnienia. Dla fal elektromagnetycznych jako zaburzenie traktować należy zmiany wektorów natężenia pola elektrycznego $\boldsymbol{E}$ i indukcji pola magnetycznego, które umożliwiają przepływ energii na duże odległości bez przepływu masy. W ośrodkach zupełnie jednorodnych fale rozchodzą się prostoliniowo ze stałą prędkością. Kierunek rozchodzenia się fali nazywany jest *promieniem*. Prędkość fali definiowana jest jako stosunek odległości $s$, o jaką przesunie się zaburzenie wzdłuż promienia, do czasu $t$, w którym to przesunięcie następuje:

$$v = \frac{s}{t}\,. \tag{1.1}$$

### 1.1.1. Fale harmoniczne

Falą harmoniczną nazywamy taką falę, dla której zaburzenie w każdym punkcie ośrodka jest zaburzeniem harmonicznym, tzn. zależy sinosoidalnie od czasu:

$$\psi(x,t) = A\sin[k(x \pm vt)] \tag{1.2}$$

lub

$$\psi(x,t) = A\cos[k(x \pm vt)]\,, \tag{1.3}$$

gdzie:

$A$ — amplituda,

$k$ — liczba falowa.

Liczba falowa jest bezpośrednio związana długością fali $\lambda$ następującą zależnością:

$$k = \frac{2\pi}{\lambda}\,. \tag{1.4}$$

Podstawiając do funkcji falowej $x = 0$ otrzymamy opis ruchu drgań harmonicznych w początku układu współrzędnych:

$$\psi(t) = A\sin k(\pm vt)] = A\sin(\pm kvt) = \pm A\sin kvt\,. \tag{1.5}$$

Odczytując ze wzoru (1.5) iloczyn $kv$ jako częstość kołową otrzymamy:

$$k = \frac{\omega}{v} = \frac{2\pi}{Tv}, \tag{1.6}$$

gdzie:

$$\omega = \frac{2\pi}{\lambda} v. \tag{1.7}$$

Przyjmując w (1.2) $t = 0$ otrzymamy przestrzenny kształt zaburzenia w chwili początkowej

$$x = A \sin kx = A \sin \frac{2\pi}{Tv} x, \tag{1.8}$$

który jest harmoniczną funkcją zmiennej $x$.

Wykorzystując powyższe oznaczenia można nadać funkcji opisującej falę harmoniczną następujące postacie, [17]:

$$\psi(x,t) = \begin{cases} A \sin k(x \pm vt) \\ A \sin 2\pi(\frac{x}{\lambda} \pm \frac{t}{T}) \\ A \sin 2\pi(k' x \pm \frac{t}{T}) \\ A \sin(kx \pm \omega t) \\ A \sin 2\pi \frac{t}{T}(\frac{x}{v} \pm t) \\ A \sin \omega(\frac{x}{v} \pm t) \end{cases} \tag{1.9}$$

Funkcję falową możemy też przedstawić w postaci zespolonej

$$\psi = A \mathrm{e}^{j(kx \pm \omega t)} \tag{1.10}$$

lub w bardziej ogólnej postaci

$$\psi = A \mathrm{e}^{j(kx \pm \omega t + \varphi)}, \tag{1.11}$$

gdzie $\varphi$ jest stałą fazową, zaś $j$ jednostką urojoną.

### 1.1.2. Fale podłużne i poprzeczne

W czasie rozchodzenia się fali mechanicznej cząstki ośrodka wykonują drgania względem swych położeń równowagi. Jeżeli kierunek tych drgań odbywa się w tym samym kierunku, w jakim rozchodzi się fala, to falę tą nazywamy *falą podłużną*. Jeżeli natomiast kierunek drgań odbywa się prostopadle do kierunku rozchodzenia się fali, to nazywamy ja *fala poprzeczną*. W literaturze spotyka się również terminologię — *fala dylatacyjna* i *fala skrętna*. Fale sprężyste w płynach są najczęściej falami podłużnymi. Fale poprzeczne występują głównie (obok fal podłużnych) w ciałach stałych, ale mogą też istnieć w lepkich cieczach.

### 1.1.3. Prędkość fazowa fali

Argument harmonicznej funkcji falowej

$$\psi(x,t) = A \sin[k(x - v \cdot t)] = A \cdot \sin(kx - \omega t) \tag{1.12}$$

nazywa się *fazą fali* i opisywany jest zależnością:

$$\varphi = kx - \omega t\,. \tag{1.13}$$

W przypadku ogólnym może istnieć faza początkowa $\varphi$ różna od zera, wyrażona następująco:

$$\varphi = kx - \omega t + \varphi_0\,. \tag{1.14}$$

Pochodna fazy względem czasu opisuje *częstość kołową* fali z dokładnością do znaku

$$\frac{\partial \varphi}{\partial t} = -\omega\,, \tag{1.15}$$

a względem położenia—*liczbę falową*:

$$\frac{\partial \varphi}{\partial x} = k = \frac{2\pi}{\lambda}\,. \tag{1.16}$$

Prędkość fali jest określona stosunkiem częstości kołowej fali i liczby falowej:

$$-\frac{\frac{\partial \varphi}{\partial t}}{\frac{\partial \varphi}{\partial x}} = -\frac{\partial x}{\partial t} = \frac{\omega}{k} = \frac{2\pi\lambda}{T2\pi} = \frac{\lambda}{T} = v\,. \tag{1.17}$$

A zatem podstawowy wzór na prędkość fali harmonicznej ma postać:

$$v = \frac{\lambda}{T}\,, \tag{1.18}$$

gdzie:

$v$ — prędkość fali,
$T$ — okres fali,
$\lambda$ — długość fali.

### 1.1.4. Nakładanie się i interferencja fal

Interferencja jest to zjawisko nakładania się dwu lub więcej fal harmonicznych o tej samej długości i zbliżonej amplitudzie, prowadzące do powstania ustalonego w czasie przestrzennego rozkładu obszarów wzmocnienia i osłabienia fali. Warunkiem otrzymania interferencji jest spójność źródeł fal. Wynikiem interferencji jest pojawienie się naprzemian leżących obszarów wzmocnień i osłabień, obszary takie nazywa się *prążkami interferencyjnymi*. Przykład fali jednowymiarowej, biegnącej w kierunku osi $x$

$$\psi_1(x,t) = A\sin(kx - \omega t) \tag{1.19}$$

oraz takiej samej fali biegnącej w kierunku przeciwnym

$$\psi_2(x,t) = A\sin(kx + \omega t) \tag{1.20}$$

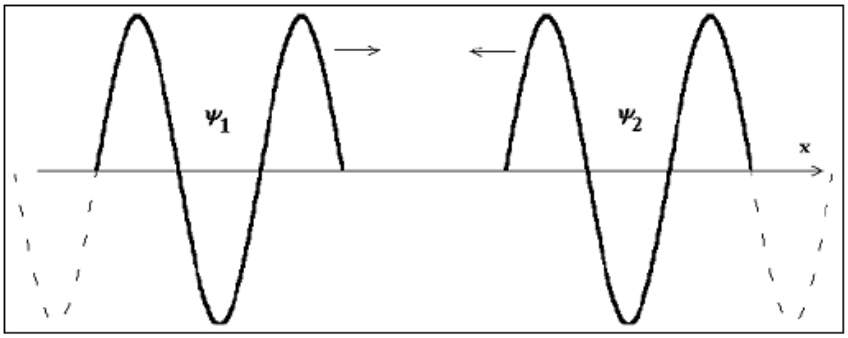
pokazano na Rys. 1.1.

W wyniku nałożenia się (sumowania) fal $\psi_1$ i $\psi_2$ otrzymamy falę wypadkową:

$$\psi = \psi_1 + \psi_2 = A(\sin(kx - \omega t) + \sin(kx + \omega t))\,. \tag{1.21}$$

Korzystając ze wzoru na sumę sinusów otrzymamy:

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}\,. \tag{1.22}$$




Rysunek 1.1. Dwie fale ($\psi_1$ i $\psi_2$) biegnące w przeciwnych kierunkach

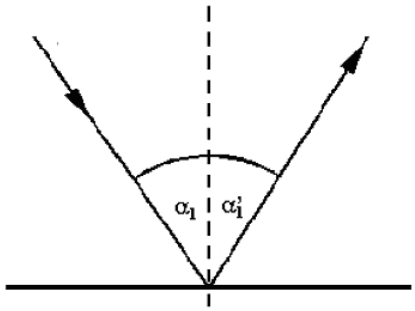
uzyskując ostatecznie:

$$\psi(x,t) = 2A\sin(kx)\cos(\omega t)\,. \tag{1.23}$$

Amplituda drgań wypadkowych zależy od położenia $x$ (inaczej niż dla fali biegnącej) i jej moduł zmienia się od zera do $2A$. Drgania występujące w różnych miejscach różnią się fazą o $\pi$ lub nie różnią się w ogóle. Tego rodzaju ruch w ośrodku nazywa się *falą stojącą*. Kwadrat amplitudy fali stojącej jest proporcjonalny do energii drgań elementów ośrodka. Miejsca zerowe $\sin(kx) = 0$, o zerowej amplitudzie drgań nazywają się *węzłami fali stojącej*. Miejsca ekstremów $\sin(kx) \pm 1$, o maksymalnej amplitudzie drgań, nazywają się *strzałkami fali stojącej*.

### 1.1.5. Odbicie i załamanie fal

Zjawisko załamania lub odbicia fal występuje wówczas, gdy fala dociera do granicy dwóch ośrodków. Kąt padania zdefiniowany jest jako kąt zawarty pomiędzy prostą prostopadłą do granicy ośrodków wyprowadzoną w punkcie, w którym do granicy dociera promień fali a tym promieniem, czyli kierunkiem rozchodzenia się fali. Kąt odbicia jest to kąt zawarty pomiędzy prostą prostopadłą do granicy ośrodków wyprowadzoną w punkcie, w którym do granicy dociera promień fali a kierunkiem rozchodzenia się fali odbitej. Prawo odbicia mówi, że kąt padania, kąt odbicia i prosta prostopadła do granicy ośrodków leżą w jednej płaszczyźnie oraz kąt padania jest równy kątowi odbicia $\alpha_1 = \alpha_1^{'}$. Zjawisko odbicia fali od granicy ośrodka pokazano na Rys. 1.2.




Rysunek 1.2. Odbicie fali od granicy ośrodka

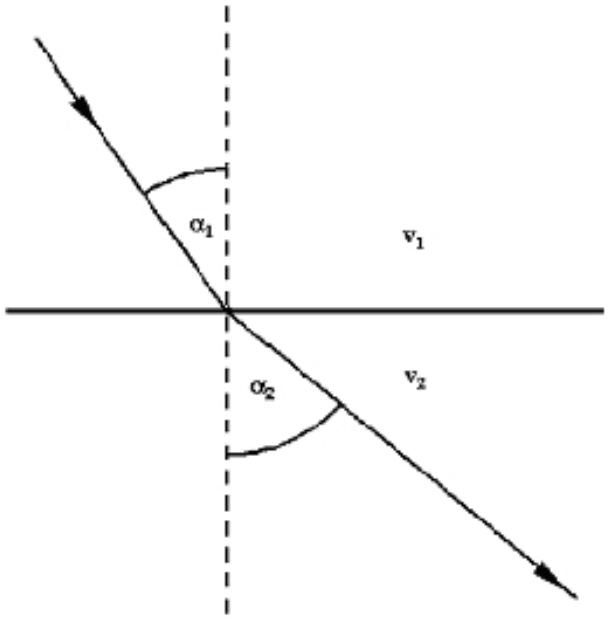
Prawo załamania mówi nam, że stosunek sinusa kąta padania do sinusa kąta załamania jest dla danych dwóch ośrodków wielkością stała i równą stosunkowi prędkości fal w tych ośrodkach.

$$\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{v_1}{v_2} = n_{12}\,, \tag{1.24}$$

gdzie:

$v_1, v_2$ — prędkość fal w ośrodkach, na granicy których występuje załamanie,
$n_{12}$ — współczynnik załamania fali.

Oba te kąty i normalna do powierzchni rozdziału leżą w jednej płaszczyźnie. Przy przejściu do innego ośrodka zmienia się prędkość fali i jej długość, nie zmienia się natomiast częstotliwość fali. Zjawisko załamania fali na granicy dwóch środków pokazano na Rys. 1.3.




Rysunek 1.3. Załamanie fali na granicy dwóch środków

## 1.2. Pole dźwiękowe

Pole akustyczne (dźwiękowe) jest to obszar wypełniony falami akustycznymi. Zakładając, że są to fale płaskie, podłużne, biegnące z szybkością $u$ w kierunku osi $x$, to wychylenie cząstki powietrza w chwili $t$ w punkcie $x$ jest opisane zależnością:

$$\xi = A\sin\omega\left(t - \frac{x}{u}\right) = A\sin 2\pi\gamma\left(t - \frac{x}{u}\right)\,, \tag{1.25}$$

gdzie:

$\omega$ — pulsacja,
$\gamma$ — częstotliwość,
$A$ — amplituda.

Różniczkując równanie (1.25) otrzymamy zależność opisującą szybkość drgań

$$v = v_0 \cos\omega\left(t - \frac{x}{u}\right), \tag{1.26}$$

gdzie: $v_0 = A\omega$ — amplituda szybkości drgań.

W przypadku dźwięku, rozchodzącego się w powietrzu, czy wodzie, występują zmiany (zaburzenia) gęstości i ciśnienia ośrodka. Zmiana ciśnienia wywołana drganiami cząsteczek, w przybliżeniu liniowym opisana jest wzorem:

$$p = p_0 + A\varphi\omega u \cdot \cos\omega\left(t - \frac{x}{u}\right), \tag{1.27}$$

gdzie:

$p_0$ — ciśnienie powietrza w obecności fali akustycznej,

$\varphi$ — gęstość powietrza.

Amplituda zmian ciśnienia jest wyrażana współczynnikiem funkcji kosinus:

$$\Delta p_0 = A\omega\varphi u = v_0\varphi u\,. \tag{1.28}$$

### 1.2.1. Amplituda

Związana jest maksymalnym wychyleniem fali, które ma związek z głośnością. Amplituda jest, zatem maksymalnym modułem zaburzenia. Ruch okresowy, jest z natury rzeczy ograniczony i zmienna go opisująca zawiera się między wartościami skrajnymi $q_{\min} \leq q \leq q_{\max}$. Biorąc pod uwagę, że ruch odbywa się symetrycznie, to istnieje określona wartość $q$, której odpowiada środek symetrii. W środku symetrii występuje $q = 0$, a wartości skraje są sobie równe i mają przeciwne znaki:

$$\begin{aligned} q_{\min} &= -q_m\,, \\ q_{\max} &= q_m\,. \end{aligned} \tag{1.29}$$

Maksymalną wartość $q_m$ okresowo zmieniającej się wartości $q$ nazywamy jej amplitudą. Możemy zapisać, że $-q_m \leq q \leq q_m$, a zatem

$$|q| \leq q_m\,. \tag{1.30}$$

Dla przykładu należy zaznaczyć, że dźwięki w zakresie 2 kHz do 4 kHz, są odbierane najsilniej przez ludzkie ucho, ale granica ta zmienia się z wiekiem.

### 1.2.2. Częstotliwość

Częstotliwości jest liczbą zdarzeń lub cykli określonego zjawiska okresowego w jednostce czasu. Miarą wysokości dźwięku jest częstotliwość fali, im większa częstotliwość fali tym wyższy jest dźwięk. Jednostką miary częstotliwości jest *Hertz* (1 Hz). Jeden *Hertz* odpowiada jednemu okresowi na sekundę. Podstawowy wzór na częstotliwość fali dźwiękowej $f$ wynika ze wzoru na prędkość (dowolnej) fali harmonicznej:

$$v = \lambda f\,. \tag{1.31}$$

Jest on właściwy zarówno dla dźwięku w powietrzu, jak i w ciałach stałych oraz cieczach. Po podzieleniu obu stron przez długość fali ($\lambda$ — *lambda*) otrzymamy właściwy wzór na częstotliwość:

$$f = \frac{v}{\lambda}, \tag{1.32}$$

gdzie:

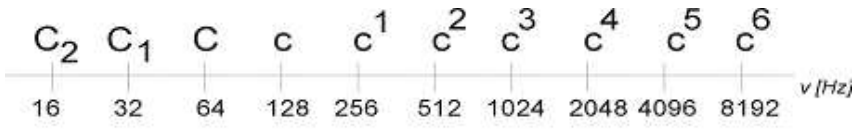
$f$ — częstotliwość dźwięku (Hz),

$\lambda$ — długość fali dźwiękowej.

Powietrze jest wypełnione falami, których ucho człowieka nie jest w stanie usłyszeć, ponieważ nie znajdują się one w zakresie *częstotliwości słyszalnych*. Pełen zakres słyszalności człowieka waha się w granicach od 20 Hz do 20000 Hz — zakres ten nazywa się *pasmem akustycznym*. Jednak w miarę starzenia się człowiek traci zdolność odbierania dźwięków o skrajnych częstotliwościach — zwłaszcza wysokich. Często mówi się o częstotliwości dźwięku jakby było oczywiste, że jest ona jasno określona. W rzeczywistości tylko jeden typ dźwięku ma dobrze określoną (dokładnie „pojedynczą") częstotliwość. Dźwięk taki jest falą harmoniczną (sinusoidalną) i nazywa się *tonem*. *Ton* jest dźwiękiem prostym, który charakteryzuje się przebiegiem sinusoidalnym o ściśle określonej częstotliwości, amplitudzie i fazie. Ton to również określenie składowej harmonicznej. Każdy dźwięk składa się z tonów. Instrumenty muzyczne wytwarzają dźwięki składające się z nieskończonej ilości tonów prostych o różnym natężeniu i częstotliwości będącej wielokrotnością tonu podstawowego (tworzących szereg harmoniczny). Barwa dźwięku w głównej mierze zależy od natężenia występujących w nim tonów lub pasm częstotliwości.

### 1.2.3. Wysokość dźwięku

Jest określana przez porównanie z *tonem wzorcowym* o częstotliwości 440 Hz (określanym jako $a^1$ — a razkreślne). Rozpatrując *szereg diatoniczny* (szereg następujących po sobie dźwięków muzycznych w obrębie oktawy) odległość między tonami mierzona jest ilością stopni nazywana *interwałem*. W *stroju naturalnym* podstawowe interwały są zdefiniowane następująco: mała sekunda, wielka sekunda, mała tercja, wielka tercja, kwarta czysta, kwinta czysta, seksta mała, seksta wielka, septyma mała, septyma wielka, oktawa. Wybierając dźwięk podstawowy i dodając dźwięk oddalony o interwały sekundy wielkiej, tercji wielkiej, kwarty czystej, kwinty czystej, seksty wielkiej, septymy wielkiej oraz oktawy tworzymy *gamę (skalę) durową*. Skalę durową można następnie przedłużać w górę lub w dół i ustalać tonację. Kolejne dźwięki gamy *C-dur* mają nazwy *c, d, e, f, g, a, h, c'* - górne *c* jest już w kolejnej, wyższej oktawie. Jeżeli dolne *c* było *„małe"* to górne będzie *razkreślne*, jeżeli dolne było razkreślne to górne będzie *dwukreślne* itd. Zakres częstotliwości w stroju naturalnym pokazano na Rys. 1.4.




Rysunek 1.4. Zakres częstotliwości muzycznych (9 oktaw) w stroju naturalnym

Wyróżniamy oktawy: *subkontra, kontra, wielką, małą, razkreślna, dwukreślną, trzykreślną* itd., i wyznaczamy dla konkretnych dźwięków — np. dla dźwięku *c*: $C_2$, $C_1$, $C$, $c$, $c^1$, $c^2$ itd.

Obok skali durowej istnieje również *skala molowa*, charakteryzująca się tym, że wielkie tercje i seksty są zastąpione *małymi tercjami* i *małymi sekstami*. Obydwa te typy (zwane w muzyce *trybami*) tworzą razem system *dur-mol*. Strój naturalny charakteryzuje się najczystszym brzmieniem, ale jest niewygodny, bo utrud-

nia *modulacje* (przejście od jednej tonacji do drugiej). Problem ten rozwiązano przez wprowadzenie *stroju temperowanego*, w którym interwał oktawy dzieli się na 12 równych interwałów o stosunku częstości, $\delta = \sqrt[12]{2} \approx 1,059$ równoważnych małej sekundzie. Kolejne interwały uzyskiwane są przez potęgowanie małej sekundy $\Delta = \delta^n$:

Tablica 1.1. Interwały w stroju temperowanym

| Interwał | n | $\Delta$ |
|---|---|---|
| wielka sekunda | 2 | $\delta^2 = 1,122$ |
| mała tercja | 3 | $\delta^3 = 1,183$ |
| wielka tercja | 4 | $\delta^4 = 1,260$ |
| kwarta czysta | 5 | $\delta^5 = 1,335$ |
| kwarta zwiększona | 6 | $\delta^6 = 1,414$ |
| kwinta czysta | 7 | $\delta^7 = 1,498$ |
| mała seksta | 8 | $\delta^8 = 1,587$ |
| wielka seksta | 9 | $\delta^9 = 1,682$ |
| mała septyma | 10 | $\delta^{10} = 1,782$ |
| wielka septyma | 11 | $\delta^{11} = 1,888$ |
| oktawa | 12 | $\delta^{12} = 2$ |

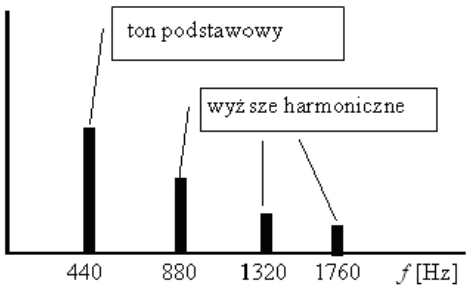
Skala złożona z kolejnych 12 małych sekund (*półtonów*) w stroju temperowanym (*gama chromatyczna*) nosi nazwę *dwunastotonowej*, [17].

### 1.2.4. Widmo dźwięku

Widmo przebiegu okresowego o okresie $T$ ma postać prążków odpowiadających poszczególnym składowym harmonicznym o określonych częstotliwościach. Wysokość każdego prążka reprezentuje amplitudę poszczególnej składowej harmonicznej. Zależność amplitud składowych od częstotliwości nosi nazwę *widma częstotliwościowego*. Widmo częstotliwości sygnałów periodycznych ma charakter dyskretny, sygnałów nieperiodycznych — charakter ciągły. Każdy ton ma w swoim widmie tylko jedną częstotliwość. Każdy dźwięk inny niż ton można matematycznie rozłożyć na składniki będące tonami. Tego rodzaju rozkład przeprowadzany jest za pomocą operacji matematycznej zwanej *analizą fourierowską* (szczegółowo opisaną w dalszej części niniejszej pracy), natomiast wykres zawartości poszczególnych tonów w całym dźwięku nazywany jest *widmem tego dźwięku*. Dźwięki emitowane przez instrumenty muzyczne składają się zazwyczaj z kilku lub nawet kilkunastu tonów. Wszystkie razem składają się na jakość (barwę) dźwięku. Spośród nich jest jeden szczególnie ważny, który określa wysokość dźwięku. Jest to tak zwany *ton podstawowy* nazywany również *pierwszą harmoniczną*. To właśnie pierwsza harmoniczna określa wysokość dźwięku.

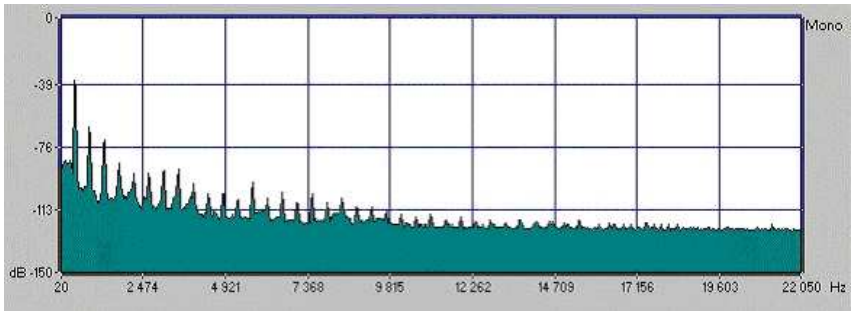
Poza tonem podstawowym w widmie dźwięku występują zazwyczaj tzw. *wyższe harmoniczne*, które mają znaczący wpływ na *barwę dźwięku*. Wszystkie te tony brzmią jednocześnie, jednak mogą zanikać z różną intensywnością. Ton podstawowy 440 Hz wraz z towarzyszącymi wyższymi harmonicznymi przedstawia Rys. 1.5.

Dla większości instrumentów muzycznych wyższe harmoniczne mają częstotliwości będące wielokrotnością częstotliwości tonu podstawowego — im więcej jest harmonicznych tym bogatsza jest barwa dźwięku. Odzwierciedla się to w tym, że wszystkie dobrze zestrojone ze sobą różne instrumenty grając ten sam dźwięk „a" (czyt. „a razkreślne"), wydają ten sam ton podstawowy o częstotliwości 440 Hz, jednak




Rysunek 1.5. Ton podstawowy wraz z towarzyszącymi wyższymi harmonicznymi

różnią się „dodatkami", czyli zawartością wyższych harmonicznych o częstotliwościach 440 Hz, 880 Hz, 1320 Hz, 1760 Hz, 2200 Hz itd. Dźwięk „C" ma częstotliwość 261,6 Hz. Jednak w typowym widmie dźwięku skrzypiec można odszukać nie tylko ton 261,6 Hz, ale i ton „C" o oktawę wyższy 523,3 Hz (druga harmoniczna), a także ton 784 Hz (trzecia harmoniczna), odpowiadający już tonowi podstawowemu dźwięku „G". Na Rys. 1.6 pokazano widmo a razkreślnego gitary akustycznej. Na wykresie wyraźnie widoczny jest ton podstawowy oraz wyższe harmoniczne bogato występujące w widmach instrumentów akustycznych.




Rysunek 1.6. Widmo dźwięku a razkreślnego (440 Hz) dla gitary akustycznej

### 1.2.5. Długość fali dźwiękowej

Związek częstotliwości fali z jej długością przedstawia się następującą zależnością:

$$L = cT = c/f\,, \tag{1.33}$$

gdzie:

$c$ — prędkość rozchodzenia się fali (m/s),
$T$ — okres (s),
$f$ — częstotliwość drgań (Hz).

Prędkość dźwięku jest uzależniona od gęstości materiału, w którym się rozchodzi oraz modułu sprężystości tego materiału. Na przykład wykorzystując zależność (1.33) istnieje możliwość obliczenia, że fala o częstotliwości np. 100 Hz ma długość równą (331,8/100) ok. 3,32 metra przy temperaturze 0°C i ciśnieniu normalnym.

### 1.2.6. Natężenie fali dźwiękowej

W związku z tym, że ucho ludzkie zbiera tylko tę energię z obszaru, jaki samo zajmuje, dla wrażenia głośności najistotniejsza jest energia skupiająca się na jednostce powierzchni w danej jednostce czasu. Natężenie fali dźwiękowej obliczane jest jako energia fali przepływająca przez jednostkę powierzchni (ustawionej prostopadle do kierunku fali) w jednostce czasu. Innymi słowy można powiedzieć, że wielkość wyznaczana jako energia fali dźwiękowej dzielona przez czas i powierzchnię, przez którą ta energia przenika nazywana jest *natężeniem fali dźwiękowej* i opisana zależnością:

$$I = \frac{W}{tS} = \frac{P}{S}\,. \tag{1.34}$$

gdzie:

$I$ — natężenie fali dźwiękowej [W/m$^2$],
$t$ — czas [s],
$S$ — pole powierzchni, na którą pada energia dźwiękowa [m$^2$],
$P$ — moc fali dźwiękowej [W],
$W$ — energia niesiona przez fale.

Minimalna wartość natężenia fali dźwiękowej, którą człowiek może jeszcze usłyszeć wynosi $10^{-12}$ W/m$^2$.

Ucho ludzkie logarytmuje natężenie dźwięku, co powoduje, że dwa razy większe natężenie dźwięku odpowiada zwiększeniu głośności o wielkość proporcjonalną do „logarytmu z dwóch". Dlatego wprowadza się jednostkę zwaną poziomem natężenia dźwięku, wyznaczaną ze wzoru:

$$B = 10 \log \frac{I}{I_0}\,, \tag{1.35}$$

gdzie:

$B$ — poziom natężenia dźwięku [dB],
$I$ — natężenie badanej fali dźwiękowej [W/m$^2$],
$I_0$ — natężenie tzw. „progu słyszalności", czyli wielkości równej $10^{-12}$ W/m$^2$.

Ponieważ logarytm z jedynki ma wartość zero, więc natężenie progu słyszalności ma poziom natężenia 0 dB. Z kolei bardzo głośny słyszalny dźwięk ma poziom głośności w okolicy 100 dB, natomiast wartość 120 dB jest określana jako *próg bólu*. Typowy audiogram człowieka z zaznaczonymi granicami bólu i słyszalności przedstawia Rys. 1.7.

### 1.2.7. Dudnienie

Jeżeli drgania składowe $y_1$ i $y_2$ różnią się amplitudami, częstościami i fazami to przy założeniu, że $A_i \geq 0$ otrzymamy:

$$y = y_1 + y_2 = A_1 \sin \omega_1 t + A_2 \sin(\omega_2 t + \vartheta)\,. \tag{1.36}$$

Przekształcając argument

$$\omega_2 t + \vartheta = \omega_1 t - ((\omega_1 - \omega_2)t - \vartheta) \tag{1.37}$$

i podstawiając go ponownie do wzoru na sumę wychyleń, otrzymamy:

$$y = A_1 \sin \omega_1 t + A_2 \sin(\omega_1 t - ((\omega_1 - \omega_2)t - \vartheta))\,. \tag{1.38}$$

Następnie zależność (1.34) można doprowadzić do postaci

$$y = A\sin(\omega_1 t - \varphi)\,. \tag{1.39}$$

Stałe $A$ i $\varphi$ znajdziemy rozwijając wyrażenia

$$y = A_1\sin\omega_1 t + A_2\sin\omega_t t\cos((\omega_1-\omega_2)t-\vartheta)) - A_2\cos\omega_1 t\sin((\omega_1-\omega_2)t-\vartheta)\,, \tag{1.40}$$

$$y = A\sin\omega_1 t\cos\varphi - A\cos\omega_1 t\sin\varphi \tag{1.41}$$

i porównując współczynniki przy $\sin\omega_1 t$ i $\cos\omega_1 t$

$$A_1 + A_2\cos((\omega_1-\omega_2)t-\vartheta) = A\cos\varphi\,, \tag{1.42}$$

$$A_2\sin((\omega_1-\omega_2)t-\vartheta) = A\sin\varphi\,. \tag{1.43}$$

Amplituda drgań wypadkowych zależy od czasu

$$|A| = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos((\omega_1-\omega_2)t-\vartheta)} \tag{1.44}$$

i zmienia się od wartości $A_1 + A_2$, gdy cosinus pod pierwiastkiem jest równy 1, do $A_1 - A_2$, gdy jest równy $-1$ (dla $A_1 = A_2 = A$ od $2A$ do 0). Okres zmienności wynosi

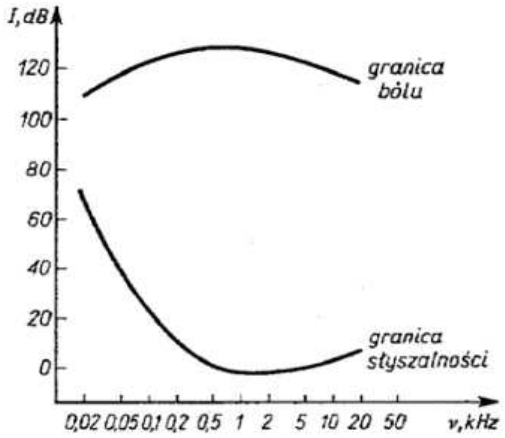
$$T = \frac{2\pi}{\omega_1 - \omega_2}\,, \tag{1.45}$$

a częstość

$$v = \frac{1}{T} = \frac{\omega_1}{2\pi} - \frac{\omega_2}{2\pi} = v_1 - v_2\,. \tag{1.46}$$

Jeżeli częstości drgań leżą w zakresie słyszalności, (20 Hz–20 kHz) to jest słyszalny tzw. *ton kombinacyjny* o częstości $v$, który przy małej różnicy częstości (ok. 16 Hz) przechodzi w rytmiczne powtarzanie się maksimów amplitudy, zwane *dudnieniami*. *Wielkość* $v_d$ nazywana jest wówczas *częstością dudnień*

$$v_d = v_1 - v_2\,. \tag{1.47}$$




Rysunek 1.7. Typowy audiogram dla człowieka o prawidłowym słuchu

### 1.2.8. Zniekształcenia dźwięku

Źródłem wszelkiego rodzaju zniekształceń są niedoskonałe składniki urządzeń. Najogólniej zniekształcenia dzielimy na dwie grupy: liniowe i nieliniowe. O zniekształceniach liniowymi mówimy wówczas, gdy dane urządzenie nie przenosi całości pasma doprowadzonego sygnału lub tłumi niektóre z jego składowych. Najbardziej wyrazistym tego przykładem może być głos w słuchawce telefonicznej lub prosty odbiornik radiowy, którego głośnik nie potrafi odtworzyć częstotliwości zbyt niskich ani wysokich. Charakterystyka częstotliwościowa danego urządzenia wykazuje jak przenoszone jest pełne pasmo akustyczne. Najkorzystniejsza jest płaska charakterystyka częstotliwościowa, a więc taka, która nie tłumi ani nie wzmacnia żadnych częstotliwości składowych. W praktyce jednak rzadko spotykamy urządzenia z płaską charakterystyką częstotliwościową. Kolejnymi przykładami zniekształceń dźwięku są zniekształcenia fazowe. Zniekształcenia te pojawiają się wówczas, gdy poszczególne częstotliwości składowe sygnału przechodzą przez dane urządzenie z różną prędkością, co powoduje powstawanie różnic czasowych między nimi.

Zniekształcenia nieliniowe nazywane są również zniekształceniami amplitudowymi. Spowodowane jest to tym, że sygnał na wyjściu urządzenia zawiera dodatkowe składowe, których nie było w sygnale wejściowym. Przyczyną powstawania zniekształceń nieliniowych są nieliniowe zależności prądowo-napięciowe elementów (np. tranzystorów). W wyniku oddziaływania sygnału na element nieliniowy sygnał ulega zniekształceniu, np. przez obcięcie części wierzchołków sygnału sinusoidalnego.

## 1.3. Sygnał cyfrowy

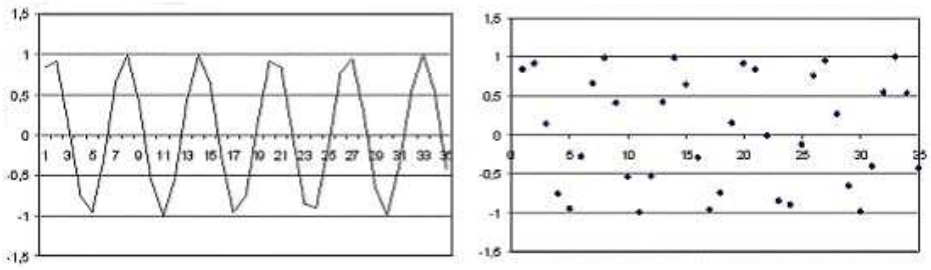
Sygnał cyfrowy $x_k(n)$ to sygnał ciągły czasu dyskretnego $x(n)$, w którym dokonano kwantowania wartości sygnału, tzn. zaokrąglono wartości rzeczywiste do najbliższych liczb całkowitych. Sygnały cyfrowe generuje się za pośrednictwem przetworników analogowo-cyfrowych (A/C), w których przeprowadza się równocześnie dyskretyzację czasu i kwantowanie wartości sygnałów analogowych (ciągłych). W wyniku działania przetwornika A/C sygnał przyjmuje skończoną liczbę określonych wartości tzn., że każdy przedział wartości rzeczywistych otrzymuje swojego jednego reprezentanta.

W cyfrowym zapisie audio stopniowe i płynne zmiany stanu fali zachodzące w czasie są opisywane w drodze pobierania próbek dźwięku w ściśle ustalonych odstępach czasowych. Postępowanie takie nazywane jest *próbkowaniem*. Celem tego procesu, jest zapisanie analogowego przebiegu fali dźwiękowej w postaci kodu binarnego. Przykładem sygnału cyfrowego jest sygnał audio zapisany na płycie CD.

### 1.3.1. Próbkowanie sygnałów analogowych

Próbkowanie jest to nic innego jak badanie sygnału. Częstotliwość próbkowania jest więc badaniem danego sygnału w określonych (stałych) odstępach czasu. Jeżeli mówimy na przykład o częstotliwości próbkowania 44 100 Hz to oznacza, że sygnał jest badany 44 100 razy w ciągu 1 sekundy. W celu opisania sygnału zbiorem próbek konieczne jest wybranie odstępu próbkowania $t_s$. Częstotliwość próbkowania $f_s$ określa się zależnością:

$$f_s = \frac{1}{t_s}\,. \tag{1.48}$$



Rysunek 1.8. Postać analogowa i dyskretna sygnału

Na Rys. 1.8 pokazano postacie czasowe sygnału ciągłego oraz jego postać po próbkowaniu.

Zwiększanie częstotliwości próbkowania sygnału ciągłego doprowadza do uzyskania większej ilości próbek, a co za tym idzie bardziej precyzyjnego opisu badanego sygnału. W celu osiągnięcia dokładnej reprezentacji cyfrowej badanego sygnału analogowego minimalna częstotliwość próbkowania powinna być, co najmniej dwa razy wyższa od najwyższej częstotliwości sygnału oryginalnego. Ta minimalna częstotliwość próbkowania jest nazywana *częstotliwością Nyquista.*

$$f_s \geq 2f_{\max}\,. \tag{1.49}$$

gdzie:

$f_s$ — częstotliwość próbkowania,

$f_{max}$ — najwyższa częstotliwość badanego sygnału.

Jeżeli rozważany sygnał jest sumą kilku sygnałów o przykładowych częstotliwościach: 45 Hz, 7 Hz, 12 kHz i 8 kHz wówczas według zależności (1.49) częstotliwość próbkowania powinna wynosić co najmniej $2 \times 12\,\text{kHz}$=24 kHz. W przypadku, gdy kryterium Nyquista nie jest spełnione pojawia się niejednoznaczność sygnału, nazywana *aliasingiem.*

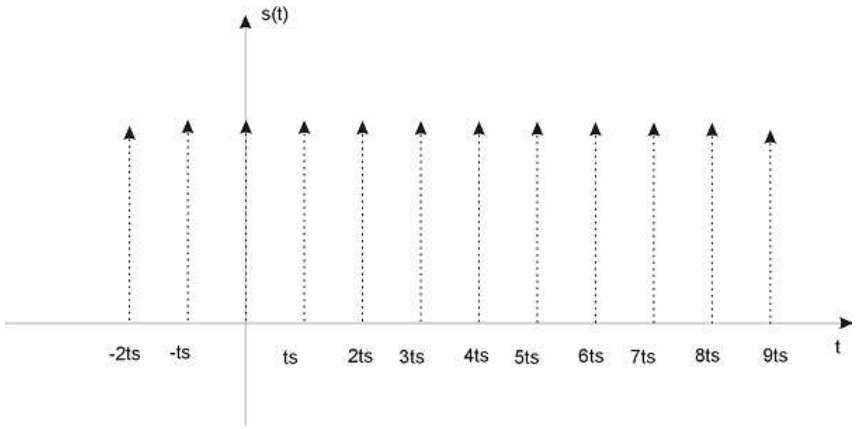
Próbkowanie jest przeprowadzane przy pomocy układu składającego się z: filtru dolnoprzepustowego, układu próbkującego, przetwornika analogowo-cyfrowego i generatora taktującego pracę całości. W module próbkującym w równych odstępach czasu mierzona jest wartość próbki i nadawana wynikowi pomiaru postać liczby dziesiętnej. Opuszczając układ próbkujący sample (próbki sygnału analogowego) docierają do układu analogowo cyfrowego, w którym następuje proces ich kwantowania, czyli wyrównywania lub zaokrąglenia. Następnie skwantowana informacja zostaje zakodowana, zależnie od długości słowa cyfrowego (8, 16 lub 24 znaki na bajt). Proces ten jest bezpośrednio związany z rozdzielczością kodowania sygnału.

### 1.3.2. Matematyczna reprezentacja próbkowania

Z matematycznego punktu widzenia, próbkowanie sygnału jest mnożeniem wejściowego przebiegu analogowego przez deltę *Diraca* — nazywaną również funkcją impulsową. Delta Diraca jest funkcją uogólnioną i można ją zdefiniować jako impuls, którego pole równe jest jedności, co wyrażone jest za pomocą wzoru:

$$\int_{-\infty}^{\infty} \delta(t)dt = 1. \tag{1.50}$$

Przykładem wyznaczenia pola konkretnego przebiegu fali może być obliczenie pola impulsu prostokątnego pokazanego na Rys. 1.9, [22].




Rysunek 1.9. Ciąg funkcji o jednakowych odstępach

Wyznaczenie pola konkretnego przebiegu fali obliczane jest wg zależności:

$$\int_{-\infty}^{\infty} A\delta(t)dt = A\,. \tag{1.51}$$

Z powyższego wzoru wynika, że pole ograniczone wykresem funkcji $A\delta(t)$ oraz osią czasu wynosi 1. Przebieg sygnału próbkowanego będzie się składał z ciągu funkcji impulsowych oddzielonych od siebie dokładnie okresem próbkowania $t_s$. Funkcje próbkującą $s(t)$ można zapisać jako sumę poszczególnych funkcji impulsowych:

$$s(t) = \sum_{n=\infty}^{n=-\infty} \delta(t - nt_s)\,. \tag{1.52}$$

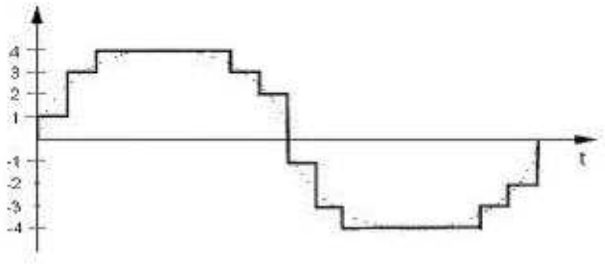
Wzięcie splotu przebiegu analogowego $f(t)$ z powyższym przebiegiem próbkującym doprowadza do pojawienia się na wyjściu ciągu impulsów polach równych amplitudom sygnału $f(t)$. Przebieg próbkowanego $y(t)$ jest prostym mnożeniem funkcji $s(t)$ przez wejściową funkcję analogową $f(t)$ i wyrażony jest wzorem:

$$y(t) = \sum_{n=\infty}^{n=\infty} f(t)\,\delta(t - nt_s)\,. \tag{1.53}$$

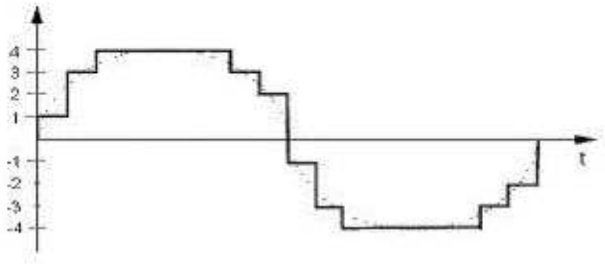
### 1.3.3. Rozdzielczość kodowania

Analizowany przebieg powinien być zapisany w systemie binarnym, tzn. w postaci zer i jedynek. Pojęcie rozdzielczości kodowania wiąże się z ilością wykorzystanych bitów do opisu fali dźwiękowej. Na Rys. 1.10 przedstawiono obraz przebiegu opisany 1 bitem.

Im większa ilość bitów zastosowana do zapisu dźwięku tym dokładniejsze odwzorowanie przebiegu. Jeżeli przebieg prezentowany na Rys. 1.10 zostanie zapisany w rozdzielczości 3-bitowej ($2^3$=8) wówczas jego postać czasowa będzie taka jak na Rys. 1.11:




Rysunek 1.10. Fala sinusoidalna zapisana w drodze samplingu 1-bitowego




Rysunek 1.11. Fala sinusoidalna zapisana w drodze samplingu 3-bitowego

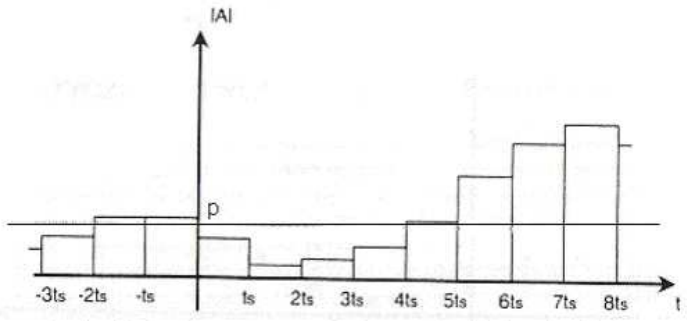
Uzyskany wykres jest bardziej precyzyjny niż przebieg pokazany na Rys. 1.10, jednak ciągle jest to za mało by oszukać ludzkie ucho. W celu dokonania wierniejszego zapisu dźwięku należy zwiększyć liczbę bitów. Standard zapisu audio CD to 16 bitów — a zatem 65 536 możliwych poziomów dźwięku. Zapis 16-bitowy jest w zupełności wystarczającym zapisem, by ucho ludzkie nie wykryło żadnej różnicy między dźwiękiem analogowym (idealną falą) a cyfrowym.

### 1.3.4. Kwantyzacja

Badając sygnał cyfrowy mamy do dyspozycji ciąg próbek o dowolnej amplitudzie. Czas dla badanego sygnału dźwiękowego zmienia charakter — z natury ciągłej przechodzi na charakter dyskretny. Zmierzając konsekwentnie do postaci cyfrowej, czyli dyskretnej amplitudy, mamy dyskretny czas. Idea sygnału cyfrowego polega na wyeliminowaniu stanów pośrednich. Mając stany 0 i 1 nie uwzględniamy stanu np. 0,3 — staje on się logicznym stanem 0. Podobnie stan 0,8521 staje się logiczną jedynką. Jeżeli badamy sygnał dyskretny o amplitudzie analogowej to oznacza, że wartość liczbowa jednej dowolnej próbki jest liczbą rzeczywistą z pewnego przedziału, a zatem wartości poszczególnych próbek mogą mieć nieskończenie dużą precyzję (np. $12,3457212\ldots$). Można rozważać, jaka liczba miejsc po przecinku jest właściwa dla zapisu danych w systemach komputerowych. Należy jednak brać pod uwagę, że w komputerze najefektywniej są przetwarzane i zapisywane liczby w systemie dwójkowym. Problem określenia sensownej precyzji pozostaje jednak dalej. Zapisując w systemie binarnym dowolny sygnał można przyjąć, że będą brane pod uwagę tylko te zbiory liczb dwójkowych, które spełniają kryterium $2^N$, a $N$ jest liczbą naturalną.

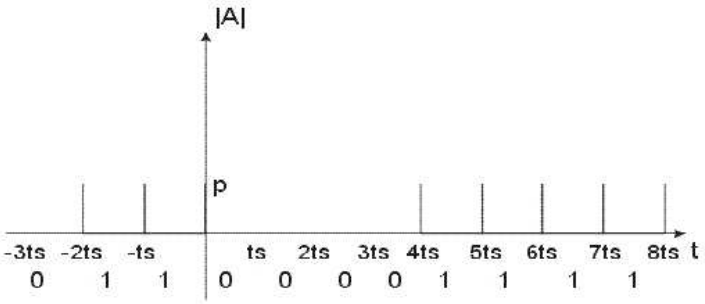
Kwantyzacja realizowana jest przez przetwornik analogowo-cyfrowy (przetwornik A/C) i polega na przyporządkowaniu określonych odcinków amplitudowych sygnału

wejściowego ustalonym przedziałom. W trakcie tego procesu dochodzi do zaokrąglenia wartości sygnału, co pociąga za sobą błędy nazywane *szumem kwantyzacji*. Jeżeli np. decydujemy się tylko na dwa takie przedziały, możemy określić pewien poziom $p$, który traktowany będzie jako punkt decyzyjny. Próbki, których wartości amplitudy znajdują się powyżej danego punktu decyzyjnego będą należały do 1, natomiast próbki poniżej linii $p$ — do 0. Przykład ten pokazano na Rys. 1.12, [22].




Rysunek 1.12. Przebieg sygnału uzyskany w wyniku pracy układu próbkującego z zaznaczonym punktem decyzyjnym $p$

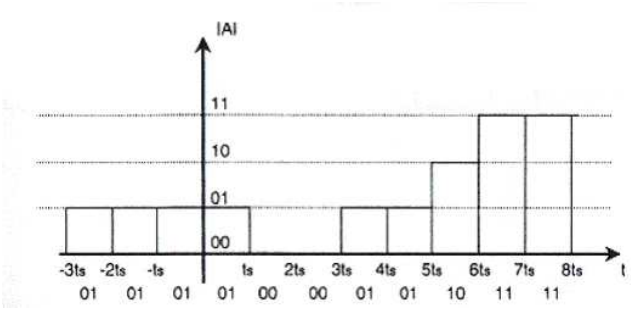
W konsekwencji takiej operacji na wyjściu układu pojawia się cyfrowy binarnych zer i jedynek, co zilustrowano na Rys. 1.13, [22].



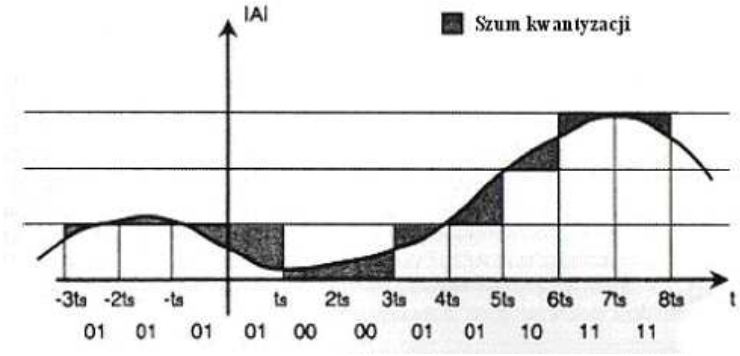

Rysunek 1.13. Przebieg z Rys. 1.12 po kwantyzacji

Rozbudowując opisany przykład, możemy użyć dwa bity doprowadzając do powstania czterech jednakowo od siebie oddalonych poziomów kwantowania oznaczonych jako 00, 01, 10, 11. W efekcie przypisujemy sygnał wejściowy czterem przedziałom amplitudowym, co pokazano na Rys. 1.14, [22].

Łatwo wywnioskować, że im większa ilość przedziałów tym dokładniejszy opis wejściowego sygnału analogowego. W systemach DSP (Digital Signal Processing) najczęściej stosuje się przetwarzanie A/C na 10–12 bitach, co oznacza, że sygnał wejściowy opisywany jest z dokładnością do $2^{10}$ (1024) lub $2^{12}$ (4096) poziomów. Podczas porównań kwantowanych poziomów sygnału z sygnałem wejściowym można dostrzec zaistniałe błędy, które wywołują wspomniany już wcześniej *szum kwantyzacji*. Zjawisko to zilustrowano na Rys. 1.15, [22].




Rysunek 1.14. Przebieg sygnału po kwantyzacji 2-bitowej




Rysunek 1.15. Próbkowanie sygnału skwantowanego — szum kwantyzacji to pole zaczernione

### 1.3.5. Zaszumienie sygnału

Szum akustyczny jest dźwiękiem, którego widmo jest w większości zakresu słyszalności zrównoważone, tzn. nie występują w nim gwałtowne maksima, które mogłyby być słyszalne jako dźwięczące rezonanse o określonej wysokości tonu. Miarą poziomu szumów jest parametr oznaczony jako $S/N$ (Signal to Noise Ratio), czyli stosunek szumów do sygnału. Parametr ten mierzony jest w decybelach. Przed przystąpieniem do szczegółów związanych z precyzyjnym opisem zjawiska zaszumienia sygnału należy zdefiniować funkcje korelacji $R(\cdot)$. Dla stacjonarnych sygnałów losowych definiuje się je w następujący sposób, [21]:

1. Dla sygnałów ciągłych:

$$R_{xx}(\tau) = \mathrm{E}\left[x(t)x(t-\tau)\right] \tag{1.54}$$

2. Dla sygnałów dyskretnych:

$$R_{xx}(m) = \mathrm{E}\left[x(n)x(n-m)\right] \tag{1.55}$$

gdzie:
E[·] — wartość oczekiwana,
$\tau$ — przesunięcie czasowe.

W powyższych wzorach $x(t)$ i $x(n)$ należy traktować jako niezależne zmienne losowe.

Podczas analizy częstotliwościowej sygnałów losowych stosuje się *funkcję gęstości widmowej mocy*. Jest ona zdefiniowana jako transformata Fouriera (szczegółowy opis transformaty Fouriera znajduje się w dalszej części niniejszej pracy) funkcji autokorelacji, [21]:

$$P_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)\mathrm{e}^{-j2\pi f\tau}d\tau, \quad P_{xx}(f) = \sum_{m=-\infty}^{\infty} R_{xx}(m)\mathrm{e}^{-j2\pi(f/f_{pr})m}. \tag{1.56}$$

Dla powyższych wyrażeń prawdziwe są odwrotne zależności:

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} P_{xx}(f)\mathrm{e}^{j2\pi f\tau}df, \quad R_{xx}(m) = \frac{1}{f_{pr}}\int_{-f_{pr}/2}^{f_{pr}/2} P_{xx}(f)\mathrm{e}^{j2\pi(f/f_{pr})m}df. \tag{1.57}$$

Para równań (1.54) oraz (1.55) nosi nazwę równań *Wienera-Chinczyna* dla sygnałów ciągłych i dyskretnych. Ponieważ funkcja autokorelacji jest symetryczna względem $\tau = 0$ ($R_{xx}(\tau) = R_{xx}(-\tau)$), $P_{xx}(f)$ jest rzeczywiste, ponieważ tylko symetryczne względem zera funkcje $\cos(2\pi ft)$ są potrzebne do rozwinięcie harmonicznego $R_{xx}(\tau)$. Można wykazać, [21], że dla ciągłych sygnałów stacjonarnych funkcja gęstości widmowej mocy jest równa:

$$\begin{aligned} P_{xx}(f) &= \lim_{T\to\infty} \mathrm{E}\left[\frac{1}{2T}\left|\int_{-T}^{T} x(t)\mathrm{e}^{-j2\pi ft}dt\right|^2\right] = \lim_{T\to\infty} \mathrm{E}\left[\frac{1}{2T}\left|X_T(f)\right|^2\right] \\ &= \lim_{T\to\infty} \mathrm{E}\left[\frac{1}{2T}X_T(f)X_T^*(f)\right] \end{aligned} \tag{1.58}$$

gdzie:

- $\mathrm{E}[\cdot]$ — wartość oczekiwana,
- $X_T$ — transformata Fouriera fragmentu sygnału z przedziału czasowego $[-T, T]$ (czyli widmo amplitudy tego fragmentu),
- $1/(2T)\left|X_T(f)\right|^2$ — periodogram (posiada taki sam wymiar jak funkcja widmowa gęstości mocy).

Postać dla dyskretnych sygnałów:

$$\begin{aligned} P_{xx}(f) &= \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{2N+1}\left|\sum_{n=-N}^{N} x(n)e^{-j\frac{2\pi nf}{f_{pr}}}\right|^2\right] \\ &= \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{2N+1}\left|X_N(f)\right|^2\right] = \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{2N+1}X_N(f)X_N^*(f)\right] \end{aligned} \tag{1.59}$$

W tym przypadku *periodogram* zdefiniowany jest jako $1/(2N+1) \times |X_N(f)|^2$.

Szum nazywa się *białym*, jeśli jego $P_{xx}(f)$ jest stałe i nie zależy od częstotliwości. Jeżeli tak nie jest, to szum nazywany jest *szumem kolorowym.* Szczególnym przypadkiem szumu kolorowego jest idealny szum *dolnopasmowy,* dla którego funkcja $P_{xx}(f)$ ma kształt prostokątny, tzn. ma wartość stałą, różną od zera dla częstotliwości z przedziału $(-f_{\max}, f_{\max})$ oraz równą zero poza tym przedziałem. Innymi przykładami szumu kolorowego jest szum *różowy* i *niebieski.* Dla szumu różowego funkcja $P_{xx}(f)$ maleje 6 decybeli na oktawę, natomiast dla szumu niebieskiego — rośnie 6 decybeli na oktawę.

ROZDZIAŁ 2

# Analiza dźwięku

## 2.1. Transformata Fouriera

Podstawowym narzędziem do obliczania i analizowania parametrów widmowych jest transformata Fouriera $X(f)$ sygnału ciągłego $x(t)$. Jest to najbardziej popularna i wydajna procedura spotykana w dziedzinie cyfrowego przetwarzania sygnałów. W wyniku zastosowanej procedury otrzymujemy widmo amplitudowe lub gęstość widmową mocy. Dla sygnału ciągłego transformatę Fouriera wyznacza się z zależności, [23]:

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt\,, \tag{2.1}$$

gdzie $x(t)$ to sygnał ciągły w dziedzinie czasu.

Transformacja przekształca dziedzinę czasu w dziedzinę widma. W nagraniach cyfrowych dziedzina czasu zostaje poddana *dyskretyzacji*. Zamiast ciągłej funkcji $x(t)$ otrzymywany jest sygnał $x(nT)$, gdzie $T$ jest okresem próbkowania. Dyskretna transformata Fouriera (ang. DFT — *Discrete Fourier Transform*) $X(k)$ dla okna czasowego o długości $N$ zdefiniowana jest na ciągu próbek $x(0), \ldots, x((N-1)T)$ za pomocą wzoru:

$$X(k) = \sum_{n=0}^{N-1} x(nT)\mathrm{e}^{-jk\Omega nT}, \quad k = 0, 1, \ldots, N-1, \tag{2.2}$$

gdzie:
$\Omega = 2\pi/NT$,
$T$ — okres próbkowania.

## 2.2. Przeciek częstotliwości

Właściwość DFT, znana jako przeciek widma powoduje, że wyniki DFT stanowią tylko aproksymację rzeczywistych widm oryginalnych sygnałów wejściowych, poddanych próbkowaniu. DFT ogranicza się do operowania na skończonych zbiorach $N$ wartości wejściowych, próbkowanych z częstotliwością $f_s$. W wyniku tej procedury uzyskujemy $N$-punktową transformatę, której dyskretne wartości wyjściowe są związane z kolejnymi wartościami częstotliwości $f_m$. Wynika z tego, że każdemu

prążkowi widma amplitudowego można przypisać kolejny numer $m$, oraz częstotliwość $f_m$, co przedstawia poniższy wzór:

$$f_m = m\frac{f_s}{N} = f_r m\,, \tag{2.3}$$

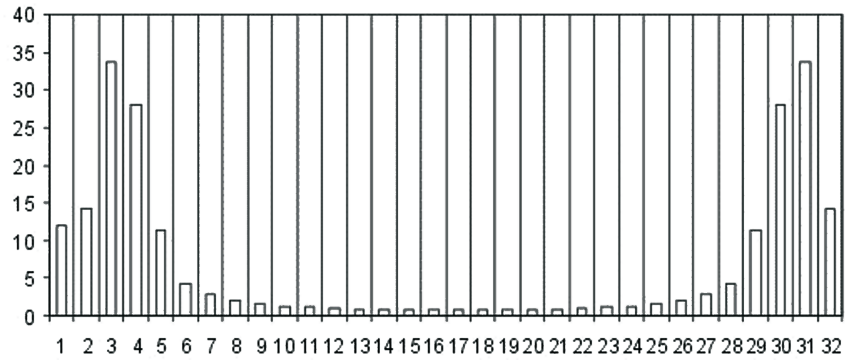
gdzie:
$f_m$ — częstotliwość w widmie dla prążka o numerze m,
$N$ — liczebność próby (liczba próbek poddanych DFT lub FFT),
$f_s$ — częstotliwość próbkowania,
$m$ — indeks (zmieniający się w granicach od 1 do $N/2-1$),
$f_r$ — rozdzielczość widma ($f_r f_s/N$).

DFT zwraca prawidłowe wyniki tylko wówczas, gdy ciąg danych wejściowych zawiera energię rozłożoną dokładnie przy częstotliwościach, dla których dokonywana jest analiza i określona zależnością (1.59), będących całkowitymi wielokrotnościami częstotliwości podstawowej $f_s/N$. Jeżeli w sygnale wejściowym występuje składowa o częstotliwości pośredniej (np. $1.3 \times f_s/N$) to energia sygnału tej częstotliwości zostanie podzielona między częstotliwościami, $mf_s/N$ — dla których wyznaczono wartości DFT. Oznacza to, że energia dowolnego sygnału wejściowego, którego częstotliwość nie jest dokładnie równa częstotliwości, dla której jest wyznaczany dany prążek DFT, przecieka do wszystkich innych wyznaczanych prążków DFT.

Przeciek częstotliwości może zilustrować przykład, gdy analizie poddano sygnał składający się z sumy dwóch składowych 2.5 Hz i 4 Hz o amplitudzie odpowiednio 3 i 0.5. Przebieg spróbkowano częstotliwością 32 Hz. Na podstawie próbek obliczono 32 składowych sinusoidalnych. Badany sygnał opisano wyrażeniem:

$$y(n) = 3\sin\left(2\pi 2.5\frac{n}{32}\right) + 0.5\sin\left(2\pi 4\frac{n}{32}\right)\,. \tag{2.4}$$

Podczas obliczeń DFT otrzymano wynik dla $n = 4$, wykorzystując (2.4) dla $f_{n=4} = 4$ Hz. Przyglądając się natomiast częstotliwości 2.5 Hz dochodzimy do wniosku, że nie istnieje takie całkowite $n$ spełniające zależność (2.1). Przedstawione zjawisko zilustrowano na Rys. 2.1.




Rysunek 2.1. Widmo amplitudowe sygnału z widocznym przeciekiem częstotliwości

W rzeczywistych sygnałach, przy określonej częstotliwości próbkowania nawet bardzo duża liczba próbek nie gwarantuje pominięcia zjawiska przecieku częstotliwości. Zjawisko to ma bardzo ścisły związek z okienkowaniem.

## 2.3. Dyskretne okna czasowe

Metoda okien czasowych powoduje redukcję przecieku częstotliwości przez zminimalizowanie listków bocznych bez konieczności poszerzania okna — a więc zwiększenia obliczeń w DFT. Transformata Fouriera operuje na danych dyskretnych o określonej (skończonej) długości. Wybranie określonego fragmentu danych o długości $N$ oznacza, że sygnał wejściowy na tym odcinku został przemnożony przez 1, natomiast na pozostałych przez 0. Jest to równoznaczne z przemnożeniem sygnału przez okno prostokątne o szerokości $N$ i wysokości 1. Operacja taka (nazywana okienkowaniem) można opisać zależnością:

$$v(n) = w(n)s(n)\,, \tag{2.5}$$

gdzie:
$s(n)$ — sygnał wejściowy,
$v(n)$ — sygnał otrzymany w wyniku okienkowania,
$w(n)$ — funkcja okna.

Okienkowanie jest realizowane za pośrednictwem wykonania operacji splotu transformaty Fouriera z sygnałem funkcji okna w dziedzinie widma. Zatem sygnał otrzymany w wyniku okienkowania jest wynikiem iloczynu dwóch sygnałów a jego widmo jest równe splotowi widm sygnałów mnożonych. Prowadzi to do przecieków widma, tzn. do pojawienia się listków bocznych. Poprzez wprowadzenie okna o wartościach dążących do zera na brzegach przedziału $[0, N]$ można zmniejszyć wysokość listków bocznych. Odbywa się to jednak kosztem poszerzenia listka głównego.

Różnice można zobrazować analizując widmo poszczególnych znanych funkcji okien. Równania najpopularniejszych okien przedstawiono w Tab. 2.1. Są to okna, których kolejne próbki są wymnażane z kolejnymi próbkami analizowanego sygnału.

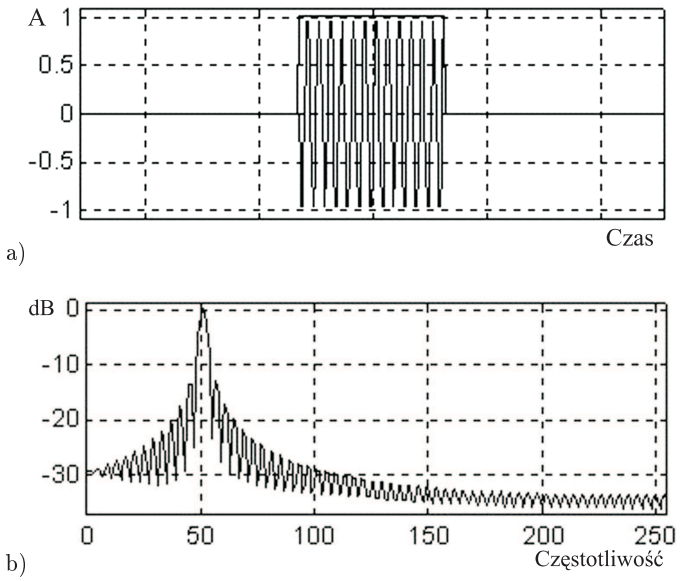
Tablica 2.1. Wybrane funkcje okienkowe

| Nazwa okna | Definicja funkcji okna $w(n)$ $0 \leq n \leq N-1$ |
|---|---|
| Prostokątne | $1$ |
| Trójkątne (Barletta) | $1 - \frac{2\lvert n-(N-1)/2\rvert}{N-1}$ |
| Hanninga (Hanna) | $\frac{1}{2}(1 - \cos(\frac{2\pi n}{N-1}))$ |
| Hamminga | $0.54 - 0.46\cos(\frac{2\pi n}{N-1})$ |
| Blackmana | $0.42 - 0.5\cos(\frac{2\pi n}{N-1}) + 0.08\cos(\frac{4\pi n}{N-1})$ |

Zwiększenie wartości $N$ spowoduje zmniejszenie szerokości listka głównego widma (niezależnie od rodzaju okna), natomiast nie ma on wpływu na tłumienie listków bocznych. Jednak w funkcji częstotliwości listki boczne dłuższego okna szybciej zanikają.

Na Rys. 2.2 pokazano okno *prostokątne* odpowiadające przedziałowi próbkowania o skończonym czasie, oraz jego widmo.

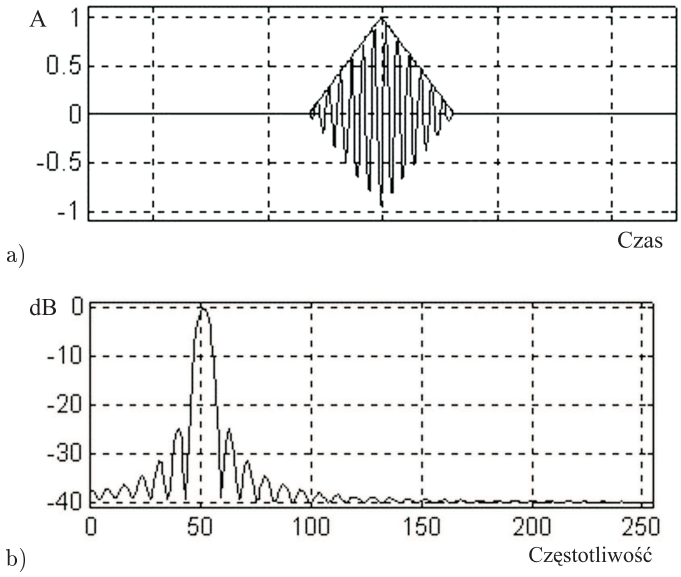
Widoczne jest na Rys. 2.2b, że w oknie prostokątnym zjawisko przecieku zostało filtrowane na poziomie $-15\,$dB.Charakterystyka amplitudowa okna prostokątnego stanowi „miarę", jakiej używa się, aby oszacować odpowiedź impulsową innego okna, przez porównanie jej z oknem prostokątnym. Aby uzyskać lepsze efekty minimalizacji




Rysunek 2.2. Okno prostokątne: (a) postać czasowa, (b) widmo

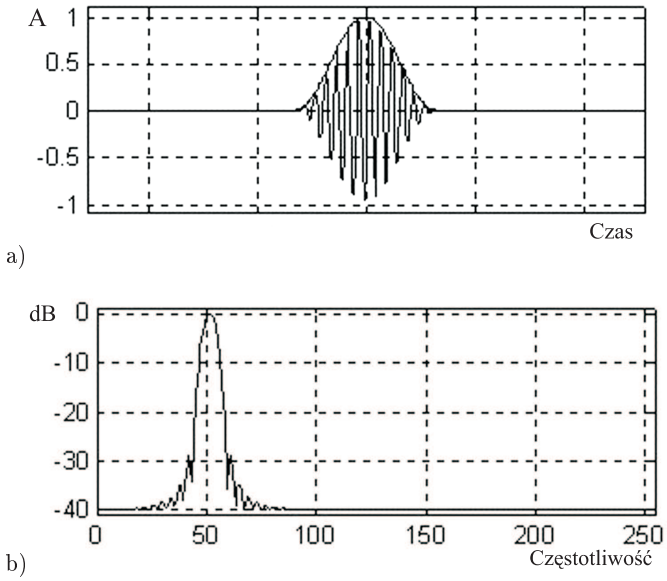
przecieku widma korzystnie jest użyć funkcji okna innego niż prostokątne. Zmniejszenie listków bocznych jest wyraźnie widoczne w oknie *trójkątnym (Bartletta)*, co pokazano na Rys. 2.3.

Z analizy Rys. 2.3b wynika, że okno *trójkątne* ma zmniejszone poziomy listków bocznych do ok. 25 dB. Wadą tego okna jest zdecydowanie szerszy listek główny, co może spowodować, że dwie składowe znajdą się w jego polu i mogą sugerować istnienie dwóch bliskich składowych, a w efekcie odtwarzania sygnału dudnienie. Szerokość listka głównego okna *Barletta* jest zazwyczaj ok. dwa razy większa, niż szerokość listka głównego okna *prostokątnego*.




Rysunek 2.3. Okno trójkątne: (a) postać czasowa, (b) widmo

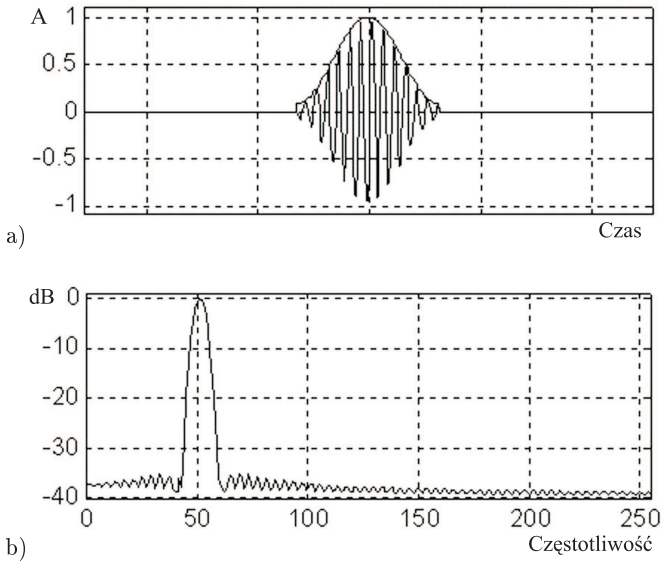
Dalsze zmniejszenie się poziomu pierwszego listka bocznego oraz gwałtowny spadek listków bocznych jest wyraźnie widoczne w charakterystyce amplitudowej okna *Hanninga*. Okno to posiada łagodniejszy kształt (związany z cosinusem) oraz lepszą dynamikę sięgającą ok. 32 dB, co zilustrowano na Rys. 2.4.




Rysunek 2.4. Okno Hanninga: (a) postać czasowa, (b) widmo

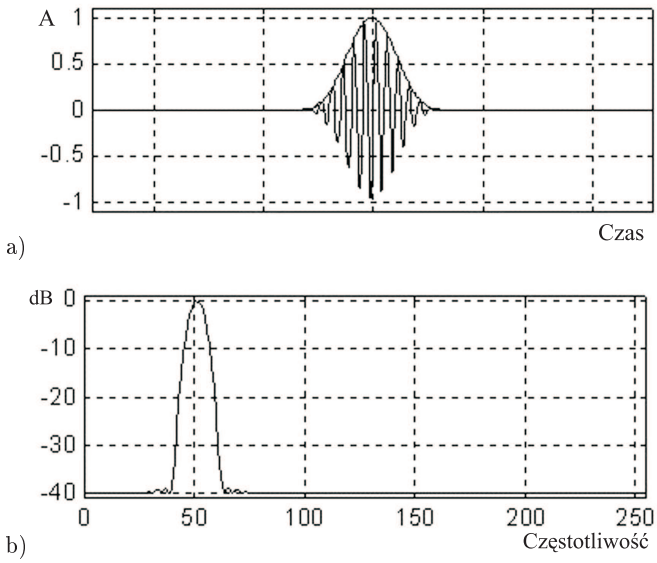
W odróżnieniu od okna *Hanninga*, charakterystyka okna *Hamminga* pokazuje mniejsze poziomy listka pierwszego. Listki boczne tego okna opadają wolniej niż listki okna *Hanninga* co pokazano na Rys. 2.5.

Fakt wolniejszego opadania listków świadczy o mniejszym przecieku widma w odległości ok. 3–4 prążków od prążka środkowego. Przeciek dla około 12 prążków od prążka środkowego jest mniejszy dla okna *Hanninga*, niż dla okna *Hamminga*, [21].




Rysunek 2.5. Okno Hamminga: (a) postać czasowa, (b) widmo

Większą dynamikę tłumienia niż opisane wyżej okno *Hanninga* posiada funkcja okna *Blackmana*. Okienkowanie przy pomocy tej funkcji doprowadza do zwiększenia dynamiki do ok. 40 dB. Jednak w stosunku do okien *Hanninga* i *Hamminga* widać zwiększenie listka głównego. Charakterystykę amplitudową okna *Blackmana* przedstawiono na Rys. 2.6a.




Rysunek 2.6. Okno Blackmana: (a) postać czasowa, (b) widmo

W literaturze cyfrowego przetwarzania sygnałów przedstawiono wiele różnych funkcji okien. Jest ich na tyle wiele, że zostały one nazwane nazwiskiem prawie każdego autora działającego w obszarze cyfrowego przetwarzania sygnałów. Można stwierdzić, że wybór okna stanowi kompromis pomiędzy poszerzeniem listka głównego, poziomami pierwszego listka bocznego, oraz tego jak szybko maleją listki boczne wraz ze wzrostem częstotliwości, [21]. Wykorzystanie poszczególnych funkcji okien zależy głównie od zastosowań, których jest wiele.

## 2.4. Transformata falkowa

Transformata falkowa (*wavelet transform*) jest jednym z najpopularniejszych i najdynamiczniej rozwijających się narzędzi analizy częstotliwościowej sygnałów niestacjonarnych (zmiennych czasie, impulsowych), [21]. Jej najważniejszą zaletą jest duża rozdzielczość w dziedzinie częstotliwości i czasu. Transformata ta, podobnie jak transformata Fouriera, stanowi reprezentację sygnału w postaci superpozycji pojedynczych składowych, z tym że składowe te nie są tonami prostymi, ale pakietami przebiegów czasowych nazwanymi falkami (ang. *wavelets*). Falki powstają przez przeskalowanie tego samego podstawowego kształtu przebiegu czasowego zwanego *falką prototypową* lub *falką macierzystą*. Stanowią one zestaw funkcji analizujących, wykorzystywanych do dekompozycji widmowej sygnału. Na skutek skalowania otrzymuje się falki o zmiennym zasięgu w skali czasu, co umożliwia precyzyjne określenie położenia wysokoczęstotliwościowych składowych widma w czasie. Falki analizujące są tworzone z falki prototypowej $g(t)$ poprzez jej przeskalowanie sprowadzające się

do jej zwężenia lub rozszerzenia oraz przesunięcie w dziedzinie czasu, co przedstawia poniższa zależność, [24]:

$$g_{b,a}(t) = \frac{g}{a}\left(\frac{t-b}{a}\right), \tag{2.6}$$

gdzie:
$g_{b,a}(t)$ — funkcja analizująca,
$g(t)$ — funkcja prototypowa,
$a$ — współczynnik rozszerzenia,
$b$ — parametr przesunięcia czasowego.

Zależność (2.6) określa dokładność lokalizacji w czasie, natomiast częstotliwość jest szacowana przez dokonanie transformacji Fouriera funkcji $g_{b,a}(t)$, [24]:

$$\bar{g}_{b,a}(\omega) = \sqrt{a}\bar{g}(a\omega)\mathrm{e}^{j\omega b}, \tag{2.7}$$

gdzie $\bar{g}(\omega)$ to transformata Fouriera funkcji $g(t)$.

Przykład funkcji macierzystej zilustrowano posługując się przeskalowaną serią falek Moleta, co wyraża zależność:

$$g_{b,a}(x) = \frac{1}{a}\exp\left(-\frac{(x-b)^2}{2a^2}\right)\cos\left(\frac{k(x-b)}{a}\right), \tag{2.8}$$

gdzie $k$ to liczba całkowita.

Ogólny wzór na transformatę $S(b,a)$ sygnału $s(t)$ dla przebiegów ciągłych wyraża zależność:

$$S(b,a) = \int g_{b,a}(t)\, s(t)dt\,. \tag{2.9}$$

Wersja dyskretna transformaty falkowej (ang. DWT — *Discrete Wavelet Transform*) wyrażona jest zależnością, [24]:

$$\mathrm{DWT}(a,n) = \frac{1}{\sqrt{a}}\sum_k h\left(\frac{k}{a}-n\right)x(k)\,, \tag{2.10}$$

gdzie:
$k$ — indeks czasu,
$h(k)$ — funkcja prototypowa,
$x(k)$ — analizowany sygnał.

ROZDZIAŁ 3

# Charakterystyka wybranych instrumentów

Instrumenty muzyczne zostały poddane klasyfikacji ze względu na różne cechy. Na przykład ze względu na sposób wzbudzania drgań źródła dźwięku dokonano podziału instrumentów na:

1. dęte,
2. smyczkowe,
3. szarpane,
4. uderzane.

Podział ten został zakwestionowany przez Carla Sachsa, który jako podstawowe kryterium przyjął właściwości źródła dźwięku zmieniając tym samym główny podział instrumentów muzycznych, [19], [20].

## 3.1. Idiofony

Instrumenty perkusyjne *samobrzmiące*, które ze względu na swoją budowę nie dają możliwości wydobycia precyzyjnych tonów, koniecznych do odtwarzania melodii. Jest to grupa instrumentów muzycznych, w których źródłem dźwięku jest ciało stałe mające niezmienną naturalną sprężystość. Najczęściej źródłem dźwięku jest cały idiofon — stąd nazwa *samobrzmiące*. Wysokość dźwięku jest uzależniona od właściwości fizycznych elementu drgającego — przede wszystkim od jego masy. Idiofony zostały sklasyfikowane ze względu na:

1. Sposób wywołania wibracji:
   - uderzane pałeczką lub prętem (np. gong, ksylofon),
   - zderzane o siebie (np. talerze, kastaniety),
   - pocierane smyczkiem, szczotka lub dłonią (harmonijka szklana),
   - szarpane wypustkami mechanizmu lub dłonią (np. pozytywka, drumla).
2. Kształt źródła dźwięku:
   - płytowe (np. talerze, gong),

- sztabkowe (np. ksylofon),
- rurowe (np. dzwony rurowe),
- prętowe (np. trójkąt).

## 3.2. Membranofony

Instrumenty perkusyjne membranowe, w których źródłem dźwięku jest drgająca membrana wykonana ze skóry lub błony rozpiętej na cylindrycznym, stożkowym lub innym korpusie tworzącym zarazem pudło rezonansowe instrumentu. Wibracja membrany jest pobudzana np. przez uderzenia. Membranofony są podzielone ze względu na:

1. Właściwości akustyczne:
   - o nieokreślonej wysokości dźwięku (szmery, szumy, dźwięki rozproszone),
   - o określonej wysokości dźwięku (np. kotły).
2. Sposób pobudzania membrany:
   - uderzane (np. bęben mały i wielki),
   - pocierane (np. bęben obręczowy),
   - dęte (np. mirliton).

## 3.3. Aerofony

Jest to grupa instrumentów muzycznych, w których źródłem dźwięku jest drgający słup powietrza, zamknięty w przestrzeni rezonansowej, pobudzony do wibracji za pomocą zadęcia. Część instrumentu zamykająca słup powietrza nazywana jest piszczałką i od jej długości zależy wysokość dźwięku. Barwa dźwięku zależy od materiału, z którego wykonano piszczałkę, jej kształtu i menzury. Wysokość dźwięku zmieniana jest przez zamykanie lub otwieranie otworów bocznych, zmianę wysokości piszczałki lub zmianę ciśnienia powietrza. Praktyczny podział aerofonów (nazywanych również instrumentami dętymi) rozróżnia trzy grupy instrumentów:

- drewniane (np. klarnet, fagot),
- blaszane (np. trąbka puzon),
- klawiszowe (np. organy).

## 3.4. Elektrofony

Instrumenty muzyczne, w których dźwięk wytwarzany jest za pośrednictwem drgań elektrycznych. Podział elektrofonów jest uzależniony od sposobu wytwarzania drgań. Rozróżnia się dwie grupy tych instrumentów:

- elektromechaniczne instrumenty muzyczne — w których wytwarzanie drgań odbywa się na drodze elektromechanicznej (np. organy Hammonda),
- elektroniczne instrumenty muzyczne — w których drgania wytwarzane są na drodze elektrycznej (np. syntezator, sampler).

W elektrofonach wyróżnia się następujące podstawowe zespoły:

- generator elektryczny — wytwarza drgania elektryczne,
- układ korekcyjny wpływający na barwę dźwięku,
- układ sterujący umożliwiający grę na instrumencie,
- wzmacniacz zwiększający amplitudę drgań,
- układ nagłaśniający.

## 3.5. Chordofony

Grupa strunowych instrumentów muzycznych, w których rolę źródła dźwięku pełnią drgające struny. Najogólniej chordofony dzielą się na dwie grupy:

- szyjkowe — szyjka umożliwia skracanie strun (np. skrzypce, gitara),
- bezszyjkowe — z każdej struny można wydobyć tylko jeden dźwięk (np. harfa, fortepian).

Kolejny podział jest uzależniony od sposobu wydobywania dźwięku.

- Instrumenty smyczkowe — struna wprowadzana jest w stan wibracji za pomocą smyczka przesuwanego po strunie (*legato, detaché*) lub przez uderzanie w nią włosiem (*spiccato*). Poza tym struna może być pobudzana przez uderzenie jej drzewcem smyczka (*col legno*), szarpnięcie struny palcem (*pizzicato*) oraz szarpnięcie struny tak silne, by uderzyła o gryf (*klang*).
- Instrumenty szarpane — pobudzenie struny do drgań odbywa się przez jej szarpnięcie gołymi rękami lub tzw. *piórkiem*.
- Instrumenty młoteczkowe — wibracja strun wzbudzana jest uderzeniem młoteczka. Młoteczki mogą stanowić część mechanizmu instrumentu (np. fortepian) lub być trzymane w dłoniach (np. cymbały).

Dla chordofonów podstawowa wysokość tonu charakteryzuje się częstotliwością wynikającą z zależności, [19]:

$$V_0 = \frac{1}{2d}\sqrt{\frac{T}{P}}, \tag{3.1}$$

gdzie:
$d$ — długość struny,
$T$ — siła naciągu,
$P$ — gęstość materiału.

ROZDZIAŁ 4

# Parametryzacja dźwięków muzycznych

## 4.1. Parametryzacja w dziedzinie czasu

W celu właściwego opisu postaci czasowej sygnału dźwiękowego konieczne jest zdefiniowanie grupy parametrów. Deskryptory opisujące obwiednię dźwięku wyrażane są poprzez stosunek czasu trwania poszczególnych faz przebiegu do czasu trwania całego dźwięku, [9], [13]. Stosowane są również metody analizy wybranego fragmentu postaci czasowej. Istotnym problemem jest wyznaczenie momentu początku dźwięku, celem wyeliminowania możliwych zakłóceń towarzyszących podczas procesu rejestrowania sygnału. Wyznaczenie momentu rozpoczęcia dźwięku w opisie sygnałów prostokątnych lub impulsowych jest oparte o model zakładający osiągnięcie 10% maksymalnej amplitudy, [2], [10]. W ten sam sposób wyznaczany jest moment zakończenia przebiegu. Stan quasi-ustalony jest definiowany jako osiągnięcie, co najmniej 75% maksymalnej amplitudy a w trakcie jego trwania dopuszcza się dziesięcioprocentowe odchylenie amplitudy dźwięku, [2].

Przebieg sygnału muzycznego w funkcji czasu można opisać za pomocą wybranej grupy parametrów:

- $l_{tn}$— logarytm czasu narastania dźwięku, [10], [15]

$$l_{tn} = \log(t_{\max} - t_{pp}), \tag{4.1}$$

gdzie:
$t_{max}$ — czas osiągnięcia maksymalnej amplitudy dźwięku,
$t_{pp}$ — czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie początkowym;

- $l_{tk}$—logarytm czasu wybrzmiewania dźwięku

$$l_{tk} = \log(t_{pk} - t_{\max}), \tag{4.2}$$

gdzie:
$t_{pk}$ — czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie końcowym,
$t_{max}$ — czas osiągnięcia maksymalnej amplitudy dźwięku;

- $Tp$—stosunek czasu trwania transjentu początkowego do czasu trwania całego dźwięku, [16];
- $Qu$—stosunek czasu trwania stanu quasi-ustalonego do czasu trwania całego dźwięku, [16];
- $T_k$—stosunek czasu trwania transjentu końcowego do czasu trwania całego dźwięku, [16];
- $ZC$—gęstość przejść przez zero sygnału (zero crossings).

## 4.2. Parametryzacja w dziedzinie widma

Rozkład amplitud drgań harmonicznych w zależności od częstości tworzy *widmo dźwięku* decydujące o jego barwie. Zawiera ono bardzo wiele szczegółów, a zatem do celów automatycznej klasyfikacji instrumentów muzycznych konieczna jest jego parametryzacja, [17]. Podstawą przeprowadzenia parametryzacji widma są transformaty Fouriera, falkowa, cepstrum czy Wigner-Ville'a.

Wyznaczenie środka ciężkości (nazywanego również jasnością dźwięku) widma jest bardzo ważnym elementem procesu odszukania wektora cech. Zastosowanie jasności dźwięku jako parametru syntezy pozwala na uzyskanie dźwięku posiadającym widmo zbliżone do widma dźwięków naturalnych, [7]. Środek ciężkości widma wyznaczany jest według zależności:

$$Br = \frac{\sum_{i=0}^{n} A(i)i}{\sum_{i=0}^{n} A(i)}, \tag{4.3}$$

gdzie:
$i$ — częstotliwość i-tego prążka widma,
$A(i)$ — amplituda i-tej składowej.

Rozkład środka ciężkości dla grupy 8 instrumentów przedstawiono na Rys. 4.1

Istnieje możliwość scharakteryzowania widma w oparciu o pojęcie momentów widmowych. Moment widmowy k-tego rzędu opisywany jest wzorem:

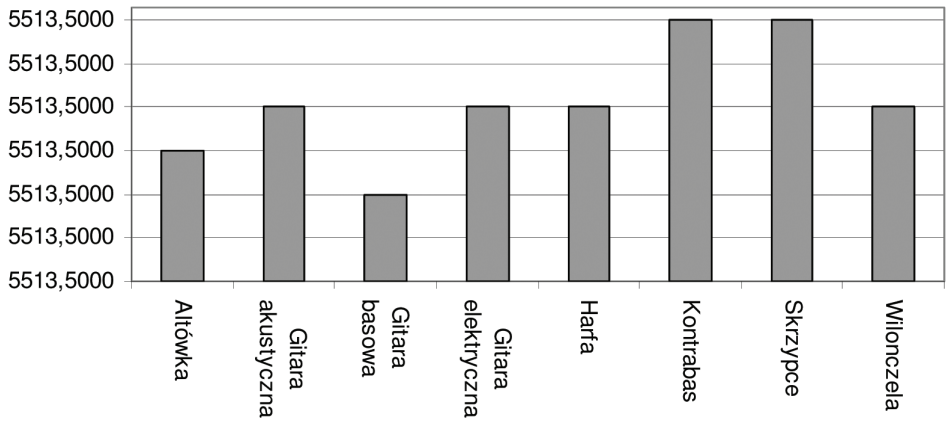
$$m_k = \sum_{i=0}^{\infty} A(i)i^k, \tag{4.4}$$

gdzie:
$i$ — częstotliwość i-tego prążka widma,
$A(i)$ — amplituda i-tej składowej.

Kolejnym ważnym parametrem jest pojęcie momentów centralnych definiowanych według wyrażenia:

$$m_k = \sum_{i=0}^{\infty} A(i)(i - Br)^k, \tag{4.5}$$

gdzie:
$Br$ — środek ciężkości widma.




Rysunek 4.1. Rozkład środka ciężkości dla wyciętego okna (n=11 025) nuty c razkreślne

Parametryzacje widma można również przeprowadzać wykorzystując informacje o ilości prążków harmonicznych przy założeniach, że:

- $i$ — indeks kolejnego prążka harmonicznego,
- $i_{max}$ — indeks prążka o maksymalnej amplitudzie,

  gdzie:

$$i_{max} \cdot i \leq n, \quad i = 2, 3, 4, \ldots$$

  $n$ — ilość prążków widma.

Ważna cechą widma są również tzw. *formanty widma,* czyli maksima obwiedni widma. Mają one wpływ na charakterystykę barwy brzmienia instrumentu. Charakterystyka ta jest związana z częstotliwością drgań własnych układu rezonansowego instrumentu a więc też z jego wielkością, [3], [14].

Podczas procesu parametryzacji widma przydatne mogą się okazać parametry statystyczne. Jednym z nich jest średnia arytmetyczna $L_n$ [dB] poziomu amplitudy $n$-tej składowej dla M ramek. Zależność tą wyraża się wzorem, [1]:

$$L_n = \frac{1}{M} \sum_{i=0}^{M-1} A(i)\,, \tag{4.6}$$

gdzie:

$A(i)$ — amplituda i-tej harmonicznej [dB].

Funkcja średniokwadratowa opisuje amplitudę składowych harmonicznych według zależności, [16]:

$$P_n = \sqrt{\frac{1}{M} \sum_{i=0}^{M-1} A_n^2(i)}\,, \tag{4.7}$$

gdzie:

$A_n^2(i) = A_0^2 10^{\left[\frac{L_n(i)}{10}\right]}$,

$A_0$ — amplituda odniesienia, odpowiadająca poziomowi 0 dB.

Średni poziom amplitudy $n$-tej składowej [dB] wyraża się wówczas wzorem, [16]:

$$A_n = 10 \log \left( \frac{A_n^2}{A_0^2} \right) . \tag{4.8}$$

Innym parametrem statystycznym może być suma procentowych zmian częstotliwości $n-$tej harmonicznej w stosunku do częstotliwości podstawowej, [1]:

$$\Delta\% = \frac{1}{M} \sum_{i=1}^{M-1} \Delta F_i(m)[\%] , \tag{4.9}$$

gdzie $\Delta F_i$ to procentowa różnica zmiany częstotliwości i-tej harmonicznej w stosunku do częstotliwości podstawowej.

Istotne informacje o widmie sygnału dźwiękowego można uzyskać analizując parametry zawartości składowych parzystych Ev i nieparzystych Od, [9]. Stosunek energii zawartej w prążkach parzystych i nieparzystych opisany jest zależnościami:

$$\mathrm{Ev} = \frac{\sqrt{\sum_{i=1}^{M} A_{2i}^2(i)}}{\sqrt{\sum_{j=1}^{N} A_j^2(j)}} , \tag{4.10}$$

$$\mathrm{Od} = \frac{\sqrt{\sum_{i=2}^{L} A_{2i-1}^2(i)}}{\sqrt{\sum_{j=1}^{N} A_j^2(j)}} , \tag{4.11}$$

gdzie:

$M$ = N/2,

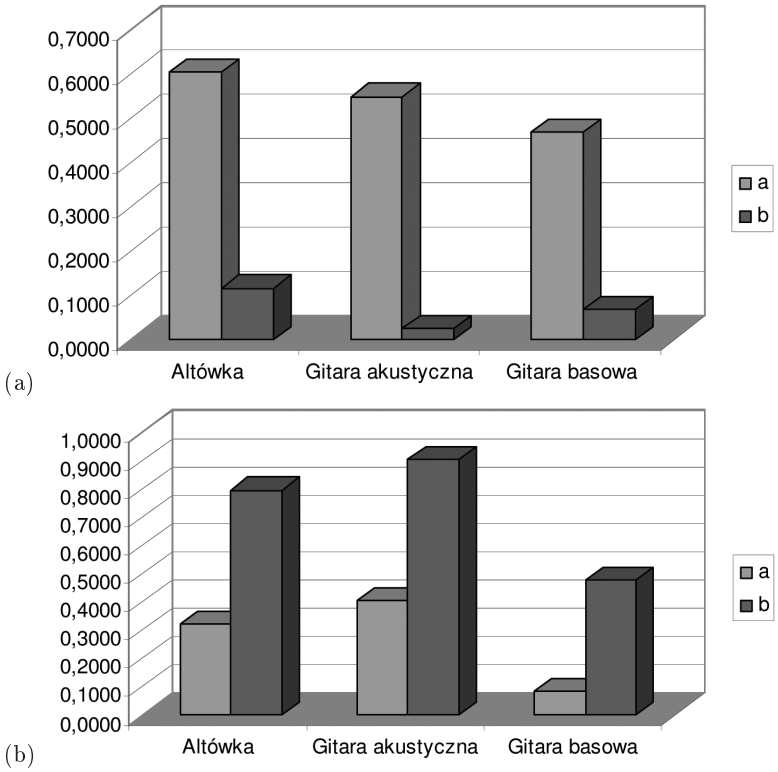
$L$ = N/2+1,

$N$ — długość okna.

Porównanie rozkładu energii w prążkach parzystych i nieparzystych dla grupy 3 instrumentów i nut a małe i b małe pokazano na Rys. 4.2.

Kolejnym parametrem opisującym widmo jest zawartość *składowych podharmonicznych*. Odpowiadają one zwykle częstotliwości właściwej dla 1/2 częstotliwości podstawowej. Jest to częste zjawisko występujące podczas gry na instrumentach dętych określane mianem *przedęć oktawowych*. Zawartość podharmonicznych w widmie odpowiadającą przedęciu oktawowemu opisywana jest zależnością, [16]:

$$f_{1/2} = \frac{\sum_{i=0}^{M} \left[ A\left(\frac{1}{2} f_1 + i f_1\right) \right]^2}{\sum_{j=1}^{N} \left[ A(j f_1) \right]^2} , \tag{4.12}$$




Rysunek 4.2. Rozkład energii w prążkach parzystych (a) i nieparzystych (b)

gdzie:
$A$ — amplituda,
$f_1$ — częstotliwość podstawowa,
$M, N$ — liczba składowych nieprzekraczających częstotliwości Nyquista.

Bardzo ciekawą i niosącą dużo informacji o widmie dźwięku jest grupa parametrów *tristimulus*, [18]. Parametry te charakteryzują widmo w kontekście zawartości składowej podstawowej, średnich oraz wyższych składowych. Całkowita głośność $N$ dźwięku może być wyrażona jako:

$$N = N_1 + N_2^4 + N_5^n \,, \tag{4.13}$$

gdzie:
$n$ — liczba dostępnych składowych,
$N_1$ — głośność podstawowej składowej,
$N_2^4$ — głośność składowych od 2 do 4,
$N_5^n$ — głośność składowych od 5 do n.

Można zdefiniować parametry *tristimulus* z wykorzystaniem współrzędnych $(x, y, z)$ — wówczas parametry te będą wyrażane zależnościami:

$$x = \frac{N_5^n}{N} \,, \tag{4.14}$$

$$y = \frac{N_2^4}{N} \,, \tag{4.15}$$

$$z = \frac{N_1}{N} \,. \tag{4.16}$$

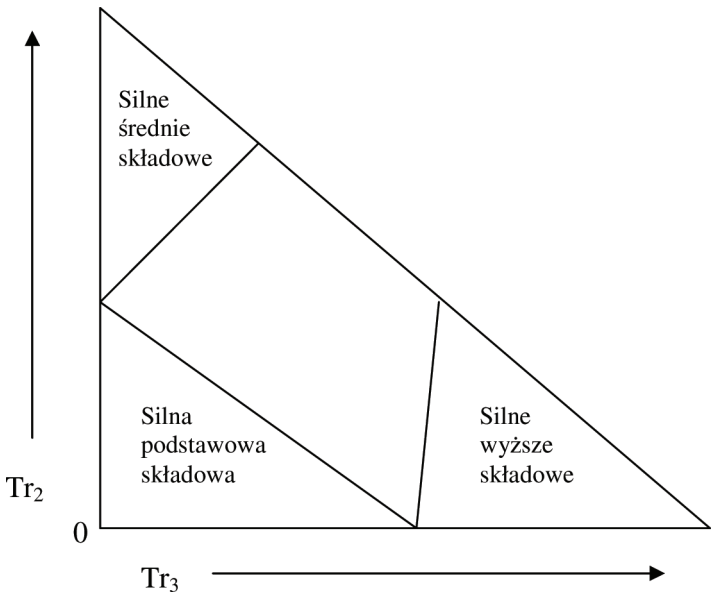
Stosowana jest również zmodyfikowana wersja parametrów *tristimulus*, która jest wyrażana zależnościami, [8]:

$$Tr_1 = \frac{A(1)^2}{\sum_{i=1}^{n} A(i)^2}, \tag{4.17}$$

$$Tr_2 = \frac{\sum_{i=2}^{4} A(i)^2}{\sum_{i=1}^{n} A(i)^2}, \tag{4.18}$$

$$Tr_3 = \frac{\sum_{i=5}^{n} A(i)^2}{\sum_{i=1}^{n} A(i)^2}. \tag{4.19}$$

Wykorzystując grupę parametrów *tristimulus* można rozróżnić dźwięki wykorzystując zawartość grup harmonicznych w widmie, co przedstawiono na Rys. 4.3.




Rysunek 4.3. Diagram parametrów tristimulus, [18]

Kolejnym podejściem do parametryzacji przebiegów dźwiękowych jest parametryzacja różnic między kolejnymi ramkami przebiegu. Procedurę tą wykonuje się zarówno w funkcji widma jak i w funkcji czasu [12], z uwzględnieniem zmian widma w czasie. Przykładem może być obserwacja czasu trwania wybrzmiewania nuty dla niższych i wyższych składowych, [11].

Parametryzacja widma może również przebiegać w oparciu o stosowane metody graficzne. Ciekawym parametrem stosowanym do danych dźwiękowych jest wymiar fraktalny (Hausdorffa) obwiedni amplitudy. Dla zbioru $X$, będącego podzbiorem zupełnej przestrzeni Euklidesowej (przestrzeni z określonym iloczynem skalarnym),

wymiar Hausdorffa definiowany jest wzorem, [16]:

$$D_H(X) = \lim_{r \to 0} \left( -\frac{\log N(r)}{\log r} \right), \tag{4.20}$$

gdzie:

$N(r)$ — najmniejsza liczba kul otwartych o promieniu r potrzebnych do pokrycia zbioru $X$.

W związku z tym, że obwiednia amplitudy widma jest krzywą ciągłą o skończonej długości, jej wymiar fraktalny wynosi 1. Dlatego w celu parametryzacji stosowany jest przybliżony wymiar fraktalny $d_{fr}$ obwiedni widma opisany zależnością, [16]:

$$d_{fr} = -\frac{\log N(\Delta s)}{\log \Delta s}, \tag{4.21}$$

gdzie $\Delta s$ to długość boku kwadratowego oczka siatki pokrywającej płaszczyznę wykresu.

Parametr ten dla ustalonej długości okna analizowanego przebiegu oraz ustalonego $\Delta s$ będzie przybierał różne wartości, które są uzależnione od szybkości zbieżności wymiaru $d_{fr}$ do 1 przy $\Delta s \to 0$.

Opisane przykłady parametrów wykorzystywanych do celów parametryzacji przebiegów dźwiękowych są jedynie wybraną grupą deskryptorów. Na podstawie wybranej grupy parametrów zostanie stworzony wektor cech, który umożliwi automatyczna klasyfikacje instrumentów szarpanych lub smyczkowych z artykulacją pizzicato.

ROZDZIAŁ 5

# Bazy danych i system zarządzania bazą danych

## 5.1. Bazy danych — pojęcia ogólne

Najprostsza i najbardziej elementarna definicja bazy danych mówi, że jest to zbiór danych istniejący przez długi czas zorganizowany przez *system baz danych* DBMS (*database management system)*, [26]. Celem takiego magazynu jest przechowywanie danych związanych z pewnym zbiorem zadań organizacyjnych. Od prawidłowo funkcjonującej bazy danych oczekuje się kilku podstawowych cech:

1. *Współdzielenie danych*, co oznacza, że składowane dane nie są przechowywane wyłącznie dla celów użytkowych jednej osoby. Zakłada się, że baza danych będzie używana przez więcej niż jedną osobę nawet w tym samym czasie. Na przykład w supermarkecie informacje o sprzedaży produktów są dostępne w systemie informacyjnym jak i zarządzania sprzedażą.

2. *Integracja danych* zakłada, że baza danych jest zbiorem informacji nie mającym niepotrzebnie powtarzających się lub zbędnych informacji. Oznacza to, że celem jest przechowywanie jednego logicznego elementu danych tylko w jednym miejscu.

3. *Integralność danych* jest konsekwencją współdzielenia danych sprowadzającą się do zapewnienia, aby baza danych dokładnie odzwierciedlała obszar analizy danych. Oznacza to, że zmiany dokonane po jednej stronie związku istniejącym w rzeczywistym świecie między obiektami będą dokładnie odzwierciedlone w zmianach dokonanych na innych stronach w tym związku.

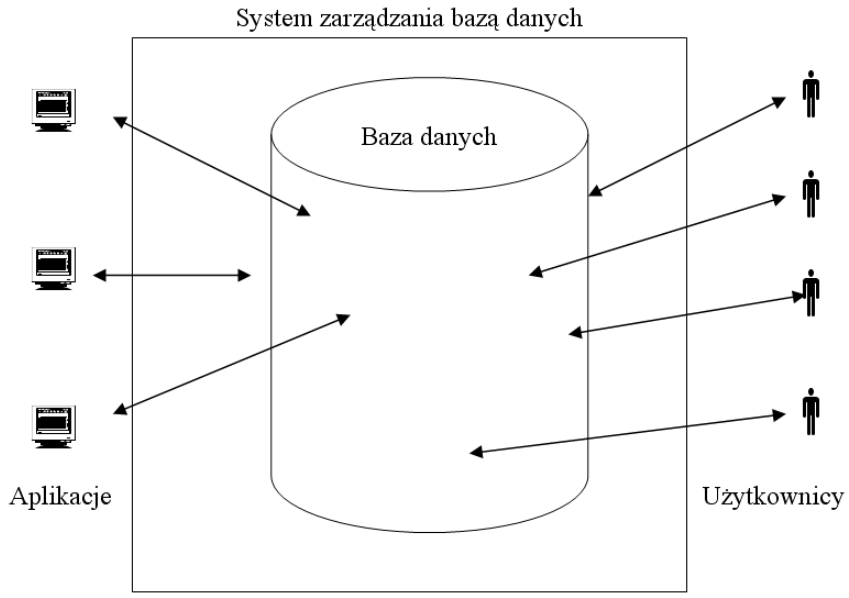
4. *Bezpieczeństwo danych* jest ściśle związane z integralnością danych, która nie może istnieć bez stosownego bezpieczeństwa danych. Najczęściej realizowane jest to za pośrednictwem określenia z pewną szczegółowością zbioru upoważnionych użytkowników w odniesieniu do całej lub pewnej części bazy danych. Przykładem może być system umożliwiający wpływy finansowe na konta klientów z wykluczeniem możliwości dokonywania zmian na tych kontach.

5. *Abstrakcja danych* oznacza, że baza danych jest abstrakcją świata rzeczywistego. Sprowadza się to do faktu, że baza danych może być traktowana jako model rzeczywistości, w którym informacje są próbą reprezentowania właściwości niektórych obiektów w świecie rzeczywistym, [27].

6. *Niezależność danych* odzwierciedla się w dążeniach do osiągnięcia sytuacji, w której organizacja danych jest niewidoczna dla użytkowników i programów użytkowych korzystających z danych, [27]. Na przykład, jeżeli dokonywane są zmiany w programie aplikacyjnym, to nie powinno mieć to wpływu na strukturę danych używanych przez ten program.

## 5.2. Systemy zarządzania bazą danych

System zarządzania bazą DBMS (*database management system)* danych jest zorganizowanym zbiorem narzędzi umożliwiającym dostęp i zarządzanie jedną lub kilkoma bazami danych. Wynika z tego, że system zarządzania bazą danych jest skomputeryzowanym systemem przechowywania rekordów, czyli jest to skomputeryzowany system, którego zasadniczym zadaniem jest przechowywanie informacji i udostępnianie jej na każde życzenie. Na Rys. 5.1 przedstawiono uproszczony schemat systemu zarządzania bazą danych z uwzględnieniem zasadniczych elementów, [25].

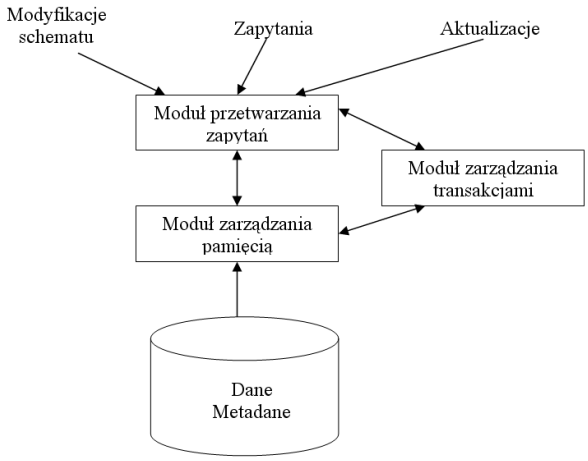



Rysunek 5.1. Uproszczony schemat systemu zarządzania bazą danych

### 5.2.1. Składowe DBMS

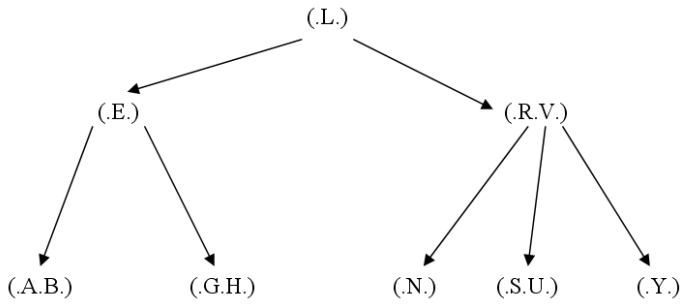
W każdym systemie zarządzania bazą danych występują istotne elementy, których uporządkowane współdziałanie gwarantuje stabilna pracę DBMS. Na Rys. 5.2 przedstawiono główne elementy systemu zarządzania bazą danych, [26].

Na Rys. 5.2 widać blok reprezentujący miejsce przechowywania danych i *metadanych*, które opisują strukturę danych (np. typy poszczególnych atrybutów nazwy relacji itp.). System DBMS obsługuje również w tym bloku *indeksy* danych, które są pewną strukturą danych pomagającą w szybkim odnajdywaniu określonych informacji. Podstawowa idea *indeksu prostego* polega na zastosowaniu dodatkowego pliku o dwóch polach, dodawanego do systemu bazy danych. Pierwsze pole indeksu zawiera




Rysunek 5.2. Główne elementy DBMS

posortowaną listę logicznych wartości kluczy, drugie pole — listę adresów bloków dla wartości kluczy. Na indeksie wykonywane jest przetwarzanie wykorzystujące algorytm wyszukiwania binarnego, [27]. Bardziej rozpowszechniona strukturą indeksów w systemach baz danych są *B-drzewa (drzewo wyważone)*, oparte na hierarchicznej strukturze danych. Na przykład analizując plik, w którym logiczne wartości klucza są literami alfabetu, to każdy rekord w indeksie składa się z jednej lub więcej wartości klucza poprzeplatanych wskaźnikami do innych rekordów w drzewie. Drzewo nazywamy *drzewem wyważonym* (B-drzewo, gdzie „B"oznacza „balanced") jeśli jego liście znajdują się w tej samej odległości od korzenia drzewa. Przykład B-drzewa przedstawiono na Rys. 5.3, w którym wartości klucza są właściwe dla liter alfabetu, [27].




Rysunek 5.3. B-drzewo

Kolejnym głównym elementem DBMS jest również *moduł zarządzania pamięcią*, który ma za zadanie wybierać właściwe dane z pamięci i (jeżeli zaistnieje taka potrzeba) dostosować je do wymagań modułów wyższych poziomów systemu. W prostych systemach baz danych jest tym samym, czym jest system plików systemu operacyjnego. Poza tym moduł ten bezpośrednio zarządza przestrzenią na dysku. Moduł zarządzania pamięcią składa się z:

1. *Modułu zarządzania plikami*, który przechowuje dane o miejscu zapisania plików na dysku,

2. *Modułu zarządzania buforami*, który obsługuje pamięć operacyjną.

Moduł zarządzania plikami przekazuje bloki danych z dysku, natomiast moduł zarządzania buforami wybiera w pamięci operacyjnej strony, które zostaną przydzielone dla wybranych bloków.

**Moduł przetwarzania zapytań** obsługuje nie tylko zapytania, ale również aktualizacje danych i metadanych. Funkcją tego modułu jest wybranie najlepszego sposobu wykonania zadanych operacji oraz wydaniu poleceń do modułu zarządzania pamięcią, który je wykona. Polecenia są wyrażane zazwyczaj w języku wysokiego poziomu (np. w języku SQL). Najtrudniejszą operacją jest tzw. *optymalizacja zapytań*, tzn. dobór dobrego algorytmu, którego wykonanie gwarantuje właściwą odpowiedź w możliwie najkrótszym czasie.

**Moduł zarządzania transakcjami** jest odpowiedzialny za spójną pracę systemu baz danych. Powinien on zagwarantować, że przetwarzane jednocześnie zapytania nie będą sobie wzajemnie przeszkadzały, oraz powinien zabezpieczyć dane przed ich utratą — nawet, jeśli nastąpi awaria systemu. Moduł ten ściśle współdziała z modułem obsługi zapytań, ponieważ musi mieć dostęp do szczegółów dotyczących danych, na których przetwarza się bieżące zapytania. System DBMS umożliwia użytkownikowi łączenie jednego lub więcej zapytań, bądź modyfikacji, w *transakcje*, która jako grupa poleceń przeznaczona jest do wykonania razem w jednym ciągu i traktowana jest jako operacja jednostkowa, [26]. Aby transakcja została przeprowadzona poprawnie moduł zarządzania transakcjami musi przeprowadzić ją zgodnie z określonymi właściwościami poprawnej transakcji:

1. *Niepodzielność (atomicity)* oznacza, że cała transakcja powinna być przeprowadzona w całości lub wcale, co za tym idzie, że odrębne elementy transakcji nie zostaną uwzględnione przez system bazy danych. Doskonałym przykładem jest operacja na kontach bankowych, której system nie może uwzględnić w przypadku pobrania gotówki z bankomatu z pominięciem zapisu tego faktu na koncie użytkownika.

2. *Spójność (consistency)* przeprowadzonych transakcji oznacza, że dane muszą zaspokajać oczekiwania użytkowników systemu zarządzania bazą danych. Tą właściwość transakcji można zilustrować przykładem systemu rezerwacji miejsc hotelowych, który nie może dopuścić do sytuacji, że jeden pokój hotelowy zostanie zarezerwowany dla więcej niż jednego klienta w tym samym czasie. Moduł zarządzania transakcjami musi zagwarantować, że po zakończeniu przetwarzania transakcji baza danych spełnia wszystkie warunki niesprzeczności, [26].

3. *Izolacja (isolation)* transakcji jest związana z sytuacją, gdy dwie transakcje są przetwarzane jednocześnie a ich działania nie mogą na siebie wzajemnie wpływać. Na przykład, gdy w sklepie internetowym zostały zgłoszone dwie oferty zakupu produktu, a na stanie magazynowym znajduje się ostatni egzemplarz to tylko jedno żądanie powinno zostać obsłużone. Zabezpieczy to system przed sprzedażą jednego egzemplarza produktu kilku osobom.

4. *Trwałość (durability)* zapewnia, że w momencie zakończenia przeprowadzanej transakcji jej wynik nie może być utracony — nawet w przypadku awarii systemu.

Równie istotnymi elementami składowymi systemu DBMS są trzy rodzaje wejść do systemu:

1. *Zapytania*, które spełniają funkcję pytania o dane. Mogą one występować w dwojakiej formie:
   (a) Poprzez interfejs zapytań bezpośrednich — realizowany zazwyczaj za pośrednictwem języka SQL.
   (b) Poprzez interfejs programów użytkowych wywołujący procedury DBMS tworzące zapytania do bazy danych.
2. *Aktualizacje*, rozumiane jako zbiór operacji odpowiadających za zmiany danych. Podobnie jak zapytania można je przeprowadzać z poziomu interfejsu zapytań bezpośrednich lub interfejsu programów użytkowych. Również mogą być przeprowadzane z wykorzystaniem języka SQL na podstawie poleceń dynamicznych *insert, update* itp.
3. *Modyfikacje schematu*, przeprowadzane tylko przez specjalnie uprawniona osobę pełniącą funkcję *administratora* systemu baz danych. Użytkownik posługujący się uprawnieniami administracyjnymi może zmieniać schemat bazy danych rozumiany jako zbiór obiektów, których właścicielem jest jeden użytkownik bazy danych (np. tabele, indeksy, klastry itp.), [28].

## 5.3. Relacyjny model danych

Model relacyjny jest sposobem patrzenia na dane, jest to przepis na sposób reprezentowania danych (za pomocą tabel) oraz na sposób manipulowania taką reprezentacją danych. Ściślej mówiąc, model relacyjny dotyczy trzech aspektów danych, a mianowicie: struktury danych, integralności danych i operowania danymi, [25]. W relacyjnym modelu danych jest tylko jedna struktura danych — tabela (nazywana również *relacją*). Konstrukcja tabeli w ramach bazy danych powinna spełniać następujące warunki:

1. Każda tabela w bazie danych ma jednoznaczną nazwę.
2. Każda kolumna (*atrybut)* ma jednoznaczną nazwę w ramach jednej tabeli.
3. Wszystkie wartości w jednej kolumnie musza być tego samego typu.
4. Porządek układu kolumn w ramach tabeli nie jest istotny.
5. Każdy wiersz (*krotka)* w tabeli musi być różny. Nadmiarowość *(redundancja)* wierszy nie jest dozwolona w obrębie jednej tabeli.
6. Porządek wierszy nie ma znaczenia.
7. Każde pole leżące na przecięciu kolumny i wiersza w tabeli powinno zawierać tzw. wartość *atomową* — tzn. niepodzielną, elementarną.

Nazwa relacji (tabeli) oraz jej zbiór atrybutów nazywa się *schematem relacji.* Zapisywany jest on za pomocą nazwy relacji, po której wypisana jest lista atrybutów ujęta w nawiasy okrągłe. Schemat przykładowej relacji *ksiazki* zostanie, zatem zapisany:

*ksiazki(tytul, cena, oprawa, IDwydawnictwa, IDautora)*

W każdej relacji powinien być zdefiniowany *klucz główny (primary key)*, reprezentowany przez jedną lub więcej kolumn, w których wartości jednoznacznie identyfikują każdy wiersz tabeli. Klucz główny jest wybierany ze zbioru tzw. *kluczy kandydujących*, które mogą wystąpić jako identyfikator wierszy w tabeli. Każdy klucz główny (również kandydujący) musi posiadać dwie właściwości:

1. Unikatową wartość w obrębie tabeli *(unique).*
2. Nie może mieć wartości *null.*

Podstawową jednostką danych w modelu relacyjnym są elementy danych nierozkładalnych (atomowych), a zbiór elementów danych tego samego typu nazywamy *dziedziną.* Dziedziny są więc zbiorami wartości, z których pochodzą elementy pojawiające się w atrybutach relacji. W nawiązaniu do nowoczesnych języków programowania, dziedzina to nic innego jak *typ danych* zilustrowany poniższym przykładem:

*type marka = (fiat, peugeot, ford, suzuki)*
*var samochod: marka;*

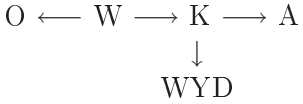
Łączenie danych przechowywanych w różnych tabelach odbywa się za pośrednictwem *kluczy obcych (foreign key).* Klucz obcy jest kolumną (lub grupą kolumn), która czerpie swoje wartości z tej samej dziedziny, co klucz główny tabeli powiązanej z nią w bazie. Wyrażenie struktury relacji może być sprecyzowane jako ciąg szablonów złożonych z nazw relacji, atrybutów i deklaracji kluczy głównych i obcych, [27]. Poniższy przykład ilustruje fragment schematu bazy biblioteki uwzględniającej dziedzinę danych, zbiór osób wypożyczających, zarejestrowane wypożyczenia oraz książki:

***Domains***
*PK_wypozycz: Integer*
*PK_osoby: Integer*
*PK_ksiazki: Integer*
*PK_autora: Integer*
*PK_wydawnictwa: Integer*
*Data_wyp: Date*
*Data_zwr: Date*
*Data_ur: Date*
*Wydano: Date*
*Nazw: Character*
*Im: Character*
*Miasto: Character*
*N_ulicy: Character*
*Kod: Character*
*Tyt: Character*
*Cena: Decimal (7,2)*
***Relation*** *OSOBY*
***Attributes***
*IDosoby: PK_osoby*

*Nazwisko: Nazw*
*Imie: Im*
*Data_ urodze: Data_ ur*
*Miejscowosc: Miasto*
*Ulica: N_ ulicy*
*Kod_ pocz: Kod*
***Primary Key*** *IDosoby*
***Relation*** *WYPOZYCZALNIA*
***Attributes***
*IDwypozycz: PK_ wypozycz*
*IDosoby: PK_ osoby*
*IDksiazki: PK_ ksiazki*
*Od_ dnia: Data_ wyp*
*Do_ dnia: Data_ zwr*
***Primary Key*** *IDwypozycz*
***Foreign Key*** *IDosoby references OSOBY*
***Foreign Key*** *IDksiazki references KSIAZKI*
***Relation*** *KSIAZKI*
***Attributes***
*IDksiazki: PK_ ksiazki*
*Idautora: PK_ autora*
*IDwydaw: PK_ wydawnictwa*
*Rok_ wydania: Wydano*
*Tytul: Tyt*
*Cena: Cena*
***Primary Key*** *IDksiazki*
***Foreign Key*** *IDautora references AUTORZY*
***Foreign Key*** *IDwydaw references WYDAWNICTWA*

Wartość klucza obcego jest referencją (odwołaniem) do krotki, która zawiera wartość odpowiadającą mu w atrybucie klucza głównego. Bardzo istotną sprawą jest zapewnienie, aby baza danych nie zawierała żadnych niedopuszczalnych wartości klucza obcego (np. wartości niezwiązanej z kluczem głównym tabeli nadrzędnej). Problem ten nazywany jest problemem *integralności referencyjnej (referential constraint)*, [25]. Rozważając przedstawiony wyżej fragment schematu bazy biblioteka możemy przedstawić istniejące więzy referencyjne za pomocą *diagramu referencyjnego* zilustrowanego na Rys. 5.4:
Na Rys.5.4 tabele są oznaczone:

O — tabela *OSOBY*,
W — tabela *WYPOZYCZALNIA*,
K — tabela *KSIAZKI*,
A — tabela *AUTORZY*,
WYD — tabela *WYDAWNICTWA*.




Rysunek 5.4. Diagram referencyjny fragmentu bazy biblioteka

Każda strzałka symbolizuje, że w tabeli, z której ona wychodzi jest zdefiniowany klucz obcy *(FK)*, który odwołuje się do klucza głównego tabeli, na którą wskazuje.

### 5.3.1. Algebra relacyjna

Algebra relacyjna składa się ze zbioru operatorów, takich jak *złączenie*, których argumentami są relacje i które w wyniku również zwracają relacje (tabele). W pracy [29] twórca relacyjnego modelu zarządzania bazami danych, E.F. Codd zdefiniował tzw. „oryginalną" algebrę obejmującą zbiór ośmiu operatorów, podzielonych na dwie grupy:

1. Tradycyjne operatory operacji na zbiorach: *suma, przecięcie, różnica, iloczyn kartezjański.*
2. Operatory relacyjne: *restrykcja, rzut, złączenie* i *iloraz.*

**Suma** Według matematycznej definicji suma dwóch zbiorów jest to zbiór wszystkich elementów należących do jednego lub drugiego zbioru. Suma jest definiowana zależnością, [31]:

$$A \cup B = \{x : x \in A \text{ lub } x \in B \text{ lub x należy do obu zbiorów}\} \tag{5.1}$$

Ponieważ relacja jest zbiorem (zbiorem krotek) jest możliwe utworzenie sumy dwóch relacji. Uzyskany wynik jest zbiorem wszystkich krotek występujących w jednej relacji bądź obu relacjach. Wynika z tego, że sumą *(union)* dwóch relacji zgodnych typów $A \cup B$ *(A UNION B)* jest relacja mająca ten sam nagłówek co $A$ bądź $B$, treść natomiast złożoną złożona ze zbioru wszystkich krotek należących do $A$, $B$ lub obu tych relacji, [25].

Jeżeli w bazie danych *biblioteka* w relacjach *AUTORZY* i *WYDAWNICTWA* występują krotki:

(a) A

| **Miejscowość** |
|---|
| Kraków |
| Warszawa |

(b) WYD

| **Miejscowość** |
|---|
| Warszawa |
| Bydgoszcz |
| Warszawa |

To wówczas relacja *A UNION WYD* jest zbiorem miast, w których mieszkają autorzy książek lub, w których znajdują się wydawnictwa. Należy również podkreślić, że powtarzające się krotki zostaną wyeliminowane. A zatem:

| *A UNION WYD* |
|---|
| **Miejscowość** |
| Kraków |
| Warszawa |
| Bydgoszcz |

**Przecięcie** Przecięcie zbiorów $A \cap B$ definiowane jest zależnością, [31]:

$$A \cap B = \{x : x \in A \wedge x \in B\} \tag{5.2}$$

Przecięciem *(intersection)* dwóch relacji zgodnych typów $A$ i $B$ *(A INTERSECT B)* jest relacja mająca ten sam nagłówek co $A$ bądź $B$, treść natomiast jest złożona

ze zbioru wszystkich krotek $z$ należących zarówno do relacji $A$, jak i do $B$, [25]. Odnosząc w/w stwierdzenie do przykładu bazy *biblioteka A INTERSECT WYD* jest relacja zawierająca miasta, w których mieszkają autorzy książek, i w których znajdują się wydawnictwa. A zatem:

| *A INTERSECT WYD* |
|---|
| **Miejscowość** |
| Warszawa |
| Warszawa |
| Warszawa |

**Różnica** Dla danych zbiorów $A$ i $B$ różnica $A\backslash B$ jest zdefiniowana jako zbiór obiektów należących do $A$ i nie należących do $B$. Różnica zbiorów opisywana jest zależnością:

$$A\backslash B = \{x : x \in A \wedge x \notin B\} = \{x \in A : x \notin B\}. \tag{5.3}$$

Jest to, zatem zbiór powstały przez usunięcie ze zbioru $A$ tych wszystkich elementów zbioru $A$, które należały też do $B$, [31].

Różnicą *(difference)* dwóch relacji $A$ i $B$ zgodnych typów, w takiej kolejności *(A MINUS B)* jest relacja mająca ten sam nagłówek co $A$ lub $B$, natomiast jej treść jest złożona ze zbioru wszystkich krotek $z$ należących do relacji $A$, a nie należących do $B$. Przeprowadzając różnicę na opisywanych relacjach bazy *biblioteka (A MINUS WYD)* otrzymujemy listę miast, w których mieszkają autorzy książek i w których nie ma wydawnictw. A więc:

| *A MINUS WYD* |
|---|
| **Miejscowość** |
| Kraków |

Natomiast sytuacja odwrotna przedstawia się:

| *WYD MINUS A* |
|---|
| **Miejscowość** |
| Bydgoszcz |

**Iloczyn kartezjański** Jest uporządkowaną parą $(x, y)$ elementu $x$ ze zbioru $X$ i elementu $y$ ze zbioru $Y$. Element $x$ jest pierwszym elementem pary uporządkowanej *(poprzednikiem)*, $y$ jest drugim elementem *(następnikiem)* i kolejność tych elementów jest istotna. Oznacza to, że $(x_1, y_1) = (x_2, y_2)$ wtedy i tylko wtedy, gdy $x_1 = x_2$ i $y_1 = y_2$. Iloczyn kartezjański *(produkt)* zbiorów $X$ i $Y$ wyrażony jest jako, [31]:

$$X \times Y = \{(x, y) : x \in X \wedge y \in Y\} \tag{5.4}$$

Iloczyn kartezjański dwóch relacji jest zbiorem uporządkowanych par krotek z uwzględnieniem domknięcia. Oznacza to, że iloczyn kartezjański w algebrze relacyjnej jest rozszerzoną postacią operatora, w którym każda uporządkowana para krotek zostaje zastąpiona pojedynczą krotką, powstałą z wykorzystaniem konkatencji (*coalescing*) obu krotek. W opisywanym przykładzie konkatenacja oznacza sumę w sensie teorii zbiorów, a nie w sensie algebry relacyjnej. Znaczy to, że jeżeli są dwie krotki:

$$\{\langle A1 : a1\rangle, \langle A2 : a2\rangle, \ldots, \langle Am : am\rangle\},$$

oraz

$$\{\langle B1 : b1\rangle, \langle B2 : b2\rangle, \ldots, \langle Bm : bm\rangle\},$$

to ich konkatencją jest pojedyncza krotka

$$\{\langle A1 : a1\rangle, \langle A2 : a2\rangle, \ldots, \langle Am : am\rangle, \langle B1 : b1\rangle, \langle B2 : b2\rangle, \ldots, \langle Bm : bm\rangle\}.$$

Poza tym występuje również wymaganie, aby relacja wynikowa miała właściwie utworzony nagłówek, [25]. Związane jest to z sytuacją, gdy relacje mają identyczne atrybuty. Doszłoby wówczas do wyniku, w którym nagłówek relacji wyjściowej byłby opisywany przez dwie identyczne nazwy (związane z nagłówkami relacji wejściowych), a zatem nie spełniałby podstawowych wymagań. W przypadku konieczności utworzenia iloczynu kartezjańskiego dwóch relacji, które mają identyczne nazwy atrybutów konieczne jest zastosowanie operatora *RENAME* w celu ich zmiany.

Definiujemy iloczyn kartezjański (*Cartesian product*) dwóch relacji $A$ i $B$ *(A TIMES B)*, gdzie $A$ i $B$ nie mają wspólnych nazw atrybutów, jako relację z nagłówkiem będącym konkatencją dwóch nagłówków $A$ i $B$ i treścią będącą zbiorem krotek $t$ takich, że krotka $t$ jest konkatencją krotki $a$ należącej do relacji $A$ i krotki $b$ należącej do relacji $B$, [25].

**Restrykcja** Restrykcja (*restrict*) jest tożsama z pojęciem *θ-restykcji*, w którym „$\theta$" jest dowolnym, prostym skalarnym operatorem porównania. W operacji złączenia naturalnego (wykorzystywanego w bazach danych znacznie częściej niż iloczyn kartezjański) należy realizować połączenie relacji przez utworzenie par krotek, które odpowiadają sobie w określony sposób. Oznacza to, że operacja naturalnego złączenia dwóch relacji $A$ oraz $B$, polega na połączeniu w pary tych krotek, które maja identyczne wartości dla określonych atrybutów, [26]. Opisywane złączenie restrykcji *(złączenia*$\theta$ określający zadany warunek) relacji $A$ oraz $B$ oznaczamy symbolem $A \triangleright \triangleleft_\theta B$ lub $A\theta B$. Operator restrykcji daje „poziomy" podzbiór danej relacji, to znaczy, że jest nim taki podzbiór krotek, które spełniają zadany warunek. Operacja restrykcji realizowana z poziomu języka SQL jest definiowana w klauzuli *WHERE* umożliwiając realizację porównań z wykorzystaniem zdefiniowania dowolnego wyrażenia logicznego.

**Rzut** Rzut (*projekcja*) relacji $R$ na $X, Y, \ldots, Z$ gdzie $X, Y, \ldots, Z$ są atrybutami $R$, jest relacją z nagłówkami $\{X, Y, \ldots, Z\}$ i treścią, będąca zbiorem wszystkich krotek $\{X : x, Y : y, \ldots, Z : z\}$ takich, że krotka ta występuje w relacji $R$ z wartością $x$ dla atrybutu $X$, $y$ dla $Y, \ldots, z$ dla $Z$. A zatem operator rzutu wyznacza pionowy podzbiór danej relacji, [25]. Wynika z tego, że za pośrednictwem operatora projekcji otrzymujemy podzbiór krotek relacji $R$ uzyskany przez eliminację wszystkich atrybutów niewymienionych na liście rzutowania. Składnia operatora projekcji jest następująca:

$$PROJECT\ \langle\text{nazwa tabeli}\rangle[\langle\text{lista kolumn}\rangle] \rightarrow \langle\text{tabela wynikowa}\rangle$$

**Złączenie** W algebrze relacyjnej operacja złączenia (*join*) relacji występuje w wielu odmianach. Operator złączenia po stronie operuje dwoma relacjami jako argumentami wejściowymi zwracając jedną relację wynikową. Najbardziej popularnym (i najczęściej używanym w praktyce) złączeniem jest *złączenie naturalne*, które jest iloczynem kartezjańskim z uwzględnieniem selekcji, po którym występuje rzut. Rozważając dwie relacje $X$ i $Y$ o schematach $X(a, b)$ oraz $Y(i, j, b)$ złączenie naturalne można opisać regułą:

$$J(a, b, i, j) \leftarrow X(a, b)\ \text{AND}\ Y(i, j, b)$$

Operacja złączenia jest łączna i przemienna. W przypadku złączenia naturalnego trzech relacji prawdziwe jest:

$$(A \text{ JOIN } B) \text{ JOIN } C$$

oraz

$$A \text{ JOIN } (B \text{ JOIN } C).$$

Jest możliwy również uproszczony zapis do postaci:

$$A \text{ JOIN } B \text{ JOIN } C.$$

Ponadto wyrażenia:

$$A \text{ JOIN } B.$$

oraz

$$B \text{ JOIN } A$$

są równoważne.

W celu uniknięcia wszelkich niedogodności związanych z interpretacją składowanych danych wykorzystuje się również inny wariant złączeń relacji, a mianowicie *złączenia zewnętrzne*. Ten rodzaj złączeń znajduje zastosowanie w sytuacji, gdy pewna krotka (lub zbiór krotek) $t$ relacji $R$ nie pasuje do żadnej krotki relacji $S$, to w wyniku złączenia naturalnego nie będzie informacji o jej istnieniu. Może to doprowadzić do zgubienia ważnych informacji, np. wyłonienie z grupy autorów książek w relacji $A$ takich, którzy nie napisali żadnej publikacji spośród wszystkich przechowywanych w relacji $K$. Wykorzystując mechanizm złączenia zewnętrznego uzyskujemy relacje oparte na złączeniu naturalnym z uwzględnieniem krotek z relacji $R$ i nie pasujących z relacji $S$ (tzw. krotkami *wiszącymi*). W związku z tym, że w relacji wynikowej muszą występować wartości atrybutów dwóch relacji, zatem w dołączonej krotce brakujące atrybuty są uzupełniane wartościami *NULL*. Występują trzy typy złączeń zewnętrznych, [27]:

- *Lewostronne złączenie zewnętrzne*, które zachowuje nie pasujące wiersze z tabeli będącej pierwszym argumentem operatora złączenia.
- *Prawostronne złączenie zewnętrzne* będące odwrotnością złączenia lewostronnego.
- *Obustronne złączenie zewnętrzne* zwracające listę zarówno krotek tworzących pary w relacji $R$ i $S$ oraz listę krotek *wiszących* — zarówno po stronie relacji $R$ jaki i relacji $S$.

**Iloraz** Jeżeli mamy dwie relacje $R$ i $S$ i relacje te mają nagłówki

$$\{X1, X2, \ldots, Xm, Y1, Y2, \ldots, Yn\}$$

oraz

$$\{Y1, Y2, \ldots, Yn\}$$

to oznacza, że atrybuty $Y1$, $Y2, \ldots, Yn$ są wspólne, a relacja $R$ ma ponadto atrybuty $X1$, $X2, \ldots, Xm$, $S$ natomiast nie ma żadnych innych atrybutów. Relacje $R$ i $S$

reprezentują odpowiednio dzielną i dzielnik. Załóżmy ponadto, że odpowiadające sobie atrybuty są określone na tej samej dziedzinie. Przyjmując, że $\{X1, X2, \ldots, Xm\}$ oraz $\{Y1, Y2, \ldots, Yn\}$ stanowią atrybuty złożone $X$ i $Y$, to wówczas iloraz $R$ przez $S$

$$R \text{ DIVIDEBY } S$$

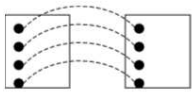
jest relacją z nagłówkiem $\{X\}$ i treścią złożoną ze wszystkich krotek $\{X : x\}$ takich, że krotka $\{X : x, Y : y\}$ występuje w $R$ dla wszystkich krotek $\{Y : y\}$ występujących w $B$. Mówiąc inaczej, wynik składa się z tych wartości $X$ z relacji $R$, których odpowiadające wartości $Y$ (w $R$) obejmują wszystkie wartości $Y$ z $S$, [25].

### 5.3.2. Związki

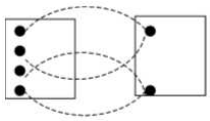
Związek jest zdefiniowany jako „połączenie (*association*) między encjami" [33], które są rozumiane jako pewien aspekt świata rzeczywistego, który można odróżnić od innych aspektów świata rzeczywistego. A zatem *encja* może być obiektem fizycznym (takim jak dom, samochód itp.) lub zdarzeniem występującym w świecie rzeczywistym (np. sprzedaż nieruchomości). Związek może występować między wydziałami a pracownikami zatrudnionymi w danym wydziale firmy. Wynika z tego, że w jednym oddziale firmy może być zatrudnionych $n$ pracowników. W teorii baz danych stwierdza się, że *encje* objęte danym związkiem są uczestnikami tego związku a liczba uczestników związku jest nazywana jego stopniem (liczebnością). Jeżeli $R$ będzie typem związku, w którym występuje jako uczestnik encja $E$ i każda instancja $E$ bierze udział w przynajmniej jednej instancji $R$, to mówi się, że uczestnictwo $E$ w $R$ jest wymagane (*total*), w przeciwnym razie jest ono opcjonalne, [25]. Wynika z tego, że opcjonalne uczestnictwo *encji* w związku jest wówczas, gdy istnieje przynajmniej jedna instancja *encji*, która nie bierze udziału w związku. Jeżeli wszystkie instancje *encji* biorą udział w związku, wówczas uczestnictwo jest uznawane wymagane.

Liczebność związków dotyczy liczby instancji, które biorą udział w danym związku. Rozróżniamy typy powiązań:

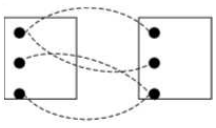
- powiązania jedno jednoznaczne typu 1 do 1 („1 - 1")



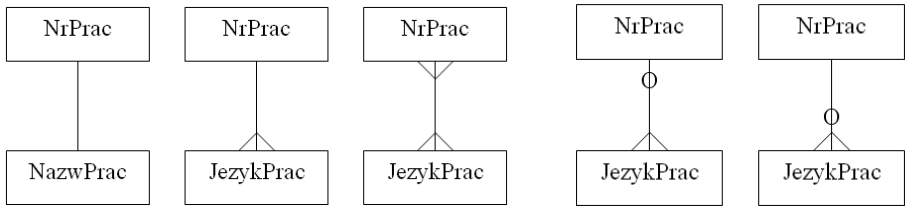
- powiązania jednoznaczne 1 do n („1 - n" )



- powiązania wieloznaczne n do m („n - m")



Opisywane modele danych są reprezentowane graficznie jako diagramy związków encji (*diagramy E-R*), na których zaznaczona jest zarówno liczebność jak i opcjonalność. Przykładowy fragment diagramu E-R przedstawiono na Rys. 5.5.




Rysunek 5.5. Uczestnictwo w związkach: a) liczebność, b) opcjonalność

### 5.3.3. Zależności funkcyjne i niefunkcyjne (wielowartościowe)

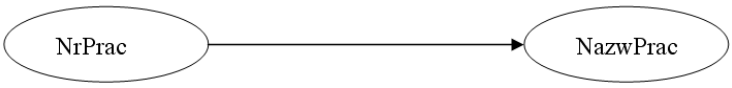
Najważniejszy rodzaj więzów związany z modelem relacyjnym dotyczy więzów jednoznaczności, nazywany również *zależnością funkcyjną*. Mówimy, że jeśli dwie krotki relacji $R$ są zgodne dla atrybutów $A_1, A_2, \ldots, A_n$ (tzn. obie krotki mają takie same wartości składowych dla wymienionych atrybutów), to muszą być również zgodne w pewnym innym atrybucie $B$. Taki rodzaj zależności zapisywany jest formalnie w postaci $A_1, A_2, \ldots, A_n \rightarrow B$, i oznacza, że $A_1, A_2, \ldots, A_n$ określają funkcyjnie $B$ na przykład:

$$\begin{aligned} A_1 A_2 \ldots A_n &\rightarrow B_1, \\ A_1 A_2 \ldots A_n &\rightarrow B_2, \\ &\ldots\ldots\ldots\ldots\ldots \\ A_1 A_2 \ldots A_n &\rightarrow B_m, \end{aligned}$$

to taki zbiór zależności przedstawiany jest również jako, [26]:

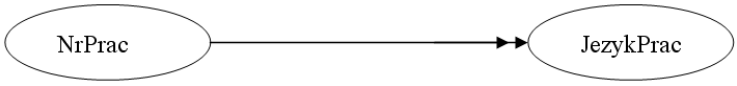
$$A_1 A_2 \ldots A_n \rightarrow B_1 B_2 \ldots B_m \,.$$

A zatem zbiór atrybutów $A$ jest zależny funkcyjnie od zbioru $B$, gdy z każdą konfiguracją wartości atrybutów z $A$ jest związana, co najwyżej jedna konfiguracja wartości z $B$. Przykładem jest sytuacja, gdy nazwisko pracownika jest funkcyjnie zależne od numeru pracownika. Opisywany przykład przedstawiono na Rys. 5.6.




Rysunek 5.6. Przykład zależności funkcyjnej

Podczas procesu modelowania bazy danych nie wszystkie zależności są zależne funkcyjnie. Kolejną (bardzo często występującą) zależnością jest *zależność niefunkcyjna (wielowartościowa)*. Mówimy, że element danych $A$ jest niefunkcyjnie zależny od elementu danych $B$, jeżeli dla każdej wartości elementu danych $B$ istnieje ograniczony zbiór wartości elementu danych $A$, [27]. Innymi słowy: zbiór atrybutów $A$ jest zależny wielowartościowo od zbioru $B$, gdy z każdą konfiguracją wartości atrybutów z $A$ jest związany zbiór konfiguracji wartości z $B$ niezależnie od wartości pozostałych atrybutów. Sytuacja taka może być zilustrowana przykładem, gdy jeden pracownik włada kilkoma językami obcymi, co przedstawiono na Rys. 5.7.




Rysunek 5.7. Przykład zależności wielowartościowej

### 5.3.4. Normalizacja

Podstawowe reguły normalizacji baz danych zostały przedstawione przez E.F. Codd'a w 1970 roku i wyrażone jak trzy postacie normalne baz danych: pierwsza, druga i trzecia. Dalszy proces przekształcania projektu bazy danych doprowadził do zdefiniowania postaci normalnej Boyce'a-Codda oraz czwartej i piątej postaci normalnej. Podstawowym celem procesu normalizacji baz danych (nazywanym ogólnie *normalizacją)* jest wyeliminowanie redundacji, czyli nadmiarowości. Mówimy, że dana relacja jest w określonej postaci normalnej, jeżeli spełnia zadane warunki określone dla danej postaci normalnej (*NF—normal form*). Jest istotne, że relacja będąca w pewnej postaci normalnej może zostać przekształcona na zbiór relacji w bardziej oczekiwanej postaci pamiętając przy tym, że procedura ta jest odwracalna, [30]. Znaczy to, że wynik procedury normalizacji można odwrócić do stanu wcześniejszego bez utraty danych. Jeżeli w zbiorze tabelarycznym występuje redundancja danych to wówczas mówimy, że jest on nieznormalizowanym zbiorem danych. A zatem poszczególne postacie normalne można zdefiniować następująco, [25], [27]:

1. Pierwsza postać normalna dotyczy powtarzających się grup danych. Mówimy, że relacja jest w pierwszej postaci normalnej (1NF) wtedy i tylko wtedy, gdy wszystkie jej dziedziny zawierają wartości atomowe (niepodzielne).

2. Druga postać normalna dotyczy zależności funkcyjnych od części klucza złożonego. Mówimy, że relacja jest w drugiej postaci normalnej (2NF) wtedy i tylko wtedy, gdy jest w $1FN$ oraz każdy niekluczowy atrybut jest w pełni funkcyjnie zależny od klucza głównego.

3. Trzecia postać normalna ma na celu wykluczenie zależności przechodnich pomiędzy danymi. Relacja jest w trzeciej postaci normalnej (3NF) wtedy i tylko wtedy, gdy jest w $2NF$ i każdy niekluczowy atrybut jest w sposób nieprzechodni zależny od klucza głównego.

4. Definicja postaci normalnej Boyce'a-Codda stanowi sumę $1NF$, $2NF$ i $3NF$. Relacja jest w postaci normalnej Boyce'a-Codda (BCNF) wtedy i tylko wtedy, gdy każdy jej atrybut zależy funkcyjnie tylko od jej klucza głównego.

5. Dana relacja $R$ jest w czwartej postaci normalnej (4NF) wtedy i tylko wtedy, gdy jest w trzeciej postaci normalnej i wielowartościowa zależność zbioru atrybutów $Y$ od $X$ pociąga za sobą funkcjonalną zależność wszystkich atrybutów tej relacji od $X$.

6. Dana relacja jest w piątej postaci normalnej (5NF) wtedy i tylko wtedy, gdy nie istnieje jej rozkład odwracalny na zbiór mniejszych tabel.

## 5.4. Obiektowy model danych

Obiektowy model danych jest modelem, w którym wykorzystano takie cechy obiektowości jak: pojęcie klasy i obiektów klasy, hermetyzacja, mechanizm identyfikacji obiektów, dziedziczenie, przeciążanie funkcji, identyfikatory obiektów (OID) i inne. Podstawowym celem modelu obiektowego jest bezpośrednie odwzorowanie obiektów i powiązań miedzy nimi, wchodzących w skład aplikacji, na zbiór obiektów i powiązań w bazie danych. Mechanizmy obiektowe zwiększają niezależność danych od aplikacji poprzez przeniesienie procedur obsługi danych (w postaci *metod*) do systemu zarządzania bazą danych. Obiektowe bazy danych (OBD) łączą własności obiektowości i obiektowych języków programowania z możliwościami systemów bazodanowych. Rozszerzają możliwości obiektowych języków programowania czyniąc z nich narzędzie do łatwego i efektywnego tworzenia systemów baz danych zmniejszając stopień złożoności kodu programu. Tworzone obiekty OBD otrzymują unikalne identyfikatory niezmienne w czasie, które są wykorzystywane przez inne obiekty w celu definiowania powiązań z tymi obiektami. Zatem OBD składa się z obiektów, powiązanych pewna liczbą mechanizmów abstrakcji.

### 5.4.1. Obiekty

Obiekt jest pakietem danych (przechowywanych w atrybutach obiektu) oraz procedur *(metod)* uaktywnianych przez komunikaty przekazywane między obiektami. Obiektowy model danych dostarcza środki do realizacji tożsamości obiektów. Jest to możliwość rozróżnienia dwóch obiektów o takich samych cechach. Jest istotne, że definicja obiektu obejmuje zarówno aspekt strukturalny, jak i aspekt zachowania. Wynika z tego, że model obiektowy daje możliwość projektowania nie tylko struktury bazy danych, ale również sposobu użycia tej struktury. Pojęcie obiektu jest rozumiane w dwóch znaczeniach, [34]:

1. Na etapie analizy obiekt oznacza składową dziedziny problemu posiadającą tożsamość, stan i zachowanie.

2. Na etapie projektowania i implementacji pojęcie to oznacza konstrukcję języka programowania łączącą dane i metody.

W obiektowych bazach danych poszczególnym obiektom przypisany jest identyfikator (*object identity—OID*). Identyfikator ten posiada cechę *unique, co* oznacza, że dwa obiekty nie mogą mieć tego samego identyfikatora. Poza tym żaden obiekt nie może mieć więcej niż jednego OID. Obiekty elementarne, takie jak liczba całkowita są samoidentyfikujące się, tzn. same są swoimi identyfikatorami. OID (nazywane też adresami pojęciowymi) można wykorzystywać w innych miejscach bazy danych jako wskaźniki do oznaczania poszczególnych obiektów, [25]. Identyfikator ten w odniesieniu do baz danych jest czymś bardziej złożonym niż np. wskaźnik w języku C++. Jest on ciągiem bitów pozwalających na dokładne umiejscowienie obiektu w pamięci drugiego lub trzeciego poziomu na jednym z wielu różnych komputerów (np. w przypadku rozproszonych baz danych). Poza tym, ponieważ dane maja charakter trwały, więc właściwa wartość identyfikatora jest trwała przez cały czas istnienia obiektu.

### 5.4.2. Trwałość obiektu

Bardzo istotną usługą dostarczaną przez obiektowe systemy zarządzania bazami danych jest trwałość obiektu (*object persistence*) rozumianą jako pewne zachowanie (lub grupa zachowań) umożliwiające stały dostęp do obiektu oraz zachowanie jego stanu pomiędzy kolejnymi wywołaniami. Obiekty, które nie są stale dostępne nazywane są *obiektami ulotnymi*. W zależności od rozwiązań (aplikacji), niektóre obiekty zmieniają swój stan z ulotnego na trwały. Obiekty trwałe są przechowywane na dysku i pozostają w stałej gotowości do wywołania, natomiast ulotne rezydują w pamięci operacyjnej RAM. Zadaniem obiektowego systemu zarządzania bazą danych (OBDMS) jest właściwe administrowanie wymianą stanu obiektów — głównie za pośrednictwem OID. OBDMS posiada procedury odzyskiwania zapewniające, że trwałe obiekty przetrwają sytuacje krytyczne — np. awarie systemu. Poza tym OBDMS dba, aby obiekty mogły przechodzić z jednego stanu w drugi, mając swoją reprezentację w pamięci RAM, a także na dysku.

### 5.4.3. Klasy

Klasa obiektów jest typem danych o dowolnej wewnętrznej złożoności. Innymi słowy klasa stanowi połączenie typu oraz jednej lub kilku funkcji (metody), które można wykonywać na obiektach danej klasy. Jeżeli obiekt jest instancją czegoś, to klasa obiektów jest grupą podobnych obiektów, a więc obiektów o tej samej grupie cech. Wynika z tego, że *Bydgoszcz* jest obiektem a *miasto* jest klasą obiektów, do której on należy. W kontekście obiektowych baz danych, klasy obiektów definiują schemat bazy danych — główny temat dziedziny projektowania baz danych. Poza tym obiekty definiują zawartość bazy danych — główny temat dziedziny implementacji baz danych. Obiektowy model danych rozróżnia trzy rodzaje klas, [35]:

1. *Klasy sterujące* — odpowiadają za sterowanie przebiegiem działania programu. Do zadań tej klasy należy rozpoczęcie wykonywania programu, wykrywanie wyboru opcji z menu oraz wykonanie właściwego kodu programu w odpowiedzi na żądanie użytkownika.

2. *Klasy encji* — służą do tworzenia obiektów, które zarządzają danymi. Klasy te opisują np. ludzi, konkretne obiekty oraz zdarzenia (np. przeprowadzony wykład). W najprostszym ujęciu obiektowy model danych zbudowany jest na podstawie reprezentacji zależności między obiektami utworzonymi z obiektów encji.

3. *Klasy interfejsowe* — są odpowiedzialne za wprowadzanie i wyprowadzanie informacji.

4. *Klasy kontenerowe* — służą do zgromadzenia i zarządzania wieloma obiektami utworzonymi na podstawie klasy tego samego typu. Klasy te również nazywane są *agregacjami.*

Klasy encji nie wykonują własnych operacji wejścia-wyjścia, co oznacza, że informacje wprowadzane (np. z klawiatury) obsługiwane są przez obiekty interfejsu. Jeżeli obiekty encji wchodzą w skład bazy danych wówczas obsługą operacji I/O zajmuje są DBMS.

### 5.4.4. Metody

Z klasami są powiązane funkcje nazywane *metodami*. Metoda klasy $C$ ma co najmniej jeden argument, jest nim obiekt klasy $C$. Jeżeli zdefiniowano klasę, której typem jest zbiór liczb całkowitych, to metoda może na przykład służyć do wyliczenia zbioru potęgowego danego zbioru, [26]. Zespół operatorów lub funkcji, które można zastosować do obiektów danego typu określany jest jako zbiór *metod*. Metodę wywołuje się za pomocą *komunikatów* sprowadzających się do wywołania danej funkcji. Przykładem komunikatu może być poniższy zapis wysłany do oddziału $A$ proszący o zatwierdzenie zakupu $Z$ :

$$A \text{ ZATWIERDZ}(Z).$$

W konwencjonalnym języku programowania w/w komunikat mógłby mieć postać:

$$\text{ZATWIERDZ}(A, Z).$$

Poza metodami zdefiniowanymi przez programistę większość klas ma kilka wspólnych metod. Należą do nich, [35]:

1. *Konstruktory* — są to metody cechujące się tą sama nazwą jak klasy, do których należą. Konstruktor jest wykonywany zawsze w momencie tworzenia obiektu na podstawie klasy. Wynika z tego, że zwykle zawiera instrukcje inicjujące zmienne obiektu.

2. *Destruktory* — są to metody wykonywane, gdy obiekty są niszczone. Służą one do zwolnienia zasobów systemowych (na przykład pamięci głównej) przydzielonych obiektowi.

3. *Akcesory* — znane również jako *metody get*. Zwracają innemu obiektowi wartość prywatną atrybutu. Stosowane często w celu umożliwienia zewnętrznym obiektom uzyskania dostępu do ukrytych danych.

4. *Mutatory* — nazywane są również *metodami set*. Wykorzystywane są do zapisu nowej wartości atrybutu. Jest to typowy sposób, w jaki zewnętrzne obiekty mogą modyfikować ukryte dane.

Metody są przypisywane do klas na etapie ich definiowania. Poniżej zaprezentowano deklarację metody *PrzyznajStypendium* tworzącej tylko jej sygnaturę (nazwę) oraz typ jej argumentu. W celu definicji treści metody wykorzystywany jest standardowy język programowania.

***Create Class*** *Student*
***Attributes***
*Imie: String,*
*Nazwisko: String,*
*Stypendium: Decimal (8,2)*
***Relatonships***
*ZapisanyNa KierStud*
***Methods***
*PrzyznajStypendium (wartosc: Decimal)*

Ważną cechą klasy jest jej zdolność do posiadania *przeciążonych (overloaded)* metod, czyli metod o takiej samej nazwie, lecz argumentach różnego typu. W definicji klas stosuje się również przeciążone konstruktory, które mogą pobierać dane w sposób bezpośredni, z pliku lub kopiować je z innego obiektu (tzw. konstruktory kopiujące). Zaletą przeciążania metod jest możliwość konstruowania spójnego interfejsu.

### 5.4.5. System typów

Obiektowy model danych dostarcza spory wybór typów, które z powodzeniem wykorzystywane są w obiektowym modelu baz danych. Stosowane są typy podstawowe, obejmujące liczby całkowite, rzeczywiste itp., poza tym tworzone są nowe typy pochodne. W celu tworzenia pochodnych typów wykorzystywane są *konstruktory typów*, umożliwiające tworzenie, [26]:

1. *Rekordów (record structures)* — jeżeli dane są typy $T_1, T_2, \ldots, T_n$, a odpowiadające im pola mają przypisane kolejno identyfikatory $f_1, f_2, \ldots, f_n$, to można utworzyć nowy typ danych przez strukturę rekordu złożonego z $n$ składowych, którego $i$-ta składowa jest typu $T_i$ i można się do niej odwołać przez $f_i$.

2. *Kolekcji (collection types)* — istnieje możliwość utworzenia z określonego typu $T$ typu pochodnego przez zastosowanie w tym celu *operatora kolekcji* (np. tablice, listy i zbiory). Odzwierciedla się to w fakcie, że jeżeli typ bazowy jest określony jako np. liczba całkowita, to można stworzyć kolekcje: *tablica typu całkowitego, lista typu całkowitego* lub *zbiór typu całkowitego*.

3. *Referencji (references types)* — do typu $T$ jest typem, którego wartości są odpowiednie do przechowywania wartości typu $T$. W systemach baz danych (szczególnie bazach rozproszonych) referencja obejmuje informacje związane z systemową nazwą komputera, numer dysku, numer bloku oraz pozycje w bloku, w którym jest przechowywana wskazywana wartość.

Operatory kolekcji i rekordów są często zwielokrotniane i nakładane na siebie. Doprowadza to do tworzenia coraz bardziej złożonych typów danych. Przykładem ilustrującym złożony typ może być zdefiniowany rekord przechowujący informacje o samochodach, którego pierwsza składowa o nazwie *marka* jest tablicą znaków, natomiast druga jest zbiorem elementów typu całkowitego (a zatem typem pochodnym) i nazywa się *NrSilnika*.

### 5.4.6. Abstrakcyjne typy danych

Abstrakcyjnym typem danych *(abstract data type—ADT)* jest typ obiektu, który określa dziedzinę wartości i zbiór operacji, zaprojektowanych do działania na tych wartościach. Jest terminem stosowanym do klasy (klasa jest implementacją ADT) i określa, że klasę można traktować jak typ danych, który może być przypisany do atrybutów. Podstawową cechą ADT jest ukrywanie informacji, a szczegóły implementacji są ukryte przed wyższymi poziomami systemu użytkowego. Jeżeli implementacja ADT zostanie zmieniona, to nie ma to żadnego wpływu na wyższe warstwy systemu, ponieważ komunikują się one z ADT tylko za pomocą zbioru abstrakcyjnych operacji tworzących jego interfejs, [27]. Obiekty te stosuje się jako składowe krotek, a nie jako krotkę. Mają one jednak strukturę krotki i zazwyczaj zawierają składowe. Schemat definicji ADT przedstawiono poniżej, [26]:

*Create Type* ⟨*nazwa typu*⟩ *(*
*Lista atrybutów oraz ich typów*
*Opcjonalna deklaracja funkcji = and* ⟨*dla deklarowanego typu*⟩
*Deklaracja funkcji (metod) typu );*

Opcjonalne deklaracje operatorów porównania maja postać:

*Equals* ⟨nazwa funkcji implementującej równość obiektów⟩.

### 5.4.7. Hermetyzacja (*encapsulation*)

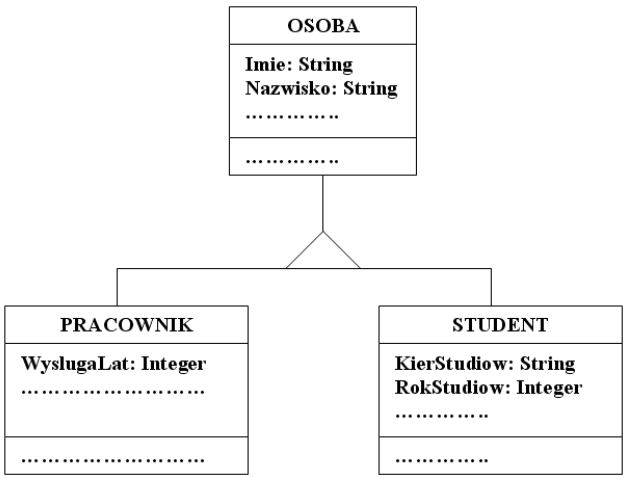
Polega na ukrywaniu przed innymi obiektami wartości danych, opisujących obiekt oraz sposób wykonywania przez obiekt jego operacji. Wynika z tego, że dostęp do danych jest możliwy tylko za pomocą odpowiednich funkcji (metod) dostępu. Rozróżniamy dwie koncepcje hermetyzacji:

1. *Hermetyzacja ortodoksyjna* — stanowiąca o tym, że wszelkie operacje, jakie można wykonać na obiekcie, są określone przez metody do niego przypisane (znajdujące się w jego klasie i nadklasach). Bezpośredni dostęp do atrybutów obiektu jest niemożliwy.

2. *Hermetyzacja ortogonalna* — w tym przypadku dowolny atrybut oraz metoda obiektu mogą występować jako *prywatne* (niedostępne z zewnątrz obiektu), bądź też publiczne (dostępne bez konieczności używania metod).

### 5.4.8. Hierarchia klas

Układ klas w systemie obiektowym tworzy tzw. hierarchię klas (*class hierarchy*). Oznacza to, że dla pewnej klasy $A$ może istnieć inna klasa (jedna lub więcej) $B$, znajdująca się na niższym poziomie, która jest uszczegółowieniem (specjalizacją) klasy $A$. Natomiast klasa $A$, będąca na wyższym poziomie w hierarchii, jest uogólnieniem (generalizacją) klasy (klas) $B$. Wynika z tego, że klasa obiektów $B$ jest podklasą obiektów $A$ — lub inaczej, klasa obiektów $A$ jest nadklasą klasy obiektów $B$ — wtedy i tylko wtedy, gdy każdy obiekt z klasy $B$ jest równocześnie obiektem z klasy $A$ („$B$ ISA $A$"). Obiekty z klasy $B$ dziedziczą zmienne instancji oraz metody stosujące się do klasy $A$, [25]. Sprowadza się to do tego, że użytkownik może korzystać z obiektu $B$ we wszystkich tych miejscach, w których jest dozwolony obiekt $A$. Przykład hierarchicznego układu klas przedstawiono na Rys. 5.8:




Rysunek 5.8. Hierarchiczny układ klas

Istotną cechą wspólną dla wszystkich systemów obiektowych jest to, że klasa może posiadać dowolną liczbę podklas.

### 5.4.9. Dziedziczenie

Nierozerwalnie z pojęciem hierarchii klas związane jest zagadnienie *dziedziczenia*. W odniesieniu do Rys. 5.8 można stwierdzić, że klasy PRACOWNIK i STUDENT dziedziczą pola i metody z klasy OSOBA. Zależności między klasa bazowa i jej klasami pochodnymi można wyrazić za pomocą określenia *„jest"* — co oznacza (na podstawie Rys. 5.8), że stwierdzamy *„pracownik jest osobą"*. Podklasa dziedziczy wszystkie właściwości swojej nadklasy, w związku z tym każdy atrybut lub związek nadklasy staje się automatycznie elementem podklasy. Istnieją dwa główne typy dziedziczenia:

1. *dziedziczenie struktury* — podklasa dziedziczy atrybuty swojej nadklasy,
2. *dziedziczenie zachowania* — podklasa dziedziczy metody swojej nadklasy.

W oprogramowaniu systemów informatycznych mogą zachodzić sytuacje, w których klasy posiadają tylko jedną nadklasę — mówimy wówczas o dziedziczeniu pojedynczym (*single inheritance*). W sytuacjach, gdy klasa dziedziczy od więcej niż jednej klasy bazowej, to wówczas zachodzi tzw. dziedziczenie wielokrotne (*multiple inheritance*), [27]. Jeżeli systemy umożliwiają dziedziczenie wielokrotne, wówczas klasy tworzą zakorzeniony spójny skierowany graf acykliczny, nazywany *kratą klas*, [36].

Istnieje możliwość zastosowania różnych metod o tej samej nazwie do różnych klas. W sytuacjach, gdy metody dziedziczone są przez podklasy od klasy bazowej i posiadają identyczne nazwy można skorzystać z *polimorfizmu,* czyli napisania różnego kodu źródłowego dla danych metod. Wielką zaletą polimorfizmu jest fakt, że można w przypadku wszystkich podklas tej samej klasy bazowej oczekiwać metod o tej samej nazwie i tym samym typie operacji wyjściowych. Doprowadza to do tego, że każda podklasa może wykonywać pewne operacje w zależności od swoich potrzeb operując tą samą nazwą metody.

### 5.4.10. Algebra obiektowa

Założeniem algebry obiektowej jest stworzenie matematycznej podstawą semantyki obiektowych języków zapytań. Algebra ta wzoruje się na algebrze relacji opisywanej w sekcji 5.3.1 niniejszej rozprawy. W odróżnieniu od algebry relacyjnej, w algebrze obiektowej operatory działają na zbiorach obiektów i zwracają zbiory obiektów. Pomimo krytyki [38], [39] istniejących propozycji algebr obiektowych związanych z zarzutami niespójności koncepcyjnej, wysokiego stopnia skomplikowania, niedostatecznej uniwersalności oraz zbyt luźnych związków z rygorystyczną matematyką istnieje wiele prób stworzenia właściwej algebry. Jedną z takich prób jest algebra AQUA, [37], [40]. W algebrze tej wyrażenia są reprezentowane przez *termy*, które mogą być zmienną, stałą, symbolem funkcji lub lambda abstrakcją danej formy. Jeżeli określimy, że dany term posiada nazwę *Nazwisko* i przyjmuje postać:

$$Nazwisko = \mathbf{apply}(\lambda(p)\,\mathbf{select\_field}(name(p))(Osoby)$$

gdzie **select_field** to symbol funkcji a $\lambda(p)\,\mathbf{select\_field}(name)(p)$ — lambda abstrakcja.

Oznacza to, że zwraca on podzbiór nazwisk ze zbioru *Osoby*. Operator **apply** zawarty w powyższym wyrażeniu odnosi się do grupy operatorów zbiorowych, które zostały opisane w dalszej części niniejszej rozprawy. Predykaty są funkcjami zwracającymi typ *boolean* przekazywanymi jako parametry do takich operatorów zapytań jak: *select*, *join* czy *exists* — tworząc w rezultacie nowy obiekt algebry. Operatory algebry obiektowej zostały podzielone na grupy, [40]:

1. operatory zbiorowe,
2. operatory multizbiorowe,
3. operatory innych typów.

Celem zwiększenia czytelności definicji operatorów definiowanych w dalszej części pracy przedstawiono wykorzystywane oznaczenia:

- $A$ oraz $B$ — wejściowe zbiory lub multizbiorów,
- $R$ — oznaczenie wyjściowego zbioru lub multizbioru,
- $a$ — element wejściowy zbioru lub multizbioru A,
- $f$, $g$ oraz $h$ — reprezentacja funkcji,
- $id$ — funkcja tożsamościowa,
- $p$ — predykat,
- $T$ — typ wynikowy operatora,
- $\langle\,\rangle$ — oznaczenie krotek,
- $L$ — nazwa pola krotki,
- $a/L$ — wartość krotki $a$ minus pole oznaczone $L$.

**Operatory zbiorowe** Operatory zbiorowe *(set operators)* podzielone zostały na kilka grup. Pierwsza grupa tych operatorów nazywana jest *zbiorem operatorów jednoargumentowych*. Zostały one zdefiniowane według poniższej listy:

$$apply(f)(A) = \{f(a) : a \in A\}\,, \tag{5.5}$$

$$select(p)(A) = \{a : a \in A, p(a)\}\,, \tag{5.6}$$

$$exists(p)(A) = \exists a \in A p(a)\,, \tag{5.7}$$

$$forall(p)(A) = \forall a \in A p(a)\,, \tag{5.8}$$

$$mem(a)(A) = a \in A, \tag{5.9}$$

$$fold(u, f, \oplus)(A) = \begin{cases} u, & A = \emptyset\,, \\ \bigoplus\limits_{a \in A} f(a), & A \neq \emptyset\,. \end{cases} \tag{5.10}$$

Kolejna reprezentacja operatorów zbiorowych to grupa *operatorów dwuargumentowych*. Podczas stosowania operatora dwuargumentowego nie jest wymagane, żeby rozpatrywane zbiory były tego samego typu. Wystarczy, aby ich elementy miały przynajmniej jeden wspólny nadtyp, ponieważ domyślna równość tego nadtypu jest używana do porównań. Jest istotne, że operatory dwuargumentowe wykorzystują dodatkowy argument $T$, określający typ wynikowy. Na przykład typem wynikowym operatora *union* musi być nadtyp zbiorów wejściowych. W odniesieniu do operatora *intersect* typ wynikowy może być nadtypem typów wejściowych lub jednym z tych

typów. Operator *diff* wymaga, aby typ wynikowy był typem pierwszego zbioru wejściowego lub jego nadtypem. Definicje operatorów dwuargumentowych przedstawiono poniżej:

$$union(T)(A,B) = \{x : x \in A \vee x \in B\}, \tag{5.11}$$

$$intersect(T)(A,B) = \{x : x \in A \wedge x \in B\}, \tag{5.12}$$

$$diff(T)(A,B) = \{x : x \in A \wedge \neg(x \in B)\} \tag{5.13}$$

Kolejna grupę operatorów tworzą tzw *zbiorowe operatory przekształcające* zdefiniowane poniżej:

$$set(a) = \{a\}, \tag{5.14}$$

$$choose(A) = \text{jakieś } a \in A, \tag{5.15}$$

$$group(f)(A) = \{(f(a), \text{eqclass}(a)) | a \in A\},$$
$$\text{gdzie: eqclass}(a) = \{a' | a' \in A, f(a) = f(a')\}, \tag{5.16}$$

$$dup\_clim(eq)(A) = R \subseteq A \text{ taki, że}$$
$$\forall x, y \in R, \neg(eq(x,y)) \wedge \forall x \in A, \exists y \in R \text{ taki, że } eq(x,y), \tag{5.17}$$

$$nest(L)(A) = \{tup\_concat(a/L,$$
$$\langle L : \{b.L | b \in A \wedge b/L = a/L\}\rangle) | a \in A\}, \tag{5.18}$$

$$unnest(L)(A) = \{tup\_concat(a/L, \langle L : s\rangle) | a \in A \wedge s \in a.L\}, \tag{5.19}$$

$$convert(A) = A \text{ (jako multizbiór).} \tag{5.20}$$

Definicje *operatorów złączenia:*

$$join(p,f)(A,B) = \{f(a,b) | a \in A, b \in B, p(a,b)\}, \tag{5.21}$$

$$tup\_join(p)(A,B) = join(p, tup\_concat)(A,B), \tag{5.22}$$

$$outer\_join(p,f,g,h,T)(A,B) = \{f(a,b) | a \in A, b \in B, p(a,b)\}$$
$$\cup \{g(a) | a \in A, \forall b \in B \neg p(a,b)\}$$
$$\cup \{h(b) | b \in B, \forall a \in A \neg p(a,b)\} \tag{5.23}$$

Definicja operatorów złączenia w kontekście baz danych jest bardzo istotna, a zatem poświęcono im kilka słów komentarza. Operator *join* przyjmuje jako parametr funkcję, umożliwiając przez to definiowanie funkcji „łączącej" *(konkatenacji)*. Pozostałe operatory złączenia są podobnymi uogólnieniami wymagającymi predykatu i funkcji. Typ wynikowy $T$ operatora *outer_join* musi być nadtypem typów wynikowych funkcji $f$, $g$ oraz $h$, po to, aby umożliwić powiązanie rezultatów funkcji.

Ostatnim operatorem zbiorowym jest operator tzw. *najmniej stałego punktu.* Odnośnie tego operatora należy przyjąć założenia w stosunku do funkcji $f$. Jest ona funkcją $T \to T$, gdzie $T$ jest typem zbioru $A$, a także musi być ona monotoniczna. Definicja operatora *LFP (least fixed point operator)* jest następująca:

$$LFP(T,f)(A) = \bigcup_{i=0}^{\infty}(f^i(A)), \text{ gdzie } f^0(A) = \emptyset. \tag{5.24}$$

**Operatory multizbiorowe** Na multizbiorach przeprowadzane są praktycznie te same operacje (jak w przypadku zbiorów) z tą różnicą, że multizbiór może zawierać wielokrotne wystąpienia tego samego elementu. Multizbiór oznaczany jest $\{^*e_1, e_2, \ldots, e_n{}^*\}$, a $e_i$ oznacza dany jego element. Z pojęciem multizbioru bardzo ściśle związane jest również pojęcie *krotności elementu*. Jest to liczba wystąpień elementu w obrębie danego multizbioru. Zapis $|A|_a$ oznacza krotność elementu $a$ w multizbiorze A. Definiuje się również liczność multizbioru $|A|$, oznaczającą całkowitą ilość elementów, z uwzględnieniem wszystkich powtórzeń. Operatory uwzględnione w niniejszym podrozdziale operują na multizbiorach, a zatem typami wejściowymi i wyjściowymi są multizbiory i zostały zdefiniowane następująco:

$$\begin{aligned} \mathit{union}(T)(A,B) = R \text{ taki, że } \forall x \in R.|R|_x = \max(|A|_x, |B|_x) \\ \text{również } \forall y \text{ takiego, że } ((y \in A) \vee (y \in B)).y \in R \end{aligned} \tag{5.25}$$

$$\begin{aligned} \mathit{additive\_union}(T)(A,B) = R \text{ taki, że } \forall x \in R.|R|_x = |A|_x + |B|_x \\ \text{również } \forall y \text{ takiego, że } ((y \in A) \vee (y \in B)).y \in R \end{aligned} \tag{5.26}$$

$$\begin{aligned} \mathit{diff}(T)(A,B) = R \text{ taki, że } \forall x \in R.|R|_x = \max(0, |A|_x - |B|_x) \\ \text{również } \forall y \text{ takiego, że}((y \in A) \wedge (\neg(y \in B))).y \in R) \end{aligned} \tag{5.27}$$

$$\mathit{multiset}(a) = \{^*a^*\} \tag{5.28}$$

**Operatory innych typów** Poza operatorami opisanymi powyżej algebra obiektowa obsługuje wiele innych operatorów umożliwiających prace z bazami danych opartych na modelu obiektowym. Na przykład operacja, *union(U, lab, e)* tworzy instancję unii $U$ i inicjalizuje jej zawartość jako jednostkę $e$ z etykietą *lab*. Zarówno *tagcase(e)* jak i *typecase(e)* selektywnie wykonują zbiór termów bazujący albo na etykiecie, albo na typie instancji unii $e$, [40]. Istotne również są tzw. *typy funkcyjne* reprezentujące funkcję, która pobiera kilka parametrów, zwracają pojedynczą wartość. Nie istnieje wyraźny konstruktor typów funkcyjnych. Instancje typów funkcyjnych są tworzone z wykorzystaniem typowanych wyrażeń lambda, [37]. W opisywanej algebrze obiektowej występują również *abstrakcyjne typy danych*. Można je określić jako typy zbiorowe, których elementy są dostępne tylko poprzez użycie specjalnych funkcji, nazywanych *interfejsem*. Funkcje są osiągalne poprzez operator *invoke(I, f)*, który wykonuje funkcję $f$ na instancji $I$.

### 5.4.11. Relacyjno-obiektowy model danych

Relacyjno-obiektowy model danych jest koncepcją, w której wykorzystuje się możliwości modelu relacyjnego i poszerzenie ich o cechy obiektowe. Systemy baz danych oparte na tym modelu powinny umożliwiać operacje przeprowadzane zgodnie z tradycyjnym podejściem relacyjnym baz danych *(„jeden-do-wielu", „wiele-do-wielu"),* oraz także relacje *„jest"* wynikające bezpośrednio z dziedziczenia. W odróżnieniu od systemów relacyjnych tutaj relacje definiowane są przez umieszczenie wewnątrz obiektu identyfikatorów *(OID* – por p. 5.4.1 niniejszej rozprawy) powiązanych z nim obiektów. W różnych systemach OID mają różne znaczenie. Mogą być dowolnie ustalonymi wartościami lub informacjami potrzebnymi do zlokalizowania obiektu w pliku bazy. W tym podejściu realizacja relacji między danymi związana jest z dwoma aspektami:

1. Gdy dany obiekt jest elementem bazy danych jego identyfikator nie może się zmieniać, [35]. Jeżeli warunek ten nie zostanie spełniony, wówczas DBMS nie może odwoływać się do rekordów powiązanych — OID będą wskazywały niewłaściwe dane (obiekty).

2. Filtrowanie obiektowych baz danych odbywa się w powiązaniu o relację wcześniej zdefiniowane (przez zarejestrowanie ich w atrybutach OID powiązanych obiektów), co ogranicza elastyczność działań użytkownika/programisty. Istotne jest jednak to, że filtry oparte na zdefiniowanych ścieżkach relacji miedzy obiektami dają szybsze wyniki zapytań. Związane jest to z faktem, że identyfikacja obiektu trwa szybciej niż złączenia tabel wykorzystywane w modelu relacyjnym.

W systemach obiektowych baz danych relacje interpretowane są w następujący sposób:

1. *Relacja „jeden-do-wielu"*: W modelu obiektowym (w przeciwieństwie do relacyjnego modelu danych) atrybuty mogą przyjmować postacie wielowartościowe (zwane *zbiorami*), co jest istotne w przypadku reprezentacji relacji *„wiele"*. Związane jest to z definicją relacji po stronie *„wiele"* odnoszącej się do atrybutu klasy, w którym jest przechowywany OID obiektu nadrzędnego *(rodzica)*. Klasa obiektów macierzystych ma atrybut, który przechowuje identyfikatory powiązanych z nią obiektów, [35]. Należy zaakcentować fakt, że w systemach obiektowych baz danych relacja *„jeden-do-jednego"* jest specyficznym przypadkiem relacji *„jeden-do-wiele"*, w której stopień relacji wynosi po obu stronach jeden.

2. *Relacja „wiele-do-wielu"*: W związku z tym, że w obiektowej bazie danych obiekty mogą mieć atrybuty wielowartościowe istnieje możliwość bezpośredniego zestawienia relacji *wiele-do-wiele"*. Relacje te są realizowane przez definiowanie atrybutu (w każdej klasie uczestniczącej w związku) zawierającego zbiór wartości innej klasy, z którą został powiązany.

3. *Relacja „jest" (ISA)*: Relacja ta oparta jest na paradygmacie związanym z realizowaniem dziedziczenia w obiektowym modelu danych. Relacja „jest", określana również mianem *generalizacja-specjalizacja* tworzy hierarchie dziedziczenia, w której podklasy są szczególnymi typami ich klasy bazowej, [35].

4. *Relacja „rozszerza"*: W przypadku tej relacji dochodzi do rozszerzenia definicji podklasy względem klasy bazowej.

5. *Relacja „całość-część"*: Relacja ta odnosi się do danej klasy powiązanej z obiektami innych klas wchodzących w jej skład. Ilustrację klasy „całość-część" może stanowić fragment bazy danych z powiązanymi obiektami pochodzącymi od z klasy *produkt* z obiektami pochodzącymi od klas *części* oraz *podzespoły*.

W relacyjno-obiektowym modelu danych bardzo ważnym zagadnieniem jest *integralność relacji*, która gwarantuje zgodność identyfikatorów obiektów po obu stronach relacji. Jeżeli ma istnieć powiązanie między autorami i książkami, wówczas DBMS musi obsłużyć mechanizm zapewniający, że jeśli OID klasy *autorzy* zostanie wstawiony do obiektu klasy *książki*, wówczas OID obiektu *książki* zostanie wstawiony do obiektu klasy *autorzy*. Integralność ta (szeroko stosowana w obiektowych bazach

danych) określana jest mianem *relacji odwrotnych (inverse relationship)*. Klasa autorzy powinna posiadać atrybut *autorzy.ksiazka* (zdefiniowany jako zbiór) i jednocześnie klasa książki powinna zostać wyposażona w atrybut *ksiazka.autor*. Instrukcja określająca lokalizacje OID może mieć postać:

*parent: autorzy*
*inverse is autorzy.ksiazka*

dla klasy *ksiazki,* oraz

*childern: (set) ksiazki*
*inverse is ksiazka.autor*

dla klasy *autorzy*.

## 5.5. Obiektowe i relacyjne bazy danych — zestawienie technologii

Niewątpliwym atutem systemów obiektowych baz danych jest dobre przystosowanie do zarządzania danymi, które posiadają złożoną strukturę, tworzącą zagnieżdżoną hierarchię oraz dynamicznie zmieniający rozmiar. Relacyjne bazy danych natomiast znajdują zastosowanie w operowaniu danymi prostymi, niezagnieżdżonymi, które agregowane są w tablicach. Powoduje to, że dla niektórych zastosowań struktury relacyjne są zbyt sztywne, aby zostały wprowadzone do praktycznej realizacji. Pociąga to za sobą potrzebę nowych rozwiązań obsługujących niezagospodarowane technologie dające pole obiektowym systemom zarządzania bazami danych. Zastosowania te są realizowane z powodzeniem w kontekście danych multimedialnych lub dziedzin reprezentujących niesformalizowane struktury danych — takich jak zastosowania pełnotekstowe lub hurtownie danych. Zagadnienie to ma odzwierciedlenie w ocenie wydajności opisywanych technologii baz danych, co jest uzależnione od rodzaju przetwarzanych informacji. Jest wiele stanowisk głoszących, że obiektowe bazy danych są coraz bardziej wydajnymi systemami dzięki zastosowaniu technik przemiany wskaźników, indeksów ścieżkowych [38] oraz mechanizmów indeksacji i buforowania, które pozwalają na przeniesienie przetwarzania danych na stronę klienta. Podstawowym niedociągnięciem w kontekście optymalizacji w systemach relacyjnych jest kosztowna operacja złączeń przeprowadzana podczas filtrowania danych. W związku z tym, że w systemach obiektowych złączenia są realizowane z wykorzystaniem wskaźników zagadnienie optymalizacji daje korzystniejsze efekty dla tych systemów. Zagadnieniem bardzo istotnym jest przewaga obiektowych baz danych nad relacyjnymi, odnosząca się do bardzo szerokiej współpracy z obiektowymi językami programowania, które zdominowały obecna technologie informatyczną.

Najczęściej natomiast kierowane zarzuty w stosunku do obiektowych baz danych sprowadzają się do faktu, że wykorzystanie wskaźników w strukturach obiektowych cofa rozwój do czasów systemów sieciowych. Poza tym zagadnienia związane z obiektowością znajdują się bardzo często w fazie laboratoryjno-teoretycznej, co powoduje, że wiele technologii jest mało stabilnych.

## 5.6. Dane multimedialne i multimedialne bazy danych

### 5.6.1. Dane multimedialne — pojęcia ogólne

Dane multimedialne są zdefiniowane jako dane reprezentujące pewne sygnały zakodowane w jednym lub wielu mediach, spośród których co najmniej jedno medium nie jest alfanumeryczne. Z praktycznego punku widzenia definicja danych multimedialnych odnosi się do zakodowanych sekwencji audio, filmów wideo (zazwyczaj uzupełnione o sekwencje audio i opcjonalnie tekst), dokumentów multimedialnych (np. prezentacji), obrazów satelitarnych oraz pozostałych rozmaicie zakodowanych dokumentów. Obrazy statyczne występują w postaci dwuwymiarowej tablicy punktów (zwanych pikselami). Podstawowe charakterystyki opisujące parametry obrazów zapisanych w postaci cyfrowej to: format, wymiary, liczba kolorów i metoda kompresji. Podstawowe cechy opisujące cyfrowe dane audio to: format, typ kodowania, liczba kanałów, częstotliwość próbkowania, rozmiar próbki, metoda kompresji i czas trwania. Dane wideo charakteryzują się takimi parametrami jak: format, typ kodowania, liczba klatek na sekundę, rozmiar klatki, metoda kompresji, liczba kolorów, czas trwania oraz częstotliwość strumienia bitów.

W obecnych rozwiązaniach stosowane są media ukierunkowane na ludzkie zmysły — głównie wzrok i słuch. Wspólną cechą tych wszystkich form zapisu jest rozmiar, który jest znacznie większy niż danych tradycyjnych (np. typ całkowity). Doprowadza to do kłopotów związanych z zarządzaniem i przechowywaniem danych multimedialnych, co wymusza stosowanie różnych metod kompresji danych. Celem kompresji jest zredukowanie rozmiaru danych, co w efekcie doprowadzi do zmniejszenia wymogów związanych z pojemnością pamięci masowych oraz przepustowości kanałów transmisyjnych. Metody kompresji multimedialnych można podzielić na metody bezstratne (*lossless compression*), np. format `wav` oraz stratne (*lossy compression*), np. formaty `mp3`, `mpeg`. W przypadku metod bez utraty jakości, po dekompresji otrzymywana jest postać identyczna z postacią źródłową (poddaną kompresji). Metody z utratą jakości wykorzystują niedoskonałość ludzkich zmysłów, a zatem utrata pewnej części danych nie jest rozpoznawalna przez człowieka. Wynika z tego, że kluczem do oszczędnego zapisu danych multimedialnych jest wiedza o sposobie percepcji danego typu przekazu przez człowieka. W przypadku obrazu, wykorzystuje się fakt niedostrzegania przez oko ludzkie niewielkich różnic pomiędzy kolorami sąsiadujących punktów obrazu oraz większej wrażliwości na zmiany luminancji niż chrominancji. W przypadku dźwięku, na zmniejszenie ilości danych pozwala efekt maskowania, polegający na zagłuszaniu przez głośne pasmo częstotliwości pasm cichych, nie tylko równoczesnych, ale także poprzedzających je i następujących po nich w czasie, [55].

### 5.6.2. Multimedia w systemach baz danych

Podstawowym zadaniem stawianym systemom multimedialnych baz danych jest wyszukiwanie obiektów spełniających kryteria zadane przez użytkownika, czyli innymi słowy przetwarzanie zapytań multimedialnych, [56].

W kontekście multimedialnych baz danych można wyróżnić dwa typy zapytań:

1. zapytania o metadane,
2. wyszukiwanie na podstawie zawartości.

Metadane stanowią alfanumeryczny opis obiektu multimedialnego i dotyczą jego parametrów (np. rozmiar, metoda kompresji) lub zawartości (np. autor, wykonawca, tytuł, słowa kluczowe). Realizacja zapytań o metadane, takich jak np. „wyszukaj wszystkie melodie zapisane w formacie `mp3` o czasie trwania krótszym niż 4 minuty" lub „znajdź wszystkie filmy, w których główną rolę odgrywa aktor X, i są zapisane w formacie `mpeg`", nie wykracza poza funkcjonalność systemów relacyjnych. Ogromnym obecnie wyzwaniem, stojącym przed twórcami systemów zarządzania multimedialnymi bazami danych, jest wyszukiwanie ze względu na zawartość, np. „wyszukaj obiekty (obrazy, utwory muzyczne) podobne do określonego", czy też „melodie, w której dominująca jest altówka". Rozwiązanie tego problemu jest związane z opracowaniem złożonych algorytmów analizy zawartości obiektów, celem ekstrakcji jego cech (ten właśnie problem jest przedmiotem badań niniejszej rozprawy). Atrakcyjność tego rozwiązania, wynika głównie z tego, że wprowadzanie metadanych przy dodawaniu kolejnych obiektów do bazy jest uciążliwe, a ponadto nie wszystkie właściwości można łatwo opisać słownie (np. barwa dźwięku). Wyszukiwanie ze względu na zawartość znajduje zastosowanie np. przy wykrywaniu plagiatów, projektowaniu mody, przeszukiwaniu wirtualnych galerii sztuki, czy też projektowaniu wnętrz. Składowanie danych multimedialnych w bazach danych realizowane jest za pomocą:

- przechowywania lokalnie w bazie danych w postaci dużych obiektów binarnych (BLOB),
- przechowywania w lokalnym systemie plików, dostępnych z poziomu bazy danych jako tzw. duże obiekty plikowe (BFILE),
- danych dostępnych pod określonym adresem URL, udostępnianych przez serwer WWW,
- danych udostępnianych przez specjalistyczne serwery np. dostarczające dane audio lub wideo strumieniowo.

ROZDZIAŁ 6

# Podstawowe algorytmy klasyfikujące

## 6.1. Algorytm minimalno-odległościowy

Aby system komputerowy przeprowadził prawidłowe rozpoznanie dwóch obiektów cyfrowych dowolnych klas konieczne jest wytypowanie cech, które pozwolą na wytypowanie oczekiwanego obiektu. Jeżeli na podstawie badanej populacji dwóch obiektów pochodzących z różnych klas (w badaniach uwzględniono populacje ok. 30 próbek), porównamy wyniki uzyskane na podstawie analizy nieregularności widma określanej zależnością:

$$\mathrm{Ir} = \log\left(20\sum_{k=2}^{N-1}\left|\log\frac{A_k}{\sqrt[3]{A_{k-1}\cdot A_k\cdot A_{k+1}}}\right|\right), \tag{6.1}$$

to każdy obiekt będzie reprezentowany przez wektor cech, [41]:

$$\mathbf{x} = \mathbf{a}^T. \tag{6.2}$$

W tym przypadku zbiór uczący $U$ jest zbiorem 30 par wartości cech danego obiektu, [41]:

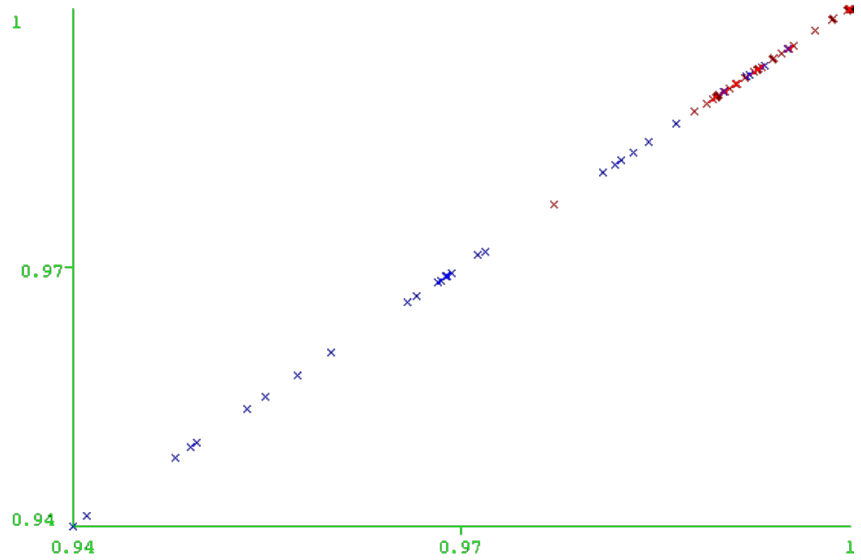
$$U = \left\{(\mathbf{x}_i, t_i) : i = 1\ldots 30,\ t_i \in \{1,2\},\ \mathbf{x}_i = (\mathbf{a}_i)^T\right\}. \tag{6.3}$$

Na Rys. 6.1 przedstawiono tzw. *wykres rozrzutu* właściwy dla zbioru uczącego $U$ w przestrzeni cech. Na wykresie pokazano rozrzut nieregularności widma dla dwóch klas — gitara akustyczna (kolor niebieski), gitara elektryczna (kolor czerwony).

Kolejnym etapem jest zdefiniowanie a potem wyznaczenie na podstawie zbioru uczącego *funkcji dyskryminacyjnych*. W minimalno-odległościowym algorytmie klasyfikacji każda klasa jest reprezentowana przy pomocy *prototypu punktowego* $m_i$, który określa środek klasy. Prototyp $m_i$ jest opisywany zależnością, [41]:

$$m_i = \frac{1}{N_i}\sum_{\mathbf{x}\in U_i}\mathbf{x}, \quad i = 1,\ldots c \tag{6.4}$$

gdzie:
$c$ — ilość klas,
$U_i$ — zbiór wektorów zbioru uczącego $U$,
$N_i$ — liczebność zbioru $U_i$.




Rysunek 6.1. Wykres rozrzutu nieregularności widma dla gitary akustycznej (kolor niebieski) i gitary elektrycznej (kolor czerwony)

Cechą algorytmu minimalno-odległościowego jest to, że zalicza wektor $\mathbf{x}$ do określonej klasy, jeżeli odległość wektora $\mathbf{x}$ od prototypu tej klasy jest najmniejsza. Aby wyprowadzić wzór opisujący funkcje dyskryminacyjne $g_i$ klasyfikatora minimalno-odległościowego rozpatrzymy funkcję postaci, [41]:

$$g_i(x) = -\|\mathbf{x} - \mathbf{m}_i\|, \quad i = 1, \ldots, c, \tag{6.5}$$

której wartości są równe co do wartości bezwzględnej normie różnicy wektorów $\mathbf{x}$ i $\mathbf{m}_i$. Przyjmując normę euklidesową, dla której zachodzi:

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} \tag{6.6}$$

otrzymujemy (pomijając pierwiastek), [41]:

$$g_i(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)^T(\mathbf{x} - \mathbf{m}_i) = -\mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{m}_i + \mathbf{m}_i^T\mathbf{x} - \mathbf{m}_i^T\mathbf{m}_i. \tag{6.7}$$

W związku z tym, że pierwszy składnik wyrażenia (6.7) nie zależy od numeru klasy (jest taki sam w każdej funkcji) może zostać pominięty, co prowadzi do uproszczonej postaci wzoru na funkcje dyskryminacyjne klasyfikatora minimalno-odległościowego:

$$g_i(\mathbf{x}) = 2\mathbf{m}_i^T\mathbf{x} - \mathbf{m}_i^T\mathbf{m}_i = \mathbf{w}_i^T\mathbf{x} + \mathbf{w}_i^0, \quad i = 1, \ldots, c \tag{6.8}$$

Ostatnim etapem jest sformułowanie sposobu przydziału wektora $\mathbf{x}$ do danej klasy. Wykorzystując postać funkcji dyskryminacyjnej (6.8) można zdefiniować minimalno-odległościowy algorytm klasyfikacji $\psi_{\text{mo}}$, [41]:

$$\psi_{\text{mo}}(\mathbf{x}) = i \quad \text{jeżeli} \quad \underset{\substack{j=1,\ldots,c \\ j\neq i}}{\forall} g_i(\mathbf{x}) < g_j(\mathbf{x}). \tag{6.9}$$

Algorytm klasyfikacji przypisuje wektorowi $\mathbf{x}$ tę klasę $i$, dla której odpowiadająca jej funkcja $g_i$ przyjmuje wartość najmniejszą lub, równoważnie, do tej klasy, dla której

odległość wektora $\mathbf{x}$ od środka $\mathbf{m}_i$ tej klasy jest najmniejsza. Natomiast powierzchnia rozdzielająca w przestrzeni cech obiekty z klasy $i$ od obiektów z klasy $j$ jest opisana zależnością, [41]:

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0. \tag{6.10}$$

Biorąc pod uwagę postać funkcji (6.8) otrzymujemy równanie powierzchni decyzyjnej klasyfikatora minimalno-odległościowego, które wyrażane jest zależnością:

$$g_{ij}(\mathbf{x}) = 2(\mathbf{m}_i^T - \mathbf{m}_j^T)\mathbf{x} + \mathbf{m}_j^T\mathbf{m}_j - \mathbf{m}_i^T\mathbf{m}_i. \tag{6.11}$$

## 6.2. Metody reprezentacji obiektów

Reprezentacja obiektów jest stworzona na podstawie zdefiniowanego odwzorowania zbioru obiektów $O$ w określony zbiór ich reprezentacji, nazywany *przestrzenią reprezentacji* lub *przestrzenią cech*. W niniejszym podrozdziale przedstawiono podstawowe metody reprezentacji obiektów: metodę wektorową i strukturalną.

### 6.2.1. Metoda wektorowa

Metoda ta uwzględnia opis obiektu za pomocą grupy cech, tzn. mierzalnych wielkości poddających się obserwacji lub pomiarowi, [41]. Charakter tych cech jest zdeterminowany przez rodzaj badanego obiektu i uzależniony od techniki pomiarowej. Cechy są reprezentowane przez konkretne wartości i ujęte w postaci $d$-wymiarowego *wektora cech*:

$$\mathbf{x} = (x_1, \ldots, x_d)^T. \tag{6.12}$$

Wektor cech stanowi opis obiektu, który w procesie klasyfikacji jest jedynym źródłem informacji na jego temat. Zbiór wszystkich wartości, z których składa się wektor cech obiektów, nazywa się *przestrzenią cech*. Bardzo istotnym zagadnieniem jest taki dobór cech, aby obiekty pochodzące z różnych klas posiadały różne wartości dla danego wyboru cech. Obiekty z tej samej klasy powinny mieć wartości zbliżone, umożliwiające właściwą zdolność rozpoznawalności danej klasy. Cechy badanych obiektów mogą być rodzaju *jakościowego* lub *ilościowego*. Oznacza to, że cechy jakościowe (a wśród nich również niemierzalne) opisują klasę obiektów w sposób niejednoznaczny i nie są wyrażane w postaci liczb (np. wartości logiki Boole'a). Cechy rodzaju ilościowego wyrażane są za pomocą liczb w określonej (ogólnie przyjętej) skali (np. wzrost, stan konta bankowego itp.). Cechy mierzalne można podzielić na dwie podgrupy:

1. cechy ciągłe — dziedziną są liczby rzeczywiste z pewnego przedziału,
2. cechy dyskretne — dziedziną są liczby całkowite z pewnego przedziału.

Cechy możemy określać w skalach wartości, [41]:

1. skala nominalna — właściwa dla cech jakościowych i umożliwiająca stwierdzenie czy cechy są równe lub różne ( np. przyjmowanie wartości logicznych),
2. skala porządkowa — pozwala ustalić porządku w zbiorze wartości cech (np. poniżej normy, w normie itp.),
3. skala liczbowa — dziedziny wartości cechy są zdefiniowane na liczbowych skalach pomiarowych.

### 6.2.2. Metoda strukturalna

Metoda ta charakteryzuje się cechą reprezentowania struktury obiektów, jako jego złożenia z prostych obiektów składowych. Badany obiekt jest charakteryzowany przez obiekty składowe do momentu otrzymania tzw. *składowej pierwotnej*, która stanowi niezależny element i nie podlega dalszemu podziałowi. Składowe pierwotne *(prymitywy)* definiowane są w zależności od zastosowania. Bardzo istotne jest określenie prawidłowego sposobu złożenia prymitywów, co realizowane jest przez właściwą identyfikacje wzajemnych relacji, jakie zachodzą między nimi w obiekcie złożonym. Istnieją dwa rodzaje reprezentacji:

1. struktury symboliczne, w których obiekty i modele klas reprezentowane są za pomocą takich struktur jak ciąg, drzewo czy graf i umożliwiają określenie w jawny sposób relacji pomiędzy składowymi obiektu;

2. gramatyki, w którym gramatyka stanowi mechanizm generujący wszystkie wystąpienia obiektów danej klasy.

Zbiór wszystkich możliwych reprezentacji strukturalnych obiektów tworzy *przestrzeń opisów strukturalnych*.

## 6.3. Klasyfikatory

### 6.3.1. Reguła decyzyjna Bayesa

*Przestrzenią obserwacji* nazywamy zbiór wszystkich możliwych wartości wektora cech, natomiast zbiór wszystkich numerów klas *przestrzenią decyzyjną*. Wartości numerów klas są losowane zgodnie z rozkładem scharakteryzowanym *prawdopodobieństwami a priori* klas $P(j)$ oraz warunkowych gęstości $f(x|j)$ cech w klasach [42]. W sytuacji, gdy na wejściu generatora obserwacji podany zostanie parametr, którego wartość odpowiada wylosowanemu numerowi klasy $j$, to zostanie wygenerowana na wyjściu *obserwacja* (wektor cech) $\mathbf{x}$, zgodnie z rozkładem cech odpowiadającym klasie $j$. Wykorzystując zbiór uczący jako źródło informacji o nieznanych wielkościach dokonuje się na jego podstawie estymacji a następnie wstawienia w miejsce rzeczywistych wartości w funkcjach dyskryminacyjnych algorytmu optymalnego, [43]. Reguła decyzyjna o charakterystyce $\psi$ podejmuje na podstawie obserwacji $\mathbf{x}$ decyzję $i = \psi(\mathbf{x})$. Oznacza to, że zostaje przyporządkowana danej obserwacji $\mathbf{x}$ numer klasy $i$. Mówimy wówczas o stracie $L_{ij}$ wyliczanej przez moduł strat:

$$0 \leq L_{ij} < \infty \quad i, j = 1, \ldots, c, \tag{6.13}$$

$$\psi_e^0(x) = i \Leftrightarrow \underset{\substack{j=1,\ldots,c \\ j \neq i}}{\forall} \sum_{k=1}^{c} L_{ik} \hat{P}(k) \hat{f}(x|k) < \sum_{k=1}^{c} L_{jk} \hat{P}(k) \hat{f}(x|k), \tag{6.14}$$

lub dla przypadku prostej funkcji strat:

$$\psi_e^0(x) = i \Leftrightarrow \underset{\substack{j=1,\cdots,c \\ j \neq i}}{\forall} \hat{f}(x|i) \hat{P}(i) > \hat{f}(x|j) \hat{P}(j) \tag{6.15}$$

gdzie: $\hat{P}(j)$, $\hat{f}(x|j)$ — estymatory odpowiednich wielkości.

Stosując różne estymatory można otrzymać różne klasyfikatory. Z (6.14) oraz (6.15) wynika, że ogólna postać funkcji dyskryminacyjnych empirycznego klasyfikatora Bayesa wyrażana jest następująco:

$$g_i(x) = \sum_{k=1}^{c} L_{ik}\hat{P}(k)\hat{f}(x|k) \quad i = 1, \ldots, c, \tag{6.16}$$

a dla przypadku prostej funkcji strat:

$$g_i(x) = \hat{f}(x|i)\hat{P}(i), \quad i = 1, \ldots, c. \tag{6.17}$$

Wszystkie empiryczne klasyfikatory Bayesa wykorzystują ogólną postać funkcji dyskryminacyjnych klasyfikatora optymalnego z wstawionymi oszacowaniami prawdopodobieństw a priori. Estymacji prawdopodobieństw a priori $P(j)$ klas dokonuje się przez udział poszczególnych klas w zbiorze uczącym, [41]:

$$\hat{P}(i) = \frac{N_i}{N}, \quad N = \sum_{j=1}^{c} N_j, \quad i = 1, \ldots, c, \tag{6.18}$$

gdzie: $N_i$ — ilość wektorów z klasy $i$ w zbiorze uczącym.

### 6.3.2. Klasyfikator k-NN

Prawdopodobieństwo zdarzenia $P$ określające, że wektor cech $\mathbf{x}$ wylosowany z określonej populacji, o rozkładzie scharakteryzowanym nieznaną gęstością $f(\mathbf{x})$ znajdzie się w obszarze zainteresowań $R$ w przestrzeni wynosi:

$$P = \int_R f(\mathbf{x}')d\mathbf{x}', \tag{6.19}$$

Zakładając, że $N$ wektorów $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ wylosowano z populacji o rozkładzie $f(\mathbf{x})$ to prawdopodobieństwo, że $k$ spośród tych $N$ wektorów znajdzie się w obszarze zainteresowań $R$ można określić za pomocą średniej frakcji wektorów, jakie są zawarte w obszarze $R$ i wynosi, [44]:

$$P \cong \frac{k}{N}. \tag{6.20}$$

Jeżeli jednak założymy, że obszar $R$ jest tak mały, że pozwala na założenie stałości w nim funkcji gęstości $f(\mathbf{x})$, to wówczas:

$$\int_R f(\mathbf{x}')d\mathbf{x}' \cong vf(\mathbf{x}), \tag{6.21}$$

gdzie: $v$ — objętość obszaru $R$.

Korzystając z (6.20) oraz (6.21) otrzymujemy:

$$P \cong vf(\mathbf{x}) = \frac{k}{N} \Rightarrow f(\mathbf{x}) \cong \frac{k}{Nv}. \tag{6.22}$$

Uzyskano w ten sposób ogólny wzór na estymator funkcji gęstości $f(\mathbf{x})$ w punkcie $\mathbf{x}$ przestrzeni cech, [41]:

$$f(\mathbf{x}) \cong \frac{k}{Nv}, \tag{6.23}$$

gdzie: $v$ — objętość obszaru $R$ dookoła punktu $\mathbf{x}$ przestrzeni cech.

W metodzie k-NN, w celu estymacji gęstości $f(\mathbf{x})$ w punkcie $\mathbf{x}$ przestrzeni na podstawie $N$-elementowego zbioru uczącego z dowolnej klasy ustala się pewną liczbę $k$, a następnie wyznacza minimalny obszar zainteresowań $R$ w przestrzeni cech (wokół punktu $\mathbf{x}$). Wyznaczony obszar $R$ zawiera wyspecyfikowaną liczbę $k$ wektorów zbioru uczącego. Wyliczając objętość $v$ określonego obszaru $R$ podstawiamy wartości $k$ i $v$ do wzoru (6.23). Pozwala to na oszacowanie gęstości w punkcie $\mathbf{x}$. Zależność (6.23) można przedstawić w postaci:

$$f(\mathbf{x}|i) = \frac{k_i}{N_i v}, \tag{6.24}$$

gdzie:

$N_i$ — liczba wszystkich wektorów $i$-tej klasy w zbiorze uczącym,

$k_i$ — liczba wektorów pośród $k$ najbliższych sąsiadów wektora $\mathbf{x}$.

Należy zwrócić uwagę na fakt, że zachodzi:

$$k = \sum_{j \in I} k_j \,. \tag{6.25}$$

Wykorzystując (6.24), oszacowania prawdopodobieństw a priori (6.17) oraz regułę Bayesa można otrzymać oszacowania prawdopodobieństw $i$-tej klasy w punkcie $\mathbf{x}$, [41]:

$$P(i|\mathbf{x}) = \frac{f(\mathbf{x}|i)P(i)}{\sum_{j=1}^{c} f(\mathbf{x}|j)P(j)} = \frac{\frac{k_i}{vN_i}\frac{N_i}{N}}{\sum_{j=1}^{c}\left(\frac{k_j}{vN_j}\frac{N_j}{N}\right)} = \frac{k_i}{k}\,. \tag{6.26}$$

Z (6.26) wynika, że funkcje dyskryminacyjne klasyfikatora k-NN można zapisać w postaci:

$$g_i(\mathbf{x}) = \frac{k_i}{k}\,, \tag{6.27}$$

gdzie:

$k_i$ — ilość wektorów zbioru uczącego z klasy $i$-tej,

$k$ — liczba najbliższych sąsiadów wektora $\mathbf{x}$ (zbioru uczącego).

Klasyfikator k-NN zalicza klasyfikowany obiekt do takiej klasy, która jest najliczniej reprezentowana wśród $k$ najbliższych mu wektorów zbioru uczącego. Klasyfikator ten wymaga określenia liczby $k$ oraz zdefiniowanie terminu *„bliskości”*, tj. wyboru metryki w przestrzeni cech dla obliczenia odległości. Na określenie właściwej liczby $k$ nie ma jednoznacznych przepisów, [41]. Należy liczbę $k$ najbliższych sąsiadów uzależnić od ilości wektorów zbioru uczącego $N$, aby zachodziły warunki:

$$\lim_{N\to\infty} k(N) = \infty, \quad \lim_{N\to\infty} \frac{k(N)}{N} = 0. \tag{6.28}$$

Dokonując klasyfikacji badanych klas obiektów (bez względu na wykorzystywany klasyfikator) wyniki klasyfikacji przedstawiane są z wykorzystywaniem tzw. *macierzy przekłamań*. Załóżmy, że badamy procent rozpoznawalności klasy $X$ i $Y$ przyjmując populację np. 30 próbek z wykorzystaniem klasyfikatora np. k-NN. Przykładowy wynik klasyfikacji przedstawiony zostanie w postaci macierzy przekłamań, którą zilustrowano w Tab. 6.1:

Wyniki przedstawione w Tab. 6.1 należy interpretować w taki sposób, że próbki pochodzące z klasy $X$ system poprawnie rozpoznał 75% populacji, a 25% zakwalifikował do klasy $Y$ co oznacza, że zostały one błędnie zinterpretowane. W przypadku próbek pochodzących z klasy $Y$ system poprawnie zinterpretował 85% tej populacji,

Tablica 6.1. Macierz przekłamań klasyfikacji obiektu A i B

| a | b | | | ← | classified |
|---|---|---|---|---|---|
| 75 | 25 | | a | = | X |
| 15 | 85 | | b | = | Y |

natomiast błędnie zostało rozpoznanych 15%. W dalszej części niniejszej rozprawy wyniki klasyfikacji będą przedstawiane za pomocą macierzy przekłamań.

W Tab. 6.2 przedstawiono macierz przekłamań dla klasyfikacji skrzypiec i kontrabasu. Wykorzystano klasyfikator k-NN dla $k = 5$.

Tablica 6.2. Macierz przekłamań dla klasyfikacji skrzypiec i kontrabasu

| a | b | | | ← | classified |
|---|---|---|---|---|---|
| 99.33775 | 0.662252 | | a | = | skrzypce |
| 2.702703 | 97.2973 | | b | = | kontrabas |

### 6.3.3. Drzewa decyzyjne

Drzewa decyzyjne są podstawową metodą indukcyjnego uczenia się maszyn. Są one strukturą drzewiastą, której każdy węzeł odpowiada przeprowadzeniu testu na wartości atrybutu, zaś każdy liść zawiera decyzję o klasyfikacji przykładu. Pociąga to za sobą, że metoda ta związana jest z analizą przykładów, przy czym każdy z nich opisywany jest przez zestaw atrybutów, gdzie każdy atrybut może przyjmować różne wartości. Innymi słowy drzewem decyzyjnym jest graf-drzewo, którego korzeń jest stworzony przez wybrany atrybut, a poszczególne gałęzie reprezentują wartości tego atrybutu. Oznacza to, że z węzła wewnętrznego wychodzi tyle gałęzi, ile jest możliwych wyników testu odpowiadającemu temu węzłowi. Każda z nich prowadzi do poddrzewa służącego do klasyfikacji tych obiektów, dla których ten test ma określony wynik.

Wyznaczenie kategorii przykładu (tzn. klas, podzbioru przykładów) polega na przejściu ścieżki od korzenia drzewa do jednego z liści, przez wykonywanie w odwiedzanych kolejno węzłach odpowiednich testów i przemieszczanie się w dół, wzdłuż gałęzi odpowiadających uzyskiwanym wynikom testów. Liście mogą zawierać dowolne kategorie, co czyni drzewa decyzyjne przydatnymi do reprezentowania pojęć wielokrotnych, [45]. Można stwierdzić, że tworzenie drzewa decyzyjnego związane jest z wyborem atrybutu etykietującego korzeń drzewa, a następnie w ten sam sposób etykietowane są poddrzewa. Następnie każda wychodząca krawędź etykietowana jest poszczególnymi wartościami wybranego atrybutu. W liściach drzewa reprezentowane są obiekty o tej samej wartości atrybutu etykietującego. Jeżeli w bazie znajdują się obiekty niezgodne, to w niektórych liściach mogą być przechowywane obiekty z różnych klas. Dla każdego liścia podawana jest informacja o ilości $N$ obiektów klasyfikowanych w danym liściu, oraz o obiektach sklasyfikowanych błędnie (jeżeli takie istnieją). Po zakończeniu budowy drzewa dokonywane jest jego przycinanie (ang. *post-pruning*), [16].

Przycinanie drzewa polega na zastąpieniu jego wybranych poddrzew przez liście. Tego typu uproszczenie spowoduje na ogół pogorszenie dokładności klasyfikacji dla zbioru trenującego, ale może dawać lepsze efekty dla danych spoza tego zbioru.

Uzyskane po przycięciu drzewo będzie mniejsze i prostsze, co daje lepszą czytelność dla człowieka, oszczędność pamięci i większą efektywność obliczeniową procesu klasyfikacji (drzewo takie posiada krótsze ścieżki), [45].

Przykład wyniku klasyfikacji gitary akustycznej i gitary elektrycznej z wykorzystaniem drzew decyzyjnych pokazano w Tab. 6.3

Tablica 6.3. Macierz przekłamań dla klasyfikacji gitary elektrycznej i gitary akustycznej

| warunek | b | | | ← | classified |
|---|---|---|---|---|---|
| 86.2069 | 13.7931 | | a | = | gitara akustyczna |
| 9.375 | 90.625 | | b | = | gitara elektryczna |

### 6.3.4. Tablice decyzyjne

Tablice decyzyjne służą do graficznej, bardzo prostej prezentacji decyzji podjętej w zaistniałych warunkach. Pomijają one pozostałe elementy procesu decyzyjnego (nie określają sposobu ani adresata decyzji oraz procesów wprowadzania i wyprowadzania danych). Podstawę ich budowy stanowią związki przyczynowo skutkowe (jeżeli x to y). Można więc stwierdzić, że tablica decyzyjna (TD) jest pewna strukturą opisu zbioru związanych ze sobą reguł decyzyjnych.

$$\mathrm{TD} = (U, C, D, V, f), \tag{6.29}$$

gdzie:

$$C, D \subset A; \quad C \neq \phi, D \neq \phi; \quad C \cup D = A; \quad C \cap D = \phi,$$

$C$ — atrybuty warunkowe,
$D$ — atrybuty decyzyjne,
$f$ — funkcja decyzyjna $f : U \times A \to V$ taka, że $\underset{\substack{\mathbf{x}\in U \\ a\in A}}{\forall} f(\mathbf{x}, a) \in V_a$,
$A$ — niepusty skończony zbiór atrybutów,
$U$ — niepusty zbiór (uniwersum), którego elementy nazywane są obiektami,

$$U = \{u_1, \ldots, u_n\},$$

$$V = \bigcup_{a\in A} V_a; \quad \text{gdzie } V_a \text{ to dziedzina atrybutu } a \in A.$$

Przykład tablicy decyzyjnej przedstawiono w Tab. 6.4

W tablicy decyzyjnej (6.29) każdy obiekt $u \in U$ można opisać (przedstawić) w postaci zdania warunkowego w postaci: „*jeżeli warunek to decyzja*" co doprowadza do tego, że jest traktowany jak *reguła decyzyjna*. W tablicy (6.29) *regułą decyzyjną* nazywamy funkcję $g : C \cup D \to V$, jeżeli istnieje $x \in U$ taki, że $g = f_x$. Obcięcie $g$ do $C(g|C)$ oraz $g$ do $D(g|D)$ nazywamy odpowiednio warunkami, oraz decyzjami reguły decyzyjnej, [46]. Odwołując się do przykładu przedstawionego w Tab 6.4 można wyprowadzić przykładowe reguły:

- **jeżeli** $(a_1 = 32)$ **i** $(a_2 = 123)$ **to** $(f = 0)$,
- **jeżeli** $(a_1 = 112)$ **i** $(a_2 = 223)$ **to** $(f = 1)$.

Tablica 6.4. Przykładowa postać tablicy decyzyjnej

| | A | | | | | |
|---|---|---|---|---|---|---|
| TD | | $a_1$ | $a_2$ | ... | | $f$ |
| | | | | | | |
| $u_1$ | | 12 | 222 | ... | | 1 |
| $u_2$ | | 32 | 123 | ... | | 0 |
| $u_3$ | | 45 | 6 | ... | | 1 |
| $u_4$ | | 112 | 223 | ... | | 1 |
| ... | | ... | ... | ... | | ... |
| $u_{500}$ | | 71 | 82 | ... | | 0 |
| ... | | ... | ... | ... | | ... |

Reguła deterministyczna w tablicy decyzyjnej jest wówczas, gdy równość atrybutów warunkowych implikuje równości atrybutów decyzyjnych. Można to wyrazić za pomocą opisu zależności dla obiektów tablicy decyzyjnej:

$$\underset{\substack{x,y\in U\\ x\neq y}}{\forall}\left(\underset{c\in C}{\forall}(f(x,c)=f(y,c))\Rightarrow\underset{d\in D}{\forall}(f(x,d)=f(y,d))\right). \tag{6.30}$$

Niedeterministyczna reguła w tablicy decyzyjnej jest wówczas, gdy równość atrybutów warunkowych nie implikuje równości atrybutów decyzyjnych. Można to wyrazić następującą zależnością dla obiektów tablicy decyzyjnej:

$$\underset{\substack{x,y\in U\\ x\neq y}}{\exists}\left(\underset{c\in C}{\forall}(f(x,c)=f(y,c))\wedge\underset{d\in D}{\exists}(f(x,d)\neq f(y,d))\right). \tag{6.31}$$

Tablica decyzyjna jest deterministyczna (spójna), gdy reguły w niej zawarte są deterministyczne. W przeciwnym wypadku tablica decyzyjna jest niedeterministyczna (źle określona, niespójna). Wynik klasyfikacji altówki i wiolonczeli z wykorzystaniem tablic decyzyjnych pokazano w Tab. 6.5.

Tablica 6.5. Macierz przekłamań dla klasyfikacji altówki i wiolonczeli

| **warunek** | **b** | | | ← | **classified** |
|---|---|---|---|---|---|
| 93.33333 | 6.666667 | | a | = | altowka |
| 13.33333 | 86.66667 | | b | = | wiolonczela |

## 6.3.5. Podstawy teorii zbiorów przybliżonych

Teoria zbiorów przybliżonych została opracowana przez Prof. Zdzisława Pawlaka na początku lat osiemdziesiątych. Jest wykorzystywana jako narzędzie do analizy oraz redukcji zbiorów danych. Przechowywanie danych może odbywać się z wykorzystaniem szeregu struktur, natomiast sposób prezentacji powinien posiadać takie podstawowe cechy jak: uniwersalność (gromadzenie i przechowywanie zbiorów różnorodnych danych), oraz efektywność (umożliwienie łatwego sposobu analizy zarejestrowanych danych). W związku z powyższym w praktyce wykorzystuje się tablicowy sposób reprezentacji danych – a zatem zbiór danych jest przedstawiony w postaci tablicy. Przyjmując $SI = (U, A, V, f)$ jako system informacyjny i niech $B \subseteq A$, to $P \subseteq U$jest zbiorem $B$-dokładnym ($B$-definiowalnym) wtedy, gdy jest on skończoną

sumą zbiorów $B$-elementarnych. Każdy zbiór, który nie jest skończoną suma zbiorów *B-elementarnych* jest zbiorem *B-przybliżonym*, [50]. Jeżeli rozpatrzymy dwa zbiory: $X_1 = \{1, 2, 3, 5\}$ oraz $X_2 = \{3, 4, 5, 6\}$ to możemy stwierdzić, że:

- Zbiór $X_1$ jest zbiorem $A_1$-*dokładnym*, ponieważ jest skończoną suma zbiorów $A_1$-*elementarnych*: $X_1 = \{\{1\} \cup \{2, 5\} \cup \{3\}\}$.
- Zbiór $X_2$ jest zbiorem $A_1$-*przybliżonym*, ponieważ nie jest skończoną sumą zbiorów $A_1$-*elementarnych* (obiekty 2 i 5 należą do zbioru *B-elementarnego*, natomiast zbiór $X_2$ zawiera tylko obiekt 5 a nie zawiera obiektu 2).
- Zbiór $X_1$ jest zbiorem $A_2$-*przybliżonym*, ponieważ nie jest skończoną sumą zbiorów $A_2$-*elementarnych* (obiekty 3 i 4 należą do jednego zbioru *C-elementarnego*, natomiast zbiór $X_1$ zawiera tylko obiekt nr 3, a nie zawiera obiektu nr 4).
- Zbiór $X_2$jest zbiorem $A_2$-*przybliżonym*, ponieważ nie jest skończoną sumą zbiorów $A_2$-*elementarnych* (obiekty 1,2 i 5 należą do jednego zbioru *C-elementarnego*, natomiast zbiór $X_2$ zawiera tylko obiekt nr 5, a nie zawiera obiektów nr 1 i 2).

Jeżeli przedstawiony wcześniej system $SI$ jest systemem informacyjnym takim, że $B \subseteq A$ oraz $X \subseteq U$to:

- *B-dolnym przybliżeniem* (aproksymacją) zbioru $X$ w systemie $SI$ nazywamy zbiór:
$$\underline{B}X = \{x \in U : I_{SI,B}(x) \subseteq X\}, \tag{6.32}$$
- *B-górnym przybliżeniem* (aproksymacją) zbioru $X$ w systemie $SI$ nazywamy zbiór:
$$\overline{B}X = \{x \in U : I_{SI,B}(x) \cap X \neq \phi\}, \tag{6.33}$$
- *B-pozytywnym obszarem* zbioru $X$ w systemie informacyjnym $SI$ nazywamy zbiór:
$$POS_B(X) = \underline{B}X, \tag{6.34}$$
- *B-brzegiem* (granicą) zbioru $X$ w systemie informacyjnym $SI$ nazywamy zbiór:
$$BN_B(X) = \overline{B}X - \underline{B}X, \tag{6.35}$$
- *B-negatywnym obszarem* zbioru $X$ w systemie informacyjnym $SI$ nazywamy zbiór:
$$NEG_B(X) = U - \overline{B}X, \tag{6.36}$$

Powyższe definicje (6.33) – (6.36) sugerują wnioski:

- $\underline{B}X \subseteq X \subseteq \overline{B}X$,
- Zbiór $X$ jest *B-dokładny,* gdy: $\underline{B}X = \overline{B}X \Leftrightarrow BN_B(X) = \phi$,
- Zbiór $X$ jest *B-przybliżony,* gdy: $\underline{B}X \neq \overline{B}X \Leftrightarrow BN_B(X) = \phi$, [46].

## 6.4. Przykładowe metody eksperymentalne

Przeprowadzając testy związane z określeniem przynależności obiektów do klasy bardzo ważne jest określenie błędu rzeczywistego. Przeprowadza się to w odniesieniu do zbioru przykładów testowych (*zbioru testowego*), który nie był użyty w procesie uczenia klasyfikatora. Proces uczenia klasyfikatora przeprowadza się na tzw. *zbiorze treningowym*. Przykłady testowe klasyfikowane są za pomocą nauczonego na zbiorze treningowym klasyfikatora. Porównanie decyzji klasyfikatora umożliwia ocenę poprawności eksperymentu, i jest podstawą do definiowania miar błędu klasyfikacji. Najczęściej stosuje się *łączny błąd klasyfikacji,* wyrażany zależnością, [41]:

$$e = \frac{n_{\text{blad}}}{n_{\text{ltest}}}, \tag{6.37}$$

gdzie:

$n_{\text{blad}}$ — liczba błędnie sklasyfikowanych przykładów testowych,
$n_{\text{ltest}}$ — liczba przykładów testowych.

Poza tym stosuje się tzw. *sprawność klasyfikatora,* która określana jest jako uzupełnienie do jedynki łącznego błędu klasyfikacji. Miara ta jest opisywana zależnością:

$$s = \frac{n_{\text{ppr}}}{n_{\text{ltest}}} = 1 - e, \tag{6.38}$$

gdzie:

$n_{\text{ppr}}$ — liczba poprawnie sklasyfikowanych przykładów testowych.

Bardzo istotnym zagadnieniem jest optymalne wyznaczenie wartości parametrów klasyfikatora, czyli jego *walidacji.* Wykorzystywany jest tzw. *zbiór walidacyjny,* który najczęściej zostaje pozyskany z wejściowego zbioru danych i za pomocą określonych metod podzielony na część *uczącą* (tzw. *zbiór treningowy)* i część *testową* (tj. *zbiór walidacyjny).* Zbiór walidacyjny służy do oceny stopnia nauczenia systemu na zborze treningowym. Jest wiele metod podziału zbioru. Do celów przeprowadzanych badań zastosowano niektóre z nich.

### 6.4.1. Metoda *holdout*

Jest to metoda, która polega na jednokrotnym podziale zbioru na część treningową i testową. Najczęściej podział ten odbywa się z wykorzystaniem proporcji 2/3—część ucząca, 1/3—część testowa. Podczas prowadzonych badań zdecydowano się uwzględnić podział, który dla części testowej oscylował od 20 % do 40 % populacji zbioru wejściowego. W Tab. 6.5 pokazano różnice wyniku klasyfikacji altówki i wiolonczeli przy pomocy tablicy decyzyjnej z wykorzystaniem dla części testowej 35 % populacji zbioru wejściowego (Tab. 6.6) oraz 25 % (Tab. 6.7).

Tablica 6.6. Macierz przekłamań dla klasyfikacji altówki i wiolonczeli przy podziale zbioru 65 %:35 %

| **a** | **b** | | | ← | **classified** |
|---|---|---|---|---|---|
| 75 | 25 | | a | = | altowka |
| 11, 11111 | 88, 88889 | | b | = | wiolonczela |

Tablica 6.7. Macierz przekłamań dla klasyfikacji altówki i wiolonczeli przy podziale zbioru 75 %:25 %

| a | b | | | ← | classified |
|---|---|---|---|---|---|
| 77,77778 | 33,33333 | | a | = | altowka |
| 0 | 100 | | b | = | wiolonczela |

## 6.4.2. Metoda *k-krotnej walidacji krzyżowej*

W metodzie tej zbiór wejściowy zbiór danych $U$ zostaje losowo podzielony na $k$ równolicznych podzbiorów $U(i)$ dla $i = 1, \ldots, k$. W $i$-tej iteracji, zbiór $(U - U(i))$ jest traktowany jako zbiór uczący klasyfikatora. Zbiór $U(i)$ jest traktowany jako zbiór testowy. Całościowy błąd klasyfikacji $e$ w tej metodzie jest określany jako wartość średnia z błędów $e_i$ estymowanych w każdej iteracji. Błąd klasyfikacji jest opisywany wyrażeniem:

$$e = \frac{1}{k} \sum_{i=1}^{k} e_i \tag{6.39}$$

Parametr $k$ jest dobierany ze względu na rozmiar zbioru uczącego $U$. Podczas prowadzonych badań bardzo często przyjmuje się wartość $k = 10$. Jeżeli mamy do czynienia z małymi zbiorami danych wartość $k$ może być większa, co w granicznym przypadku $k = N$ daje metodę *leave-one-out.* W kolejnych iteracjach metody *leave-one-out* zbiór $U(i)$ tworzy pojedynczy obiekt zbioru uczącego $U$, którym testuje się nauczony za pomocą $U - U(i)$ klasyfikator. Ogólny błąd klasyfikacji $e$ jest obliczany jako średnia błędów $e_i$ estymowanych w każdej iteracji—co wyrażono wzorem (6.39).

Duże wartości $k$ w metodzie *k-krotnej walidacji krzyżowej* powodują, że uzyskany estymator błędu posiada małe obciążenie, lecz dużą wariancję. Małe wartości parametru $k$ powodują natomiast, że estymator posiada duże obciążenie i małą wariancję, [41].

Poniżej pokazano wyniki (macierze przekłamań) klasyfikacji z wykorzystaniem metody *k-krotnej walidacji krzyżowej* dla różnego $k$. Klasyfikacji poddano altówkę, wiolonczele, gitarę elektryczną i gitarę akustyczną wykorzystując drzewa decyzyjne.

Tablica 6.8. Macierz przekłamań dla k=29

| a | b | c | d | | | ← | classified |
|---|---|---|---|---|---|---|---|
| 86,2069 | 13,7931 | 0 | 0 | | a | = | gitara_akustyczna |
| 6,25 | 78,125 | 3,125 | 12,5 | | b | = | gitara_akustyczna |
| 3,333333 | 3,333333 | 86,66667 | 6,666667 | | c | = | altowka |
| 0 | 0 | 10 | 90 | | d | = | wiolonczela |

Tablica 6.9. Macierz przekłamań dla k=10

| a | b | c | d | | | ← | classified |
|---|---|---|---|---|---|---|---|
| 82,75862 | 13,7931 | 0 | 3,448276 | | a | = | gitara_akustyczna |
| 9,375 | 75 | 3,125 | 12,5 | | b | = | gitara_akustyczna |
| 6,666667 | 6,666667 | 76,66667 | 10 | | c | = | altowka |
| 0 | 6,666667 | 10 | 83,33333 | | d | = | wiolonczela |

Tablica 6.10. Macierz przekłamań dla k=5

| a | b | c | d | | | ← | classified |
|---|---|---|---|---|---|---|---|
| 93, 10345 | 6, 896552 | 0 | 0 | | a | = | gitara_akustyczna |
| 12, 5 | 75 | 3, 125 | 9, 375 | | b | = | gitara_akustyczna |
| 6, 666667 | 16, 66667 | 73, 33333 | 3, 333333 | | c | = | altowka |
| 0 | 0 | 20 | 80 | | d | = | wiolonczela |

## 6.5. Przygotowanie danych testowych

Przed przystąpieniem do procesu klasyfikacji istotną rolę odgrywa właściwy dobór cech ze zbioru wszystkich dostępnych. Często się zdarza, że zbiór dostępnych cech osiąga wartość kilkudziesięciu lub nawet kilkuset—co nie jest optymalnym stanem. W kontekście procesu klasyfikacji pożądane jest wyselekcjonowanie grupy cech, które przynoszą najkorzystniejszy efekt podczas klasyfikacji danych. Doprowadza to do tego, że przed przystąpieniem do faktycznej segregacji należy przeprowadzić proces redukcji wektora cech. Zbyt duża liczba cech w wektorze cech powoduje wzrost ilości wolnych parametrów w klasyfikatorze, koniecznych do oszacowania, a więc zarazem wzrost złożoności klasyfikatora. Proces ten zwiększa niebezpieczeństwo przeuczenia, a w konsekwencji spadku zdolności uogólniających klasyfikatora. Oznacza to, że podczas procesu klasyfikacji często okazuje się, że zbyt duży wektor cech ostatecznie doprowadza do gorszej rozpoznawalności obiektów. Dla zilustrowania opisywanego stanu przeprowadzono test dla 4 klas (gitara akustyczna, gitara elektryczna, altówka, wiolonczela) z uwzględnieniem wektora cech o różnych rozmiarach. Wykorzystano drzewa decyzyjne oraz metodę k-krotnej walidacji krzyżowej ($k = 10$). Proces selekcji cech przeprowadzony był z wykorzystaniem algorytmów genetycznych—co zostanie opisane w dalszej części niniejszej rozprawy. Dla wektora cech 51 elementowego rozpoznawalność wynosiła 63.5556 %—macierz przekłamań klasyfikacji przedstawiono w Tab. 6.11:

Tablica 6.11. Macierz przekłamań dla wektora cech 51 elementowego

| a | b | c | d | e | f | g | h | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|
| 37,5 | 0 | 31,25 | 12,5 | 0 | 12,5 | 0 | 6,25 | a | = | harfa |
| 0 | 86,2069 | 6,896552 | 6,896552 | 0 | 0 | 0 | 0 | b | = | gitara akustyczna |
| 9,375 | 6,25 | 68,75 | 12,5 | 0 | 3,125 | 0 | 0 | c | = | gitara elektryczna |
| 0 | 3,571429 | 14,28571 | 78,57143 | 0 | 0 | 3,571429 | 0 | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 96,66667 | 3,333333 | 0 | 0 | e | = | altówka |
| 10 | 0 | 0 | 0 | 0 | 73,33333 | 3,333333 | 13,33333 | f | = | skrzypce |
| 0 | 0 | 0 | 0 | 3,333333 | 0 | 96,66667 | 0 | g | = | kontrabas |
| 6,666667 | 0 | 0 | 0 | 0 | 20 | 6,666667 | 66,66667 | h | = | wiolonczela |

Dla wektora cech 25 elementowego rozpoznawalność wynosiła 65.3333 % — macierz przekłamań pokazano w Tab. 6.12.

Z powyższych przykładów jasno wynika, że skrócenie wektora cech o 49,02 % zwiększyło rozpoznawalność danych klas o 1,7774 %. Wypływa stąd wniosek, że w praktyce może wystąpić zjawisko spadku sprawności klasyfikatora w miarę dodawania nowych cech.

Równie istotnym zagadnieniem związanym z przygotowaniem danych testowych jest *normalizacja danych*. Proces ten przeprowadza się w celu wyrównania wpływu poszczególnych cech, które w sposób znaczący różnią się pod względem wielkości

Tablica 6.12. Macierz przekłamań dla wektora cech 25 elementowego

| a | b | c | d | e | f | g | h | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|
| 12,5 | 6,25 | 50 | 0 | 6,25 | 18,75 | 6,25 | 0 | a | = | harfa |
| 0 | 79,31034 | 10,34483 | 6,896552 | 0 | 0 | 3,448276 | 0 | b | = | gitara akustyczna |
| 6,25 | 3,125 | 62,5 | 9,375 | 6,25 | 6,25 | 0 | 6,25 | c | = | gitara elektryczna |
| 0 | 3,571429 | 10,71429 | 78,57143 | 0 | 0 | 7,142857 | 0 | d | = | gitara basowa |
| 0 | 3,333333 | 0 | 3,333333 | 80 | 3,333333 | 6,666667 | 3,333333 | e | = | altówka |
| 6,666667 | 3,333333 | 10 | 0 | 0 | 63,33333 | 0 | 16,66667 | f | = | skrzypce |
| 0 | 0 | 10 | 0 | 10 | 10 | 63,33333 | 6,666667 | g | = | kontrabas |
| 0 | 3,333333 | 3,333333 | 0 | 0 | 16,66667 | 16,66667 | 60 | h | = | wiolonczela |

bądź zakresem przyjmowanych wartości. Doprowadza to do tego, że cechy o większych wartościach mają większy wpływ w różnorodnych kryteriach stosowanych w algorytmach klasyfikacji, niż cechy o porównywalnie małych wartościach. Metody normalizacji danych sprowadzają się do liniowego lub nieliniowego skalowania danych do odpowiedniego zakresu. Jest wiele metod normalizacji, a jedną z nich jest normalizacja każdego z $d$ wymiarów, [41]:

$$x_{ij} = \frac{x^*_{ij} - m_j}{\sigma_j}\,, \tag{6.40}$$

$$\sigma_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x^*_{ij} - m_j\right)^2, \tag{6.41}$$

$$m_j = \frac{1}{N}\sum_{i=1}^{N} x^*_{ij}\,, \qquad j=1,2,\ldots,d\ , \tag{6.42}$$

gdzie:

$x_{ij}$ — wartość j-tej współrzędnej i-tego wektora cech po normalizacji,
$x^*_{ij}$ — przed normalizacją.

Tak znormalizowane cechy posiadają zerową wartość średnią i jednostkową wariancję.

Można również normalizować cechy przez sprowadzenie wartości dla wszystkich wymiarów do przedziału $[0,1]$ przez odjęcie najmniejszej wartości na każdej osi i podzielenie jej przez zakres wartości na danej osi:

$$x_{ij} = \frac{x^*_{ij} - \min\limits_j\{x_{ij}\}}{\max\limits_j\{x_{ij}\} - \min\limits_j\{x_{ij}\}}\,. \tag{6.43}$$

Kolejnym przypadkiem normalizacji jest normalizacja do wartości średniej danej cechy:

$$x_i = \frac{x^*_i}{śr_i}\,, \tag{6.44}$$

gdzie: $śr_i$ — średnia wartość i-tej cechy.

Normalizując cechy altówki do wartości średniej zanotowano istotne zmiany wartości dla odchylenia standardowego poszczególnych cech. Przykładowo, dla nieregularności widma odchylenie standardowe przed normalizacja wynosiło $\sigma' = 0.0254816$ — natomiast po normalizacji $\sigma' = 0.005135$. W przypadku cechy *zero crossing* odchylenie standardowe przed normalizacją $\sigma' = 17.455233$ — natomiast po normalizacji $\sigma' = 0.564214$.

## 6.6. Wykorzystanie algorytmów genetycznych do selekcji cech

Jak już wspomniano wcześniej prawidłowa selekcja cech może doprowadzić do korzystniejszych wyników procesu klasyfikacji. Zatem podczas procesu selekcji dokonuje się wyboru $d$ elementowego podzbioru cech spośród $D$ elementowego, wejściowego zbioru cech ($d < D$) tak, aby zoptymalizować pewną funkcję kryterialną $J(\cdot)$, [41]:

$$\begin{bmatrix} x_1 \\ \dots \\ \dots \\ x_D \end{bmatrix} \rightarrow \begin{bmatrix} x_{i_1} \\ \dots \\ x_{i_d} \end{bmatrix} \quad \{x_{i_1}, \dots, x_{i_d}\} = \arg\max_{d, i_d} [J(\{x_i | i = 1, \dots, D\})] \,. \tag{6.45}$$

W celu wyczerpującego przeszukiwania wszystkich możliwych podzbiorów $D$ cech wejściowych dla zadanej wartości konieczne jest przyjęcie odpowiedniej strategii. Podczas procesu selekcji cech konieczne jest ustalenie dwóch elementów:

1. *Funkcji kryterialnej*, która ocenia właściwość podzbiorów cech. Wyróżnia się dwie grupy funkcji kryterialnych. Pierwszą z nich stanowią funkcje wykorzystujące tzw. *miary informacyjne*, których zadaniem jest ocena stopnia separowalności klas. W drugiej grupie w charakterze funkcji kryterialnej stosuje się klasyfikator, za pomocą którego dokonuje się oceny podzbioru cech na podstawie sprawności klasyfikowania.

2. *Strategii przeszukiwania* dla odpowiedniego wyboru podzbiorów cech. Możemy tu wyróżnić trzy podgrupy algorytmów selekcji cech.

   a) algorytmy wykładnicze—wśród których najczęściej stosowanym jest algorytm *branch&bound* , [47], w którym ogranicza się przeszukiwanie całej przestrzeni rozwiązań przez stosowanie ograniczenia na wartość funkcji kryterialnej. Wymaga on spełnienia przez funkcje kryterialną $J(.)$ *warunku monotoniczności*, który określa, że dla każdych dwóch podzbiorów cech $X_1, X_2$ takich, że $X_1 \subset X_2$, wartości funkcji kryterialnych spełniają warunek $J(X_1) < J(X_2)$. Sprowadza się to do tego, że skuteczność (właściwość) podzbioru cech wzrasta w miarę dodawania nowych cech.

   b) algorytmy sekwencyjne—istotą ich jest sekwencyjne dodawanie lub usuwanie cech w procesie przeszukiwania. Algorytmy te maja tendencje do uzyskiwania rozwiązań stanowiących minima lokalne, [41].

   c) algorytmy stochastyczne—stosuje się w nich element losowości dla zapobiegania otrzymania rozwiązania stanowiącego minimum lokalne. Przykładem tej grupy algorytmów są algorytmy genetyczne.

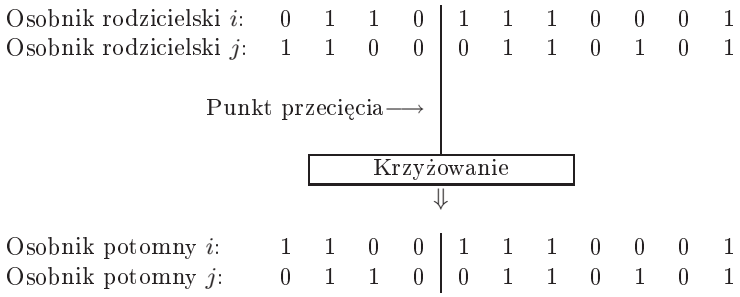
Zadaniem selekcji cech jest wytypowanie tych cech, które są najbardziej istotne z punktu widzenia ich przydatności w procesie klasyfikacji. W celu przeprowadzenia selekcji cech można wykorzystać algorytmy genetyczne, które znajdują bardzo szerokie zastosowanie wśród badaczy zajmujących się automatyczną klasyfikacja obiektów. Algorytm genetyczny jest procedurą optymalizacyjną, która określa sposób przeszukiwania przestrzeni odpowiednio zakodowanych rozwiązań. W algorytmach

genetycznych występuje zjawisko równoległego poszukiwania wielu rozwiązań, które tworzy tzw. *populację*—a zatem algorytm genetyczny operuje na populacji jednostek, [48]:

$$P(n) = \{x_1^n, x_2^n, \ldots, x_R^n\} \tag{6.46}$$

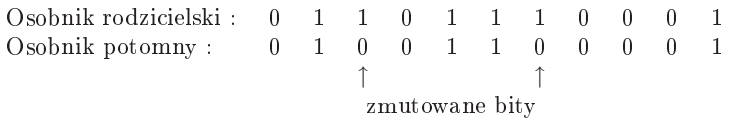
gdzie:
$n$ — numer generacji,
$R$ — rozmiar populacji.

Pojedyncze rozwiązanie tzw. punkt w przestrzeni rozwiązań nazywane jest *chromosomem*, który w przypadku kodowania binarnego jest $n$-elementowym wektorem genów, złożonym z zer i jedynek. Algorytm genetyczny dokonuje transformacji populacji chromosomów w sposób analogiczny do procesu naturalnej ewolucji. Podczas reprodukcji chromosomy o lepszym przystosowaniu są powielane z większym prawdopodobieństwem niż osobniki gorzej przystosowane. Proces ten doprowadza do ukierunkowania w lepszą stronę rozwiązań. W celu przeprowadzenia reprodukcji stosuje się operacje *krzyżowania* i *mutacji*. Krzyżowanie chromosomów rodzicielskich polega na ich przecięciu i utworzenie dwóch potomnych. Proces ten zilustrowano na Rys. 6.2.




Rysunek 6.2. Operacja krzyżowania

Mutacja chromosomów wykonywana jest z odpowiednim prawdopodobieństwem dla każdego genu osobno i polega na zmianie wartości bitu na przeciwny. Proces mutacji przedstawiono na Rys. 6.3.




Rysunek 6.3. Operacja mutacji

Sztuczna ewolucja chromosomów trwa do momentu aż zostanie spełniony warunek zakończenia, którym może być wyszukanie chromosomu o odpowiednio dużej wartości funkcji przystosowania.

ROZDZIAŁ 7

# Przygotowanie danych eksperymentalnych i przyjęcie metodologii badań

## 7.1. Zaproponowana grupa instrumentów wykorzystywana w badaniach

Celem prowadzonych badań była analiza oraz próba parametryzacji dźwięków muzycznych, których źródłem są instrumenty zaliczane do grupy *chordofonów*. Dodatkowo postanowiono zawęzić obszar zainteresowań do artykulacji *pizzicato*, co pozwala skupić się na rozwiązaniu bardzo wąskiego problemu — zarazem nie często podejmowanego jako oddzielne zagadnienie. Podczas realizacji badań związanych z automatyczną klasyfikacją instrumentów muzycznych często do rozważań włącza się klasy instrumentów pochodzących z różnych grup (charakterystykę wybranych instrumentów opisano w rozdziale 3), co w niektórych przypadkach znacznie ułatwia proces klasyfikacji. Wydaje się sprawą oczywistą, że łatwiej odszukać charakterystyczne cechy pozwalające odróżnić instrumenty dęte od młoteczkowych, gdyż wtedy uzyskuje się wysoką rozpoznawalność badanych klas. Jeżeli natomiast swoją uwagę badawczą skupimy np. na skrzypcach i altówce, to stosowane deskryptory mogą okazać się mniej efektywne niż tego oczekiwano — wynika to z dużego podobieństwa obu instrumentów. Zadanie, które zostało podjęte podczas realizacji niniejszych badań, ma na celu odszukanie takiego wektora cech, który pozwoli w zadowalający sposób dokonać automatycznej klasyfikacji tej wąskiej grupy instrumentów z uwzględnieniem tylko artykulacji *pizzicato* (pominięto atak na strunę smyczkiem). Ponadto zdecydowano się wybrać do badań próbki pochodzące tylko z czterech oktaw:

1. wielkiej (A 110 Hz),
2. małej (a 220 Hz),
3. razkreślnej ($a^1$ 440 Hz),
4. dwukreślnej ($a^2$ 880 Hz).

Źródłem decyzji związanej z wyborem oktaw jest chęć uwzględnienia takich próbek, które znajdują swoją reprezentację w określonej oktawie dla wszystkich badanych

klas instrumentów — skrzypce nie posiadają zdolności artykulacji tak niskich dźwięków jak kontrabas (np. contra) — a zatem rozpoznawalność w tej oktawie dla tych dwóch klas staje się zagadnieniem elementarnym.

Ostatecznie do badań zdecydowano się przeznaczyć 820 monofonicznych (16 bitów, częstotliwość próbkowania 44.1 kHz) próbek dźwięków zawierających się w zakresie w/w oktaw. Zakres częstotliwości $f$ badanych próbek leży w granicach $65.41\,\mathrm{Hz} < f < 987.77\,\mathrm{Hz}$. W trakcie badań analizowano pojedyncze dźwięki do momentu naturalnego wybrzmiewania nuty. Bazę dźwięków skompletowano wykorzystując:

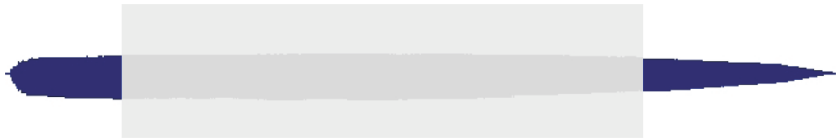
1. Efekt sesji nagraniowej zorganizowanej w studio nagraniowym Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych w Warszawie;
2. Skorzystano z darmowej bazy dźwięków udostępnionej przez The University of Iowa Electronic Music Studios (`http://theremin.music.uiowa.edu/MIS.html`);
3. Skorzystano ze zbiorów dr Alicji Wieczorkowskiej — za udostępnienie próbek autor rozprawy składa serdeczne podziękowania.

## 7.2. Fizyczne cechy próbek dźwięków badanych klas instrumentów

Analizując przebiegi czasowe aerofonów lub elektrofonów możemy zaobserwować trzy podstawowe składowe dźwięku:

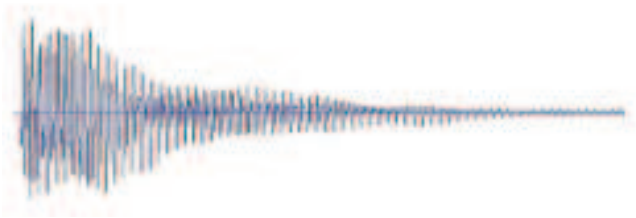
1. czas narastania dźwięku — tzw. transjent początkowy,
2. stan quasi-ustalony,
3. czas wybrzmiewania nuty — tzw. transjent końcowy.

Przykład przebiegu zawierającego wszystkie w/w składowe pokazano na Rys. 7.1, [53]:



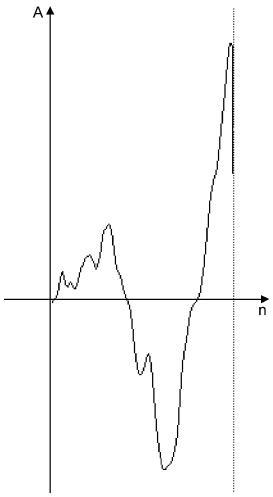
Rysunek 7.1. Przykład wykresu postaci czasowej dźwięku waltorni, a razkreślne (440 Hz). Zacieniowano stan quasi-ustalony.

Dokonując wnikliwej analizy czasowej i widmowej każdego fragmentu przebiegu można odszukać cechy istotne dla procesu automatycznej klasyfikacji. W przypadku analizy sygnałów pochodzących z grupy instrumentów strunowych, z wykorzystaniem artykulacji *pizzicato*, można dokonywać parametryzacji tylko transjentu początkowego oraz końcowego — stan quasi-ustalony w tym przypadku nie występuje (co jest cechą charakterystyczną tych instrumentów).



Rysunek 7.2. Przykład wykresu postaci czasowej dźwięku altówki, a razkreślne (440 Hz)

Skupiając się na poszczególnych składowych w/w przebiegu, można zauważyć, że transjent początkowy jest bardzo krótki, a co za tym idzie zbyt ubogi w kontekście parametryzacji. W przypadku altówki (dźwięk $a^1$) transjent początkowy składa się z zaledwie ok. 130 próbek i nie stanowi jednego okresu przebiegu (podobną sytuację zaobserwowano analizując dźwięki pozostałych klas instrumentów przeznaczonych do badań). Transjent początkowy altówki a razkreślne pokazano na Rys. 7.3.




Rysunek 7.3. Transjent początkowy — altówka, dźwięk $a^1$

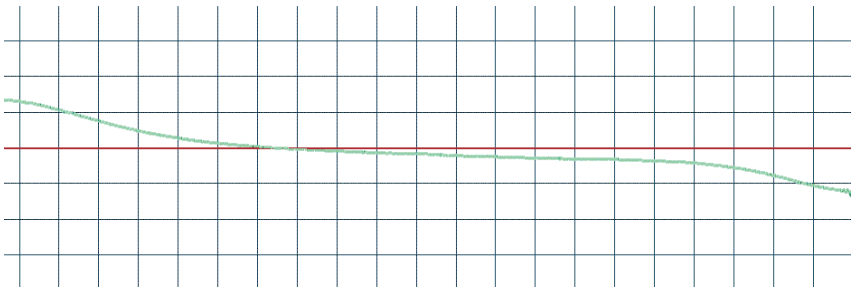
Stwierdzono, że taka charakterystyka transjentu początkowego jest właściwa dla chordofonów z artykulacją *pizzicato*. W związku z tym w trakcie prowadzonych badań zrezygnowano z analizy tej części przebiegu, skupiając się na badaniu tylko transjentu końcowego (rozpoczynając od wartości maksymalnej amplitudy).

## 7.3. Zaproponowana metodologia badań

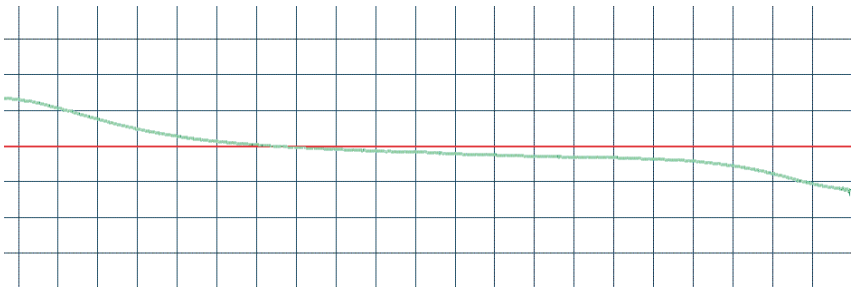
Jak już wspomniano w rozdziale 4, proces parametryzacji może odbywać się z wykorzystaniem zarówno postaci czasowej jak i widmowej danego przebiegu. W trakcie prowadzonych badań zdecydowano się włączyć do wektora cech zarówno deskryptory czasowe jak i widmowe.

### 7.3.1. Wykorzystane deskryptory funkcji czasu

W kontekście parametryzacji instrumentów muzycznych z artykulacją *pizzicato* wykorzystanie deskryptorów postaci czasowej badanego przebiegu może wzbudzać polemikę. Jak już wspomniano wcześniej w celu analizy przebiegu konieczne jest pobranie okna o określonej długości. Powstaje zatem pytanie: jak długie okno czasowe należy pobrać, aby otrzymać efektywne deskryptory postaci czasowej? Jeżeli założymy, że do analizy zostanie przeznaczone okno zbyt długie (np. 1/10 sekundy), to może się okazać, że próbka dźwięku będzie krótsza niż zadane okno — taka sytuacja może się wydarzyć, gdy muzyk użyje techniki gry *staccato*. Jeżeli jednak do analizy przeznaczymy zbyt krótkie okno czasowe, to może się okazać, że nie zawiera ono informacji mogących się przyczynić do poprawienia skuteczności klasyfikacji. Na Rys. 7.4.a i 7.4.b pokazano fragmenty przebiegów czasowych o długości 1/100 sekundy (441 sampli) gitary basowej oraz kontrabasu dla dźwięku dis$^2$:



Rysunek 7.4. a) Fragment przebiegu czasowego gitary basowej



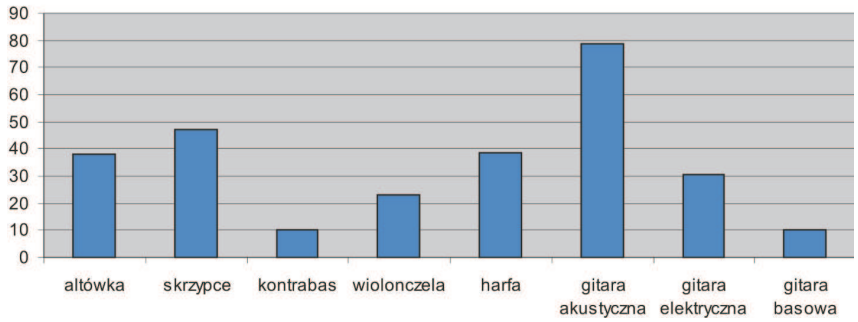
Rysunek 7.4. b) Fragment przebiegu czasowego kontrabasu

Na podstawie powyższych przykładów łatwo dojść do wniosku, że podczas analizy postaci czasowej zastosowanie okna o długości 1/100 sekundy nie przynosi oczekiwanych efektów automatycznej klasyfikacji.

W celu opisu postaci czasowej sygnału dźwiękowego zdecydowano się zatem wykorzystać dwa klasyczne (bardzo szeroko stosowane w procesie klasyfikacji) parametry czasowe:

1. $ZC$ — (zero crossing) gęstość przejść przez zero osi OX w zadanym oknie. Na podstawie obserwacji postaci czasowej wszystkich zgromadzonych próbek, do analizy zdecydowano się wybrać okno o długości, n=1 500 (ok. 1/30 sekundy) próbek rozpoczynając od wartości *max*, a więc wykluczono transjent początkowy przebiegu. Stwierdzono, że okno o takim rozmiarze jest wystarczająco długie, aby zastosować je do wszystkich badanych próbek dźwięku. Należy

jednak pamiętać, że podczas analizy badano przebiegi do naturalnego wybrzmiewania — a zatem nie uwzględniano techniki *staccato*. Średni rozkład wartości parametru $ZC$ dla wszystkich badanych klas instrumentów pokazano na Rys. 7.5.




Rysunek 7.5. Średni rozkład wartości parametru $ZC$

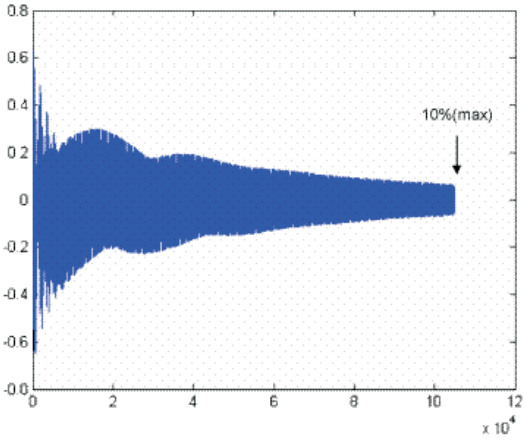
2. $l_{tk}$ — logarytm czasu wybrzmiewania dźwięku wyrażony zależnością:

$$l_{tk} = \log(t_{pk} - t_{\max}), \tag{7.1}$$

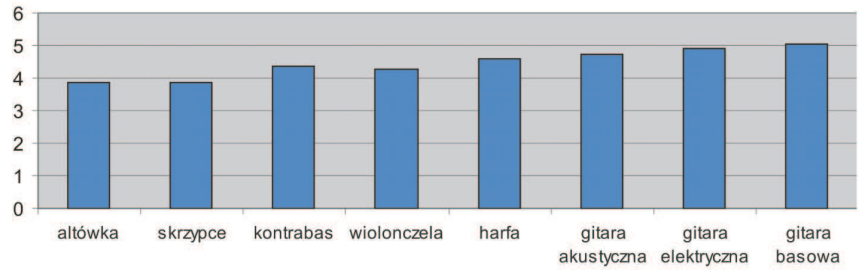
gdzie:
$t_{max}$ — czas osiągnięcia maksymalnej amplitudy dźwięku,
$t_{pk}$ — czas osiągnięcia progu 10% maksymalnej amplitudy dźwięku w transjencie końcowym.

Badając czas wybrzmiewania nuty analizowano fragment przebiegu rozpoczynając od wartości maksymalnej amplitudy do osiągnięcia wartości mniejszej niż 10% wartości maksymalnej amplitudy przebiegu. Założenie takie jest szeroko przyjmowane w teorii sygnałów, [2]. Przykład pobranego okna (gitara akustyczna — c małe 131.12 Hz) pokazano na Rys. 7.6.

Średni rozkład parametru $t_{lk}$ dla badanych klas instrumentów pokazano na Rys. 7.7.




Rysunek 7.6. Przykład pobranego fragmentu przebiegu stanowiącego podstawę do obliczenia wartości parametru $l_{tk}$




Rysunek 7.7. Rozkład parametru $t_{lk}$ dla badanych klas instrumentów

Z powyższego wykresu łatwo odczytać, że parametr $t_{lk}$ może wykazać się skutecznością za wyjątkiem altówki i skrzypiec (czego można było się się spodziewać z uwagi na dalekie podobieństwo tych instrumentów). Przykład wyniku klasyfikacji ośmiu instrumentów z wykorzystaniem tylko parametru $t_{lk}$ (klasyfikator k-NN, metoda holdout, podział zbioru 80%:20%) pokazano w poniższej macierzy przekłamań. Ogólna rozpoznawalność dla tego testu wynosiła 46.6667%.

Tablica 7.1. Macierz przekłamań dla parametru $t_{lk}$ — klasyfikator k-NN, metoda holdout

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 0 | 25 | 0 | 0 | 25 | 0 | 0 | | **a** | = | **harfa** |
| 0 | 40 | 40 | 0 | 0 | 0 | 20 | 0 | | **b** | = | **gitara akustyczna** |
| 0 | 14.3 | 42.9 | 14.3 | 0 | 0 | 14.3 | 14.3 | | **c** | = | **gitara elektryczna** |
| 0 | 33.3 | 33.3 | 33.3 | 0 | 0 | 0 | 0 | | **d** | = | **gitara basowa** |
| 0 | 0 | 0 | 0 | 33.3 | 33.3 | 0 | 33.3 | | **e** | = | **altówka** |
| 0 | 0 | 0 | 0 | 0 | 66.7 | 0 | 33.3 | | **f** | = | **skrzypce** |
| 10 | 0 | 20 | 10 | 10 | 20 | 30 | 0 | | **g** | = | **kontrabas** |
| 0 | 0 | 0 | 0 | 0 | 0 | 28.6 | 71.4 | | **h** | = | **wiolonczela** |

### 7.3.2. Deskryptory postaci widmowej

Widmo zawiera bardzo wiele szczegółów, a zatem do celów automatycznej klasyfikacji instrumentów muzycznych konieczna jest jego parametryzacja. W celu odszukania wektora cech widmowych wybranych instrumentów przeprowadzono szereg badań związanych z odczytem wartości dla ogólnie stosowanych deskryptorów (np. środek ciężkości widma). Celem uzyskania porównywalnych wyników dla wszystkich klas instrumentów oraz wszystkich badanych próbek dźwięku zdecydowano się wybrać do analizy „stałe" okno czasowe dla każdej próbki. Pod pojęciem *stałego okna czasowego* rozumiemy fragment przebiegu, który został pobrany zawsze w tym samym czasie oraz zawiera ta samą ilość próbek. W efekcie założenie to doprowadzi do porównywania widma takiego samego fragmentu przebiegu dla całej populacji badanych dźwięków. Przyjęto, że takie rozwiązanie umożliwi analizę i porównanie tych samych fragmentów widma, co pozwoli uzyskać wysoką skuteczność automatycznej klasyfikacji. Jak już opisano wcześniej (por 7.2), transjent początkowy zdecydowano się pominąć w trakcie analizy próbek dźwięków. Do analizy widmowej zdecydowano się przeznaczyć okno czasowe, które zostało pobrane od momentu osiągnięcia maksymalnej wartości amplitudy. Długość pobranego okna jest zdeterminowana ustaleniem

właściwej rozdzielczości widma, wyrażonej zależnością, [53]:

$$f_r = \frac{f_s}{n}, \tag{7.2}$$

gdzie:
$f_r$ — rozdzielczość widma,
$f_s$ — częstotliwość próbkowania,
$n$ — ilość próbek.

Podczas prowadzonych badań analizowano okno sygnału o długości 11 025 próbek, co oznacza, że przyjęto rozdzielczość widma $f_r = 4\,\text{Hz}$. Jeżeli zaistniała sytuacja, że badany dźwięk jest krótszy niż zadana długość okna (sytuacja taka ma miejsce w przypadku wyższych oktaw), wówczas miejsca brakujących wartości zostały uzupełnione zerami, aż do długości okna $n = 11\,025$. Wycięty fragment przebiegu postaci czasowej został poddany DFT, a jego widmo poddano szczegółowej analizie. Ponadto podczas prowadzonych badań brano pod uwagę pełną reprezentację widma, a nie tylko składowe harmoniczne. Oznacza to, że w trakcie procesu odczytania wartości poszczególnych deskryptorów uwzględniano zarówno składowe harmoniczne oraz składowe przecieku częstotliwości. Motywacją do zastosowania takiej metodologii badań jest fakt, że:

1. Wszelkie operacje wstępne na widmie wyciętego okna powodują zwiększenie złożoności algorytmicznej procesu automatycznej klasyfikacji, co nie jest korzystne w kontekście filtrowania multimedialnych baz danych.

2. Badając dźwięki, o których mowa w niniejszej rozprawie, nie można wykluczyć zjawiska przecieku widma. Oznacza to, że przeciek jest integralną składową badanego przebiegu. Założono, że z badawczego punku widzenia, włączając do analizy również zjawisko przecieku widma można otrzymać wyniki, które w większym stopniu oddają rzeczywistość.

Poniżej przedstawiono wyniki procesu klasyfikacji ośmiu klas instrumentów — w obu przypadkach zastosowano 94 elementowy (taki sam) wektor cech. W Tab. 7.2 przedstawiono macierz przekłamań dla klasyfikacji z użyciem klasyfikatora k-NN, metoda holdout (80%:20%). Wartości deskryptorów obliczono na podstawie prążków harmonicznych badanego fragmentu przebiegu. Ogólna rozpoznawalność dla tego testu wynosiła 40%.

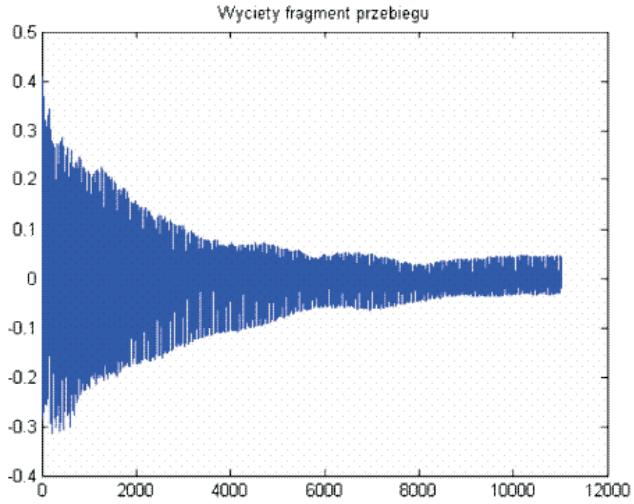
Tablica 7.2. Macierz przekłamań — wyniki uzyskane na podstawie harmonicznych widma badanego fragmentu przebiegu

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 33.3 | 33.3 | 0 | 0 | 0 | 33.3 | 0 | 0 | | **a** | = | **harfa** |
| 0 | 16.7 | 16.7 | 16.7 | 16.7 | 0 | 33.3 | 0 | | **b** | = | **gitara akustyczna** |
| 0 | 0 | 42.9 | 0 | 28.6 | 14.3 | 0 | 14.3 | | **c** | = | **gitara elektryczna** |
| 0 | 0 | 14.3 | 71.4 | 0 | 0 | 0 | 14.3 | | **d** | = | **gitara basowa** |
| 0 | 25 | 0 | 0 | 50 | 0 | 0 | 25 | | **e** | = | **altówka** |
| 50 | 0 | 0 | 25 | 0 | 25 | 0 | 0 | | **f** | = | **skrzypce** |
| 0 | 0 | 14.3 | 28.6 | 28.6 | 14.3 | 14.3 | 0 | | **g** | = | **kontrabas** |
| 28.6 | 14.3 | 0 | 0 | 0 | 0 | 0 | 57.1 | | **h** | = | **wiolonczela** |

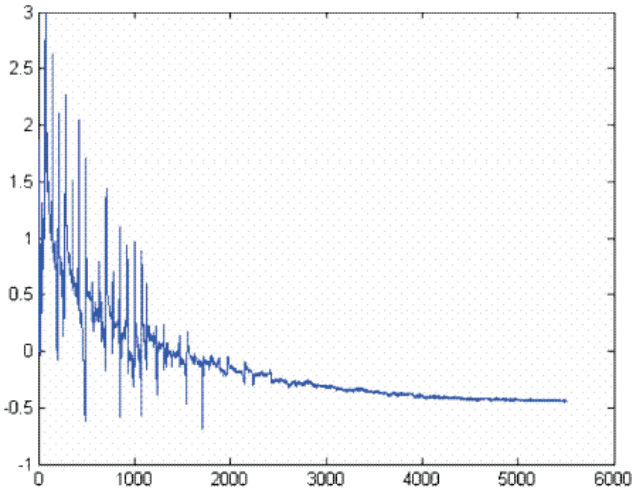
W Tab. 7.3 pokazano efekt testu przeprowadzonego wykorzystując pełną reprezentację widma. Ogólna rozpoznawalność dla ośmiu klas instrumentów wynosiła 75.6%.

Tablica 7.3. Macierz przekłamań — wyniki uzyskane na podstawie analizy pełnej reprezentacji widma badanego fragmentu przebiegu

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 0 | 0 | 0 | 0 | 0 | 25 | 0 | | **a** | = | **harfa** |
| 0 | 40 | 40 | 20 | 0 | 0 | 0 | 0 | | **b** | = | **gitara akustyczna** |
| 0 | 0 | 85.7 | 14.3 | 0 | 0 | 0 | 0 | | **c** | = | **gitara elektryczna** |
| 0 | 33.3 | 0 | 66.7 | 0 | 0 | 0 | 0 | | **d** | = | **gitara basowa** |
| 0 | 0 | 0 | 0 | 66.7 | 0 | 33.3 | 0 | | **e** | = | **altówka** |
| 0 | 0 | 0 | 0 | 16.7 | 83.3 | 0 | 0 | | **f** | = | **skrzypce** |
| 10 | 0 | 0 | 0 | 0 | 0 | 80 | 10 | | **g** | = | **kontrabas** |
| 0 | 0 | 0 | 14.3 | 0 | 0 | 0 | 85.7 | | **h** | = | **wiolonczela** |




Rysunek 7.8. Przykład postaci czasowej wyciętego fragmentu przebiegu przeznaczonego do analizy widmowej. Harfa Cis[1].




Rysunek 7.9. Widmo (przeznaczone do analizy) pobranego fragmentu przebiegu przedstawionego na Rys. 7.8

Ostatecznie zdecydowano się do dalszych badań wykorzystać pełną reprezentację widma, analizując okno o długości 11 025 próbek.

Na Rys. 7.9 zilustrowano postać widmową badanego fragmentu przebiegu. Jak widać, widmo jest przedstawione w skali logarytmicznej, co jest standardem w cyfrowym przetwarzaniu sygnałów. W trakcie prowadzenia badań próbowano dokonywać analizy widma przedstawionego w skali liniowej, ale uzyskane wyniki nie przyniosły zadowalających efektów.

W bogatej literaturze związanej z cyfrowym przetwarzaniem sygnałów można odszukać definicję różnych deskryptorów (zarówno czasowych jak i widmowych). Część z nich znajduje zastosowanie w procesie automatycznej klasyfikacji instrumentów — np. środek ciężkości widma, [15], grupa parametrów tristimulus, [18] lub rozkład energii w prążkach parzystych i nieparzystych, [16]. Podczas prowadzonych badań stwierdzono, że niektóre deskryptory (stosowane przez badaczy do parametryzacji sygnałów cyfrowych) nie znajdują zastosowania podczas procesu klasyfikacji sygnałów pochodzących od chordofonów z artykulacją *pizzicato*. Przykładem takich deskryptorów jest metoda momentów widmowych $k$-tego rzędu, wyrażona zależnością:

$$m_k = \sum_{i=0}^{\infty} A(i) i^k, \tag{7.3}$$

gdzie:
$A(i)$ — amplituda i-tej składowej,
$i$ — częstotliwość i-tego prążka widma,
oraz metoda centralnych momentów widmowych $k-$tego rzędu wyrażana:

$$m_k = \sum_{i=0}^{\infty} A(i)(i - Br)^k, \tag{7.4}$$

gdzie: $Br$ — środek ciężkości widma definiowany jako:

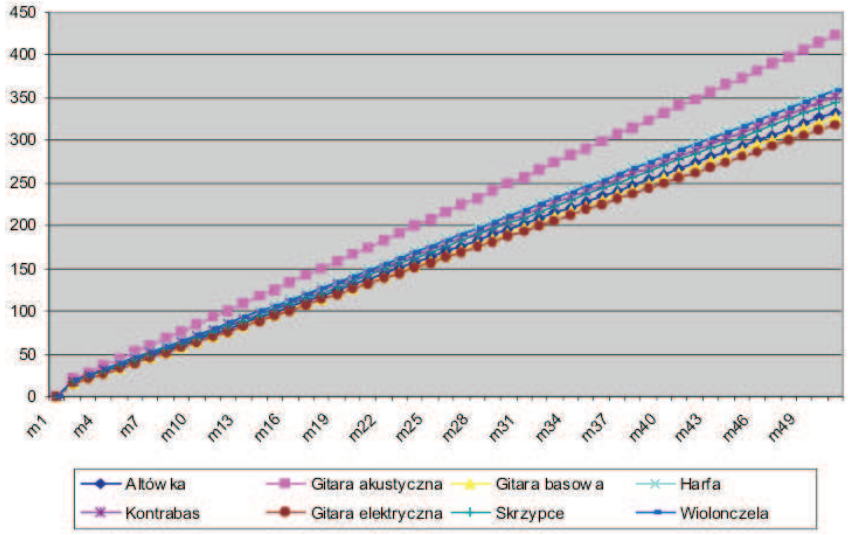
$$Br = \frac{\sum_{i=0}^{n} A(i) i}{\sum_{i=0}^{n} A(i)}. \tag{7.5}$$

Deskryptory te znajdują szerokie uznanie w analizie obrazów, ale nie dostarczają satysfakcjonujących rezultatów podczas klasyfikacji dźwięków *pizzicato*. Na Rys. 7.10 przedstawiono rozkład momentów widmowych dla ośmiu klas instrumentów, dla $m = 51$.

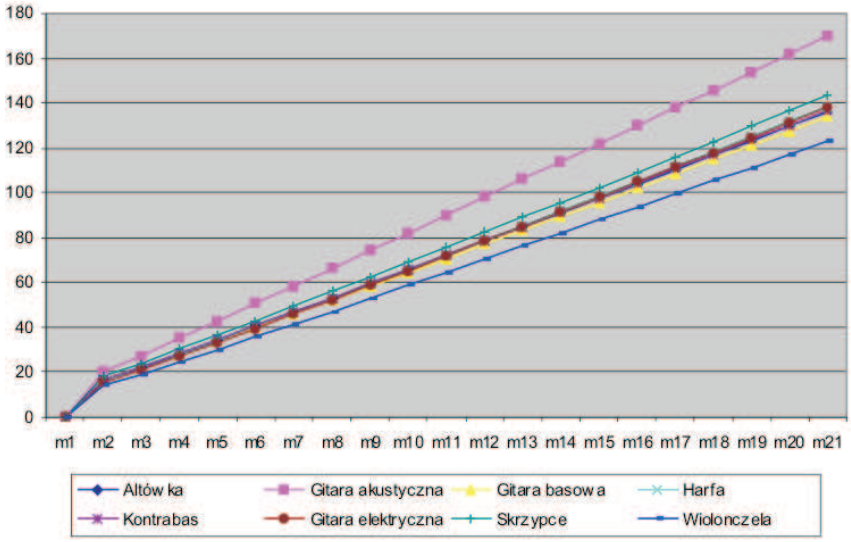
Na Rys. 7.10 pokazano, że istotne różnice w otrzymanych wynikach zanotowano dla gitary akustycznej. Pozostałe rezultaty nie przyczyniają się do podniesienia poziomu separowalności poszczególnych instrumentów. Szczególnie jest to widoczne w kontekście harfy i wiolonczeli oraz gitary basowej i gitary elektrycznej. Podobna sytuacja została zaobserwowana w przypadku centralnych momentów widmowych. Na Rys. 7.11 pokazano uzyskane wyniki dla $m = 21$.

Na Rys. 7.11 widać, że skuteczność opisywanego deskryptora jest mało zadowalająca w odniesieniu do altówki, gitary basowej, harfy, kontrabasu i gitary elektrycznej.

W dalszej części prowadzonych badań zdecydowano się zrezygnować z metody momentów widmowych oraz metody centralnych momentów widmowych.




Rysunek 7.10. Rozkład momentów widmowych dla $m = 51$




Rysunek 7.11. Rozkład centralnych momentów widmowych dla $m = 21$

Ostatecznie podczas realizacji badań zdecydowano się skorzystać z klasycznych deskryptorów (wzory wykorzystanych deskryptorów przedstawiono w rozdziale 4):

1. $Tr$ — grupa parametrów tristimulus,
2. $Ev$ — stosunek energii zawartej w prążkach parzystych,
3. $Od$ — stosunek energii zawartej w prążkach nieparzystych,
4. $Ir$ — nieregularność widma,
5. $ZC$ — zero crossing,
6. $l_{tk}$ — logarytm czasu wybrzmiewania nuty.

Korzystając tylko z w/w deskryptorów uzyskano rozpoznawalność (dla ośmiu klas instrumentów) w granicach 65%. Przykładowy wynik klasyfikacji z użyciem drzew decyzyjnych (metoda holdout 80%:20%) zaprezentowano w poniższej macierzy przekłamań.

Tablica 7.4. Macierz przekłamań dla klasyfikacji z użyciem klasycznych deskryptorów

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | | **a** | = | **harfa** |
| 0 | 40 | 0 | 0 | 0 | 0 | 40 | 20 | | **b** | = | **gitara akustyczna** |
| 0 | 0 | 57.1 | 28.6 | 14.3 | 0 | 0 | 0 | | **c** | = | **gitara elektryczna** |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | **d** | = | **gitara basowa** |
| 0 | 0 | 0 | 0 | 33.3 | 66.7 | 0 | 0 | | **e** | = | **altówka** |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | | **f** | = | **skrzypce** |
| 0 | 0 | 0 | 0 | 10 | 0 | 80 | 10 | | **g** | = | **kontrabas** |
| 0 | 0 | 0 | 0 | 42.9 | 14.3 | 0 | 42.9 | | **h** | = | **wiolonczela** |

Podczas dalszych rozważań zdecydowano, że otrzymane wyniki nie są satysfakcjonujące i zdecydowano się zaproponować alternatywne deskryptory poprawiające ogólną rozpoznawalność dźwięków muzycznych z artykulacją *pizzicato*. Zaproponowaną metodologię opisano szeroko w rozdziale 8 i 9.

ROZDZIAŁ 8

# Zaproponowana metodologia analizy postaci widmowej badanych dźwięków

Nawiązując do wektora cech opisanego w rozdziale 7.3.2, oraz wyników klasyfikacji uzyskanych dzięki jego zastosowaniu, można stwierdzić, że procent rozpoznawalności badanych klas instrumentów jest mało zadowalający. Okazuje się, że wykorzystując opisane wcześniej deskryptory, w powiązaniu z przyjętą metodologią badań, otrzymano wynik klasyfikacji oscylujący w granicach 51%–66.7% (w zależności od metody i zastosowanego algorytmu klasyfikującego). Przykładowe macierze przekłamań dla różnych metod i algorytmów klasyfikujących pokazano w Tab. 8.1–8.3.

Tablica 8.1. Macierz przekłamań dla klasyfikacji ośmiu klas instrumentów. Tablice decyzyjne, metoda holdout 80%:20%. Ogólna rozpoznawalność 51%

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 0 | 25 | 0 | 0 | 25 | 0 | 0 | | **a** | = | **harfa** |
| 0 | 20 | 20 | 40 | 20 | 0 | 0 | 0 | | **b** | = | **gitara akustyczna** |
| 0 | 0 | 85.7 | 14.3 | 0 | 0 | 0 | 0 | | **c** | = | **gitara elektryczna** |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | **d** | = | **gitara basowa** |
| 0 | 0 | 0 | 0 | 33.3 | 66.7 | 0 | 0 | | **e** | = | **altówka** |
| 0 | 0 | 0 | 0 | 16.7 | 83.3 | 0 | 0 | | **f** | = | **skrzypce** |
| 0 | 0 | 0 | 40 | 0 | 20 | 40 | 0 | | **g** | = | **kontrabas** |
| 14.3 | 0 | 14.3 | 14.3 | 0 | 42.8 | 0 | 14.3 | | **h** | = | **wiolonczela** |

Tablica 8.2. Macierz przekłamań dla klasyfikacji ośmiu klas instrumentów. k-NN, metoda $k$-krotnej walidacji krzyżowej dla $k = 10$. Ogólna rozpoznawalność 63.1%

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 18.8 | 0 | 31.3 | 12.5 | 31.3 | 0 | 0 | 6.3 | | **a** | = | **harfa** |
| 0 | 51.7 | 17.2 | 17.2 | 0 | 0 | 13.8 | 0 | | **b** | = | **gitara akustyczna** |
| 9.4 | 6.3 | 65.6 | 12.5 | 0 | 0 | 0 | 6.3 | | **c** | = | **gitara elektryczna** |
| 0 | 17.9 | 10.7 | 64.3 | 0 | 0 | 7.1 | 0 | | **d** | = | **gitara basowa** |
| 10 | 0 | 0 | 0 | 66.7 | 13.3 | 0 | 10 | | **e** | = | **altówka** |
| 3.3 | 0 | 0 | 0 | 10 | 76.7 | 0 | 10 | | **f** | = | **skrzypce** |
| 3.3 | 6.7 | 0 | 10 | 3.3 | 3.3 | 66.7 | 6.7 | | **g** | = | **kontrabas** |
| 0 | 3.3 | 6.7 | 0 | 3.3 | 10 | 3.3 | 73.3 | | **h** | = | **wiolonczela** |

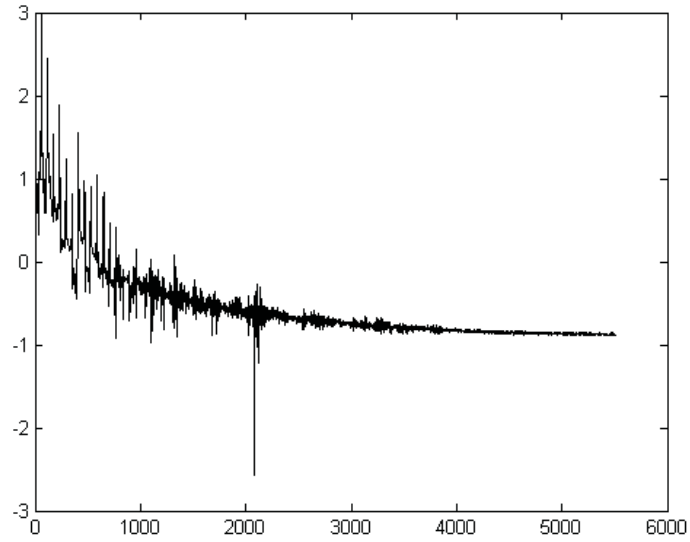
Tablica 8.3. Macierz przekłamań dla klasyfikacji ośmiu klas instrumentów. Drzewa decyzyjne, metoda holdout 80%:20%. Ogólna rozpoznawalność 66.7%

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | | a | = | harfa |
| 0 | 40 | 0 | 0 | 0 | 0 | 40 | 20 | | b | = | gitara akustyczna |
| 0 | 0 | 57.1 | 28.6 | 14.3 | 0 | 0 | 0 | | c | = | gitara elektryczna |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 33.3 | 66.7 | 0 | 0 | | e | = | altówka |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | | f | = | skrzypce |
| 0 | 0 | 0 | 0 | 10 | 0 | 80 | 10 | | g | = | kontrabas |
| 0 | 0 | 0 | 0 | 42.9 | 14.3 | 0 | 42.9 | | h | = | wiolonczela |

W celu polepszenia stopnia rozpoznawalności w zakresie badanych instrumentów zdecydowano się zaproponować nową grupę deskryptorów widmowych.

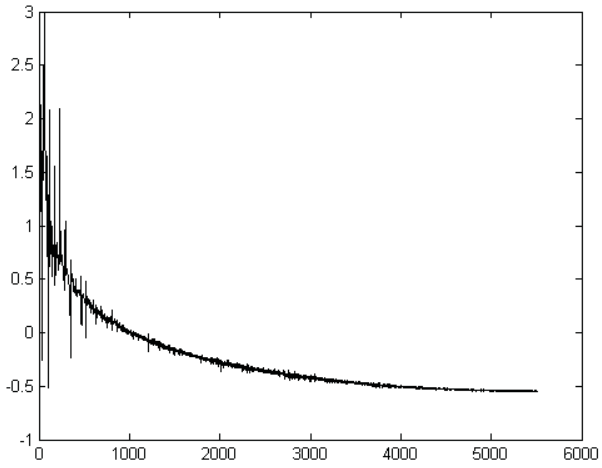
## 8.1. Wybór obszaru widma badanych instrumentów przeznaczony do dalszych badań

Analizując postacie widmowe badanych próbek stwierdzono, że nie jest konieczne skupianie uwagi badawczej na widmie o długości $N = 5\,512$ (jak wiemy widmo jest symetryczne, a zatem analizie poddaje się $\frac{1}{2}$ widma). Przyglądając się postaciom widmowym przedstawionym na Rys. 8.1–8.3 można wywnioskować, że analiza widma powyżej 1 000 próbki sprowadza się do analizy szumu, co nie jest interesujące dla prowadzonych badań.

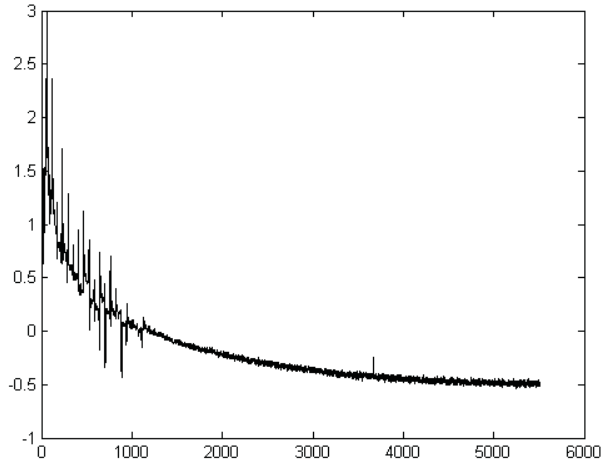



Rysunek 8.1. a) Widmo — harfa ($b^1$)

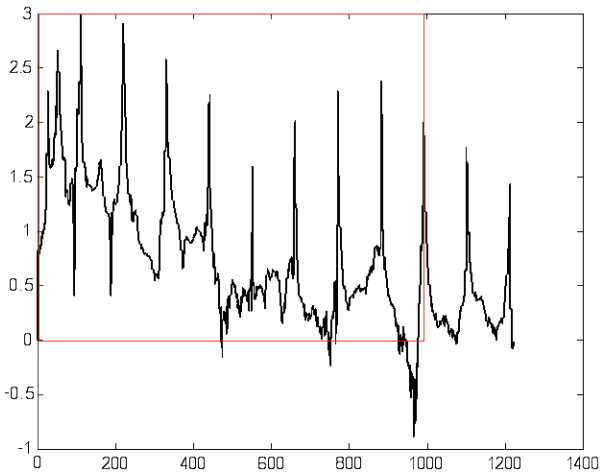
Wykorzystując powyższe spostrzeżenia zdecydowano się poddać analizie fragment widma dla $N = 1\,000$. Oznacza to, że dalszą swoją uwagę skupiono na rozkładzie częstotliwościowym do 4 kHz, a zatem poszukiwania efektywnych deskryptorów odbywały się na fragmencie widma przedstawionym na Rys. 8.4.




Rysunek 8.2. b) Widmo — kontrabas ($b^1$)




Rysunek 8.3. c) Widmo — wiolonczela ($b^1$)




Rysunek 8.4. Fragment widma przeznaczony do dalszej analizy. Zaznaczono obszar zainteresowań

W trakcie badań zdecydowano się skorzystać z powszechnie używanego i uznanego klasyfikatora *WEKA* — wykorzystując jego wewnętrzne mechanizmy związane zarówno z procesem klasyfikacji danych jak i selekcji atrybutów.

## 8.2. Analiza wybranej przestrzeni widma

W celu odszukania istotnych dla procesu klasyfikacji cech widma, zdecydowano się przeprowadzić analizę rozkładu częstotliwościowego oraz energetycznego. W tym celu dokonano podziału widma na $n$ przedziałów częstotliwościowych, w których zliczano zgromadzoną energię. Poza tym podzielono widmo na $m$ przedziałów rozkładu energetycznego. W pierwszym etapie zdecydowano się podzielić widmo na 10 równych przedziałów (kolumn) rozkładu częstotliwościowego, a zatem uzyskano przedziały o szerokości 100 próbek. Przyjęty podział oznacza, że analizowano rozkład częstotliwości ze stałą szerokością 400 Hz. Kolejnym krokiem był najlepszy dobór szerokości przedziału rozkładu energetycznego.

W rozdziale 6 zwrócono uwagę na korzystny wpływ procesu normalizacji cech. W trakcie prowadzonych badań zastosowano normalizację atrybutów do wartości średniej. Okazuje się, że normalizacja poprawia proces automatycznej klasyfikacji o ok. 4–5%. Poniżej przedstawiono macierze przekłamań dla procesu klasyfikacji zbioru znormalizowanego oraz z pominięciem procesu normalizacji. Wykorzystano drzewa decyzyjne oraz metodę $k$-krotnej walidacji krzyżowej dla $k = 15$. Do procesu klasyfikacji wykorzystano wektor klasycznych cech opisany w sekcji 7.3.2.

Tablica 8.4. Wynik klasyfikacji. Atrybuty znormalizowane — ogólna rozpoznawalność 61.4%

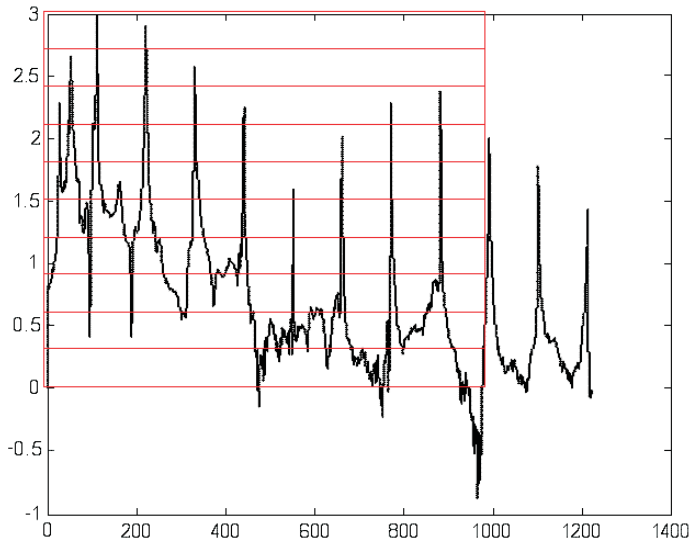
| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 58.6 | 17.2 | 3.4 | 13.8 | 3.4 | 3.4 | 0 | 0 | | **a** | = | **altówka** |
| 10.3 | 79.3 | 0 | 6.9 | 3.4 | 0 | 0 | 0 | | **b** | = | **skrzypce** |
| 0 | 3.6 | 57.1 | 14.3 | 0 | 17.9 | 0 | 7.1 | | **c** | = | **kontrabas** |
| 20.7 | 10.3 | 10.3 | 51.7 | 6.9 | 0 | 0 | 0 | | **d** | = | **wiolonczela** |
| 0 | 0 | 0 | 6.25 | 68.75 | 6.25 | 18.75 | 0 | | **e** | = | **harfa** |
| 0 | 0 | 17.2 | 0 | 3.4 | 51.7 | 6.9 | 20.7 | | **f** | = | **gitara akustyczna** |
| 0 | 0 | 0 | 6.25 | 12.5 | 3.1 | 62.5 | 15.6 | | **g** | = | **gitara elektryczna** |
| 0 | 0 | 3.6 | 3.6 | 0 | 14.3 | 14.3 | 64.3 | | **h** | = | **gitara basowa** |

Tablica 8.5. Wynik klasyfikacji. Atrybuty bez normalizacji — ogólna rozpoznawalność 57%

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 51,7 | 10,3 | 3,4 | 13,8 | 17,2 | 3,4 | 0 | 0 | | **a** | = | **altówka** |
| 17,2 | 65,5 | 0 | 17,2 | 0 | 0 | 0 | 0 | | **b** | = | **skrzypce** |
| 3,6 | 3,6 | 67,9 | 10,7 | 0 | 10,7 | 0 | 7,1 | | **c** | = | **kontrabas** |
| 10,3 | 24,1 | 3,4 | 51,7 | 6,9 | 0 | 3,4 | 0 | | **d** | = | **wiolonczela** |
| 25 | 0 | 6,25 | 0 | 37,5 | 6,25 | 25 | 0 | | **e** | = | **harfa** |
| 0 | 0 | 13,8 | 3,4 | 3,4 | 51,7 | 6,9 | 20,7 | | **f** | = | **gitara akustyczna** |
| 3,1 | 3,1 | 3,1 | 0 | 9,4 | 6,25 | 59,4 | 15,6 | | **g** | = | **gitara elektryczna** |
| 0 | 0 | 3,6 | 0 | 7,1 | 10,7 | 14,3 | 64,3 | | **h** | = | **gitara basowa** |

### 8.2.1. Analiza rozkładu energetycznego badanej przestrzeni widma

Celem odszukania optymalnej fragmentacji widma zdecydowano się przyjąć podział na warstwy o równej szerokości. Analizowano badany fragment widma z uwzględnieniem 10, 20, 30, 40, 50, 60, 70, 80, 90 i 100 warstw.




Rysunek 8.5. Przykładowy podział widma na warstwy rozkładu energetycznego

W trakcie analizy widma przy podziale 10-cio warstwowym stwierdzono, że uzyskane wyniki klasyfikacji nie przynoszą zadowalających efektów. Jako przykład może posłużyć wynik klasyfikacji dwóch klas instrumentów (altówka i skrzypce). Wynik testu informuje, że z uwzględnieniem drzew decyzyjnych oraz metody holdout 80%:20% otrzymano tylko 72.73% rozpoznawalności — do badań wykorzystano 10-cio elementowy wektor cech utworzony tylko na podstawie podziału widma na 10 warstw. Jako porównanie może posłużyć test przeprowadzony dla tych samych instrumentów (również z uwzględnieniem drzew decyzyjnych oraz metody holdout (80%:20%) z wykorzystaniem wektora cech opisanego w sekcji 7.3.2. Okazuje się, że dla klasycznych deskryptorów otrzymano 91.7% rozpoznawalności. Oznacza to, że zaproponowany podział fragmentu widma na 10 warstw energetycznych nie przynosi lepszego rezultatu.

W dalszej części przeprowadzonych eksperymentów badano skuteczność podziału widma ze względu na ilość zaproponowanych warstw. Podczas eksperymentów uwzględniono bazę próbek pochodzących od ośmiu klas instrumentów. Poza tym (na tym etapie badań) zrezygnowano z selekcji cech, co oznacza, że każdy z uwzględnionych przedziałów (warstw) został potraktowany jako pojedynczy deskryptor. Wektor cech zawierał tylko informacje wynikające z warstwowania badanego fragmentu widma.

W Tablicy 8.6 przedstawiono przykładowe wyniki, dla ośmiu klas instrumentów, z uwzględnieniem różnej ilości warstw.

Analizując wyniki zgromadzone w Tab. 8.6 można wywnioskować, że najefektywniejszy podział na przedziały energetyczne widma jest dla 40 warstw. Należy pamiętać, że podczas wyboru optymalnego podziału powinno się uwzględnić nie tylko procent rozpoznawalności klasyfikowanych obiektów, ale również ilość warstw — zbyt duża ilość deskryptorów przyczyni się do zwiększenia złożoności algorytmicznej podczas procesu filtrowania baz danych.

Tablica 8.6. Przykładowe wyniki klasyfikacji (dla ośmiu klas instrumentów) z uwzględnieniem różnych algorytmów klasyfikujących

| Ilość warstw | Drzewa decyzyjne | | Tablice decyzyjne | | k-NN | |
|---|---|---|---|---|---|---|
| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k-$krotnej walidacji krzyżowej | metoda holdout | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
| 10 | 47.25% dla $k = 5$ | 50% dla 80%:20% | 47.25% dla $k = 10$ | 40% dla 66%:34% | 46.3% dla $k = 20$ | 45.45% dla 60%:40% |
| 20 | 50% dla $k = 10$ | 48.03% dla 80%:20% | 47.4% dla $k = 15$ | 47.8% dla 70%:30% | 49.1% dla $k = 10$ | 48.4% dla 75%:25% |
| 30 | 49.1 dla $k = 20$ | 51.6% dla 75%:25% | 51.97 dla $k = 20$ | 51.3% dla 66%:34% | 49.2% dla $k = 10$ | 50.6% dla 80%:20% |
| 40 | 52.4% dla $k = 15$ | 55.7% dla 65%:35% | 53.8% dla $k = 15$ | 51.1% dla 80%:20% | 54.7% dla $k = 20$ | 57.8% dla 80%:20% |
| 60 | 51.4% dla $k = 10$ | 50% dla 80%:20% | 47.1% dla $k = 15$ | 47.4% dla 80%:20% | 49.5% dla $k = 15$ | 48% dla 80%:20% |
| 80 | 53.7% dla $k = 15$ | 53.9% dla 80%:20% | 47.5% dla $k = 5$ | 45.4% dla 80%:20% | 48.55% dla $k = 15$ | 48.2% dla 70%:30% |
| 100 | 53.1% dla $k = 5$ | 51.3% dla 80%:20% | 44.7% dla $k = 15$ | 43.4% dla 80%:20% | 49.2% dla $k = 15$ | 42.1% dla 70%:30% |

W dalszym procesie badawczym zdecydowano się również zaproponować podział wybranego fragmentu widma na warstwy różnej szerokości. Na podstawie obserwacji postaci widmowych badanych próbek, do dalszej analizy zaproponowano podział na 7 warstw o szerokościach:

- 1 warstwa 0–0.09
- 2 warstwa 0.09–0.3
- 3 warstwa 0.3–0.75
- 4 warstwa 0.75–1.2
- 5 warstwa 1.2–1.8
- 6 warstwa 1.8–2.4
- 7 warstwa 2.4–3

Przykładowe wyniki klasyfikacji ośmiu instrumentów z wykorzystaniem w/w szerokości warstw przedstawiono w Tab. 8.7:

Z dotychczasowych rozważań związanych z analizą rozkładu energetycznego w poszczególnych warstwach widma wynika, że dobór szerokości warstw został przyjęty z wykorzystaniem intuicyjnego kryterium. W dalszych badaniach zdecydowano się

Tablica 8.7. Przykładowe wyniki klasyfikacji (dla 8 klas instrumentów) z uwzględnieniem różnych algorytmów klasyfikujących oraz uwzględnieniem różnej szerokości warstw

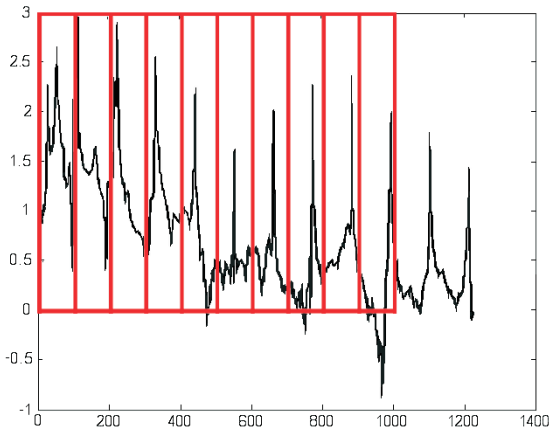
| Ilość warstw | Drzewa decyzyjne | | Tablice decyzyjne | | k-NN | |
|---|---|---|---|---|---|---|
| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
| 7 | 46.7% dla $k = 20$ | 42.2% dla 80%:20% | 45.8% dla $k = 15$ | 47.4% dla 75%:25% | 48% dla $k = 15$ | 57.8% dla 80%:20% |

ustalić optymalną szerokość poszczególnych warstw oraz zaproponować takie kryterium doboru, które znajdzie swoje uzasadnienie. Zaproponowane kryterium doboru szerokości warstw opisano szerzej w rozdziale 9.

### 8.2.2. Analiza rozkładu częstotliwościowego badanej przestrzeni widma

Tak jak wspomniano na początku niniejszego rozdziału, poza analizą rozkładu energetycznego, zdecydowano się dokonać analizy widma z uwzględnieniem poszczególnych przedziałów częstotliwościowych. Kontynuując analizę wybranego fragmentu widma zdecydowano się określić, w jakim stopniu akumulacja energii w poszczególnych przedziałach częstotliwościowych ma wpływ na polepszenie rozpoznawalności badanych klas instrumentów muzycznych. W pierwszym etapie badań rozkładu częstotliwościowego zdecydowano się podzielić dziedzinę na 10 kolumn, co oznacza, że analizowano rozkład częstotliwościowy ze stałą szerokością 400 Hz (100 próbek). Przykładowy podział widma na kolumny zilustrowano na Rys. 8.6.

Okazuje się, że porównując wynik klasyfikacji z wykorzystaniem rozkładu energetycznego (dla $n = 10$, gdzie $n =$ ilość warstw) z wynikiem klasyfikacji uzyskanym na podstawie rozkładu częstotliwościowego (dla $n = 10$, gdzie $n =$ ilość kolumn) można mieć nadzieję, że analiza rozkładu częstotliwościowego dostarczy istotnych informacji dla procesu segregacji badanych instrumentów. Odwołując się do przykładowego




Rysunek 8.6. Przykładowy podział widma na kolumny rozkładu częstotliwościowego

wyniku testu wykorzystującego drzewa decyzyjne oraz metodę holdout (80%:20%) uzyskano rezultat ogólnej rozpoznawalności (dla 8 klas instrumentów) na poziomie 55.4% — a więc o 5.4% więcej niż przy analizie rozkładu energetycznego widma (por. Tab. 8.6). Wykorzystując tę informację zdecydowano się dokonać wnikliwej analizy wpływu podziału badanego fragmentu widma na kolumny w kontekście poprawienia skuteczności automatycznej klasyfikacji instrumentów. Oczekiwano, że wynik przeprowadzonych testów wskaże przybliżoną ilość kolumn, która powinna być rozpatrywana w dalszym procesie analizy. Zdecydowano się podzielić analizowaną przestrzeń widma na 5, 8, 10, 20 oraz 25 kolumn. Kryterium takiego podziału jest konieczność uzyskania liczby całkowitej dla szerokości analizowanej kolumny — co jest zdeterminowane ilością próbek.

Tablica 8.8. Przyjęta szerokość kolumn rozkładu częstotliwościowego widma

| Ilość kolumn | 5 | 8 | 10 | 20 | 25 |
|---|---|---|---|---|---|
| Szerokość kolumny (ilość próbek) | 200 | 125 | 100 | 50 | 40 |

Podobnie jak w przypadku analizy przedziałów energetycznych widma, zdecydowano się pominąć proces z selekcji cech. Zastosowany wektor cech zawierał tylko deskryptory wynikające z podziału na określoną ilość kolumn. Wyniki eksperymentów przedstawiono w Tab. 8.9 — w celu lepszej interpretacji wyników pokazano rezultaty dla tych samych algorytmów klasyfikujących, jakie zastosowano dla analizy rozkładu energetycznego (por. Tab. 8.6).

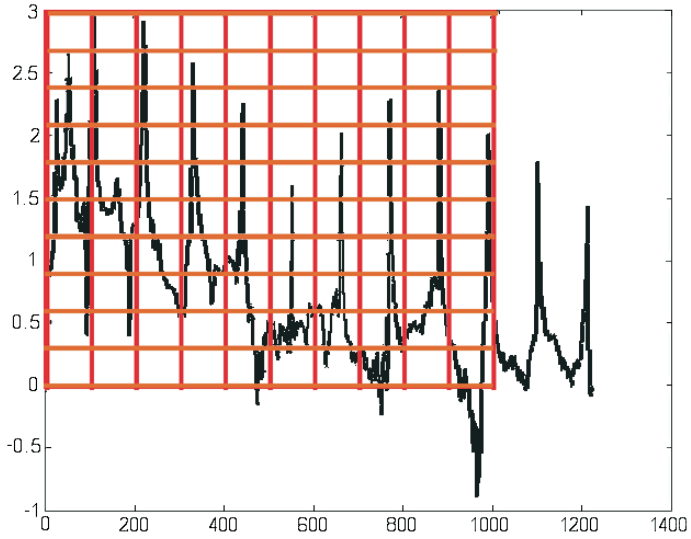
Analizując wyniki zgromadzone w Tab. 8.9 można przyjąć, że wykorzystując do celów klasyfikacji tylko deskryptory wynikające z rozkładu częstotliwościowego należy się skupić na 10-cio kolumnowym podziale widma. Wyniki oscylujące poniżej 50% należy traktować jako nieprzydatne dla dalszych rozważań. Oznacza to, że nie warto się skupiać na 20 i 25 kolumnowym podziale widma. Można zatem wnioskować, że analiza zbyt wąskich przedziałów częstotliwościowych nie przynosi oczekiwanych efektów.

Tablica 8.9. Przykładowe wyniki klasyfikacji (dla 8 klas instrumentów) z uwzględnieniem różnych algorytmów klasyfikujących

| Ilość kolumn | Drzewa decyzyjne | | Tablice decyzyjne | | k-NN | |
|---|---|---|---|---|---|---|
| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
| 5 | 43.1% dla $k = 15$ | 63.2% dla 75%:25% | 37.8% kla $k = 10$ | 35.3% dla 70%:30% | 44% dla $k = 20$ | 48.9% dla 80%:20% |
| 8 | 39.5% dla $k = 10$ | 40.4% dla 75%:25% | 29.8% dla $k = 20$ | 29.8% dla 75%:25% | 49.3% dla $k = 15$ | 48.9% dla 80%:20% |
| 10 | 56.7% dla $k = 10$ | 51.3% dla 70%:30% | 52.4% dla $k = 20$ | 52.6% dla 75%:25% | 54.4% dla $k = 15$ | 48% dla 80%:20% |
| 20 | 44.9% dla $k = 15$ | 42.2% dla 80%:20% | 36% dla $k = 5$ | 35.6% dla 80%:20% | 48% dla $k = 15$ | 45.6% dla 75%:25% |
| 25 | 45.3% dla $k = 20$ | 51.4% dla 70%:30% | 37.3% dla $k = 5$ | 40% dla 80%:20% | 44.9% dla $k = 15$ | 48.9% dla 80%:20% |

### 8.2.3. Analiza rozkładu energii w poszczególnych fragmentach badanej przestrzeni widma — metoda siatki

W dalszej części badań zdecydowano się poszukać korzystnych wyników stosując połączenie opisywanych wcześniej metod. Oznacza to, że zaproponowano metodę analizy zależną od ilości zgromadzonej energii w poszczególnych fragmentach widma. Do realizacji tego założenia zdecydowano się zdefiniować macierz, na postawie metod opisywanych wcześniej (kolumn i warstw), agregującą energię zgromadzoną w badanym fragmencie widma. Na Rys. 8.7 przedstawiono przykładową siatkę utworzoną z $n = 10$ kolumn i $m = 10$ warstw.




Rysunek 8.7. Siatka $10 \times 10$

Celem opisywanej metody jest ekstrakcja istotnych dla procesu automatycznej klasyfikacji pewnych fragmentów analizowanej przestrzeni widma. Wykorzystując wcześniej zaproponowane podziały widma zdecydowano się zdefiniować kilka propozycji siatek oraz zbadać efektywność tej metody. Zdecydowano się jednak zrezygnować z podziału 20 i 25 kolumnowego. Pominięcie tego podziału w procesie tworzenia macierzy zdeterminowane było zbyt długim wynikowym wektorem cech. Na przykład w przypadku 20 kolumn i 40 warstw zostałby utworzony 800 elementowy wektor cech — co zdecydowanie wydłużyłoby proces przeszukiwania multimedialnych baz danych. Przykładowe wyniki badań (dla podziału 10-cio kolumnowego) przedstawiono w Tab. 8.10.

W powyższych badaniach wykorzystano 2 siatki utworzone z 10 kolumn oraz 10 i 20 warstw widma. W rezultacie otrzymano 2 wektory cech (odpowiednio 100 i 200 elementowy), które posłużyły do klasyfikacji ośmiu klas instrumentów (z pominięciem selekcji cech). Analizując wyniki zgromadzone w Tab. 8.10 łatwo dojść do wniosku, że konstrukcja siatki z wykorzystaniem większej ilości warstw widma nie przynosi oczekiwanego rezultatu. Należy również pamiętać, że poza zadowalającym rezultatem rozpoznawalności obiektów należy zwrócić uwagę na złożoność wektora cech. W przypadku konstrukcji siatki $10 \times 20$ otrzymana ilość deskryptorów może okazać się na tyle obciążająca system klasyfikujący, że proces filtrowania baz danych będzie obarczony zbyt dużą złożonością. W kolejnym etapie poszukiwań zdecydowano się wykorzystać siatkę skonstruowaną na bazie 10 kolumn oraz

Tablica 8.10. Przykładowe wyniki klasyfikacji dla metody siatki (dla ośmiu klas instrumentów) z uwzględnieniem różnych algorytmów klasyfikujących

| Ilość kolumn ($k = 10$) | Drzewa decyzyjne | | Tablice decyzyjne | | k-NN | |
|---|---|---|---|---|---|---|
| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
| 10 warstw: siatka $10 \times 10$ | 45.3% dla $k = 10$ | 41.7% dla 65%:35% | 40.4% dla $k = 15$ | 35.3% dla 70%:30% | 41.8% dla $k = 15$ | 37.8% dla 80%:20% |
| 20 warstw: siatka $10 \times 20$ | 48.4% dla $k = 20$ | 37.8% dla 80%:20% | 36% dla $k = 10$ | 31.1% dla 80%:20% | 42.7% dla $k = 15$ | 37.8% dla 80%:20% |

7 warstw widma. Wykorzystano różną szerokość warstw widma, o czym pisano we wcześniejszym fragmencie niniejszego rozdziału. Przykładowe wyniki klasyfikacji przedstawiono w Tab. 8.11.

Tablica 8.11. Przykładowe wyniki klasyfikacji dla metody siatki (dla ośmiu klas instrumentów) z uwzględnieniem różnych algorytmów klasyfikujących

| Ilość kolumn ($k = 10$) | Drzewa decyzyjne | | Tablice decyzyjne | | k-NN | |
|---|---|---|---|---|---|---|
| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
| 7 warstw: siatka $10 \times 7$ | 47.3% dla $k = 20$ | 60% dla 80%:20% | 39.3% dla $k = 5$ | 35.4% dla 65%:35% | 50.5% dla $k = 15$ | 44.6% dla 75%:25% |

Rozpatrując wyniki zgromadzone w Tab. 8.11 można zaobserwować poprawę wyniku klasyfikacji w stosunku do danych zgromadzonych w Tab. 8.10, które oscylowały poniżej granicy 50% poprawnego rozpoznania obiektów. W przypadku wykorzystania 7 warstw widma o różnej szerokości poprawiła się ogólna rozpoznawalność (np. dla drzew decyzyjnych) o ok. 10%. Wynika z tego wniosek, że w dalszej części prowadzonych badań należy się skupić na konstruowaniu siatki z wykorzystaniem mniejszej ilości warstw. Poza tym należy zrezygnować z podziału badanego fragmentu widma na warstwy o równej szerokości dla $n > 40$. Powstaje zatem pytanie związane z optymalnym doborem szerokości warstw. Rozwiązanie tego problemu zostanie zaproponowane w rozdziale 9.

## 8.3. Wykorzystanie zaproponowanych deskryptorów w połączeniu z klasycznymi atrybutami

Na podstawie przedstawionych koncepcji podziału i analizy widma uzyskano rozpoznawalność oscylującą w granicach 60% — podobnie jak w przypadku zastosowania deskryptorów opisywanych w rozdziale 7. Zdecydowano się w trakcie dalszych rozważań dokonać próby wyodrębnienia najistotniejszych (w kontekście klasyfikacji) przestrzeni badanego fragmentu widma. Poza tym zaproponowano włączenie do

otrzymanego wektora cech, deskryptorów omawianych w rozdziale 7. Jest sprawą oczywistą, że pominięcie cech mało istotnych dla klasyfikacji przestrzeni widma, pozwoli na skrócenie wektora cech, oraz zmniejszenie złożoności algorytmicznej procesu filtrowania baz danych. Do celów selekcji atrybutów wykorzystano zaimplementowane w klasyfikatorze *WEKA* algorytmy genetyczne.

### 8.3.1. Zaproponowanie zbiorów cech z uwzględnieniem selekcji

Tak jak już wspomniano już w niniejszej rozprawie, nadmierna ilość deskryptorów może wpływać na pogorszenie wyniku klasyfikacji. W związku z tym zdecydowano się wykluczyć takie atrybuty, które nie przyczyniają się do efektywności segregacji lub ją pogarszają. W pierwszym etapie badań zdecydowano się zaproponować wektor cech, który uwzględnia:

1. podział fragmentu widma na 40 równych warstw,
2. podział fragmentu widma na 7 warstw o różnej szerokości,
3. podział fragmentu widma na 10 kolumn,
4. podział fragmentu widma z wykorzystaniem siatki $10 \times 7$,
5. wektor cech opisany w rozdziale 7 (8 atrybutów).

Na podstawie powyższego zestawienia otrzymano wektor cech składający się ze 135 atrybutów. Poniżej przedstawiono wynik klasyfikacji z uwzględnieniem pełnej reprezentacji opisywanego wektora cech.

Tablica 8.12. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 135 elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 75.5% dla $k = 5$ | 76.6% dla 65%:35% |
| Tablice decyzyjne | 64.1% dla $k = 10$ | 57.6% dla 70%:30% |
| k-NN | 72.3% dla $k = 5$ | 72.3% dla 80%:20% |
| Random Forest | 75.9% dla $k = 20$ | 80.5% dla 65%:35% |

Z powyższego zestawienia widać, że analiza poszczególnych elementów widma wzbogacona 8 deskryptorami opisanymi w poprzednim rozdziale przynosi lepsze efekty procesu klasyfikacji. W dalszych rozważaniach zdecydowano się zaproponować krótsze wektory cech, co miało mocniej zaakcentować kluczowe fragmenty badanej przestrzeni widma. W kolejnych zestawieniach przedstawiono wynik klasyfikacji z uwzględnieniem selekcji cech. W Tab. 8.13. przedstawiono wyniki klasyfikacji z uwzględnieniem 36 elementowego wektora cech składającego się z:

1. grupy deskryptorów opisanych w rozdziale 7: *Tr3*, *Ev*,
2. rozkładu częstotliwościowego: kolumny 2, 5, 8,
3. rozkładu energetycznego (7 warstwy różnej szerokości): warstwa 1 i 2,

4. rozkładu energetycznego (40 warstw równej szerokości): warstwa 3, 11, 13, 31, 32, 33, 35, 39,

5. analizy rozkładu energii z wykorzystaniem metody siatki. Siatka utworzona na bazie 7 warstw i 10 kolumn. Opis współrzędnych wg indeksowania [warstwa, kolumna]: [3,2], [3,3], [3,5], [3,7], [4,1], [4,6], [4,7], [4,8], [5,1], [5,4], [5,5], [6,1], [6,4], [6,6], [6,7], [6,8], [7,1], [7,2], [7,4], [7,6], [7,7].

Tablica 8.13. Wyniki klasyfikacji ośmiu klas instrumentów z wykorzystaniem 36-cio elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 68.2% dla $k = 10$ | 72.7 dla 70%:30% |
| Tablice decyzyjne | 62.3% dla $k = 5$ | 60.2 dla 60%:40% |
| k-NN | 68.2% dla $k = 15$ | 70.5% dla 80%:20% |
| Random Forest | 73.6% dla $k = 10$ | 75.8% dla 70%:30% |

Porównując wyniki przedstawione w Tab. 8.12 i Tab. 8.13 można zauważyć, że ogólna rozpoznawalność uległa pogorszeniu — np. w przypadku metody $k$-krotnej walidacji krzyżowej. Należy jednak pamiętać, że pogorszenie wyniku klasyfikacji zostało zdeterminowane redukcją aż 99 deskryptorów. Łatwo się domyśleć, że redukcja tak dużej ilości deskryptorów doprowadzi do przyśpieszenia filtrowania baz danych, a więc pogorszenie ogólnej rozpoznawalności o średnio ok. 3% wydaje się być opłacalne. Ponadto należy również zwrócić uwagę na fakt, że pomimo redukcji cech odnotowano wzrost rozpoznawalności — w przypadku tablic decyzyjnych, metoda holdout.

Kolejny test został przeprowadzony z wykorzystaniem 15-to elementowego wektora cech składającego się z:

1. rozkład częstotliwościowy: kolumny 5, 8,

2. rozkład energetyczny (40 warstw równej szerokości): warstwa 3, 13, 35, 39,

3. analiza rozkładu energii z wykorzystaniem metody siatki. Siatka utworzona na bazie 7 warstw i 10 kolumn. Opis współrzędnych wg indeksowania [warstwa, kolumna]: [3,3], [3,5], [4,1], [4,6], [4,7], [4,8], [6,8], [7,1], [7,2].

Wyniki przedstawiono w Tab. 8.14.

Powyższe zestawienie wyników klasyfikacji potwierdza korzystny wpływ selekcji cech. Warto zwrócić uwagę na fakt, że opisywany powyżej 15-to elementowy wektor cech nie zawiera „klasycznych" deskryptorów opisywanych w rozdziale 7. Oznacza to, że zaproponowane w niniejszej rozprawie deskryptory (oraz metodologia ich pozyskania) przynosi obiecujące rezultaty. Poza tym warto zaakcentować fakt, że podział 7-mio warstwowy (dla różnych szerokości warstw) znalazł zastosowanie tylko dla tworzenia siatki, natomiast jako oddzielna grupa deskryptorów (z uwzględnieniem analizy rozkładu energetycznego) wykorzystano 40-to warstwowy podział badanego fragmentu widma. W kolejnym etapie poszukiwań wykorzystano 12-to elementowy wektor cech, składający się z:

Tablica 8.14. Wyniki klasyfikacji ośmiu klas instrumentów z wykorzystaniem 15-to elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 69.5% dla $k = 15$ | 69.3 dla 64%:36% |
| Tablice decyzyjne | 63.6% dla $k = 20$ | 60.2% dla 60%:40% |
| k-NN | 64.1% dla $k = 15$ | 62.5% dla 60%:40% |
| Random Forest | 75% dla $k = 20$ | 81.3% dla 66%:34% |

1. rozkład częstotliwościowy: kolumny 5, 8,
2. rozkład energetyczny (40 warstw równej szerokości): warstwa 3, 13, 35, 39,
3. analiza rozkładu energii z wykorzystaniem metody siatki. Siatka utworzona na bazie 7 warstw i 10 kolumn. Opis współrzędnych wg indeksowania [warstwa, kolumna]: [3,3], [3,5], [4,1], [4,6], [4,8], [7,1].

Wyniki przedstawiono w Tab. 8.15.

Tablica 8.15. Wyniki klasyfikacji ośmiu klas instrumentów z wykorzystaniem 12-to elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.5% dla $k = 20$ | 69.3% dla 66%:34% |
| Tablice decyzyjne | 64.1% dla $k = 5$ | 60.2% dla 60%:40% |
| k-NN | 66.8% dla $k = 15$ | 66.2% dla 65%:35% |
| Random Forest | 75.5% dla $k = 20$ | 72.7% dla 75%:25% |

Analizując powyższe wyniki oraz dobór deskryptorów łatwo dojść do wniosku, że istotne w kontekście automatycznej klasyfikacji instrumentów muzycznych cechy dla rozkładu częstotliwościowego są zlokalizowane w zakresie wyższych częstotliwości. Należy również zwrócić uwagę, że w porównaniu z 15-to elementowym wektorem cech zostały wykluczone tylko niektóre elementy siatki (konkretnie [4,7], [6,8], [7,2]). Deskryptory związane z rozkładem energetycznym nie zostały na tym etapie selekcji zredukowane, co oznacza, że ich dobór jest uzasadniony. Poniżej przedstawiono kolejne wyniki klasyfikacji oraz dobór wektora cech. Zaproponowano 11-to elementowy wektor cech, zredukowany o 13-tą warstwę dla podziału 40-to warstwowego. Zestawienie wyników ilustruje Tab. 8.16.

Podsumowując zaprezentowane powyżej wyniki badań można stwierdzić, że zaproponowane deskryptory są obiecujące w kontekście automatycznej klasyfikacji. Okazuje się, że istotnym obszarem widma są częstotliwości zawierające się w granicach 1.6 kHz–2 kHz oraz 2.8 kHz–3.2 kHz (5-ta i 8-ma kolumna rozkładu częstotliwo-

Tablica 8.16. Wyniki klasyfikacji ośmiu klas instrumentów z wykorzystaniem 11-to elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 68.6% dla $k = 15$ | 64.8% dla 60%:40% |
| Tablice decyzyjne | 65.5% dla $k = 5$ | 60.6% dla 70%:30% |
| k-NN | 68.6% dla $k = 15$ | 67.5% dla 65%:35% |
| Random Forest | 77.3% dla $k = 10$ | 78.2% dla 75%:25% |

ściowego). Poza tym dla rozkładu energetycznego istotne cechy są skumulowane w 3, 35 i 39 warstwie widma (dla podziału 40-to warstwowego), a więc w obszarach:

1. $0.15 \rightarrow 0.225$ (3 warstwa badanego fragmentu widma),
2. $2.55 \rightarrow 2.625$ (35 warstwa badanego fragmentu widma),
3. $2.85 \rightarrow 2.925$ (39 warstwa badanego fragmentu widma).

Wykorzystując metodę siatki stwierdzono, że korzystne efekty dostarcza siatka utworzona na bazie podziału 10-cio kolumnowego rozkładu częstotliwościowego oraz 7-mio warstwowego podziału energetycznego dla różnej szerokości warstw. Najistotniejsze cechy badanego fragmentu widma są zlokalizowane we współrzędnych siatki [3,3], [3,5], [4,1], [4,6], [4,8], [7,1] — opis współrzędnych wg indeksowania [warstwa, kolumna]. Na Rys. 8.8 pokazano istotne fragmenty widma dla procesu automatycznej klasyfikacji instrumentów muzycznych.

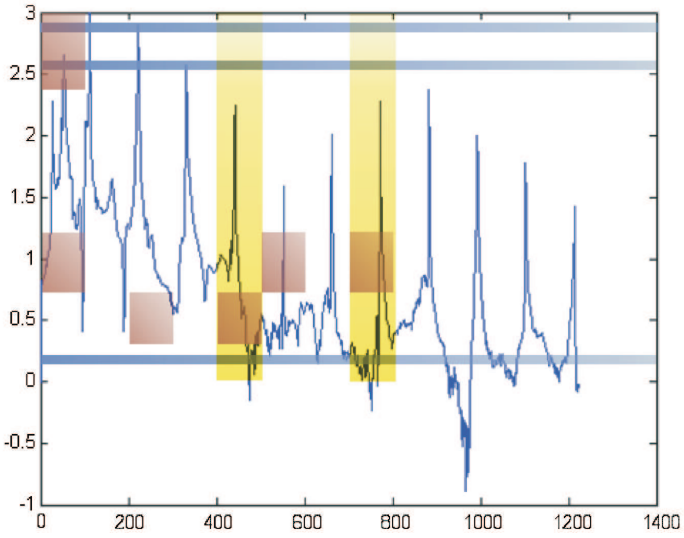
Z zestawienia wyników w Tab. 8.16 wynika, że najlepszą skuteczność uzyskano dla Random Forest, metoda holdout 75%:25%. W Tab. 8.17 przedstawiono macierz przekłamań dla tego przypadku.

Tablica 8.17. Wynik klasyfikacji. Random Forest, metoda holdout 75%:25%– ogólna rozpoznawalność 78.2%

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 0 | 25 | 0 | 0 | 25 | 0 | 25 | | **a** | = | **harfa** |
| 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | | **b** | = | **gitara akustyczna** |
| 0 | 12.5 | 62.5 | 0 | 0 | 25 | 0 | 0 | | **c** | = | **gitara elektryczna** |
| 0 | 0 | 0 | 80 | 0 | 0 | 0 | 20 | | **d** | = | **gitara basowa** |
| 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | | **e** | = | **altówka** |
| 0 | 0 | 0 | 0 | 0 | 55.6 | 0 | 44.4 | | **f** | = | **skrzypce** |
| 10 | 0 | 0 | 0 | 0 | 0 | 90 | 0 | | **g** | = | **kontrabas** |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | | **h** | = | **wiolonczela** |

Z macierzy przekłamań wynika, że zadowalająca jest rozpoznawalność dla 5-ciu klas instrumentów (gitara akustyczna, gitara basowa, altówka, kontrabas, wiolonczela). Wynik klasyfikacji dla pozostałych instrumentów nie jest zadowalający. Szczególnie problem błędnego rozpoznania jest widoczny dla harfy, która została w tylko 1/4 rozpoznana prawidłowo. Pozostałe próbki zostały zakwalifikowane do obcych

klas instrumentów: gitary elektrycznej, skrzypiec, wiolonczeli — w każdym przypadku po 25%. Można zatem wyciągnąć wniosek, że zaproponowane deskryptory są skuteczniejsze w kontekście ogólnej rozpoznawalności 8-miu klas instrumentów, natomiast w niektórych przypadkach można zanotować niezadowalającą separację pomiędzy poszczególnymi instrumentami. Wynika z tego, że należy poprawić skuteczność zaproponowanych deskryptorów skupiając się głównie na poprawieniu separowanlości badanych instrumentów. Problem ten został podjęty w następnym rozdziale.




Rysunek 8.8. Najistotniejsze fragmenty widma dla procesu automatycznej klasyfikacji instrumentów muzycznych

ROZDZIAŁ 9

# Poprawa skuteczności zaproponowanych metod dla analizy przebiegów wybranych instrumentów muzycznych

W rozdziale 8 przedstawiono wyniki uzyskane w powiązaniu z przyjętą metodologią badań. Analizując macierz przekłamań (por. Tab. 8.17, rozdział 8) dla najlepszego rezultatu automatycznej klasyfikacji można stwierdzić, że niepokojąco słaba rozpoznawalność została zarejestrowana dla niektórych instrumentów (np. dla harfy lub skrzypiec). W trakcie realizacji niniejszych badań założono, że uwaga badawcza zostanie skierowana nie tylko na uzyskanie możliwie wysokiego ogólnego stopnia rozpoznawalności instrumentów muzycznych, ale też zadowalającą separowalność rozpoznania wśród badanych klas instrumentów. W trakcie dalszych poszukiwań zdecydowano się skupić na poprawieniu skuteczności zaproponowanych deskryptorów związanych z analizą rozkładu częstotliwościowego oraz energetycznego. Głównym celem postawionym na tym etapie badań było zwiększenie stopnia rozpoznawalności poszczególnych instrumentów, zachowując (lub poprawiając) ogólny procent skuteczności automatycznej klasyfikacji badanych ośmiu klas instrumentów.

## 9.1. Zaproponowana metodologia doboru szerokości warstw

Na podstawie wyników badań zaprezentowanych w rozdziale 8 łatwo dojść do wniosku, że korzystniejszy w kontekście automatycznej klasyfikacji jest podział na 40 warstw równej szerokości (por. Tab. 9.1). Wynika z tego, że nie należy rezygnować z deskryptorów uzyskanych na drodze podziału 40-warstwowego. Poza tym ustalono, że podział fragmentu widma na równe warstwy nie musi być podziałem optymalnym. Analizując wyniki badań przedstawione w Tab. 9.2 można wnioskować, że przyjęcie różnych szerokości dla warstw fragmentu widma może dostarczyć sporo interesujących informacji — szczególnie dla k-NN. W poprzednim rozdziale zdecydowano się dobrać szerokości warstw na podstawie obserwacji postaci widmowych przebiegu, co za tym idzie dobór ten odbywał się na drodze intuicyjnej, co nie jest optymalnym kryterium. W dalszym etapie prowadzonych badań zdecydowano się zoptymalizować kryterium doboru szerokości warstw.

Zdecydowano się przyjąć założenie, że kluczowym kryterium doboru szerokości warstw jest równa ilość zgromadzonej energii w poszczególnych warstwach. Oznacza to, że szerokość warstw została tak dobrana, aby w poszczególnej warstwie fragmentu widma została zgromadzona $1/n$ część energii, gdzie $n$ jest całkowitą ilością

zgromadzonej energii. Na podstawie wyników opisanych w rozdziale 8 zdecydowano się przyjąć 4- i 7- warstwowy podział widma (warstwy różnej szerokości). Poza tym zdecydowano się wykorzystać 10- i 8-kolumnowy podział dla rozkładu częstotliwościowego. Podobnie jak w 8.2.3 zdecydowano się podjąć analizę rozkładu energii w poszczególnych fragmentach badanej przestrzeni widma.

## 9.2. Otrzymane wyniki automatycznej klasyfikacji z uwzględnieniem doboru szerokości warstw

Wykorzystując informacje zawarte w rozdziale 8 zdecydowano się zaproponować kilka wektorów cech, które zostały utworzone na drodze podziału energetycznego i częstotliwościowego badanego fragmentu widma. Poza tym włączono do zaproponowanych wektorów grupę cech opisywanych w rozdziale 7 (por. sekcja 7.3.2). Uwzględniono 10- i 8-kolumnowy oraz 4- i 7-warstwowy podział fragmentu widma. Wykorzystując opisywana metodologie doboru szerokości warstw przyjęto następujące progi:

I 4-warstwowy podział fragmentu widma:

1 warstwa: 0–0.14,
2 warstwa: 0.14–0.33,
3 warstwa: 0.33–0.66,
4 warstwa: 0.66–3.

II 7-warstwowy podział fragmentu widma:

1 warstwa: 0–0.07,
2 warstwa: 0.07–0.16,
3 warstwa: 0.16–0.27,
4 warstwa: 0.27–0.40,
5 warstwa: 0.40–0.60,
6 warstwa: 0.60–0.93,
7 warstwa: 0.93–3.

Na bazie opisywanego podziału zdecydowano się zaproponować następujące wektory cech:

### 9.2.1. Podział 10 kolumnowy z uwzględnieniem 4 warstw podziału energetycznego

**Wektor cech — 102 atrybuty.** W skład tego wektora cech zaliczono deskryptory:

1. cechy opisane w rozdziale 7, sekcja 7.3.2 (8 atrybutów);
2. 10-cio kolumnowy podział częstotliwościowy (10 atrybutów);
3. 4 warstwy podziału energetycznego — różna szerokość warstw (4 atrybuty);

4. 40 warstw podziału energetycznego — równa szerokość warstw (40 atrybutów);
5. siatka $4 \times 10$ (40 atrybutów).

Łącznie wyselekcjonowano 102 atrybuty. Poniżej przedstawiono wyniki testów z uwzględnieniem przykładowych algorytmów klasyfikacyjnych oraz uwzględnieniem procesu selekcji cech.

Tablica 9.1. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 105-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.2 % dla $k$=10 | 80% dla 80%:20% |
| Tablice decyzyjne | 52.9 % dla $k$=10 | 62.2 % dla 80%:20% |
| k-NN | 65.8 % dla $k$=15 | 77.8 % dla 80%:20% |
| Random Forest | 72% dla $k$=15 | 77,2% dla 75%:25% |

**Wektor cech — 28 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr3, Ev, $l_{tk}$ (3 atrybuty);
2. 10-kolumnowy podział częstotliwościowy — kol8 (1 atrybut);
3. 4 warstwy podziału energetycznego: w1, w3 (2 atrybuty);
4. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 2, 3, 7, 11, 16, 21, 23, 33, 34, 38, 40 (11 atrybutów);
5. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,1], [1,3], [1,9], [1,10], [2,5], [2,7], [2,8], [2,9], [4,3], [4,7], [4,8] (11 atrybutów).

Tablica 9.2. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 28-elementowego wektora cech

| | metoda $k$-krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 65.3 % dla $k$=5 | 67.6 % dla 70%:30% |
| Tablice decyzyjne | 57.3 % dla $k$=10 | 68.9 % dla 80%:20% |
| k-NN | 61.3 % dla $k$ =10 | 57.8 % dla 80%:20% |
| Random Forest | 73.3 % dla $k$=10 | 82.2 % dla 80%:20% |

**Wektor cech — 19 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr3, Ev, $l_{tk}$ (3 atrybuty);
2. 4 warstwy podziału energetycznego: w1, w3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 2, 11, 16, 21, 23, 33, 34, 38, 40 (9 atrybutów);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,3], [1,9], [2,9], [4,3], [4,7] (5 atrybutów).

Tablica 9.3. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 19-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 69.3 % dla $k$=15 | 77.8 % dla 80%:20% |
| Tablice decyzyjne | 60% dla $k$ =5 | 70.2 % dla 75%:25% |
| k-NN | 55.6 % dla $k$ =20 | 53.3 % dla 80%:20% |
| Random Forest | 72.4 % dla $k$=5 | 73.3 % dla 80%:20% |

**Wektor cech — 16 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr3, $l_{tk}$ (2 atrybuty);
2. 4 warstwy podziału energetycznego: w1, w3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 2, 11, 21, 23, 33, 34, 38, 40 (8 atrybutów);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,3], [1,9], [2,9], [4,3] (4 atrybuty).

Tablica 9.4. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 16-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 67.6 % dla $k$=10 | 71.1 % dla 80%:20% |
| Tablice decyzyjne | 59.6 % dla $k$=20 | 53.3 % dla 80%:20% |
| k-NN | 55.6 % dla $k$=5 | 53.3 % dla 80%:20% |
| Random Forest | 71.6 % dla $k$=15 | 75.6 % dla 80%:20% |

**Wektor cech — 15 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr3, $l_{tk}$ (2 atrybuty);
2. 4 warstwy podziału energetycznego: w3 (1 atrybut);
3. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 2, 11, 21, 23, 33, 34, 38, 40 (8 atrybutów);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,3], [1,9], [2,9], [4,3] (4 atrybuty).

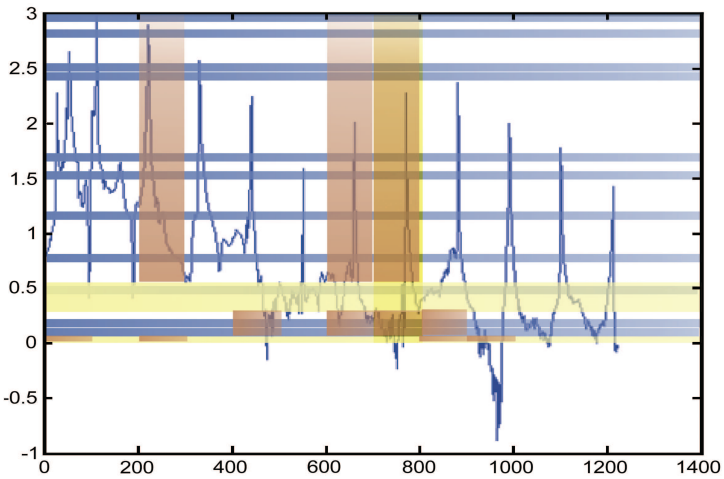
Tablica 9.5. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 15-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 68% dla $k$=20 | 71.1 % dla 80%:20% |
| Tablice decyzyjne | 59.6 % dla $k$ =20 | 53.3 % dla 80%:20% |
| k-NN | 56.4 % dla $k$=10 | 57.9 % dla 70%:30% |
| Random Forest | 74.7 % dla $k$ =15 | 71.1 % dla 80%:20% |

Analizując wyniki zgromadzone w Tab. 9.1–9.5 można stwierdzić, że dokonując podziału badanego fragmentu widma na 10 kolumn i 4 warstwy najlepszy wynik uzyskano z wykorzystaniem 28 elementowego wektora cech przy wykorzystaniu Random Forest oraz metody holdout 80%:20%. Wykorzystując wspomniany wektor cech uzyskano 82.2% ogólnej rozpoznawalności dla 8 klas instrumentów. Macierz przekłamań oraz graficzną reprezentacje najistotniejszych fragmentów widma przedstawiono poniżej.

Tablica 9.6. Macierz przekłamań dla klasyfikacji 8 klas instrumentów. Random Forest, metoda holdout 80%:20%. Ogólna rozpoznawalność 82.2%.

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | | a | = | harfa |
| 0 | 80 | 20 | 0 | 0 | 0 | 0 | 0 | | b | = | gitara akustyczna |
| 14.3 | 0 | 71.4 | 14.3 | 0 | 0 | 0 | 0 | | c | = | gitara elektryczna |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | | e | = | altówka |
| 0 | 0 | 0 | 0 | 0 | 83.3 | 0 | 16.7 | | f | = | skrzypce |
| 0 | 0 | 0 | 0 | 0 | 10 | 80 | 10 | | g | = | kontrabas |
| 0 | 0 | 0 | 0 | 14.3 | 0 | 0 | 85.7 | | h | = | wiolonczela |




Rysunek 9.1. Najistotniejsze fragmenty widma dla procesu automatycznej klasyfikacji instrumentów muzycznych dla 28 elementowego wektora cech

### 9.2.2. Podział 10 kolumnowy z uwzględnieniem 7 warstw podziału energetycznego

**Wektor cech — 135 atrybutów.** W skład tego wektora cech zaliczono deskryptory:

1. cechy opisane w rozdziale 7, sekcja 7.3.2 (8 atrybutów);
2. 10-kolumnowy podział częstotliwościowy (10 atrybutów);
3. 7 warstw podziału energetycznego — różna szerokość warstw (7 atrybutów);
4. 40 warstw podziału energetycznego — równa szerokość warstw (40 atrybutów);
5. siatka 7x10 (70 atrybutów).

Łącznie wyselekcjonowano 135 atrybuty. Poniżej przedstawiono wyniki testów z uwzględnieniem przykładowych algorytmów klasyfikacyjnych oraz uwzględnieniem procesu selekcji cech.

Tablica 9.7. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 135-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.7 % dla $k$=10 | 75.6 % dla 80%:20% |
| Tablice decyzyjne | 55.6 % dla $k$=20 | 64.4 % dla 80%:20% |
| k-NN | 64.4 % dla $k$=15 | 77.8 % dla 80%:20% |
| Random Forest | 72.4 % dla $k$=20 | 72 % dla 75%:25% |

**Wektor cech — 50 atrybutów.** W skład wektora zaliczono cechy:

1. Grupa deskryptorów opisanych w rozdziale 7: Tr1, Ev, Od (3 atrybuty);
2. 10-kolumnowy podział częstotliwościowy — kol2, kol4, kol6, kol7 (4 atrybuty);
3. 7 warstw podziału energetycznego warstwy: 3, 6 (2 atrybuty);
4. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 6, 8, 10, 12, 13, 18, 21, 22, 24, 27, 29, 30, 32, 33, 37, 38 (16 atrybutów);
5. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,2], [1,4], [1,8], [1,9], [1,10], [2,1], [2,3], [2,5], [2,6], [2,7], [2,8], [3,7], [3,8], [4,1], [4,2], [4,3], [4,10], [5,4], [5,6], [5,9], [6,4], [6,8], [7,1], [7,2], [7,6] (25 atrybutów).

Tablica 9.8. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 50-elementowego wektora

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 68.4 % dla $k$=20 | 67.1 % dla 65%:35% |
| Tablice decyzyjne | 52.9 % dla $k$=20 | 52.6 % dla 75%:25% |
| k-NN | 58.7 % dla $k$=15 | 60% dla 80%:20% |
| Random Forest | 70.7 % dla $k$=5 | 70.2 % dla 75%:25% |

**Wektor cech — 22 atrybuty.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr1 (1 atrybut);
2. 10-cio kolumnowy podział częstotliwościowy — kol4 (1 atrybut);
3. 7 warstwy podziału energetycznego warstwy: 3 (1 atrybut);
4. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 8, 13, 18, 27, 30, 32, 33, 37, 38 (9 atrybutów);
5. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,9], [2,1], [2,6], [3,7], [3,8], [4,1], [4,2], [4,3], [6,4], [6,8] (10 atrybutów);

Tablica 9.9. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 22-elementowego wektora

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 67.6 % dla $k$=10 | 67.6 % dla 70%:30% |
| Tablice decyzyjne | 53.8 % dla $k$=20 | 57.9 % dla 75%:25% |
| k-NN | 54.7 % dla $k$=10 | 52.2 % dla 60%:40% |
| Random Forest | 72.9 % dla $k$=15 | 68.9 % dla 80%:20% |

**Wektor cech — 15 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr1 (1 atrybut);
2. 10-kolumnowy podział częstotliwościowy — kol4 (1 atrybut);
3. 7 warstwy podziału energetycznego — warstwy: 3 (1 atrybut);
4. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 8, 18, 27, 30, 33, 37, 38 (7 atrybutów);
5. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 10 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [3,7], [3,8], [4,1], [4,3], [6,8] (5 atrybutów).

Tablica 9.10. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 15-elementowego wektora

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 68% dla $k$=10 | 61.4 % dla 75%:25% |
| Tablice decyzyjne | 54.2 % dla $k$=5 | 57.4 % dla 70%:30% |
| k-NN | 57.3 % dla $k$=10 | 58.8 % dla 70%:30% |
| Random Forest | 68.4 % dla $k$=15 | 70.2 % dla 75%:25% |

Wyniki zgromadzone w Tab. 9.7–9.10 wskazują, że najkorzystniejszy rezultat uzyskano wykorzystując 135-elementowy wektor cech z uwzględnieniem k-NN oraz metody holdout 80%:20%. Wykorzystując wspomniany wektor cech uzyskano 77.8% ogólnej rozpoznawalności dla 8 klas instrumentów. Macierz przekłamań dla wyżej wymienionego testu przedstawiono poniżej.

Wyniki uzyskane z rozpatrywaniem podziału 10-kolumnowego, z uwzględnieniem 7 warstw podziału energetycznego wydają się być mało zadowalające. Wniosek ten można poprzeć dwoma argumentami:

Tablica 9.11. Macierz przekłamań dla klasyfikacji 8 klas instrumentów. k-NN, metoda holdout 80%:20%. Ogólna rozpoznawalność 77.8%.

| a | b | c | d | e | f | g | h | | | ← | classified |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 0 | 0 | 0 | 25 | 25 | 0 | 0 | | a | = | harfa |
| 0 | 60 | 20 | 20 | 0 | 0 | 0 | 0 | | b | = | gitara akustyczna |
| 0 | 0 | 71.4 | 14.3 | 0 | 14.3 | 0 | 0 | | c | = | gitara elektryczna |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 66.7 | 0 | 0 | 33.3 | | e | = | altówka |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | | f | = | skrzypce |
| 0 | 0 | 0 | 10 | 10 | 0 | 80 | 0 | | g | = | kontrabas |
| 0 | 0 | 0 | 0 | 0 | 14.3 | 0 | 85.7 | | h | = | wiolonczela |

1. Najlepszy wynik jest zdeterminowany 135-elementowym wektorem cech. Uwzględnienie tak dużej ilości atrybutów znacznie zwiększy złożoność algorytmiczną procesu filtrowania multimedialnych baz danych, co jest zjawiskiem niepożądanym.

2. Opisywany wektor cech zawiera deskryptory wynikające z analizy postaci czasowej próbki dźwięku. W 7.3.1 autor niniejszej rozprawy zaakcentował swoje obawy w stosunku do efektywnego stosowania deskryptorów postaci czasowe — odnośnie do klasyfikacji instrumentów muzycznych. Jednym z celów postawionych podczas prowadzonych badań było zaproponowanie takiego wektora cech, który zostanie odszukany tylko z wykorzystaniem analizy postaci widmowej badanej próbki.

Ponadto należy zwrócić uwagę na fakt, że w dla trzech klas instrumentów (harfa, gitara akustyczna i altówka) separowalność ogólnej rozpoznawalności waha się w granicach 50%–66.7%, co nie jest zadowalającym rezultatem.

Powyższe uwagi skłaniają do rezygnacji z dalszej analizy podziału fragmentu widma, wykorzystującego fragmentację $10 \times 7$.

### 9.2.3. Podział 8 kolumnowy z uwzględnieniem 4 warstw podziału energetycznego

**Wektor cech — 92 atrybuty.** W skład tego wektora cech zaliczono deskryptory:

1. cechy opisane w rozdziale 7, sekcja 7.3.2 (8 atrybutów);

2. 8-kolumnowy podział częstotliwościowy (8 atrybutów);

3. 4 warstwy podziału energetycznego — różna szerokość warstw (4 atrybuty);

4. 40 warstw podziału energetycznego — równa szerokość warstw (40 atrybutów);

5. siatka $4 \times 8$ (32 atrybuty).

Łącznie wyselekcjonowano 92 atrybuty. Poniżej przedstawiono wyniki testów z uwzględnieniem przykładowych algorytmów klasyfikacyjnych oraz uwzględnieniem procesu selekcji cech.

Tablica 9.12. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 92-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.2 % dla $k$=15 | 67.6 % dla 70%:30% |
| Tablice decyzyjne | 56.9 % dla $k$=5 | 60% dla 80%:20% |
| k-NN | 66.2 % dla $k$=5 | 77.8 % dla 80%:20% |
| Random Forest | 73.8% dla $k$=10 | 75.6% dla 80%:20% |

**Wektor cech — 29 atrybutów.** W skład wektora zaliczono cechy:

1. Grupa deskryptorów opisanych w rozdziale 7: Tr1, Ir (2 atrybuty);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3, 6, 7 (4 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 3, 10, 11, 12, 14, 15, 20, 21, 24, 27, 28, 30, 37 (13 atrybutów);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,1], [1,7], [1,8], [2,4], [3,1], [3,5], [3,7], [3,8], [4,1], [4,3] (10 atrybutów).

Tablica 9.13. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 29-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 76% dla $k$=5 | 84.4 % dla 80%:20% |
| Tablice decyzyjne | 56.9 % dla $k$=5 | 70.6 % dla 70%:30% |
| k-NN | 62.7 % dla $k$=5 | 73.3 % dla 80%:20% |
| Random Forest | 75.6% dla $k$=15 | 77.8% dla 80%:20% |

**Wektor cech — 15 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr1 (1 atrybut);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 3, 10, 12, 21, 37 (5 atrybutów);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,7], [1,8], [3,1], [3,5], [3,7], [3,8], [4,1] (7 atrybutów).

Tablica 9.14. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 15-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 72.4 % dla $k$=10 | 73.3 % dla 80%:20% |
| Tablice decyzyjne | 58.2 % dla $k$=10 | 73.3 % dla 80%:20% |
| k-NN | 62.7 % dla $k$=10 | 67.6 % dla 70%:30% |
| Random Forest | 75.6% dla $k$=10 | 86.7% dla 80%:20% |

**Wektor cech — 13 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr1 (1 atrybut);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 12, 21, 37 (3 atrybuty);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 4 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,7], [1,8], [3,1], [3,5], [3,7], [3,8], [4,1] (7 atrybutów).

Tablica 9.15. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 13-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 71.6 % dla $k$=10 | 73.3 % dla 80%:20% |
| Tablice decyzyjne | 60.9 % dla $k$=5 | 68.4 % dla 75%:25% |
| k-NN | 61.8 % dla $k$=10 | 71.1 % dla 80%:20% |
| Random Forest | 73.8% dla $k$=10 | 82.7% dla 77%:23% |

**Wektor cech — 12 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr1 (1 atrybut);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 12, 21, 37 (3 atrybuty);
4. analiza rozkładu energii z wykorzystaniem metody siatki. Siatka utworzona na bazie 4 warstw i 8 kolumn. Opis współrzędnych według indeksowania [warstwa, kolumna]: [1,7], [3,1], [3,5], [3,7], [3,8], [4,1] (7 atrybutów).

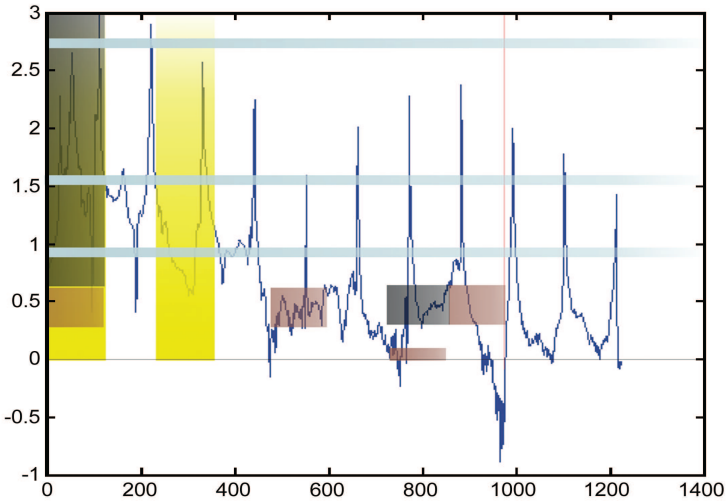
Tablica 9.16. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 12-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 71.6 % dla $k$=10 | 73.3 % dla 80%:20% |
| Tablice decyzyjne | 60.9 % dla $k$=5 | 68.4 % dla 75%:25% |
| k-NN | 56.9 % dla $k$=10 | 62.2 % dla 80%:20% |
| Random Forest | 74.7% dla $k$=15 | 89.6% dla 79%:21% |

Wyniki zgromadzone w Tab. 9.12–9.16 informują, że najkorzystniejszy rezultat uzyskano wykorzystując 12 elementowy wektor cech z uwzględnieniem Random Forestoraz metody holdout 79%:21%. Wykorzystując przedstawiony wektor cech uzyskano 89.6% ogólnej rozpoznawalności dla 8 klas instrumentów. Macierz przekłamań oraz graficzną reprezentacje najistotniejszych fragmentów widma przedstawiono poniżej.

Tablica 9.17. Macierz przekłamań dla klasyfikacji 8 klas instrumentów. Random Forest, metoda holdout 79%:21%. Ogólna rozpoznawalność 89.6%.

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 75 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | | a | = | harfa |
| 0 | 80 | 20 | 0 | 0 | 0 | 0 | 0 | | b | = | gitara akustyczna |
| 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | | c | = | gitara elektryczna |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | | e | = | altówka |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | | f | = | skrzypce |
| 0 | 0 | 0 | 0 | 0 | 0 | 81.8 | 18.2 | | g | = | kontrabas |
| 12.5 | 0 | 0 | 0 | 0 | 0 | 0 | 87.5 | | h | = | wiolonczela |




Rysunek 9.2. Najistotniejsze fragmenty widma dla procesu automatycznej klasyfikacji instrumentów muzycznych dla 12-elementowego wektora cech.

Analizując uzyskane wyniki przedstawione w macierzy przekłamań (Tab. 9.17) można zauważyć zadowalającą separację instrumentów. 50% badanych klas instrumentów zostało zinterpretowane ze 100% skutecznością, 3 klasy ze skutecznością ponad 80% oraz harfa z 75%. Należy również zaznaczyć fakt, że proces klasyfikacji przeprowadzony został z wykorzystaniem tylko deskryptorów widmowych.

### 9.2.4. Podział 8 kolumnowy z uwzględnieniem 7 warstw podziału energetycznego

**Wektor cech — 119 atrybutów.** W skład tego wektora cech zaliczono deskryptory:

1. cechy opisane w rozdziale 7, sekcja 7.3.2 (8 atrybutów);
2. 8 kolumnowy podział częstotliwościowy (8 atrybutów);
3. 7 warstw podziału energetycznego — różna szerokość warstw (7 atrybutów);
4. 40 warstw podziału energetycznego — równa szerokość warstw (40 atrybutów);
5. siatka $7 \times 8$ (56 atrybutów).

Łącznie wyselekcjonowano 119 atrybuty. Poniżej przedstawiono wyniki testów z uwzględnieniem przykładowych algorytmów klasyfikacyjnych oraz uwzględnieniem procesu selekcji cech.

Tablica 9.18. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 119-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.2 % dla $k$=20 | 82.2 % dla 80%:20% |
| Tablice decyzyjne | 57.3 % dla $k$=10 | 68.9 % dla 80%:20% |
| k-NN | 66.2 % dla $k$=10 | 75.6 % dla 80%:20% |
| Random Forest | 75.6% dla $k$=10 | 84.2% dla 75%:25% |

**Wektor cech — 34 atrybuty.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr2, Ir, ZC, $l_{tk}$ (4 atrybuty);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 2, 3, 7 (4 atrybuty);
3. 7 warstw podziału energetycznego — różna szerokość warstw, warstwy: 4, 7 (2 atrybuty);
4. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 1, 4, 7, 17, 18, 19, 23, 34, 37, 39 (10 atrybutów);
5. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,8], [2,3], [2,6], [3,4], [4,5], [4,6], [4,7], [4,8], [5,1], [5,2], [5,3], [6,1], [7,4], [7,7] (14 atrybutów).

Tablica 9.19. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 34-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 70.2 % dla $k$=20 | 82.2 % dla 80%:20% |
| Tablice decyzyjne | 60.4 % dla $k$=10 | 66.7 % dla 80%:20% |
| k-NN | 64.9 % dla $k$= 15 | 75.6 % dla 80%:20% |
| Random Forest | 77.3% dla $k$=20 | 84.2% dla 75%:25% |

**Wektor cech — 16 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr2, Ir, ZC (3 atrybuty);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 4, 17, 18, 39 (4 atrybuty);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,8], [2,3], [4,7], [5,3], [6,1], [7,4], [7,7] (7 atrybutów).

Tablica 9.20. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 16-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 67.1 % dla $k$=10 | 80% dla 80%:20% |
| Tablice decyzyjne | 54.7 % dla $k$=15 | 62.2 % dla 80%:20% |
| k-NN | 65.3 % dla $k$=5 | 71.1 % dla 80%:20% |
| Random Forest | 73.8% dla $k$=15 | 82.5% dla 75%:25% |

**Wektor cech — 13 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr2, ZC (2 atrybuty);
2. 8-kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw. Warstwy: 4, 17, 18, 39 (4 atrybuty);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,8], [4,7], [6,1], [7,4], [7,7] (5 atrybutów).

Tablica 9.21. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 13-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 67.6 % dla $k$=20 | 80% dla 80%:20% |
| Tablice decyzyjne | 60.4 % dla $k$=5 | 66.7 % dla 80%:20% |
| k-NN | 63.6 % dla $k$=5 | 66.2 % dla 70%:30% |
| Random Forest | 74.7% dla $k$=15 | 82.7% dla 77%:23% |

**Wektor cech — 12 atrybutów.** W skład wektora zaliczono cechy:

1. grupa deskryptorów opisanych w rozdziale 7: Tr2, ZC (2 atrybuty);
2. 8 kolumnowy podział częstotliwościowy — kolumny: 1, 3 (2 atrybuty);
3. 40 warstw podziału energetycznego — równa szerokość warstw, warstwy: 4, 17, 39 (3 atrybuty);
4. analiza rozkładu energii z wykorzystaniem metody siatki, siatka utworzona na bazie 7 warstw i 8 kolumn, opis współrzędnych według indeksowania [warstwa, kolumna]: [1,8], [4,7], [6,1], [7,4], [7,7] (5 atrybutów).

Tablica 9.22. Wyniki klasyfikacji 8 klas instrumentów z wykorzystaniem 12-elementowego wektora cech

| | metoda $k-$krotnej walidacji krzyżowej | metoda holdout |
|---|---|---|
| Drzewa decyzyjne | 65.8 % dla $k$=5 | 75.6 % dla 80%:20% |
| Tablice decyzyjne | 60.4 % dla $k$=5 | 66.7 % dla 80%:20% |
| k-NN | 62.7 % dla $k$=10 | 66.7 % dla 80%:20% |
| Random Forest | 74.2% dla $k$=20 | 82.2% dla 80%:20% |

Wyniki zgromadzone w Tab. 9.18–9.22 pokazują, że najwyższy stopień rozpoznawalności dla 8 klas instrumentów uzyskano wykorzystując 34-elementowy wektor cech z uwzględnieniem Random Forest oraz metody holdout 75%:25%. Wykorzystując przedstawiony wektor cech uzyskano 84.2% ogólnej rozpoznawalności dla 8 klas instrumentów. Należy również zwrócić uwagę na fakt, że identyczny procent rozpoznawalności otrzymano wykorzystując 119-elementowy wektor cech — również stosując Random Forest, holdout 75%:25%. Okazuje się zatem, że w kontekście opisywanych dwóch testów aż 85 deskryptorów nie odgrywało żadnej roli z punktu widzenia poprawy rozpoznawalności badanych klas instrumentów. Ponadto istotna kwestią jest fakt, że wykorzystując podział fragmentu widma na 7 warstw i 8 kolumn nie udało się wyeliminować deskryptora postaci czasowej (ZC). Należy zatem przyjąć,

że opisywana fragmentacja widma nie jest optymalna z punktu widzenia oczekiwanych wyników badań. Ponadto należy zwrócić uwagę na bardzo słaba rozpoznawalność dla harfy, która uzyskała zaledwie 20% poprawności w trakcie procesu klasyfikacji. Macierz przekłamań dla opisywanego testu przedstawiono poniżej.

Tablica 9.23. Macierz przekłamań dla klasyfikacji 8 klas instrumentów. Random Forest, metoda holdout 75%:25%. Ogólna rozpoznawalność 84.2%.

| **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | | | ← | **classified** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 0 | 40 | 0 | 0 | 0 | 40 | 0 | | a | = | harfa |
| 14.3 | 85.7 | 0 | 0 | 0 | 0 | 0 | 0 | | b | = | gitara akustyczna |
| 0 | 0 | 88.9 | 11.1 | 0 | 0 | 0 | 0 | | c | = | gitara elektryczna |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | | d | = | gitara basowa |
| 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | | e | = | altówka |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | | f | = | skrzypce |
| 0 | 0 | 0 | 0 | 9.1 | 0 | 72.7 | 18.2 | | g | = | kontrabas |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | | h | = | wiolonczela |

# Podsumowanie

Głównym celem tej pracy było odszukanie wektora cech umożliwiającego skuteczną klasyfikacje instrumentów strunowych (chordofonów) z artykulacją pizzicato. Założono, że poszukiwany wektor cech będzie zawierał możliwie jak najmniejszą ilość deskryptorów opisujących przebiegi badanych instrumentów muzycznych. Podjęto próbę skoncentrowania się przede wszystkim na analizie postaci widmowej fragmentu próbki dźwięku. Zdecydowano się zaproponować nową metodologie badań zakładającą analizę widma z uwzględnieniem stałej (4 Hz) rozdzielczości widma w stosunku do wszystkich badanych próbek. Ponadto zaproponowano analizę "stałego" okna czasowego dla badanej próbki. Pod pojęciem stałego okna czasowego rozumiano fragment przebiegu, który został pobrany zawsze w tym samym czasie oraz zawiera ta samą ilość próbek. Przyjęto, że realizacja tej metodologii badań doprowadzi do porównywania widma takiego samego fragmentu przebiegu, co zagwarantuje wysoką skuteczność automatycznej klasyfikacji. Poza tym przyjęto, że najistotniejsze w kontekście klasyfikacji cechy widma, zlokalizowane są w określonych przedziałach częstotliwościowych oraz energetycznych. Ponadto stwierdzono, że dla wybranej grupy instrumentów kluczowe cechy widma zlokalizowane są w przedziale częstotliwościowym 0–4 kHz. W związku z tym zaproponowano poszukiwanie cech dokonując fragmentacji określonej przestrzeni widma — co doprowadziło do podziału na warstwy, kolumny oraz zaproponowanie metody siatki. W trakcie realizacji badań zaproponowano kryterium doboru szerokości warstw oraz udowodniono, że przynosi ono zadowalające rezultaty. Definicja warstw, kolumn oraz siatki pozwoliła na uzyskanie wysokiego (blisko 90 procentowego) wyniku klasyfikacji 8 klas instrumentów pochodzących z tej samej grupy (chordofonów).

Ostatecznie zaproponowano dwa wektory cech, które zostały zdefiniowane z wykorzystaniem tylko analizy widmowej. Wektor cech składający się z 28 deskryptorów posiada zdolność 82.2 procentowej rozpoznawalności badanych klas instrumentów. Drugi z zaproponowanych zbiorów cech został utworzony na bazie 12 deskryptorów. Rozpoznawalność badanych klas, z wykorzystaniem tego wektora cech jest zbliżona do 90 % (89.6 % skuteczności), co jest zadowalającym wynikiem. Należy podkreślić, że podczas prowadzonych badań osiągnięto:

1. zbiór cech wynikający tylko z analizy przestrzeni widmowej,
2. wysoką skuteczność ogólnej rozpoznawalności badanych instrumentów,
3. zadowalającą separację instrumentów podczas procesu klasyfikacji.

Osiągnięcie wyżej wymienionych rezultatów uzasadnia słuszność przyjętych w niniejszej rozprawie tez. Poza tym dopinguje do dalszych prac, które powinny być skoncentrowane na:

1. Wykorzystaniu zaproponowanych wektorów cech do analizy dźwięków stereofonicznych z uwzględnieniem różnic w poszczególnych kanałach.

2. Zaproponowanie histogramu jako głównego kryterium doboru szerokości warstw oraz poszczególnych przestrzeni w zaproponowanej metodzie siatki.

3. Analiza wpływu artykulacji dźwięku, zdeterminowanej wyszkoleniem technicznym różnych muzyków w obrębie jednej klasy instrumentów, na skuteczność klasyfikacji.

4. Wykorzystanie zaproponowanych deskryptorów w przebiegach polifonicznych.

5. Opracowanie modułu klasyfikującego instrumenty muzyczne, współpracującego z wybranym systemem zarządzania bazą danych (np. Oracle, SQL-serwer).

# Bibliografia


[1] Shigeo Ando, Kiminori Yamaguchi, *Statistical study of spectral parameters in musical instrument tones*, J. Acoust. Soc. Am. **94** (1), 37–45, 1993.

[2] K. Blair Benson, *Audio Engineering Handbook*, McGraw Hill, 1988.

[3] Chen T., *Construction and frequency characteristics of Chinese bowed string instrument*, Proc. 15th Intern. Congres on Acoustics, Trondheim, Norway 1995, vol. III, pp. 401–404.

[4] J. Bonada, A. Loscos, P. Cano, X. Serra, *Spectral Approach to the Modeling of the Singing Voice* Presented at the 111th Convention 2001 September 21–24 New York, NY, USA.

[5] J.M. Martínez, *MPEG-7 Overview*, Klangenfurt, July 2002.

[6] X. Serra, X. Amatriain, J. Bonada, A. Loscos, *Spectral Modeling for Higher-level Sound Transformations*, Music Technology Group, Pompeu Fabra University.

[7] A. Horner, J. Beauchamp, *Synthesis of trumpet tones using a wavetable and a dynamic filter*, Audio Eng. Soc., **43** (10), 799–812, 1995.

[8] B. Kostek, A. Wieczorkowska, *Study of parameter relations in musical instrument patterns*, 100th AES Convention, Copenhagen, 1996, preprint 4173, J. Audio Eng. Soc. (Abstracts), **44**, No 7/8, p.634.

[9] B. Kostek, A. Wieczorkowska, *Parametric representation of musical sounds*, Archives of Acoustic, **22** (1), 3–26, 1997.

[10] J. Krimphoff, S. McAdams, S. Winsberg *Caracteristation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique*, Journal de physique IV, Colloque C5, J. de Physique III, 4, 3eme Congres Francais d'Acoustique, I, pp. 625–628, 1994.

[11] K.D. Martin, Y.E. Kim, *2pMU9. Musical instrument identyfication: A pattern-recognition approach*, Internet: `ftp://sound.media.mit.edu/pub/Papers/kdm-asa98.pdf`, presented at the 136th Meeting of the Acoustical Society of America, Norfolk, VA ,October 13, 1998.

[12] M. Paraskevas, J. Mourjopoulos, *A statistical study of the variability and features of audio signals: Some preliminary results*, 100th AES Convention, preprint 4256, Copenhagen 1996.

[13] P. Toiviainen, *Optimizing self-organizing timbre maps: Two approaches*, Joint International Conference 1996, College of Europe at Brugge, Belgium, 8–11 September 1996, II Int. Conf on Cognitive Musicology, pp. 264–271.

[14] Z. Żyszkowski, *Podstawy akustyki*, Wydawnictwo Naukowo-Techniczne, Warszawa 1987.

[15] B.S Manjunath, P. Salembier, T. Sikora, *Introduction to MPEG-7*, John Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex PO19 1UD, England 2002.

[16] A. Wieczorkowska, *Skuteczność rozpoznawania dźwięków instrumentów muzycznych w zależności od sposobu parametryzacji i rodzaju klasyfikatora*, Praca doktorska, Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, 1999.

[17] A. Januszajtis, *Fizyka dla politechnik, tom III, Fale.*, Wydawnictwo Naukowe PWN, Warszawa 1991, ISBN 83-09708-6.

[18] H.F. Pollard, E.V. Jansson, *A tristimulus method for the specification of musical timbre*, Acustica, **51**, 162–171, 1982.

[19] *Popularna Encyklopedia Powszechna*, tom 7, Fogra Oficyna Wydawnicza, Kraków 1995.

[20] K. Tyburek, *Rozpoznawanie zależności dźwięku instrumentów szarpanych*, IV Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim", Materiały konferencyjne, 285–289, Oprogramowanie Naukowo-Techniczne, ISBN 83-916420-1-1. Kraków 26–28 listopad 2003.

[21] T. Zieliński, *Od teorii do cyfrowego przetwarzania sygnałów*, Wydział EAIiE AGH Kraków 2002, ISBN 83-88309-55-2.

[22] C. Marven, G. Ewers, *Zarys cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 1999, ISBN 83-206-1306-X.

[23] R.G. Lyons, *Wprowadzenie do cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 2003, ISBN 83-206-1318-3.

[24] A. Czyżewski, *Dźwięk cyfrowy. Wybrane zagadnienia teoretyczne, technologia, zastosowania*, EXIT Warszawa 1998, ISBN 83-87674-08-7.

[25] C.J. Date, *Wprowadzenie do systemów baz danych*, WNT, Warszawa 2000 wyd. II.

[26] J.D. Ullman, J. Widom *Podstawowy wykład z systemów baz danych*, WNT Warszawa 1999 wyd. I.

[27] P. Beynon-Davies, *Systemy baz danych*, WNT, Warszawa 2000 wyd. II.

[28] M. Lentner, *Oracle 9i. Kompletny podręcznik użytkownika*, Wydawnictwo PJWSTK, Warszawa 2003.

[29] *Data base Systems, Courant Computer Science Symposia Series 6*, Prentice-Hall, N.J., 33–64, 1972.

[30] *Further Normalization of the Data Base Relational Model in Data Base systems, courant computer science symposia series 6*, Englewood Cliffs, N.J. Prentice-Hall, 1972.

[31] K.A.Ross, C.R.B. Wright, *Matematyka dyskretna*, Wydawnictwo naukowe PWN, Warszawa 1999.

[32] R.Elmasri, S.B. Navathe *Wprowadzenie do systemów baz danych*, Helion, Warszawa, 2005.

[33] Peter Pin-Shan Chen, *The Entity-Relationship Model — Toward a United View of Data*, ACM TODS 1, No.1, March 1976.

[34] A. Jaszkiewicz, *Inżynieria oprogramowania*, Helion, Warszawa 1997.

[35] J.L. Harrington, *Obiektowe bazy danych dla każdego*, Mikom, Warszawa 2001.

[36] W. Kim, *Wprowadzenie do obiektowych baz danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.

[37] T.W. Leung, G. Mitchell, B. Subramanian, B. Vence, S.L. Vandenberg, S.B. Zdonik, *The Aqua Data Model And Algebra*, Technical Report No. CS-93-09, March 1993.

[38] K. Subieta, *Słownik terminów z zakresu obiektowości*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[39] K. Subieta, J. Leszczyłowski, *A Critique of Object Algebras*,Institute of Computer Science, Polish Acad. Sci., Warszawa, Poland, 1995 (także: `http://www.ipipan.waw.pl/~subieta/artykuly/CritiqObjAlg.html`, wrzesień 2005)

[40] P. Józwik, M. Mazur, *Obiektowe bazy danych — przegląd i analiza rozwiązań*, praca dyplomowa AGH, Kraków 2002.

[41] K. Stąpor, *Automatyczna klasyfikacja obiektów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005.

[42] W. Greblicki, *Asymptotycznie optymalne algorytmy rozpoznawania i identyfikacji w warunkach probabilistycznych*, prace ICT Politechniki Wrocławskiej, Nr 18, seria Monografie, Nr 3, Wrocław 1974.

[43] M. Kurzyński, *Rozpoznawanie obiektów. Metody statystyczne*, Oficyna wydawnicza Politechniki Wrocławskiej. Wrocław 1997.

[44] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification and Scene Analysis*, John Wiley&Sons, New York 2000.

[45] P. Cichosz, *Systemy uczące się*, WNT, Warszawa 2000.

[46] A. Dominik *Analiza danych z zastosowaniem teorii zbiorów przybliżonych*, Praca dyplomowa magisterska, Politechnika Warszawska 2004.

[47] W. Siedlecki, J. Sklansky, *On automatic feature selection*, Int. J Pattern Recognition and Artificial Intelligence, **2** (2), 197–220, 1988.

[48] D. Goldberg, *Algorytmy genetyczne i ich zastosowania*, Wydawnictwa Naukowo-Techniczne, Warszawa 1995.

[49] T. Strąkowski, *Analiza danych medycznych z zastosowaniem metod zbiorów przybliżonych*, Praca magisterska, Politechnika Warszawska — Wydział Elektroniki i Technik Informacyjnych, Instytut Informatyki. Warszawa 2003.

[50] A.Mrózek, L. Płonka, *Analiza danych metodą zbiorów przybliżonych. Zastosowania w ekonomii, medycynie i sterowaniu*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[51] Z. Pawlak, *Rough Set.Teoretical Aspects of Reasoning About Data*, Wydawnictwo Politechniki Warszawskiej, Warszawa 1990.

[52] Z. Pawlak, *Systemy informacyjne — Podstawy teoretyczne*, Wydawnictwo Naukowo-Techniczne, Warszawa 1983.

[53] K. Tyburek, W. Cudny, W. Kosiński, *Analiza rozkładu częstotliwościowego dźwięków pizzicato*, InterPor.Lubostroń 2006.

[54] K. Tyburek, W. Cudny, W. Kosiński, *Pizzicato sound analysis of selected instruments in the frequency domain*, Image Processing & Communications, **11**(1), 53–57, 2006.

[55] J. Swacha, M. Bandosz, Ł. Radliński, *Zaawansowane multimedia na stronach www*, III Krajowa Konferencja „Multimedialne i Sieciowe Systemy Informacyjne" MISSI, Kliczków, 2002.

[56] M. Wojciechowski, Ł. Matuszczak, *Oracle interMedia na tle standardu SQL/MM i prototypowych systemów multimedialnych baz danych*, IX Konferencja PLOUG Kościelisko, Październik 2003.

Instytut Podstawowych Problemów Techniki
Polskiej Akademii Nauk

# Dodatek

do rozprawy doktorskiej

**Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato**

mgr Krzysztof Tyburek

Warszawa, Luty 2007

# Spis treści

# Wprowadzenie

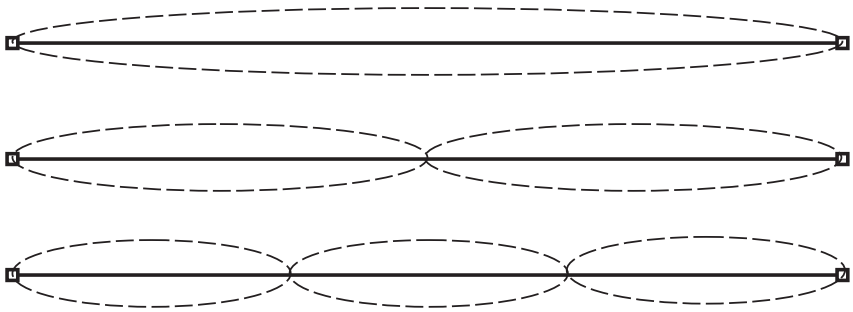
Przestawiony Dodatek powstał dla wypełnienia wymagań postawionych w pierwszym etapie recenzowania mojej rozprawy doktorskiej Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato przez Recenzentów. Niektóre z wymagań były wspólne, inne były stawiane tylko przez jednego z Recenzentów. Zamieszczona Literatura zawiera kilka nowych pozycji umieszczonych na końcu spisu.

ROZDZIAŁ 1

# Fale i ruch falowy. Dodatek

## Drganie struny

Struną nazywamy ciało wydłużone, wiotkie - tzn. nie wykazujące oporu przy zginaniu (wykonane najczęściej z metalu). Szarpnięcie napiętej struny powoduje zaburzenia (drgania poprzeczne) rozchodzące się wzdłuż struny. W związku z tym, że struna na końcach jest umocowana, to fala odbita od ośrodka gęstszego (umocowanie struny) daje po nałożeniu na falę padającą fale stojącą - co zilustrowano poniżej.



Rysunek 1.1. Fale stojące wzdłuż struny.

Na obu końcach struny powstają nieruchome węzły, a wszystkie powstałe punkty mają tę samą fazę drgań i wychylają się wszystkie równocześnie w jedna stronę (dla częstotliwości podstawowej $\nu_0$ struny). Dla wyższych tonów harmonicznych obok węzłów na końcu struny powstają wzdłuż struny węzły, dzielące ją na równe części. Liczba powstających węzłów odpowiada liczbie wyższym harmonicznym o częstotliwości $2\nu_0, 3\nu_0, 4\nu_0, ...$ itd. Prędkość rozchodzenia się fali $v$ w strunie określa zależność:

$$v = \sqrt{\frac{p}{\rho}} \tag{1.1}$$

gdzie: $p = \frac{F}{\delta}$ jest napięciem struny, $F$ - siłą napinającą, $\delta$ -przekrojem struny, $\rho$ - gęstość materiału struny. Dla częstotliwości podstawowej $\nu_0$ długość fali $\lambda_0$ jest równa podwójnej długości $l$ struny (co ilustruje Rys.1.1.). Oznacza to, że $\lambda_0 = 2l$ oraz $\nu_0 = \frac{v}{\lambda_0}$ ($v$ - prędkość przesuwania się zaburzenia wzdłuż struny). Częstotliwość podstawowa struny wyraża się zatem zależnością:

$$\nu_0 = \frac{1}{2l}\sqrt{\frac{F}{\delta\rho}} \tag{1.2}$$

Częstotliwości wyższych harmonicznych oblicza się wg zależności:

$$\nu_k = \frac{k+1}{2l}\sqrt{\frac{F}{\delta\rho}} = (k+1)\nu_0 \ , k = 1, 2, ... \tag{1.3}$$

$k$ - liczba węzłów występujących wzdłuż struny (nie licząc węzłów na końcach). Drgania podstawowe i wyższe harmoniczne tworzą tzw. *układ drgań własnych struny* [59].

## Rezonans

Wydobywanie słyszalnego dźwięku w instrumentach muzycznych, a w szczególności w instrumentach strunowych, nie byłoby możliwe bez zjawiska rezonansu (akustycznego), którym określa się zespól efektów występujących przy szybkim wzroście amlitudy drgań układu fizycznego, gdy częstość zewnętrznych drgań wymuszajacych jest zbliżona do częstości drgań własnych układu [14]. Jest to efekt przekazywania energi między układami i polega na tym, że jeśli mamy dwa układy: pudło i strunę (ogólnie elementy instrumentów), które mogą drgać, to jeśli istnieje między nimi połączenie umożliwiające propagację (rozchodzenie się) fali dźwiękowej, to drgania jednego elementu będą przekazywane innemu elementowi. O właściwym rezonansie mówimy jednak dopiero wtedy, gdy owo przekazywanie energii akustycznej osiąga największą efektywność.

Różne układy fizyczne – czytaj pudła rezenansowe instrumentów strunowych – mają różne zdolności do drgań (rezonansowych). Zdolności takie opisuje zazwyczaj tzw. krzywa rezonansowa, przedstawiająca zależność amplitudy drgań układu (tutaj pudła) od częstości drgań wymuszających (tutaj strun), jej maksimum winien wypadać przy wartości częstości drgań własnych. Charakter tej krzywej jest mocno związany z tłumieniem drgań w układzie – jego tarciem wewnętrznym: wzrost amplitudy drgań własnych układu jest tym szybszy im mniejsze są tłumienia drgań w układzie. Przykładowo, pudło skrzypiec jest złożone z kilku elemntów i to o różnych kształtach, z których każdy ma swoją charakterystykę rezonansową. Są te elementy połaczone klejem. Charkterystyka kleju też ma wpływa na przebieg tej krzywej. Dla odmiany harfa ma w zasadzie dwa pudła rezonansowe: tzw. właściwe i drugie - całą ramę. Wyznaczenie dla takiego złożonego układu krzywej rezonansowej nie jest proste. Tym bardziej, że w bardziej skomplikowanych sytuacjach, właśnie kiedy jest to układ złożony, krzywa rezonansowa może mieć kilka maksimów, odpowiadających różnym postaciom drgań w układzie.

Wydaje się, że dobre zrozumienie tego zjawiska i jego dobry opis jakościowy i ilościowy mogłby być pewną wskazówką na drodze poszukiwania odpowiednich deskryptorów zarejestrowanych dźwięków w artykulacji piccicato. Zapewne będzie to wspomagajacy krok naszych przyszłych badań i poszukiwań. W niniejszej rozprawie przyjęliśmy inny tok postępowania.

ROZDZIAŁ 2

# Charakterystyka wybranych instrumentów. Dodatek

## Chordofony.Dodatek

Chordofony Chordofony [gr. chordé = struna, phoné = dźwięk] - jest to grupa instrumentów, w których źródłem dźwięku są napięte struny. Instrumenty strunowe należą do najstarszych instrumentów muzycznych. Pierwszy znany wizerunek instrumentu strunowego pochodzi z malunków odkrytych we francuskich jaskiniach. Przedstawiają one mężczyznę grającego na jednostrunowym instrumencie za pomocą smyczka. Pierwotnie w celu wzmocnienia dźwięku pochodzącego z instrumentu strunowego używano ust, a następnie innych komór rezonansowych o naturalnym pochodzeniu. Pierwsze instrumenty strunowe przypominające współczesne ich odpowiedniki z grupy szarpanych powstały na Bliskim Wschodzie już w trzecim tysiącleciu p.n.e., skąd w wyniku ekspansji kulturowej dotarły także do Europy. Chordofony ze względu na sposób wzbudzania drgań dzielą się na:

1. Szarpane (np. palcem - harfa)
2. Uderzane (młoteczkiem, pałeczką, np. cymbały, fortepian)
3. Pocierane (smyczkowe, np. skrzypce, wiolonczela)
4. Dęte (pobudzane strumieniem powietrza, np. harfa eolska).

Kolejnym kryterium podziału jest konstrukcja instrumentu, ze względu na którą dzielimy chordofony na: 1. Łuki (np. łuk muzyczny) 2. Liry (np. lira klasyczna, chrotta) 3. Harfy i cytry, tj. chordofony bezszyjkowe (np. cytra, cymbały, fortepian) 4. Lutnie, tj. chordofony szyjkowe (np. lutnia, skrzypce, gitara).

Prototypem chordofonów był wywodzący się z myśliwskiego, łuk muzyczny z drgającą cięciwą jako źródłem dźwięku. Łuk ten zaopatrzony był w prymitywny rezonator. Szarpanie struny (cięciwy) palcami, uderzanie drewienkiem lub sztabką i pocieranie cięciwy łuku dało pierwsze podstawy do rozwoju grup chordofonów szarpanych, uderzanych i pocieranych. Na podstawie materiałów historycznych wiadomo, że już ok. 2000 roku przed Chrystusem starożytna Asyria i Babilonia znały takie instrumenty jak harfy czy liry. Ok. 1000 roku przed Chrystusem pojawiły się w Indiach chordofony smyczkowe, natomiast zastosowanie mechanizmu klawiszowego było dziełem średniowiecza.

Podczas gry na chordofonach wyróżniamy różne sposoby artykulacji dźwięku:

1. *Flażolety* - ton fletowy, ton harmoniczny dźwięku struny wydobyty przy stłumieniu tonu podstawowego i pozostałych tonów harmonicznych. Jego barwa przypomina delikatną barwę fletu. Struna pobudzona do drgań drga w sposób złożony, tzn. całą swoją długością i dzieląc się na odcinki. Drganie całą długością wytwarza ton podstawowy, natomiast drganie przy podziale na dwie, trzy i więcej części jest źródłem odpowiednio wyższych tonów harmonicznych. Flażolety naturalne wydobywa się ze struny pustej przez lekkie dotknięcie w 1/2, 1/3, 1/4, 1/5, 2/5 lub 1/6 długości struny. Są to odpowiednio flażolety oktawy, kwinty, kwarty, wielkiej tercji, wielkiej seksty i małej tercji. Nazwy tych flażoletów utworzono od interwałów, które uzyskałoby się ze struny przez normalne jej przyciśnięcie w danym punkcie. Realna wysokość flażoletu zależy od tego, który ton harmoniczny zostanie wydobyty - np. przy flażolecie kwinty dotknięcie struny w 1/3 długości powoduje powstanie trzeciego tonu harmonicznego, czyli tzw. duodecymy powyżej dźwięku struny pustej. Flażolety stosowane są przy grze na instrumentach smyczkowych, na gitarze i na harfie. Na pięciolinii oznacza się je nutami rombowymi lub kółkiem nad nutą.

2. *Smyczkowanie* - wydobywanie dźwięku z instrumentu za pomocą pociągania smyczkiem po strunie. Sposób użycia smyczka łączy się ściśle z artykulacją i może być różnorodny, dlatego wprowadzono szereg znaków i określeń, za pomocą których notuje się szczegółowo rodzaje smyczkowania. W zespołach i orkiestrach ustalenie jednakowego smyczkowania dla całej grupy wykonawców powierza się koncertmistrzowi. Istnieją dwa podstawowe sposoby pociągnięcia smyczkiem: a) Z góry do dołu (franc. *tiré*), czyli od karafułki do główki b) Z dołu do góry (franc. *Poussé*) - od główki do karafułki. Pierwszy sposób pozwala na silniejsze zaatakowanie dźwięku.

3. *Legato* - na pięciolinii zaznaczane łukiem nad grupą nut. Wykonuje się zawsze jednym pociągnięciem smyczka. Liczbę nut przypadających na to pociągnięcie określa się łukiem.

4. *Martelé, martellato* - wykonuje się pojedynczymi, krótkimi pociągnięciami smyczka. Na pięciolinii notowane znakiem (...).

5. *Sautillé, spiccato, saltato* - oznaczane na pięciolinii za pomocą kropek nad nutami. Wykonuje się środkową częścią smyczka - każdą nutę oddzielnym pociągnięciem, przy czym smyczek nie przylega do struny, lecz lekko podskakuje.

6. *Staccato* - wykonywane jednym, lecz przerywanym ruchem smyczka, oznaczane za pomocą kropek nad nutami połączonymi łukiem.

7. *Jeté, ricochet, gettato* - polega na wykonywaniu kilku dźwięków staccato jednym pociągnięciem sprężyście rzuconego na strunę smyczka, który odbija się kilkakrotnie.

8. *Flatter la corde* - miękkie i delikatne uderzenie smyczka w strunę, oznaczane za pomocą kropek nad nutami połączonymi łukiem.

9. *Con legno* - uderzanie strun drzewcem smyczka.

10. *Tremolo* - szybkie i krótkie zmienne pociągnięcie smyczkiem w celu wielokrotnego powtórzenia dźwięku.

11. *Flautando, flautato, sul tasto, sulla tastiera* - prowadzenie smyczka tuż nad strunnikiem (gryfem). Dźwięki wydobywane w ten sposób mają barwę matową, zbliżoną do dźwięków fletu.

12. *Sul ponticello, au chevalet* - prowadzenie smyczka przy podstawku, stosowane w celu osiągnięcia jasnej, metalicznej barwy.

13. *Pizzicato* - szarpnięcie struny (skrót *pizz.* szczypiąc, szarpiąc). W grze na instrumentach smyczkowych oznacza to wydobywanie dźwięku nie za pomocą smyczka, lecz szarpiąc strunę palcem - podobnie jak w instrumentach szarpanych, np. gitarze.

Pizzicata użył po raz pierwszy R. Keiser w operze Adonis (1697), później G. F. Händel (opery Agrippina 1709). N. Paganini wykonywał również pizzicato lewą dłonią przy jednoczesnym użyciu smyczka prowadzonego prawą ręką [60].

ROZDZIAŁ 3

# Przygotowanie danych eksperymentalnych. Dodatek

## Baza wybranych do analizy dźwięków

Do badań przeznaczono bazę dźwięków pochodzących z 4 oktaw:

1. wielkiej ($A$ 110 Hz)
2. małej ($a$ 220 Hz)
3. razkreślnej ($a^1$ 440 Hz)
4. dwukreślnej ($a^2$ 880 Hz).

Badane próbki dźwięków pochodziły z różnych baz dźwięków (por. punkt 7.1 rozprawy), co świadczy o tym, że autorami opisywanych próbek są różni (pochodzący z różnych części świata) muzycy. Na bazie zgromadzonych próbek wyselekcjonowano populację przeznaczoną do prowadzonych badań, uwzględniając dźwięki reprezentujące każdą z w/w oktaw. Łącznie do badań przeznaczono próbki:

- **Gitara akustyczna** (29 próbek) - zakres instrumentu $E - h^2$
  - Oktawa wielka $(A, B, G, Gis)$
  - Oktawa mała $(a, c, cis, d, f, fis, fis, h)$
  - Oktawa razkreślna $(a^1, c^1, cis^1, d^1, dis^1, dis^1, e^1, e^1, f^1, fis^1, g^1, gis^1, gis^1)$
  - Oktawa dwukreślna $(b^2, c^2, d^2, dis^2)$
- **Altówka** (27 próbek) - zakres instrumentu $c - e^3$
  - Oktawa wielka ( )
  - Oktawa mała $(a, b, e, fis, g, gis)$
  - Oktawa razkreślna $(a^1, c^1, cis^1, cis^1, d^1, d^1, e^1, f^1, f^1, fis^1, g^1, gis^1, h^1)$
  - Oktawa dwukreślna $(b^2, cis^2, d^2, e^2, f^2, g^2, gis^2, h^2)$
- **Gitara basowa** (28 próbek) - zakres instrumentu $D^1 - c^1$
  - Oktawa wielka $(A, B, E, F, Fis, G, Gis)$
  - Oktawa mała $(a, a, b, c, c, cis, cis, d, dis, dis, e, e, f, fis, fis, g, gis)$
  - Oktawa razkreślna $(c^1, d^1, dis^1, f^1)$

- Oktawa dwukreślna ( )

**Gitara elektryczna** ( 30 próbek) - zakres instrumentu $E - h^2$

- Oktawa wielka $(A, Fis, H)$
- Oktawa mała $(a, b, cis, e, f, fis, g, gis)$
- Oktawa razkreślna $(a^1, b^1, c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, g^1, gis^1)$
- Oktawa dwukreślna $(a^2, cis^2, d^2, e^2, f^2, fis^2, g^2, h^2)$

- **Harfa** (29 próbek) - zakres instrumentu $Ces^1 - ges^4$
  - Oktawa wielka ( )
  - Oktawa mała $(b, b, h, gis, g, a, h)$
  - Oktawa razkreślna $(b^1, cis^1, dis^1, e^1, e^1, fis^1, gis^1, cis^1, a^1, gis^1, c^1)$
  - Oktawa dwukreślna $(a^2, b^2, cis^2, d^2, dis^2, e^2, fis^2, g^2, c^2, gis^2)$
- **Kontrabas** ( 30 próbek) - zakres instrumentu $D - c^1$
  - Oktawa wielka $(A, B, C, D, D, Dis, E, E, F, F, Fis, Fis, Gis, Gis)$
  - Oktawa mała $(a, b, c, cis, d, dis, e, fis)$
  - Oktawa razkreślna $(c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, h^1)$
  - Oktawa dwukreślna ( )
- **Skrzypce** ( 30 próbek) - zakres instrumentu $g - c^4$
  - Oktawa wielka ( )
  - Oktawa mała $(g, a, c)$
  - Oktawa razkreślna $(a^1, a^1, b^1, c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, g^1, g^1, gis^1, gis^1)$
  - Oktawa dwukreślna $(a^2, b^2, b^2, c^2, cis^2, d^2, dis^2, e^2, f^2, fis^2, g^2, g^2, gis^2)$
- Wiolonczela ( 30 próbek) - zakres instrumentu $C - e^2$
  - Oktawa wielka $(C, A, F, Fis, G)$
  - Oktawa mała $(a, b, h, c, cis, d, dis, e, fis, g, gis, )$
  - Oktawa razkreślna $(a^1, b^1, c^1, d^1, dis^1, e^1, f^1, fis^1, g^1, gis^1)$
  - Oktawa dwukreślna $(c^2, cis^2, d^2, dis^2)$

Dobór poszczególnych próbek przeznaczonych do badań odbywał się losowo. Puste nawiasy przy niektórych oktawach oznaczają, że te dźwieki nie zostały wylosowane (lub instrument nie gra w niej).

## Zaproponowana metodologia badań. Dodatek

Już w trakcie początkowego etapu badań, analizując przebiegi nagrań, podejrzewano, że procesu ekstrakcji cech powinien być nakierowany głównie na parametryzację postaci widmowej przebiegu. Robiono też próby połączenia atrybutów wyprowadzonych z czystych przebiegów czasowych wzmocnionych atrybutami wynikającymi z zastosowania tranformat czasowo-częstotliwościowych. Próby te, częściowo

omawiane w rozprawie, nie dały wystarczajco dobrej skuteczności rozpoznania instrumentów. Podstawowe LLD deskryptory standardu MPEG-7, których wiekszość została przeanalizowana, są tutaj najlepszym przykładem

Naszym zdaniem w przestrzeni widmowej można znaleźć - ewentualnie dodatkowo zdefiniować - wystarczającą liczbę cech, które pozwolą na skuteczną parametryzację badanych klas instrumentów. Ostatecznie zdecydowano się dążyć do eliminacji deskryptorów opisujących postać czasową próbki i budowę wektora cech składającego się z deskryptorów widmowych.

Podstawą do realizacji tego pomysłu jest praktyczny aspekt związany z techniką gry na instrumentach przeznaczonych do badań. W trakcie analizy funkcji czasu dźwieku konieczne jest pobranie okna przebiegu w celu jego analizy. Powstaje zatem pytanie - jak długie powinno być okno czasowe przeznaczone do analizy? Jeżeli okno będzie zbyt krótkie, to może się okazać, że nie zawiera ono wystarczającej informacji, która może się przyczynić do skutecznej klasyfikacji instrumentów. Jeżeli natomiast poddamy analizie okno zbyt długie (np. 1/10 sekundy), to jest możliwe, że próbka dźwięku będzie krótsza niż automatycznie pobierane przez system komputerowy okno. Sytuacja taka jest szczególnie prawdopodobna w sytuacji, gdy zostanie poddany analizie materiał dźwiękowy generowany przez muzyka o wyższym stopniu wyszkolenia technicznego - szczególnie podczas gry staccato.

Jeżeli skupimy swoją uwagę tylko na postaci widmowej przebiegu, to w łatwy sposób można pominąć opisywany problem. Pobierając "stałe" okno czasowe (o długości 11025 próbek, tak jak założono w rozprawie - por rozdz. 7), tzn. fragment przebiegu, który został pobrany zawsze w tym samym czasie oraz zawiera tą samą ilość próbek: doprowadzamy do porównywania widma, takiego samego fragmentu przebiegu dla całej populacji badanych dźwięków. Ostatecznie do analizy widmowej zdecydowano się przeznaczyć okno czasowe, które zostało pobrane od momentu osiągnięcia maksymalnej wartości amplitudy. Długość okna - czyli moment zakończenie pobierania - jest zdeterminowane ustaleniem właściwej rozdzielczości widma, co opisano w (7.3.2) rozprawy. Jeżeli próbka dźwięku okaże się krótsza niż przyjęta długość okna (11025 próbek), to wówczas system pobiera fragment dźwięku rozpoczynający się od wartości $t_{max}$ i trwający do całkowitego wybrzmiewania nuty. Brakujące indeksy wektora, tzn. różnicę $r = ind_k - ind_{wybrz}$ gdzie: $ind_k$ - indeks ostatniej próbki badanego dźwięku (w przypadku prowadzonych badań przyjęto długość okna 11025 próbek), $ind_{wyburz}$ - indeks ostatniej próbki w trancjencie końcowym badanego dźwieku, uzupełniono zerami.

Wycięty fragment przebiegu postaci czasowej został poddany DFT, a jego widmo poddano analizie. Wykorzystując opisywaną metodologię doprowadzono do analizy widma zdeterminowanego zawsze tą samą rozdzielczością. Tak jak opisano w rozdz. 8 zdecydowano się skupić na rozkładzie zarówno częstotliwościowym jak i energetycznym badanego fragmentu widma. Kluczowym problemem był optymalny dobór szerokości warstw. Zdecydowano się zdeterminować go równym rozkładem energii w poszczególnych warstwach dla każdej z nut. Dla celów dalszych badań zdecydowano się określić szerokość warstw na podstawie wartości średnich wyliczonych na drodze analizy każdej nuty. Ostatecznie przyjęto następujące progi dla rozkładu energetycznego: 4 warstwowy podział fragmentu widma: 1.warstwa: 0 - 0.14; 2.warstwa 0.14 - 0.33; 3.warstwa 0.33 - 0.66; 4.warstwa 0.66 - 3.

Wykorzystując analizę fragmentu widma z wykorzystaniem metody siatki oraz metody warstw zdecydowano się badać ilość zgromadzonej energii w poszczególnych zakresach widma. Oznacza to, że określono szerokości warstw podziału energetycz-

nego widma, co opisano szerzej w punkcie 9.2 rozprawy. Na bazie podziału na wymienione warstwy oraz kolumny (również opisanych w rozdz. 9) stworzono siatkę. Zliczając energię skumulowaną w jej komórkach pozwoliło na zdefiniowanie kolejnych deskryptorów opisujących postać widmową badanego dźwięku. Deskryptory te reprezentują energię $W(p)$ w poszczególnych zakresach (warstwach, kolumnach, komórkach) zliczaną zgodnie z zależnością:

$$W(p) = \sum_{k=1}^{n_p} (A(f_k)_{max} - A(f_k))^2, \tag{3.1}$$

gdzie: $A(f_k)_{max}$ - maksymalna wartość amplitudy w danej warstwie[1] lub górna granica analizowanej warstwy, $A(f_k)$ - dolny próg (wartość amplitudy) analizowanej warstwy podziału energetycznego, $k$ - numer próbki, $n_p$ - liczba próbek w $p$-tej komórce, warstwie bądź kolumnie. Do zależności (3.1) będziemy się odwoływać w rozdziale 6 Dodatku mówiąc o nowych deskryptorach widmowych zaproponowanych w rozprawie.

Opisywana metodologia pozwoliła na zdefiniowanie kilku wektorów cech charakteryzujących się różnym stopniem ogólnej rozpoznawalności dla 8 klas instrumentów muzycznych. Ostatecznie najskuteczniejszy wektor cech wykazał zdolność 89,6% ogólnej rozpoznawalności. Szczegółową listę deskryptorów zawarto w punkcie 9.2 rozprawy.

[1] W przypadku kolumny wartości $A(f_k)_{max}$ i $A(f_k)$ wynikają z przebiegu funkcji amlitudy.

ROZDZIAŁ 4

# Parametryzacja dźwięków muzycznych. Dodatek

Tak jak wspomniano powyżej w rozdziale 4 rozprawy parametryzacja dźwięków muzycznych może odbywać się na drodze analizy sygnału zarówno w przestrzeni czasowej jak i widmowej. W licznych pracach naukowych z dziedziny *music information retrieval* (MIR), ich autorzy wprowadzają, uzasadniając celowość stosowania, swoje definicje deskryptorów stosowanych do automatycznej klasyfikacji instrumentów muzycznych.

Podstawą metodologiczną zaproponowanych deskryptorów są różne koncepcje analizy fragmentu przebiegu muzycznego w powiązaniu z ogólnie znanymi transformatami czasowo-częstotliwościowymi. Bardzo szeroko stosowanym deskryptorem opisującym postać widmową jest środek ciężkości widma. Deskryptor ten jest stosowany w procesie parametryzacji zarówno chordofonów jak i aerofonów. W swoich pracach skuteczność tego parametru podkreślają tacy autorzy jak X. Serra [61], I. Kaminskyj [62], B. Kostek i A. Czyżewski [63], G. Agostini [64], A. Wieczorkowska et all. [65]. Ponadto często stosowaną metodą analizy widma jest wyznaczenie grupy parametrów tristimulus. Parametry te pozwalają rozróżnić dźwięki w zależności od zawartości grup harmonicznych w widmie. Ponadto kształt widma można opisać za pomocą momentów widmowych $k$-tego rzędu. Problem ten został podjęty m.in. w [66]. Kolejnymi parametrami widma jest również zawartość składowych parzystych ($Ev$) i nieparzystych ($Od$) w widmie opisywana, m.in. w [67]. Należy również zaakcentować skuteczność deskryptora opisującego nieregularność widma $Ir$, wyrażoną zależnością:

$$Ir = \log \sum_{n=2}^{N-1} |20(\log A_n - \frac{1}{3}\log(A_{n+1}A_nA_{n-1}))| \tag{4.1}$$

gdzie: $A_n$ – amplituda $n$-tego prążka widma.

Parametr ten został również wyszczególniony przez w/w autorów m.in. w [65]. Powszechnie stosowanym parametrem do opisu postaci czasowej sygnału jest wartość skuteczna $RMS$ (*Root Mean Square Value*) – pierwiastek z wartości średniokwadratowej – wyrażony, w przypadku ciągłym, zależnością:

$$RMS = (\frac{1}{T}\int_0^T (x(t))^2 dt)^{1/2} \ . \tag{4.2}$$

Wykorzystanie do celów parametryzacji dźwięków muzycznych deskryptora $RMS$ zaakcentowano m.in. w pracy [68], w której poszukiwano wektora cech w celu klasy-

fikacji instrumentów muzycznych w pasażach solowych. Ponadto parametr ten został włączony do rozważań m.in. w pracy [62].

Kolejnym zagadnieniem związanym z analizą postaci widmowej jest metoda cepstralna rozumiana jako rezultat obliczania transformaty Fouriera widma sygnału w skali decybelowej. Istnieje zarówno zespolone cepstrum jak i rzeczywiste. Cepstrum rzeczywiste zdefiniowane jest jako odrotna transformata Fouriera z logarytmu modułu transformaty Fouriera samej funkcji

$$X(t') = F^{-1}(\ln|F(x(t))|) \tag{4.3}$$

Definicja ta wykorzystuje logarytm rzeczywisty, liczony jedynie na bazie widma amplitudowego. Metoda cepstralna została również zaakcentowana między innymi w pracach [62] i [69]. Autorzy w pracy [68] również wykorzystują deskryptory opisu dźwięków muzycznych MPEG–7 (standard ten został opisany w dalszej części niniejszego Dodatku), takie jak:

1. Harmonic centroid (HC)
2. Harmonic deviation (HD)
3. Harmonic spread (HS)
4. Harmonic variation (HV)
5. Log-attack-time (LAT)
6. Temporal centroid (TC)
7. Spectral centroid (SC)

Podczas procesu parametryzacji dźwięków muzycznych poza optymalnym doborem deskryptorów istotne jest zastosowanie ich do właściwego fragmentu badanego przebiegu. Parametryzacja może odbywać się na podstawie analizy transjentu początkowego, transjentu końcowego oraz stanu quasi-ustalonego. Poza tym w zależności od badanego instrumentu, każdy z w/w fragmentów dźwięku może podlegać fragmentacji na $N$ ramek, które są analizowane niezależnie. Analizie może podlegać pełen zakres częstotliwości badanej ramki lub tylko pewien jego fragment. Np. w pracy [68] autorzy skupili się na analizie sygnału w zakresie częstotliwościowym 141 - 8877 Hz. A. Wieczorkowska w swoich pracach (m.in. w pozycji [16]) stosuje zwykle podobne podziały na podstawie których dokonywana jest analiza stanu quasi-ustalonego. Dzięki takiej metodologii łatwo można zbadać rozkład energii między niskimi, średnimi i wysokimi przedziałami częstotliwości.

Należy zwrócić uwagę, że między innymi rozkład energii w poszczególnych zakresach częstotliwościowych stanowił jeden z aspektów pracy badawczej podjętej w niniejszej rozprawie. Szczegółowy opis zamieszczono rozdziale 8. Tak jak wspomniano wcześniej grupa wymienionych deskryptorów stosowana jest do opisu poszczególnych fragmentów przebiegu. Wymienieni wyżej autorzy, podejmujący próbę parametryzacji dźwięków instrumentów muzycznych, analizują najczęściej przebiegi, które posiadają stan quasi-ustalony – np. [62, 68]. Przebiegi te charakteryzują się często stosunkowo długim czasem trwania (ok. 15 sekund), a analiza stanu quasi-ustalonego pozwala zdefiniować taki wektor cech, który dostarcza wysoki procent rozpoznawalności badanych klas instrumentów. Na przykład analizując stan quasi-ustalony w łatwy sposób można zbadać obecność wibrata w analizowanej próbce dźwięku. Np.

analiza obecności wibrata została poruszona w pracach [65, 70]. Wielkość wibrata [Hz] rozumiana jest jako wartość bezwzględna różnicy między wysokością dźwięku przy maksymalnej oraz minimalnej amplitudzie w stanie quasi–ustalonym. Łatwo wywnioskować, że deskryptor opisujący obecność wibrata nie jest skuteczny w kontekście parametryzacji dźwięków instrumentów strunowych z artykulacją pizzicato, które podjęto podczas realizacji niniejszej rozprawy. Związane jest to z brakiem stanu quasi-ustalonego w dźwiękach pizzicato instrumentów strunowych.

Szerzej na temat fizycznych cech badanych próbek napisano w rozdziale 7. W pracy [68] autorzy badali 20 klas instrumentów muzycznych (zarówno chordofony jak i aerofony) uzyskując ogólną rozpoznawalnośc ponad 90% - analizowano głównie stan quasi- ustalony. Wyniki ich badań dowodzą, że najlepiej rozpoznawalnymi instrumentami są trąbka, flet, skrzypce i fortepian (wykorzystano pakiet WEKA, $k$-NN, metoda holdout 66:34). Ponadto autorzy stwierdzili, że najwięcej pomyłek w rozpoznawaniu klas instrumentów zanotowano w parze trąbką i pianino.

Najlepsze wyniki w [68] jej autorzy uzyskali w trakcie analizy próbek skrzypiec i fletu. Należy jednak zaakcentować fakt, że badane instrumenty muzyczne pochodziły z różnych grup instrumentów. Wydaje się być oczywiste i intuicyjne, że rozpoznawalność trąbki i skrzypiec i ich odróżnienie będzie zdecydowanie wyższa niż między skrzypcami i altówka. Trudnością, z jaką spotykamy się przy rozpoznawaniu tej drugiej pary instrumentów, może być fakt, że altówka pochodzi z rodziny skrzypiec (altówka jest określana jako skrzypce altowe) a co za tym idzie charakterystyka tych instrumentów jest bardzo zbliżona – tym bardziej, jeżeli uwzględnimy taką samą artykulację dźwięku. Nie można natomiast tego samego powiedzieć porównując np. skrzypce i trąbkę. W trakcie realizacji badań opisanych w niniejszej rozprawie skupiono się tylko na parametryzacji dźwięków, których źródło stanowią chordofony z artykulacją pizzicato. Oznacza to, że zdecydowano się zaproponować taki wektor cech, który bazuje na analizie transjentu końcowego oraz dostarcza zadowalający stopień rozpoznawalności.

Stwierdzono też, że ogólnie znane i stosowane (przez w/w, uznanych Autorów) deskryptory w połączeniu z zaproponowaną w rozprawie metodologią badań (por. rozdział 8) nie przynoszą zadowalających rezultatów. Doprowadziło to w efekcie do zaproponowania nowego wektora cech, który wykazuje ok. 90% skuteczności podczas klasyfikacji wybranych instrumentów strunowych.

ROZDZIAŁ 5

# Bazy danych i system zarządzania bazą danych. Dodatek

## Bazy multimedialne. Dodatek

Multimedialne bazy danych (MMBD), będące najczęściej elementem systemu rozproszonego umożliwiają zarządzanie danymi multimedialnymi. W przeciwieństwie do klasycznych baz danych (wykorzystujących relacyjny, obiektowy lub obiektowo-relacyjny model danych), MMBD nie przechowują informacji jako takich (np. nagrań dźwiękowych, fragmentów filmów lub grafiki), a jedynie informacje o danych (metadane), które stanowią podstawę jej funkcjonalności. Metadane powstają w wyniku procesu indeksowania zawartości multimedialnej. Wyszukiwanie informacji odbywa się na drodze analizy i porównań metadanych (np. rozkład harmonicznych w poszczególnych przestrzeniach częstotliwościowych, właściwe nasycenie RGB fragmentu grafiki) kwerendy wystosowanej do systemu w postaci danych multimedialnych oraz metadanych przechowywanych w MMBD. Wynik porównań (kwerendy) kierowany jest do rzeczywistego obiektu (np. fragmentu nagrania muzycznego), który może być przesłany do klienta w postaci strumienia danych. Rzeczywiste składowanie danych realizowane jest w następujących formach:

1. Wewnątrz struktur bazy danych
2. W plikach w systemie plików systemu operacyjnego
3. Na zewnętrznych serwerach przeznaczonych do składowania multimediów.

Ostatecznie można stwierdzić, że przeszukiwanie MMBD z wykorzystaniem kwerendy multimedialnej odbywa się w 3 fazach:

1. Parametryzacja szukanego fragmentu
2. Wyszukanie metadanych pasujących do zapytania
3. Informacja o wyniku wyszukiwania.

W MMBD wyróżnia się dwie metody indeksowania danych:

- Tradycyjna (etykietowanie danych) – uzależniona od słów kluczowych i opisie tekstowym danych multimedialnych. Metoda ta charakteryzuje się małymi możliwościami wyszukiwania.

- Metoda bazująca na opisie zawartości (context-based indexing) - parametry uzyskiwane na podstawie analizy zawartości, powinny tak opisywać dane, aby zagwarantować wysoką skuteczność procesu filtracji. Metoda ta znacznie zwiększa (w porównaniu do metody tradycyjnej) możliwości wyszukiwania danych.

W wyniku indeksowania danych uzyskuje się $n$-wymiarowy wektor cech, opisujący zawartość - a więc „punkt w przestrzeni $n$-wymiarowej". Wyszukiwanie podobieństwa danych sprowadza się do odszukania punktów znajdujących się możliwie najbliżej punktu odpowiadającego zapytaniu. Z praktycznego punktu widzenia, do wyszukiwania danych wykorzystuje się kombinację cech, które w połączeniu z algorytmami klasyfikującymi zwracają grupę podobnych obiektów. Rozróżnia się dwa sposoby korzystania z multimedialnej bazy danych:

- *Pull* („aktywne") - użytkownik systemu wysyła do bazy polecenie wyszukiwania w postaci kwerendy multimedialnej a serwer zwraca metadane lub pasujące dane.

- *Push* („pasywne") - użytkownik systemu określa w sposób opisowy pewne preferencje w kwerendzie (np. interesująca go tematyka filmu). System bazy danych wyszukuje i przesyła zawartość wg. zadanych preferencji (np. wszystkie filmy związane z tematyką II Wojny Światowej), bez konieczności ingerencji użytkownika.

Przykładową multimedialną bazą jest opracowana w Instytucie Informatyki Politechniki Poznańskiej baza dźwięków skrzypiec AMATI. W bazie tej zgromadzono dźwięki kilkudziesięciu instrumentów nadesłanych na Międzynarodowy Konkurs Lutniczy im. Henryka Wieniawskiego w Poznaniu, który odbył się na jesieni 2001 roku. Zebrany materiał dźwiękowy wraz z ocenami jurorów ma służyć badaniom w dziedzinach związanych z cyfrowym przetwarzaniem sygnałów muzycznych oraz słuchaniem maszynowym, np. automatyczną klasyfikacją barwy dźwięku skrzypiec, automatyczną oceną jakości ich dźwięków w zestawieniu z subiektywnymi ocenami jurorów. Dźwięki w bazie są w pełni identyfikowane na podstawie nazw plików, w których są przechowywane. Nazwa zawiera numer konkursowy skrzypiec, sposób wydobycia dźwięku (détaché, pizzicato), strunę, na której dźwięk był zagrany, kierunek ruchu smyczka oraz numer powtórzenia dźwięku.

Baza AMATI dostarcza parametry pozwalające na wstępne porównanie dźwięków tego samego i różnych instrumentów. Deskryptory te otrzymano głównie na bazie analizy amplitud harmonicznych sygnału. Amplitudy te obliczane są na podstawie widma o dużej rozdzielczości. Wśród tych parametrów znajdują się między innymi: średnia energia dźwięku, względne energie pierwszej harmonicznej, harmonicznych pasma średniego (sumy drugiej, trzeciej i czwartej harmonicznej), sumy wyższych harmonicznych (powyżej czwartej), parzystych i nieparzystych harmonicznych – w odniesieniu do energii wszystkich harmonicznych, jasność dźwięku (środek ciężkości widma) oraz trzy pierwsze współczynniki cepstralne [71].

ROZDZIAŁ 6

# Standard MPEG -7 Audio. Nowe deskryptory widmowe

Większość dotychczasowych rozwiązań związanych z wydobywaniem wiedzy bazuje na technice etykietowania przechowywanych informacji — do pewnego czasu technika ta dotyczyła również danych opisujących dźwięk. Cała procedura etykietowania jest dość pracochłonna i czasochłonna. Poza tym takie rozwiązanie nie zawsze daje rzetelny wynik — to znaczy wysyłane zapytanie nie zawsze jest zgodne z oczekiwaniami osoby (czy systemu) pytającej. Istnieje wielkie prawdopodobieństwo, że dwie zupełnie różne (binarnie) informacje dźwiękowe mogą się okazać tą samą sekwencją utworu muzycznego zagranego z różną dynamiką lub w pomieszczeniu o różnej akustyce. Kolejny problem, który występuje w procesie rozpoznawania sygnałów dźwiękowych, jest właściwa interpretacja źródła dźwięku. Rozpoznanie dźwięku pochodzącego na przykład z drgającej struny gitary może być bardzo trudne. Trudność ta najczęściej wynika z doskonałych procesorów muzycznych, za pomocą których z łatwością można "podrobić" oryginalny instrument. W obliczu pojawiających się problemów związanych z możliwością wyszukiwania informacji audio najistotniejszym zagadnieniem jest opracowanie stosownych algorytmów wyszukujących właściwy system kodowania informacji multimedialnych oraz klasyfikujący typ źródła dźwięku (na przykład instrumenty strunowe, dęte drewniane i tym podobne).

## Ogólna charakterystyka standardu MPEG-7

Drogą do rozwiązania problemu klasyfikacji i agregacji danych multimedialnych jest - posiadający certyfikat ISO - standard MPEG-7. Standard ten dostarcza szeregu podstawowych deskryptorów opisujących dźwięk. Na bazie standardu MPEG 7 stworzono nowe deskryptory rozpoznające pewne instrumenty muzyczne. MPEG-7 jest standardem definiującym język opisu zawartości obiektów multimedialnych (MM) (ang. *Multimedia Content Description Interface*). O ile poprzednie standardy (MPEG-1, MPEG-2 i MPEG-4) zajmowały się normowaniem przekazywania zawartości obiektów multimedialnych, to standard MPEG-7 pozwala na odpowiedni opis, indeksowanie, a następnie wyszukiwanie tej zawartości zgodnie z potrzebami użytkowników. Część systemowa standardu udostępnia narzędzia do opakowania tych opisów i do tworzenia postaci binarnej dla przesyłanego strumienia MPEG-7. Służy do tego język DDL z rodziny XML, który pozwala na definiowanie nowych typów deskryptorów, a także na rozszerzanie i przedefiniowanie typów już istniejących.

MPEG-7 obejmuje standardem [5]:

1. Pewne zbiory deskryptorów - (ang. *descriptor*) (D), tj. zbiory obiektów, które mogą reprezentować cechę obiektu multimedialnych (MM) zarówno w sensie składniowym jak i znaczeniowym (np. kolory, kształty, częstotliwości dźwięków), jak i cech na poziomie semantycznym (np. dane o zawodach sportowych, które zostały zarejestrowane przez kamery).

2. Pewne zbiory schematów deskryptorów (ang. *description scheme*) (DS), które stanowią zapis i znaczenie relacji między swoimi składowymi, tj. wybranymi deskryptorami, albo też mogą być rekurencyjnie schematami deskryptorów.

3. Jezyk definiowania deskryptorów i schematów deskryptorów (ang. *description definition language*)(DDL). Język DDL, będący rozszerzeniem języka XML Schema, umożliwia też rozszerzenia i modyfikacje istniejących schematów deskryptorów. Tym samym schematy deskryptorów i same deskryptory są dokumentami języka XML. Jest to część systemowa standardu, która udostępnia narzędzia do opakowania powyższych opisów (tj. deskryptorów i schematów deskryptorów) i do tworzenia postaci binarnej (por. BiM) dla przesyłanego strumienia MPEG-7. Język ten pozwala na definiowanie nowych typów deskryptorów, a także na rozszerzanie i przedefiniowanie typów już istniejących.

4. Metodę binarnego kodowania deskryptorów – BiM ( skrót pochodzi od ang. *Binary Format for MPEG-7 data*). Jest to metoda kodowania strumieniowego, spełniająca takie wymagania jak: efektywność kompresji, odporność na błędy oraz bezpośredni dostęp do swoich składowych.

Podstawowym celem standardu MPEG-7 jest stworzenie metod opisu, indeksacji i klasyfikacji obiektów (danych) multimedialnych. Dla tego celu konieczne jest stworzenie efektywnych metod wyszukiwania poszczególnych elementów czy cech danych multimedialnych. Stąd dochodzi sie do: a) metadanych: danych o danych, takich jak autor, producent, copyright; b) semantyki: obiekty, zdarzenia, ludzie; c) podziałów: regiony, segmenty; d) cech: tekstury, kolory, kontury.

Schemat opisu danego deskryptora składa się z trzech zasadniczych kroków - opisu schematu deskryptora w języku DDL, opisu struktury deskryptora w kodzie binarnym (BiM) oraz opisu semantyki deskryptora. Jeśli z powyższych trzech punktów nie wynika jednoznacznie sposób dekodowania deskryptora, to zwykle dodaje się jeszcze krok definiujący ściśle kod dekodera.

MPEG-7 wyróżnia pięć następujących kategorii schematów deskryptorów:

1. Elementy podstawowe (ang. *Basic elements*), a wśród nich:

   (a) Narzędzia tworzenia schematów (ang. *Schema tools*);
   (b) Podstawowe typy danych i struktury (ang. *Datatype & structures*);
   (c) Referencje i lokalizacje mediów (ang *Link & media localization*);
   (d) Bazowe schematy deskryptorów (ang. *Basic description schemes*) – używane przez inne schematy deskryptorów.

2. Elementy opisu i zarządzania zawartością (ang. *Content description & management*), a wśród nich:

   (a) Elementy opisu zawartości (ang *Content description*): Struktura zawartości (ang. *Structural aspects*) – cechy oparte na analizie sygnałów; znaczenie zawartości (ang. *Semantic aspects*) – informacje o zdarzeniach i obiektach.

3. Elementy zarządzania zawartością (ang. *Content management*) – mogą być dołączone również do komponentów danego obiektu MM: tworzenie i produkcja treści medialnej (ang. *Creation & production*) – twórcy, klasyfikatory, określenie odbiorców treści medialnej; typy mediów (ang. *Media*) – formaty kodowania MM, identyfikacja mediów; użycie mediów (ang. *Usage*) – prawa dostępu, jednostki uprawnione, informacje finansowe i publikacyjne.

## Podstawy standardu MPEG-7 Audio

Przechodząc do części audio warto zwrócić uwagę na znaczenie niniejszego standardu w przeszukiwaniu multimedialnych baz danych. MPEG-7 zawiera warstwę nośną niskiego poziomu narzędzi, które stosowane są ogólnie we wszelkich dźwiękach. Mechanizmy te ustalają podstawowy poziom kompatybilności wśród opisów audio i pozwalają tworzyć nowe aplikacje. Warstwa ta również integruje MPEG-7 z innymi częściami standardu. Na warstwie nośnej zbudowana jest też seria narzędzi wysokiego poziomu, które są dostosowywane do poszczególnych grup aplikacji.

Zadaniem tych grup jest przeszukiwanie i wyszukiwanie cech sygnałów. Usługa audio realizowana jest często z wykorzystaniem bazy danych skompresowanego MPEG-4 znajdującego się na jednym (lub więcej) serwerze medialnym oraz skojarzoną z nim asocjacyjną bazą metadanych MPEG-7 na serwerze kwerendowym. Dla każdego indeksowanego skompresowanego sygnału audio MPEG-4, baza metadanych MPEG-7 przechowuje pełną reprezentację melodii oraz mechanizm łączący metadane ze skojarzonym nośnikiem, na przykład adres URL. Baza danych może również przechowywać inne deskryptory opisujące, np. styl i gatunek muzyki oraz bazę dźwięków instrumentów w charakterystycznym pasażu. Wykorzystując opis sygnału za pośrednictwem deskryptorów czy schematów deskryptorów istnieje możliwość wysłania kwerendy w formie fonicznej. Kształt fali sygnału próbnego jest przenoszony na serwer zapytań (kwerend), gdzie przeprowadzany jest proces, w którym uzyskiwane są jego metadane, które są celem kwerendy. Serwer MPEG-7 wyszukuje zatem odpowiedników melodii z baz danych. Kilka najlepszych odpowiedników przenoszonych jest z powrotem do urządzenia.

W części audio tego standardu najogólniej deskryptory można podzielić na dwa podtypy:

1. deskryptory funkcji widma,
2. deskryptory funkcji czasu.

Podstawowy podział deskryptorów i schematów deskryptorów audio jest na dwie klasy:

- ogólne narzędzia niskiego poziomu (ang.*generic low–level tools*),
- narzędzia nakierowane na szczególną aplikację (ang. *application–specific tools*).

W konsekwencji tej klasyfikacji pliki dźwiękowe w standardzie MPEG-7 są opisywane za pomocą deskryptorów niskiego poziomu (*low level descriptors*)(LLD), które odnoszą się do właściwości dźwięku (czasowe, widmowe) oraz wysokiego poziomu (*high level descriptors*). Deskryptory wysokiego poziomu stanowią opis zawartości multimedialnej sporządzony pod kątem określonego zastosowania, np. opisujące barwę

dźwięku, melodię itp. Nie są to parametry uzyskiwane na podstawie analizy pliku dźwiękowego, ale np. sposób wykorzystania deskryptorów niskiego poziomu w celu umożliwienia wyszukiwania nagrań w multimedialnej bazie danych [15].

MPEG-7 dostarcza 17 podstawowych deskryptorów (niskiego poziomu)sklasyfikowanych w poszczególnych klasach:

- postawowe (ang. *Basic*),
- podstawowe widmowe (ang. *Basic Spectral*),
- parametry sygnałowe (ang. *Signal Parameters*),
- czasowy barwy dźwięku (ang. *Timbral Temporal*),
- widmowy barwy dźwięku (ang. *Timbral Spectral*),
- baza widmowa (ang. *Spectral Basis*),
- deskryptor ciszy (ang. *Silence Descriptor*).

Scharakteryzujmy pokrótce poszczególne klasy przez wyszczególnione ich składników [15, 57, 58].

- Basic – deskryptory podstawowe zawierają:
  - przebieg czasowy (*AudioWaveform*) Jest to obwiednia (górna i dolna) przebiegu czasowego spróbkowanego sygnału audio. Deskryptor ten jest używany głównie do wyświetlania przebiegu czasowego.
  - moc sygnału (*AudioPower*) Jest to wygładzony przebieg chwilowej mocy sygnału w funkcji czasu i jest używany razem z deskryptorami widmowymi.
- Basic spectral – deskryptory widmowe podstawowe zawierają:
  - obwiednię widma (*AudioSpectrumEnvelope*) Jest to obwiednia krótkoterminowego widma mocy sygnału.
  - logarytmiczną skalę częstotliwości
  - środek ciężkości widma (*AudioSpectrumCentroid*) Środek ciężkości widma mocy sygnału opisuje rozkład częstotliwościowy widma; odpowiada na pytanie czy w widmie sygnału dominują składowe nisko- czy wysokoczęstotliwościowe.
  - rozkład widma (*AudioSpectrumSpread*) Drugi moment widma mocy sygnału. Deskryptor opisuje, czy widmo sygnału jest skupione wokół środka ciężkości widma, czy też pokrywa szerszy zakres częstotliwości. Pomaga to np. odróżnić dźwięki szumowe od tonów.
  - płaskość widma (*AudioSpectrumFlatness*) Deskryptor ten opisuje płaskość widma w każdym z pasm częstotliwości. Płaskie widmo może oznaczać sygnał szumowy lub impulsowy, mała wartość parametru może wskazywać na sygnał harmoniczny.
- Signal parameters – parametry sygnału odnoszą się do sygnałów okresowych bądź quasi-okresowych i zawierają:

- częstotliwość podstawową (*AudioFundamentalFrequency*) Jest to częstotliwość podstawowa sygnału, uzyskiwana za pomocą algorytmu pitch-tracking.
- harmoniczność sygnału (*AudioHarmonicity*) Pozwala rozróżnić sygnały harmoniczne, nieharmoniczne oraz nie posiadające widma prążkowego (np. szum).

➤ TIMBRAL TEMPORAL – PARAMETRY CZASOWE BARWY DŹWIĘKU

- czas narastania dźwięku (*LogAttackTime*) Wartość skalarna opisująca logarytm (przy podstawie 10) czasu mierznego w sekundach narastania amplitudy dźwięku od ciszy do wartości maksymalnej. Pozwala opisać np. transjenty dźwięków muzycznych.
- środek ciężkości przebiegu czasowego (*TemporalCentroid*) Wartość skalarna opisująca punkt na osi czasu, w którym skupia się energia sygnału (środek ciężkości obwiedni czasowej sygnału). Pozwala rozróżnić np. dźwięki o takim samym czasie narastania, ale różnym czasie wybrzmiewania.

➤ TIMBRAL SPECTRAL – PARAMETRY WIDMOWE BARWY DŹWIĘKUzawierają:

- środek ciężkości widma (*SpectralCentroid*) Liniowa skala częstotliwości jest dzielona na zakresy, liczona jest średnia ważona częstotliwości środkowych wszystkich zakresów, ważona według mocy sygnału w każdym paśmie. Parametr ten jest skorelowany z subiektywnym wrażeniem ostrości dźwięku (*sharpness*).
- środek ciężkości harmonicznych (*HarmonicSpectralCentroid*) Średnia ważona częstotliwości prążków harmonicznych w widmie, ważona według amplitudy prążków. Dotyczy tylko składowych harmonicznych widma.
- odchyłkę harmonicznych (*HarmonicSpectralDeviation*) Różnica między amplitudami prążków harmonicznych a obwiednią widma .
- rozkład harmonicznych (*HarmonicSpectralSpread*) Rozkład harmonicznych w widmie względem środka ciężkości.
- zmienność harmonicznych (*HarmonicSpectralVariation*) Znormalizowana korelacja między amplitudami prążków harmonicznych w dwóch kolejnych ramkach sygnału. Opisuje zmienność widma harmonicznego w czasie.

➤ SPECTRAL BASIS – BAZA WIDMOWA to deskryptory służące do opisu dynamicznego 3D (trójwymiarowego) widma sygnału jako trójka: czas – częstotliwość – poziom i zawiera:

- funkcje bazowe widma (*AudioSpectrumBasis*) Zbiór funkcji bazowych, otrzymanych przez dekompozycję znormalizowanego widma mocy.
- funkcje przekształcające (*AudioSpectrumProjection*) Zbiór funkcji, które wraz z funkcjami bazowymi dostarczają informacji o dynamicznym widmie sygnału.

➤ SILENCE – CISZA Deskryptor opisujący fragment nie zawierający żadnych istotnych dźwięków. Jest użyteczny przy segmentacji dźwięku oraz jako znacznik: „nie przetwarzaj tego fragmentu".

Deskryptory niskiego poziomu (LLD) mogą być wyekstraktowane (próbkowane) z pewnego ciągu jednakowych, regularnych, przedziałów bądź z dowolnego segmentu audio. Stąd pojawia się następny podział: na deskryptory wektorowe (*AudioLLDVectorTYpe*)(np.widmo) bądź skalarne (*AudioLLDScalarTYpe*)(np. moc, częstotliwość podstawowa ).

W dalszej kolejności deskryptory powstałe na skutek próbkowania mogą, po zastosowaniu odpowiednich narzędzi, być doprowadzone do postaci skalowalnych ciągów (*scalable series*). Wtedy często mogą być reprezentowane jako wektory przechowujące różne dane zsumowane (scałkowane), takie jak: minimalne, maksymalne średnie, rozrzut wartości (ciągu deskryptorów).

Druga klasa deskryptorów powstaje po zastosowaniu operacji ekstrakcji do segmentu; przy czym przez segment rozumie się w MPEG-7, w sensie ogólnym, dowolny przedział czasowy, do którego mogą być zastosowane wymienione techniki. Ma on strukturę rekursywną w tym sensie, że sygnał audio może być w spoób hierarchiczny dekomponowany na mniejsze elementy (segmenty).

Przy wyliczaniu wartości obu typów deskryptorów przyjmuje się pewien standardowy czas próbkowania, zwany w MPEG-7 wielkością skoku (*hopSize*), który został przyjęty jako 10 $\mu s$. Autorzy standardu sugerują, aby tę wielkość (przedział) zachować przy definiowaniu nowych deskryptprów bądź jego wielkrotność całkowitą lub część powstałą z podzielenia 10 $\mu s$ przez liczbę całkowitą.

Przy tworzeniu deskryptorów widmowych zaleca się korzystać ze skali logarytmicznej [57] dla uzyskania zwartego opisu przy jednoczesnym bliskiemu człowiekowi zdolności i czułości reakcji jego ucha właśnie w skali logarytmicznej.

## Propozycja nowych deskryptorów widmowych

Celem, jaki przyświecał autorowi niniejszej rozprawy, było dążenie do utworzenia mechanizmów pozwalających na rozpoznanie źródła dźwięku ze szczególnym uwzględnienie instrumentów strunowych, które nie są wyczerpująco opisane w pracach naukowych. Postacie czasowe przebiegów instrumentów strunowych (chordofonów) charakteryzują się brakiem stanu quasi-ustalonego oraz bardzo krótkim transjentem początkowym. Oznacza to, że proces parametryzacji może odbywać się tylko z wykorzystaniem transjentu końcowego.

Bardzo istotnym powodem, dla których wszczęto poszukiwania nowych deskryptorów treści multimedialnych, jest aspekt praktyczny. Algorytmy przeszukiwania zasobów danych multimedialnych nie powinnie mieć dużej złożoności obliczeniowej. Jednoczesnie artykulacja piccicato - znacznie różniąca się od typowgo pociągnięcia smyczkiem - wymusza poszukiwanie nowych, nietypowych rozwiązań.

Przedstawione w rozdziale 9 zestawy deskryptorów skalarnych był oganiczany pod kątem ich liczności i zdolności klasyfikacji wybranych 8 instrumentów strunowych. W końcowym etapie zmniejszono ich liczność by osiągnąć zbiory kilkunastoelementowe.

Najlepszy wynik klasyfikacyjny uzyskano przy liczbie 12, por. str. 129. Wektor cech składał się z jedengo deskryptora grupy podstawowej *LLD*, a mianowicie tristimulusa Tr1 oraz 11 - w pewnym sensie - nowych deskryptorów bazowych, które są wyliczoną energią zgromadzoną bądź w poszczególnych kolumnach, warstwach bądź komórkach siatki powstałej z odpowiednich podziałów wycinka płaszczyzny widma (tj. płaszczyzny wykresu amplitudy w funkcji częstotliwości). Sposób podziału, na odpowiednie kolumny, warstwy i w konsekewencji na komórki (wycinki pewnych

warstw i kolumn, został przedstawiony w rozdziałach 8 i 9. Przepis na wyliczenie tej energii umieszczono jawnie w Dodatku zależnością (3.1).

Podsumowując, wspomniany 12-elementowy wektor cech, składający się z 12 atrybutów czysto widmowych, zawiera (por. str.129 rozprawy):

1. 1 atrybut z grupy deskryptorów opisanych w rozdziale 7 oraz 4 (równanie (4.17)) : Tr1 ;

2. 2 atrybuty z 8-kolumnowego podział częstotliwościowy: energię $W$ z kolumn 1 i 3;

3. 3 atrybuty z podziału na 40 warstw o równej szerokości: energię $W$ warstw 12, 21 i 37;

4. 6 atrybutów wynikających z rozkładu wycinka płaszczyzny widma na 4 warstwy o jednakowej[1] energii i na 8 jednakowych kolumn. Z powstałej - z wykorzystaniem tej metody - siatki wybrano 6 komórek o współrzędnych, według następujacego indeksowania [warstwa, kolumna]: [1,7], [3,1], [3,5], [3,7], [3,8], [4,1] i dla nich wyliczono skumulowana energię $W$, wszysko zgodnie z zależnością (3.1) w Dodatku.

W tekście rozprawy pomyłkowo wpisano cyfrę 7 zamiast 6 w ostatnim nawiasie, wyliczając liczbę atrybutów w punkcie 4.

[1] Równość energetyczna jest to rozumiana w sensie średnim, tj. na zbiorze trenującym klasyfikowanych próbek dźwięków.

# Literatura


[1] Shigeo Ando, Kiminori Yamaguchi, *Statistical study of spectral parameters in musical instrument tones*, J. Acoust. Soc. Am. **94** (1), 37–45, 1993.

[2] K. Blair Benson, *Audio Engineering Handbook*, McGraw Hill, 1988.

[3] Chen T., *Construction and frequency characteristics of Chinese bowed string instrument*, Proc. 15th Intern. Congress on Acoustics, Trondheim, Norway 1995, vol. III, pp. 401–404, 1995.

[4] J. Bonada, A. Loscos, P. Cano, X. Serra, *Spectral approach to the modeling of the singing voice* Presented at the 111th Convention 2001 September 21–24 New York, NY, USA, 2001.

[5] J.M. Martínez, *MPEG-7 Overview*, Klangenfurt, July 2002, ISO/IECJTC1/SC29/WG11, **N4980**, pp. 1–96.

[6] X. Serra, X. Amatriain, J. Bonada, A. Loscos, *Spectral Modeling for Higher-level Sound Transformations*, Music Technology Group, Pompeu Fabra University 2001, także w: Proceedings of the first MOSART Workshop on Current Research Directions in Computer Music, November 15-16-17, 2001, Barcelona, Spain, 2001, CD ROM.

[7] A. Horner, J. Beauchamp, *Synthesis of trumpet tones using a wavetable and a dynamic filter*, Audio Eng. Soc., **43** (10), 799–812, 1995.

[8] B. Kostek, A. Wieczorkowska, *Study of parameter relations in musical instrument patterns*, 100th AES Convention, Copenhagen, 1996, preprint 4173, J. Audio Eng. Soc. (Abstracts), **44**, No 7/8, p.634, 1996.

[9] B. Kostek, A. Wieczorkowska, *Parametric representation of musical sounds*, Archives of Acoustic, **22** (1), 3–26, 1997.

[10] J. Krimphoff, S. McAdams, S. Winsberg *Caracteristation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique*, Journal de physique IV, Colloque C5, J. de Physique III, 4, 3eme Congres Francais d'Acoustique, I, pp. 625–628, 1994.

[11] K.D. Martin, Y.E. Kim, *2pMU9. Musical instrument identyfication: A pattern-recognition approach*, Internet: `ftp://sound.media.mit.edu/pub/Papers/kdm-asa98.pdf`, presented at the 136th Meeting of the Acoustical Society of America, Norfolk, VA ,October 13, 1998.

[12] M. Paraskevas, J. Mourjopoulos, *A statistical study of the variability and features of audio signals: Some preliminary results*, 100th AES Convention, preprint 4256, Copenhagen 1996.

[13] P. Toiviainen, *Optimizing self-organizing timbre maps: Two approaches*, Joint International Conference 1996, College of Europe at Brugge, Belgium, 8–11 September 1996, II Int. Conf on Cognitive Musicology, pp. 264–271.

[14] Z. Żyszkowski, *Podstawy akustyki*, Wydawnictwo Naukowo-Techniczne, Warszawa 1987.

[15] B. S.Manjunath, P. Salembier, T. Sikora, (Eds.) *Introduction to MPEG–7. Multimedia Content Description Interface*, John Wiley & Sons, Chichester, 2002.

[16] A. Wieczorkowska, *Skuteczność rozpoznawania dźwięków instrumentów muzycznych w zależności od sposobu parametryzacji i rodzaju klasyfikatora*, Praca doktorska, Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, 1999.

[17] A. Januszajtis, *Fizyka dla politechnik, tom III, Fale.*, Wydawnictwo Naukowe PWN, Warszawa 1991, ISBN 83-09708-6.

[18] H.F. Pollard, E.V. Jansson, *A tristimulus method for the specification of musical timbre*, Acustica, **51**, 162–171, 1982.

[19] *Popularna Encyklopedia Powszechna*, tom 7, Fogra Oficyna Wydawnicza, Kraków 1995.

[20] K. Tyburek, *Rozpoznawanie zależności dźwięku instrumentów szarpanych*, IV Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim", Materiały konferencyjne, 285–289, Oprogramowanie Naukowo-Techniczne, ISBN 83-916420-1-1. Kraków 26–28 listopad 2003.

[21] T. Zieliński, *Od teorii do cyfrowego przetwarzania sygnałów*, Wydział EAIiE AGH Kraków 2002, ISBN 83-88309-55-2.

[22] C. Marven, G. Ewers, *Zarys cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 1999, ISBN 83-206-1306-X.

[23] R.G. Lyons, *Wprowadzenie do cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 2003, ISBN 83-206-1318-3.

[24] A. Czyżewski, *Dźwięk cyfrowy. Wybrane zagadnienia teoretyczne, technologia, zastosowania*, EXIT Warszawa 1998, ISBN 83-87674-08-7.

[25] C.J. Date, *Wprowadzenie do systemów baz danych*, WNT, Warszawa 2000 wyd. II.

[26] J.D. Ullman, J. Widom *Podstawowy wykład z systemów baz danych*, WNT Warszawa 1999 wyd. I.

[27] P. Beynon-Davies, *Systemy baz danych*, WNT, Warszawa 2000 wyd. II.

[28] M. Lentner, *Oracle 9i. Kompletny podręcznik użytkownika*, Wydawnictwo PJWSTK, Warszawa 2003.

[29] Data Base Systems, *Courant Computer Science Symposia Series 6*, Prentice-Hall, N.J., 33–64, 1972.

[30] *Further Normalization of the Data Base Relational Model in Data Base systems, courant computer science symposia series 6*, Englewood Cliffs, N.J. Prentice-Hall, 1972.

[31] K.A.Ross, C.R.B. Wright, *Matematyka dyskretna*, Wydawnictwo Naukowe PWN, Warszawa 1999.

[32] R.Elmasri, S.B. Navathe *Wprowadzenie do systemów baz danych*, Helion, Warszawa, 2005.

[33] Peter Pin-Shan Chen, *The Entity-Relationship Model — Toward a United View of Data*, ACM TODS 1, No.1, March 1976.

[34] A. Jaszkiewicz, *Inżynieria oprogramowania*, Helion, Warszawa 1997.

[35] J.L. Harrington, *Obiektowe bazy danych dla każdego*, Mikom, Warszawa 2001.

[36] W. Kim, *Wprowadzenie do obiektowych baz danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.

[37] T.W. Leung, G. Mitchell, B. Subramanian, B. Vence, S.L. Vandenberg, S.B. Zdonik, *The Aqua Data Model and Algebra*, Technical Report No. CS-93-09, March 1993.

[38] K. Subieta, *Słownik terminów z zakresu obiektowości*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[39] K. Subieta, J. Leszczyłowski, *A Critique of Object Algebras*,Institute of Computer Science, Polish Acad. Sci., Warszawa, Poland, 1995 (także: `http://www.ipipan.waw.pl/~subieta/artykuly/CritiqObjAlg.html`, wrzesień 2005)

[40] P. Józwik, M. Mazur, *Obiektowe bazy danych — przegląd i analiza rozwiązań*, Praca dyplomowa AGH, Kraków 2002.

[41] K. Stąpor, *Automatyczna klasyfikacja obiektów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005.

[42] W. Greblicki, *Asymptotycznie optymalne algorytmy rozpoznawania i identyfikacji w warunkach probabilistycznych*, prace ICT Politechniki Wrocławskiej, Nr 18, seria Monografie, Nr 3, Wrocław 1974.

[43] M. Kurzyński, *Rozpoznawanie obiektów. Metody statystyczne*, Oficyna Wydawnicza Politechniki Wrocławskiej. Wrocław 1997.

[44] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification and Scene Analysis*, John Wiley&Sons, New York 2000.

[45] P. Cichosz, *Systemy uczące się*, WNT, Warszawa 2000.

[46] A. Dominik *Analiza danych z zastosowaniem teorii zbiorów przybliżonych*, Praca dyplomowa magisterska, Politechnika Warszawska 2004.

[47] W. Siedlecki, J. Sklansky, *On automatic feature selection*, Int. J Pattern Recognition and Artificial Intelligence, **2** (2), 197–220, 1988.

[48] D. Goldberg, *Algorytmy genetyczne i ich zastosowania*, Wydawnictwa Naukowo-Techniczne, Warszawa 1995.

[49] T. Strąkowski, *Analiza danych medycznych z zastosowaniem metod zbiorów przybliżonych*, Praca magisterska, Politechnika Warszawska — Wydział Elektroniki i Technik Informacyjnych, Instytut Informatyki. Warszawa 2003.

[50] A.Mrózek, L. Płonka, *Analiza danych metodą zbiorów przybliżonych. Zastosowania w ekonomii, medycynie i sterowaniu*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[51] Z. Pawlak, *Rough Set.Teoretical Aspects of Reasoning About Data*, Wydawnictwo Politechniki Warszawskiej, Warszawa 1990.

[52] Z. Pawlak, *Systemy informacyjne — Podstawy teoretyczne*, Wydawnictwo Naukowo-Techniczne, Warszawa 1983.

[53] K. Tyburek, W. Cudny, W. Kosiński, *Analiza rozkładu częstotliwościowego dźwięków pizzicato*, referat na INTERPOR Conference, Lubostroń k. Bydgoszczy 2006.

[54] K. Tyburek, W. Cudny, W. Kosiński, *Pizzicato sound analysis of selected instruments in the frequency domain*, Image Processing & Communications, **11**(1), 53–57, 2006.

[55] J. Swacha, M. Bandosz, Ł. Radliński, *Zaawansowane multimedia na stronach www*, III Krajowa Konferencja „Multimedialne i Sieciowe Systemy Informacyjne" MISSI, Kliczków, 2002.

[56] M. Wojciechowski, Ł. Matuszczak, *Oracle interMedia na tle standardu SQL/MM i prototypowych systemów multimedialnych baz danych*, IX Konferencja PLOUG Kościelisko, Październik 2003.

[57] Adam T. Lindsay, Ian Burnett, Schuyler Quackenbush, Melanie Jackson, *Fundamentals of audio descriptions*, in [15], pp. 283–298.

[58] Michael A. Casey, *Sound classification and similarity*, in [15], pp. 317–331.

[59] J. Massalski, M. Massalska, *Fizyka dla inżynierów*, Tom 1, Fizyka klasyczna, Wydawnictwo Naukowo-Techniczne, Warszawa 1980.

[60] Andrzej Chodkowski(Red.), *Encyklopedia Muzyki*, Wydawnictwo Naukowe PWN, Warszawa 2001.

[61] Xavier Amatriain, Jordi Bonada, Alex Loscos, Xavier Serra, *Spectral Modeling for Higher-level Sond Transformations Music Technology Group*, Pompeu Fabra University xavier.amatriain, jordi.bonada, alex.loscos, xavier.serra@iua.upf.es, `http://www.iua.upf.es/mtg`.

[62] I. Kaminskiyj, *Automatic Recognition of Musical Instruments Using Isolated Monophonic Sounds*. Ph. D Thesis. Department of Electrical and Computer Systems Engineering. Monash University. February 2004.

[63] Kostek, B. and Czyżewski, A., *Representing Musical Instrument Sounds for their Automatic Classification*. Journal Audio Engineering Society, 49(9), 768–785, 2001.

[64] Agostini, G., Longari, M., et al., *Musical Instrument Timbres Classification with Spectral Features*, Eurasip J Appl Signal Process, 2003(1), 5–14, 2003.

[65] A. Wieczorkowska, J. Wróblewski, D. Ślęzak and P. Synak, *Problems with Automatic Classification of Musical Sounds*. Proc. of the Intelligent Information Processing and Web Mining Conference IIS: IIPWM'2003, Zakopane, Poland, Advances in Soft Computing, Springer, Berlin, Heidelberg , New York, pp. 423 - 430, 2003.

[66] R. Tadeusiewicz, *Sygnał mowy*. WKŁ, Warszawa 1988.

[67] B. Kostek and A. Wieczorkowska, *Parametric Representation of Musical Sounds*. Archives of Acoustics, **22**, 1, 1997, pp. 3-26.

[68] Christian Simmermacher, Da Deng and Stephen Cranefield, *Feature Analysis and Classification of Classical Musical Instruments*. An Empirical Study Department of Information Science, University of Otago, New Zealand (ddeng,scranefield)@infoscience.otago.ac.nz, luty 2007.

[69] M. Szczerba and A Czyżewski, *Pitch Detection Enhancement Employing Music Prediction Journal of Intelligent Information Systems*, 24:2/3, 223–251, 2005. Springer Science + Business Media, Inc. Manufactured in The Netherlands.

[70] B. Kostek, *"Computing with words" Concept Applied to Musical Information Retrieval*. Electronic Notes in Theoretical Computer Science, vol. 82, No. 4, 2003.

[71] E. Łukasik, *Multimedialna baza dźwięków skrzypiec – AMATI*. Dostępne http://www.zsi.pwr.wroc.pl/zsi/missi2002/pdf/s206.pdf , luty 2007.

Instytut Podstawowych Problemów Techniki
Polskiej Akademii Nauk

# Dodatek

do rozprawy doktorskiej

**Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato**

mgr Krzysztof Tyburek

Warszawa, Luty 2007

# Spis treści

# Wprowadzenie

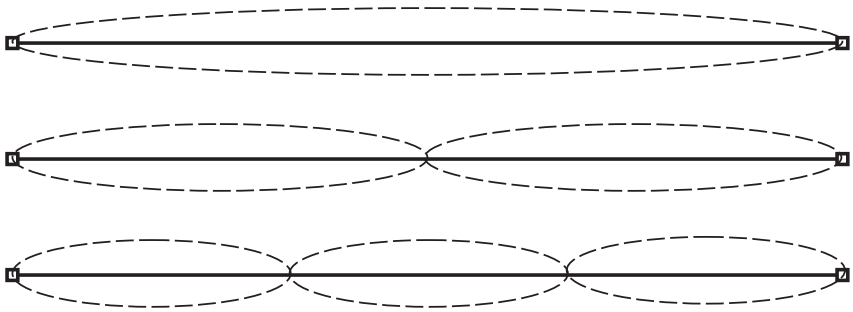
Przestawiony Dodatek powstał dla wypełnienia wymagań postawionych w pierwszym etapie recenzowania mojej rozprawy doktorskiej Klasyfikacja instrumentów strunowych w multimedialnych bazach danych ze szczególnym uwzględnieniem artykulacji pizzicato przez Recenzentów. Niektóre z wymagań były wspólne, inne były stawiane tylko przez jednego z Recenzentów. Zamieszczona Literatura zawiera kilka nowych pozycji umieszczonych na końcu spisu.

ROZDZIAŁ 1

# Fale i ruch falowy. Dodatek

## Drganie struny

Struną nazywamy ciało wydłużone, wiotkie - tzn. nie wykazujące oporu przy zginaniu (wykonane najczęściej z metalu). Szarpnięcie napiętej struny powoduje zaburzenia (drgania poprzeczne) rozchodzące się wzdłuż struny. W związku z tym, że struna na końcach jest umocowana, to fala odbita od ośrodka gęstszego (umocowanie struny) daje po nałożeniu na falę padającą fale stojącą - co zilustrowano poniżej.



Rysunek 1.1. Fale stojące wzdłuż struny.

Na obu końcach struny powstają nieruchome węzły, a wszystkie powstałe punkty mają tę samą fazę drgań i wychylają się wszystkie równocześnie w jedna stronę (dla częstotliwości podstawowej $\nu_0$ struny). Dla wyższych tonów harmonicznych obok węzłów na końcu struny powstają wzdłuż struny węzły, dzielące ją na równe części. Liczba powstających węzłów odpowiada liczbie wyższym harmonicznym o częstotliwości $2\nu_0, 3\nu_0, 4\nu_0, ...$ itd. Prędkość rozchodzenia się fali $v$ w strunie określa zależność:

$$v = \sqrt{\frac{p}{\rho}} \tag{1.1}$$

gdzie: $p = \frac{F}{\delta}$ jest napięciem struny, $F$ - siłą napinającą, $\delta$ -przekrojem struny, $\rho$ - gęstość materiału struny. Dla częstotliwości podstawowej $\nu_0$ długość fali $\lambda_0$ jest równa podwójnej długości $l$ struny (co ilustruje Rys.1.1.). Oznacza to, że $\lambda_0 = 2l$ oraz $\nu_0 = \frac{v}{\lambda_0}$ ($v$ - prędkość przesuwania się zaburzenia wzdłuż struny). Częstotliwość podstawowa struny wyraża się zatem zależnością:

$$\nu_0 = \frac{1}{2l}\sqrt{\frac{F}{\delta\rho}} \tag{1.2}$$

Częstotliwości wyższych harmonicznych oblicza się wg zależności:

$$\nu_k = \frac{k+1}{2l}\sqrt{\frac{F}{\delta\rho}} = (k+1)\nu_0 \ , k = 1, 2, ... \tag{1.3}$$

$k$ - liczba węzłów występujących wzdłuż struny (nie licząc węzłów na końcach). Drgania podstawowe i wyższe harmoniczne tworzą tzw. *układ drgań własnych struny* [59].

## Rezonans

Wydobywanie słyszalnego dźwięku w instrumentach muzycznych, a w szczególności w instrumentach strunowych, nie byłoby możliwe bez zjawiska rezonansu (akustycznego), którym określa się zespól efektów występujących przy szybkim wzroście amlitudy drgań układu fizycznego, gdy częstość zewnętrznych drgań wymuszajacych jest zbliżona do częstości drgań własnych układu [14]. Jest to efekt przekazywania energi między układami i polega na tym, że jeśli mamy dwa układy: pudło i strunę (ogólnie elementy instrumentów), które mogą drgać, to jeśli istnieje między nimi połączenie umożliwiające propagację (rozchodzenie się) fali dźwiękowej, to drgania jednego elementu będą przekazywane innemu elementowi. O właściwym rezonansie mówimy jednak dopiero wtedy, gdy owo przekazywanie energii akustycznej osiąga największą efektywność.

Różne układy fizyczne – czytaj pudła rezenansowe instrumentów strunowych – mają różne zdolności do drgań (rezonansowych). Zdolności takie opisuje zazwyczaj tzw. krzywa rezonansowa, przedstawiająca zależność amplitudy drgań układu (tutaj pudła) od częstości drgań wymuszających (tutaj strun), jej maksimum winien wypadać przy wartości częstości drgań własnych. Charakter tej krzywej jest mocno związany z tłumieniem drgań w układzie – jego tarciem wewnętrznym: wzrost amplitudy drgań własnych układu jest tym szybszy im mniejsze są tłumienia drgań w układzie. Przykładowo, pudło skrzypiec jest złożone z kilku elemntów i to o różnych kształtach, z których każdy ma swoją charakterystykę rezonansową. Są te elementy połaczone klejem. Charkterystyka kleju też ma wpływa na przebieg tej krzywej. Dla odmiany harfa ma w zasadzie dwa pudła rezonansowe: tzw. właściwe i drugie - całą ramę. Wyznaczenie dla takiego złożonego układu krzywej rezonansowej nie jest proste. Tym bardziej, że w bardziej skomplikowanych sytuacjach, właśnie kiedy jest to układ złożony, krzywa rezonansowa może mieć kilka maksimów, odpowiadających różnym postaciom drgań w układzie.

Wydaje się, że dobre zrozumienie tego zjawiska i jego dobry opis jakościowy i ilościowy mogłby być pewną wskazówką na drodze poszukiwania odpowiednich deskryptorów zarejestrowanych dźwięków w artykulacji piccicato. Zapewne będzie to wspomagajacy krok naszych przyszłych badań i poszukiwań. W niniejszej rozprawie przyjęliśmy inny tok postępowania.

ROZDZIAŁ 2

# Charakterystyka wybranych instrumentów. Dodatek

## Chordofony.Dodatek

Chordofony Chordofony [gr. chordé = struna, phoné = dźwięk] - jest to grupa instrumentów, w których źródłem dźwięku są napięte struny. Instrumenty strunowe należą do najstarszych instrumentów muzycznych. Pierwszy znany wizerunek instrumentu strunowego pochodzi z malunków odkrytych we francuskich jaskiniach. Przedstawiają one mężczyznę grającego na jednostrunowym instrumencie za pomocą smyczka. Pierwotnie w celu wzmocnienia dźwięku pochodzącego z instrumentu strunowego używano ust, a następnie innych komór rezonansowych o naturalnym pochodzeniu. Pierwsze instrumenty strunowe przypominające współczesne ich odpowiedniki z grupy szarpanych powstały na Bliskim Wschodzie już w trzecim tysiącleciu p.n.e., skąd w wyniku ekspansji kulturowej dotarły także do Europy. Chordofony ze względu na sposób wzbudzania drgań dzielą się na:

1. Szarpane (np. palcem - harfa)
2. Uderzane (młoteczkiem, pałeczką, np. cymbały, fortepian)
3. Pocierane (smyczkowe, np. skrzypce, wiolonczela)
4. Dęte (pobudzane strumieniem powietrza, np. harfa eolska).

Kolejnym kryterium podziału jest konstrukcja instrumentu, ze względu na którą dzielimy chordofony na: 1. Łuki (np. łuk muzyczny) 2. Liry (np. lira klasyczna, chrotta) 3. Harfy i cytry, tj. chordofony bezszyjkowe (np. cytra, cymbały, fortepian) 4. Lutnie, tj. chordofony szyjkowe (np. lutnia, skrzypce, gitara).

Prototypem chordofonów był wywodzący się z myśliwskiego, łuk muzyczny z drgającą cięciwą jako źródłem dźwięku. Łuk ten zaopatrzony był w prymitywny rezonator. Szarpanie struny (cięciwy) palcami, uderzanie drewienkiem lub sztabką i pocieranie cięciwy łuku dało pierwsze podstawy do rozwoju grup chordofonów szarpanych, uderzanych i pocieranych. Na podstawie materiałów historycznych wiadomo, że już ok. 2000 roku przed Chrystusem starożytna Asyria i Babilonia znały takie instrumenty jak harfy czy liry. Ok. 1000 roku przed Chrystusem pojawiły się w Indiach chordofony smyczkowe, natomiast zastosowanie mechanizmu klawiszowego było dziełem średniowiecza.

Podczas gry na chordofonach wyróżniamy różne sposoby artykulacji dźwięku:

1. *Flażolety* - ton fletowy, ton harmoniczny dźwięku struny wydobyty przy stłumieniu tonu podstawowego i pozostałych tonów harmonicznych. Jego barwa przypomina delikatną barwę fletu. Struna pobudzona do drgań drga w sposób złożony, tzn. całą swoją długością i dzieląc się na odcinki. Drganie całą długością wytwarza ton podstawowy, natomiast drganie przy podziale na dwie, trzy i więcej części jest źródłem odpowiednio wyższych tonów harmonicznych. Flażolety naturalne wydobywa się ze struny pustej przez lekkie dotknięcie w 1/2, 1/3, 1/4, 1/5, 2/5 lub 1/6 długości struny. Są to odpowiednio flażolety oktawy, kwinty, kwarty, wielkiej tercji, wielkiej seksty i małej tercji. Nazwy tych flażoletów utworzono od interwałów, które uzyskałoby się ze struny przez normalne jej przyciśnięcie w danym punkcie. Realna wysokość flażoletu zależy od tego, który ton harmoniczny zostanie wydobyty - np. przy flażolecie kwinty dotknięcie struny w 1/3 długości powoduje powstanie trzeciego tonu harmonicznego, czyli tzw. duodecymy powyżej dźwięku struny pustej. Flażolety stosowane są przy grze na instrumentach smyczkowych, na gitarze i na harfie. Na pięciolinii oznacza się je nutami rombowymi lub kółkiem nad nutą.

2. *Smyczkowanie* - wydobywanie dźwięku z instrumentu za pomocą pociągania smyczkiem po strunie. Sposób użycia smyczka łączy się ściśle z artykulacją i może być różnorodny, dlatego wprowadzono szereg znaków i określeń, za pomocą których notuje się szczegółowo rodzaje smyczkowania. W zespołach i orkiestrach ustalenie jednakowego smyczkowania dla całej grupy wykonawców powierza się koncertmistrzowi. Istnieją dwa podstawowe sposoby pociągnięcia smyczkiem: a) Z góry do dołu (franc. *tiré*), czyli od karafułki do główki b) Z dołu do góry (franc. *Poussé*) - od główki do karafułki. Pierwszy sposób pozwala na silniejsze zaatakowanie dźwięku.

3. *Legato* - na pięciolinii zaznaczane łukiem nad grupą nut. Wykonuje się zawsze jednym pociągnięciem smyczka. Liczbę nut przypadających na to pociągnięcie określa się łukiem.

4. *Martelé, martellato* - wykonuje się pojedynczymi, krótkimi pociągnięciami smyczka. Na pięciolinii notowane znakiem (...).

5. *Sautillé, spiccato, saltato* - oznaczane na pięciolinii za pomocą kropek nad nutami. Wykonuje się środkową częścią smyczka - każdą nutę oddzielnym pociągnięciem, przy czym smyczek nie przylega do struny, lecz lekko podskakuje.

6. *Staccato* - wykonywane jednym, lecz przerywanym ruchem smyczka, oznaczane za pomocą kropek nad nutami połączonymi łukiem.

7. *Jeté, ricochet, gettato* - polega na wykonywaniu kilku dźwięków staccato jednym pociągnięciem sprężyście rzuconego na strunę smyczka, który odbija się kilkakrotnie.

8. *Flatter la corde* - miękkie i delikatne uderzenie smyczka w strunę, oznaczane za pomocą kropek nad nutami połączonymi łukiem.

9. *Con legno* - uderzanie strun drzewcem smyczka.

10. *Tremolo* - szybkie i krótkie zmienne pociągnięcie smyczkiem w celu wielokrotnego powtórzenia dźwięku.

11. *Flautando, flautato, sul tasto, sulla tastiera* - prowadzenie smyczka tuż nad strunnikiem (gryfem). Dźwięki wydobywane w ten sposób mają barwę matową, zbliżoną do dźwięków fletu.

12. *Sul ponticello, au chevalet* - prowadzenie smyczka przy podstawku, stosowane w celu osiągnięcia jasnej, metalicznej barwy.

13. *Pizzicato* - szarpnięcie struny (skrót *pizz.* szczypiąc, szarpiąc). W grze na instrumentach smyczkowych oznacza to wydobywanie dźwięku nie za pomocą smyczka, lecz szarpiąc strunę palcem - podobnie jak w instrumentach szarpanych, np. gitarze.

Pizzicata użył po raz pierwszy R. Keiser w operze Adonis (1697), później G. F. Händel (opery Agrippina 1709). N. Paganini wykonywał również pizzicato lewą dłonią przy jednoczesnym użyciu smyczka prowadzonego prawą ręką [60].

ROZDZIAŁ 3

# Przygotowanie danych eksperymentalnych. Dodatek

## Baza wybranych do analizy dźwięków

Do badań przeznaczono bazę dźwięków pochodzących z 4 oktaw:

1. wielkiej ($A$ 110 Hz)
2. małej ($a$ 220 Hz)
3. razkreślnej ($a^1$ 440 Hz)
4. dwukreślnej ($a^2$ 880 Hz).

Badane próbki dźwięków pochodziły z różnych baz dźwięków (por. punkt 7.1 rozprawy), co świadczy o tym, że autorami opisywanych próbek są różni (pochodzący z różnych części świata) muzycy. Na bazie zgromadzonych próbek wyselekcjonowano populację przeznaczoną do prowadzonych badań, uwzględniając dźwięki reprezentujące każdą z w/w oktaw. Łącznie do badań przeznaczono próbki:

- **Gitara akustyczna** (29 próbek) - zakres instrumentu $E - h^2$
  - Oktawa wielka $(A, B, G, Gis)$
  - Oktawa mała $(a, c, cis, d, f, fis, fis, h)$
  - Oktawa razkreślna $(a^1, c^1, cis^1, d^1, dis^1, dis^1, e^1, e^1, f^1, fis^1, g^1, gis^1, gis^1)$
  - Oktawa dwukreślna $(b^2, c^2, d^2, dis^2)$
- **Altówka** (27 próbek) - zakres instrumentu $c - e^3$
  - Oktawa wielka ( )
  - Oktawa mała $(a, b, e, fis, g, gis)$
  - Oktawa razkreślna $(a^1, c^1, cis^1, cis^1, d^1, d^1, e^1, f^1, f^1, fis^1, g^1, gis^1, h^1)$
  - Oktawa dwukreślna $(b^2, cis^2, d^2, e^2, f^2, g^2, gis^2, h^2)$
- **Gitara basowa** (28 próbek) - zakres instrumentu $D^1 - c^1$
  - Oktawa wielka $(A, B, E, F, Fis, G, Gis)$
  - Oktawa mała $(a, a, b, c, c, cis, cis, d, dis, dis, e, e, f, fis, fis, g, gis)$
  - Oktawa razkreślna $(c^1, d^1, dis^1, f^1)$

- Oktawa dwukreślna ( )

**Gitara elektryczna** ( 30 próbek) - zakres instrumentu $E - h^2$

- Oktawa wielka $(A, Fis, H)$
- Oktawa mała $(a, b, cis, e, f, fis, g, gis)$
- Oktawa razkreślna $(a^1, b^1, c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, g^1, gis^1)$
- Oktawa dwukreślna $(a^2, cis^2, d^2, e^2, f^2, fis^2, g^2, h^2)$

- **Harfa** (29 próbek) - zakres instrumentu $Ces^1 - ges^4$
  - Oktawa wielka ( )
  - Oktawa mała $(b, b, h, gis, g, a, h)$
  - Oktawa razkreślna $(b^1, cis^1, dis^1, e^1, e^1, fis^1, gis^1, cis^1, a^1, gis^1, c^1)$
  - Oktawa dwukreślna $(a^2, b^2, cis^2, d^2, dis^2, e^2, fis^2, g^2, c^2, gis^2)$
- **Kontrabas** ( 30 próbek) - zakres instrumentu $D - c^1$
  - Oktawa wielka $(A, B, C, D, D, Dis, E, E, F, F, Fis, Fis, Gis, Gis)$
  - Oktawa mała $(a, b, c, cis, d, dis, e, fis)$
  - Oktawa razkreślna $(c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, h^1)$
  - Oktawa dwukreślna ( )
- **Skrzypce** ( 30 próbek) - zakres instrumentu $g - c^4$
  - Oktawa wielka ( )
  - Oktawa mała $(g, a, c)$
  - Oktawa razkreślna $(a^1, a^1, b^1, c^1, cis^1, d^1, dis^1, e^1, f^1, fis^1, g^1, g^1, gis^1, gis^1)$
  - Oktawa dwukreślna $(a^2, b^2, b^2, c^2, cis^2, d^2, dis^2, e^2, f^2, fis^2, g^2, g^2, gis^2)$
- Wiolonczela ( 30 próbek) - zakres instrumentu $C - e^2$
  - Oktawa wielka $(C, A, F, Fis, G)$
  - Oktawa mała $(a, b, h, c, cis, d, dis, e, fis, g, gis, )$
  - Oktawa razkreślna $(a^1, b^1, c^1, d^1, dis^1, e^1, f^1, fis^1, g^1, gis^1)$
  - Oktawa dwukreślna $(c^2, cis^2, d^2, dis^2)$

Dobór poszczególnych próbek przeznaczonych do badań odbywał się losowo. Puste nawiasy przy niektórych oktawach oznaczają, że te dźwieki nie zostały wylosowane (lub instrument nie gra w niej).

## Zaproponowana metodologia badań. Dodatek

Już w trakcie początkowego etapu badań, analizując przebiegi nagrań, podejrzewano, że procesu ekstrakcji cech powinien być nakierowany głównie na parametryzację postaci widmowej przebiegu. Robiono też próby połączenia atrybutów wyprowadzonych z czystych przebiegów czasowych wzmocnionych atrybutami wynikającymi z zastosowania tranformat czasowo-częstotliwościowych. Próby te, częściowo

omawiane w rozprawie, nie dały wystarczajco dobrej skuteczności rozpoznania instrumentów. Podstawowe LLD deskryptory standardu MPEG-7, których wiekszość została przeanalizowana, są tutaj najlepszym przykładem

Naszym zdaniem w przestrzeni widmowej można znaleźć - ewentualnie dodatkowo zdefiniować - wystarczającą liczbę cech, które pozwolą na skuteczną parametryzację badanych klas instrumentów. Ostatecznie zdecydowano się dążyć do eliminacji deskryptorów opisujących postać czasową próbki i budowę wektora cech składającego się z deskryptorów widmowych.

Podstawą do realizacji tego pomysłu jest praktyczny aspekt związany z techniką gry na instrumentach przeznaczonych do badań. W trakcie analizy funkcji czasu dźwieku konieczne jest pobranie okna przebiegu w celu jego analizy. Powstaje zatem pytanie - jak długie powinno być okno czasowe przeznaczone do analizy? Jeżeli okno będzie zbyt krótkie, to może się okazać, że nie zawiera ono wystarczającej informacji, która może się przyczynić do skutecznej klasyfikacji instrumentów. Jeżeli natomiast poddamy analizie okno zbyt długie (np. 1/10 sekundy), to jest możliwe, że próbka dźwięku będzie krótsza niż automatycznie pobierane przez system komputerowy okno. Sytuacja taka jest szczególnie prawdopodobna w sytuacji, gdy zostanie poddany analizie materiał dźwiękowy generowany przez muzyka o wyższym stopniu wyszkolenia technicznego - szczególnie podczas gry staccato.

Jeżeli skupimy swoją uwagę tylko na postaci widmowej przebiegu, to w łatwy sposób można pominąć opisywany problem. Pobierając "stałe" okno czasowe (o długości 11025 próbek, tak jak założono w rozprawie - por rozdz. 7), tzn. fragment przebiegu, który został pobrany zawsze w tym samym czasie oraz zawiera tą samą ilość próbek: doprowadzamy do porównywania widma, takiego samego fragmentu przebiegu dla całej populacji badanych dźwięków. Ostatecznie do analizy widmowej zdecydowano się przeznaczyć okno czasowe, które zostało pobrane od momentu osiągnięcia maksymalnej wartości amplitudy. Długość okna - czyli moment zakończenie pobierania - jest zdeterminowane ustaleniem właściwej rozdzielczości widma, co opisano w (7.3.2) rozprawy. Jeżeli próbka dźwięku okaże się krótsza niż przyjęta długość okna (11025 próbek), to wówczas system pobiera fragment dźwięku rozpoczynający się od wartości $t_{max}$ i trwający do całkowitego wybrzmiewania nuty. Brakujące indeksy wektora, tzn. różnicę $r = ind_k - ind_{wybrz}$ gdzie: $ind_k$ - indeks ostatniej próbki badanego dźwięku (w przypadku prowadzonych badań przyjęto długość okna 11025 próbek), $ind_{wyburz}$ - indeks ostatniej próbki w trancjencie końcowym badanego dźwieku, uzupełniono zerami.

Wycięty fragment przebiegu postaci czasowej został poddany DFT, a jego widmo poddano analizie. Wykorzystując opisywaną metodologię doprowadzono do analizy widma zdeterminowanego zawsze tą samą rozdzielczością. Tak jak opisano w rozdz. 8 zdecydowano się skupić na rozkładzie zarówno częstotliwościowym jak i energetycznym badanego fragmentu widma. Kluczowym problemem był optymalny dobór szerokości warstw. Zdecydowano się zdeterminować go równym rozkładem energii w poszczególnych warstwach dla każdej z nut. Dla celów dalszych badań zdecydowano się określić szerokość warstw na podstawie wartości średnich wyliczonych na drodze analizy każdej nuty. Ostatecznie przyjęto następujące progi dla rozkładu energetycznego: 4 warstwowy podział fragmentu widma: 1.warstwa: 0 - 0.14; 2.warstwa 0.14 - 0.33; 3.warstwa 0.33 - 0.66; 4.warstwa 0.66 - 3.

Wykorzystując analizę fragmentu widma z wykorzystaniem metody siatki oraz metody warstw zdecydowano się badać ilość zgromadzonej energii w poszczególnych zakresach widma. Oznacza to, że określono szerokości warstw podziału energetycz-

nego widma, co opisano szerzej w punkcie 9.2 rozprawy. Na bazie podziału na wymienione warstwy oraz kolumny (również opisanych w rozdz. 9) stworzono siatkę. Zliczając energię skumulowaną w jej komórkach pozwoliło na zdefiniowanie kolejnych deskryptorów opisujących postać widmową badanego dźwięku. Deskryptory te reprezentują energię $W(p)$ w poszczególnych zakresach (warstwach, kolumnach, komórkach) zliczaną zgodnie z zależnością:

$$W(p) = \sum_{k=1}^{n_p} (A(f_k)_{max} - A(f_k))^2, \tag{3.1}$$

gdzie: $A(f_k)_{max}$ - maksymalna wartość amplitudy w danej warstwie[1] lub górna granica analizowanej warstwy, $A(f_k)$ - dolny próg (wartość amplitudy) analizowanej warstwy podziału energetycznego, $k$ - numer próbki, $n_p$ - liczba próbek w $p$-tej komórce, warstwie bądź kolumnie. Do zależności (3.1) będziemy się odwoływać w rozdziale 6 Dodatku mówiąc o nowych deskryptorach widmowych zaproponowanych w rozprawie.

Opisywana metodologia pozwoliła na zdefiniowanie kilku wektorów cech charakteryzujących się różnym stopniem ogólnej rozpoznawalności dla 8 klas instrumentów muzycznych. Ostatecznie najskuteczniejszy wektor cech wykazał zdolność 89,6% ogólnej rozpoznawalności. Szczegółową listę deskryptorów zawarto w punkcie 9.2 rozprawy.

[1] W przypadku kolumny wartości $A(f_k)_{max}$ i $A(f_k)$ wynikają z przebiegu funkcji amlitudy.

ROZDZIAŁ 4

# Parametryzacja dźwięków muzycznych. Dodatek

Tak jak wspomniano powyżej w rozdziale 4 rozprawy parametryzacja dźwięków muzycznych może odbywać się na drodze analizy sygnału zarówno w przestrzeni czasowej jak i widmowej. W licznych pracach naukowych z dziedziny *music information retrieval* (MIR), ich autorzy wprowadzają, uzasadniając celowość stosowania, swoje definicje deskryptorów stosowanych do automatycznej klasyfikacji instrumentów muzycznych.

Podstawą metodologiczną zaproponowanych deskryptorów są różne koncepcje analizy fragmentu przebiegu muzycznego w powiązaniu z ogólnie znanymi transformatami czasowo-częstotliwościowymi. Bardzo szeroko stosowanym deskryptorem opisującym postać widmową jest środek ciężkości widma. Deskryptor ten jest stosowany w procesie parametryzacji zarówno chordofonów jak i aerofonów. W swoich pracach skuteczność tego parametru podkreślają tacy autorzy jak X. Serra [61], I. Kaminskyj [62], B. Kostek i A. Czyżewski [63], G. Agostini [64], A. Wieczorkowska et all. [65]. Ponadto często stosowaną metodą analizy widma jest wyznaczenie grupy parametrów tristimulus. Parametry te pozwalają rozróżnić dźwięki w zależności od zawartości grup harmonicznych w widmie. Ponadto kształt widma można opisać za pomocą momentów widmowych $k$-tego rzędu. Problem ten został podjęty m.in. w [66]. Kolejnymi parametrami widma jest również zawartość składowych parzystych ($Ev$) i nieparzystych ($Od$) w widmie opisywana, m.in. w [67]. Należy również zaakcentować skuteczność deskryptora opisującego nieregularność widma $Ir$, wyrażoną zależnością:

$$Ir = \log \sum_{n=2}^{N-1} |20(\log A_n - \frac{1}{3}\log(A_{n+1}A_nA_{n-1}))| \tag{4.1}$$

gdzie: $A_n$ – amplituda $n$-tego prążka widma.

Parametr ten został również wyszczególniony przez w/w autorów m.in. w [65]. Powszechnie stosowanym parametrem do opisu postaci czasowej sygnału jest wartość skuteczna $RMS$ (*Root Mean Square Value*) – pierwiastek z wartości średniokwadratowej – wyrażony, w przypadku ciągłym, zależnością:

$$RMS = (\frac{1}{T}\int_0^T (x(t))^2 dt)^{1/2} \ . \tag{4.2}$$

Wykorzystanie do celów parametryzacji dźwięków muzycznych deskryptora $RMS$ zaakcentowano m.in. w pracy [68], w której poszukiwano wektora cech w celu klasy-

fikacji instrumentów muzycznych w pasażach solowych. Ponadto parametr ten został włączony do rozważań m.in. w pracy [62].

Kolejnym zagadnieniem związanym z analizą postaci widmowej jest metoda cepstralna rozumiana jako rezultat obliczania transformaty Fouriera widma sygnału w skali decybelowej. Istnieje zarówno zespolone cepstrum jak i rzeczywiste. Cepstrum rzeczywiste zdefiniowane jest jako odrotna transformata Fouriera z logarytmu modułu transformaty Fouriera samej funkcji

$$X(t') = F^{-1}(\ln|F(x(t))|) \tag{4.3}$$

Definicja ta wykorzystuje logarytm rzeczywisty, liczony jedynie na bazie widma amplitudowego. Metoda cepstralna została również zaakcentowana między innymi w pracach [62] i [69]. Autorzy w pracy [68] również wykorzystują deskryptory opisu dźwięków muzycznych MPEG–7 (standard ten został opisany w dalszej części niniejszego Dodatku), takie jak:

1. Harmonic centroid (HC)
2. Harmonic deviation (HD)
3. Harmonic spread (HS)
4. Harmonic variation (HV)
5. Log-attack-time (LAT)
6. Temporal centroid (TC)
7. Spectral centroid (SC)

Podczas procesu parametryzacji dźwięków muzycznych poza optymalnym doborem deskryptorów istotne jest zastosowanie ich do właściwego fragmentu badanego przebiegu. Parametryzacja może odbywać się na podstawie analizy transjentu początkowego, transjentu końcowego oraz stanu quasi-ustalonego. Poza tym w zależności od badanego instrumentu, każdy z w/w fragmentów dźwięku może podlegać fragmentacji na $N$ ramek, które są analizowane niezależnie. Analizie może podlegać pełen zakres częstotliwości badanej ramki lub tylko pewien jego fragment. Np. w pracy [68] autorzy skupili się na analizie sygnału w zakresie częstotliwościowym 141 - 8877 Hz. A. Wieczorkowska w swoich pracach (m.in. w pozycji [16]) stosuje zwykle podobne podziały na podstawie których dokonywana jest analiza stanu quasi-ustalonego. Dzięki takiej metodologii łatwo można zbadać rozkład energii między niskimi, średnimi i wysokimi przedziałami częstotliwości.

Należy zwrócić uwagę, że między innymi rozkład energii w poszczególnych zakresach częstotliwościowych stanowił jeden z aspektów pracy badawczej podjętej w niniejszej rozprawie. Szczegółowy opis zamieszczono rozdziale 8. Tak jak wspomniano wcześniej grupa wymienionych deskryptorów stosowana jest do opisu poszczególnych fragmentów przebiegu. Wymienieni wyżej autorzy, podejmujący próbę parametryzacji dźwięków instrumentów muzycznych, analizują najczęściej przebiegi, które posiadają stan quasi-ustalony – np. [62, 68]. Przebiegi te charakteryzują się często stosunkowo długim czasem trwania (ok. 15 sekund), a analiza stanu quasi-ustalonego pozwala zdefiniować taki wektor cech, który dostarcza wysoki procent rozpoznawalności badanych klas instrumentów. Na przykład analizując stan quasi-ustalony w łatwy sposób można zbadać obecność wibrata w analizowanej próbce dźwięku. Np.

analiza obecności wibrata została poruszona w pracach [65, 70]. Wielkość wibrata [Hz] rozumiana jest jako wartość bezwzględna różnicy między wysokością dźwięku przy maksymalnej oraz minimalnej amplitudzie w stanie quasi–ustalonym. Łatwo wywnioskować, że deskryptor opisujący obecność wibrata nie jest skuteczny w kontekście parametryzacji dźwięków instrumentów strunowych z artykulacją pizzicato, które podjęto podczas realizacji niniejszej rozprawy. Związane jest to z brakiem stanu quasi-ustalonego w dźwiękach pizzicato instrumentów strunowych.

Szerzej na temat fizycznych cech badanych próbek napisano w rozdziale 7. W pracy [68] autorzy badali 20 klas instrumentów muzycznych (zarówno chordofony jak i aerofony) uzyskując ogólną rozpoznawalnośc ponad 90% - analizowano głównie stan quasi- ustalony. Wyniki ich badań dowodzą, że najlepiej rozpoznawalnymi instrumentami są trąbka, flet, skrzypce i fortepian (wykorzystano pakiet WEKA, $k$-NN, metoda holdout 66:34). Ponadto autorzy stwierdzili, że najwięcej pomyłek w rozpoznawaniu klas instrumentów zanotowano w parze trąbką i pianino.

Najlepsze wyniki w [68] jej autorzy uzyskali w trakcie analizy próbek skrzypiec i fletu. Należy jednak zaakcentować fakt, że badane instrumenty muzyczne pochodziły z różnych grup instrumentów. Wydaje się być oczywiste i intuicyjne, że rozpoznawalność trąbki i skrzypiec i ich odróżnienie będzie zdecydowanie wyższa niż między skrzypcami i altówka. Trudnością, z jaką spotykamy się przy rozpoznawaniu tej drugiej pary instrumentów, może być fakt, że altówka pochodzi z rodziny skrzypiec (altówka jest określana jako skrzypce altowe) a co za tym idzie charakterystyka tych instrumentów jest bardzo zbliżona – tym bardziej, jeżeli uwzględnimy taką samą artykulację dźwięku. Nie można natomiast tego samego powiedzieć porównując np. skrzypce i trąbkę. W trakcie realizacji badań opisanych w niniejszej rozprawie skupiono się tylko na parametryzacji dźwięków, których źródło stanowią chordofony z artykulacją pizzicato. Oznacza to, że zdecydowano się zaproponować taki wektor cech, który bazuje na analizie transjentu końcowego oraz dostarcza zadowalający stopień rozpoznawalności.

Stwierdzono też, że ogólnie znane i stosowane (przez w/w, uznanych Autorów) deskryptory w połączeniu z zaproponowaną w rozprawie metodologią badań (por. rozdział 8) nie przynoszą zadowalających rezultatów. Doprowadziło to w efekcie do zaproponowania nowego wektora cech, który wykazuje ok. 90% skuteczności podczas klasyfikacji wybranych instrumentów strunowych.

ROZDZIAŁ 5

# Bazy danych i system zarządzania bazą danych. Dodatek

## Bazy multimedialne. Dodatek

Multimedialne bazy danych (MMBD), będące najczęściej elementem systemu rozproszonego umożliwiają zarządzanie danymi multimedialnymi. W przeciwieństwie do klasycznych baz danych (wykorzystujących relacyjny, obiektowy lub obiektowo-relacyjny model danych), MMBD nie przechowują informacji jako takich (np. nagrań dźwiękowych, fragmentów filmów lub grafiki), a jedynie informacje o danych (metadane), które stanowią podstawę jej funkcjonalności. Metadane powstają w wyniku procesu indeksowania zawartości multimedialnej. Wyszukiwanie informacji odbywa się na drodze analizy i porównań metadanych (np. rozkład harmonicznych w poszczególnych przestrzeniach częstotliwościowych, właściwe nasycenie RGB fragmentu grafiki) kwerendy wystosowanej do systemu w postaci danych multimedialnych oraz metadanych przechowywanych w MMBD. Wynik porównań (kwerendy) kierowany jest do rzeczywistego obiektu (np. fragmentu nagrania muzycznego), który może być przesłany do klienta w postaci strumienia danych. Rzeczywiste składowanie danych realizowane jest w następujących formach:

1. Wewnątrz struktur bazy danych
2. W plikach w systemie plików systemu operacyjnego
3. Na zewnętrznych serwerach przeznaczonych do składowania multimediów.

Ostatecznie można stwierdzić, że przeszukiwanie MMBD z wykorzystaniem kwerendy multimedialnej odbywa się w 3 fazach:

1. Parametryzacja szukanego fragmentu
2. Wyszukanie metadanych pasujących do zapytania
3. Informacja o wyniku wyszukiwania.

W MMBD wyróżnia się dwie metody indeksowania danych:

- Tradycyjna (etykietowanie danych) – uzależniona od słów kluczowych i opisie tekstowym danych multimedialnych. Metoda ta charakteryzuje się małymi możliwościami wyszukiwania.

- Metoda bazująca na opisie zawartości (context-based indexing) - parametry uzyskiwane na podstawie analizy zawartości, powinny tak opisywać dane, aby zagwarantować wysoką skuteczność procesu filtracji. Metoda ta znacznie zwiększa (w porównaniu do metody tradycyjnej) możliwości wyszukiwania danych.

W wyniku indeksowania danych uzyskuje się $n$-wymiarowy wektor cech, opisujący zawartość - a więc „punkt w przestrzeni $n$-wymiarowej". Wyszukiwanie podobieństwa danych sprowadza się do odszukania punktów znajdujących się możliwie najbliżej punktu odpowiadającego zapytaniu. Z praktycznego punktu widzenia, do wyszukiwania danych wykorzystuje się kombinację cech, które w połączeniu z algorytmami klasyfikującymi zwracają grupę podobnych obiektów. Rozróżnia się dwa sposoby korzystania z multimedialnej bazy danych:

- *Pull* („aktywne") - użytkownik systemu wysyła do bazy polecenie wyszukiwania w postaci kwerendy multimedialnej a serwer zwraca metadane lub pasujące dane.

- *Push* („pasywne") - użytkownik systemu określa w sposób opisowy pewne preferencje w kwerendzie (np. interesująca go tematyka filmu). System bazy danych wyszukuje i przesyła zawartość wg. zadanych preferencji (np. wszystkie filmy związane z tematyką II Wojny Światowej), bez konieczności ingerencji użytkownika.

Przykładową multimedialną bazą jest opracowana w Instytucie Informatyki Politechniki Poznańskiej baza dźwięków skrzypiec AMATI. W bazie tej zgromadzono dźwięki kilkudziesięciu instrumentów nadesłanych na Międzynarodowy Konkurs Lutniczy im. Henryka Wieniawskiego w Poznaniu, który odbył się na jesieni 2001 roku. Zebrany materiał dźwiękowy wraz z ocenami jurorów ma służyć badaniom w dziedzinach związanych z cyfrowym przetwarzaniem sygnałów muzycznych oraz słuchaniem maszynowym, np. automatyczną klasyfikacją barwy dźwięku skrzypiec, automatyczną oceną jakości ich dźwięków w zestawieniu z subiektywnymi ocenami jurorów. Dźwięki w bazie są w pełni identyfikowane na podstawie nazw plików, w których są przechowywane. Nazwa zawiera numer konkursowy skrzypiec, sposób wydobycia dźwięku (détaché, pizzicato), strunę, na której dźwięk był zagrany, kierunek ruchu smyczka oraz numer powtórzenia dźwięku.

Baza AMATI dostarcza parametry pozwalające na wstępne porównanie dźwięków tego samego i różnych instrumentów. Deskryptory te otrzymano głównie na bazie analizy amplitud harmonicznych sygnału. Amplitudy te obliczane są na podstawie widma o dużej rozdzielczości. Wśród tych parametrów znajdują się między innymi: średnia energia dźwięku, względne energie pierwszej harmonicznej, harmonicznych pasma średniego (sumy drugiej, trzeciej i czwartej harmonicznej), sumy wyższych harmonicznych (powyżej czwartej), parzystych i nieparzystych harmonicznych – w odniesieniu do energii wszystkich harmonicznych, jasność dźwięku (środek ciężkości widma) oraz trzy pierwsze współczynniki cepstralne [71].

ROZDZIAŁ 6

# Standard MPEG -7 Audio. Nowe deskryptory widmowe

Większość dotychczasowych rozwiązań związanych z wydobywaniem wiedzy bazuje na technice etykietowania przechowywanych informacji — do pewnego czasu technika ta dotyczyła również danych opisujących dźwięk. Cała procedura etykietowania jest dość pracochłonna i czasochłonna. Poza tym takie rozwiązanie nie zawsze daje rzetelny wynik — to znaczy wysyłane zapytanie nie zawsze jest zgodne z oczekiwaniami osoby (czy systemu) pytającej. Istnieje wielkie prawdopodobieństwo, że dwie zupełnie różne (binarnie) informacje dźwiękowe mogą się okazać tą samą sekwencją utworu muzycznego zagranego z różną dynamiką lub w pomieszczeniu o różnej akustyce. Kolejny problem, który występuje w procesie rozpoznawania sygnałów dźwiękowych, jest właściwa interpretacja źródła dźwięku. Rozpoznanie dźwięku pochodzącego na przykład z drgającej struny gitary może być bardzo trudne. Trudność ta najczęściej wynika z doskonałych procesorów muzycznych, za pomocą których z łatwością można "podrobić" oryginalny instrument. W obliczu pojawiających się problemów związanych z możliwością wyszukiwania informacji audio najistotniejszym zagadnieniem jest opracowanie stosownych algorytmów wyszukujących właściwy system kodowania informacji multimedialnych oraz klasyfikujący typ źródła dźwięku (na przykład instrumenty strunowe, dęte drewniane i tym podobne).

## Ogólna charakterystyka standardu MPEG-7

Drogą do rozwiązania problemu klasyfikacji i agregacji danych multimedialnych jest - posiadający certyfikat ISO - standard MPEG-7. Standard ten dostarcza szeregu podstawowych deskryptorów opisujących dźwięk. Na bazie standardu MPEG 7 stworzono nowe deskryptory rozpoznające pewne instrumenty muzyczne. MPEG-7 jest standardem definiującym język opisu zawartości obiektów multimedialnych (MM) (ang. *Multimedia Content Description Interface*). O ile poprzednie standardy (MPEG-1, MPEG-2 i MPEG-4) zajmowały się normowaniem przekazywania zawartości obiektów multimedialnych, to standard MPEG-7 pozwala na odpowiedni opis, indeksowanie, a następnie wyszukiwanie tej zawartości zgodnie z potrzebami użytkowników. Część systemowa standardu udostępnia narzędzia do opakowania tych opisów i do tworzenia postaci binarnej dla przesyłanego strumienia MPEG-7. Służy do tego język DDL z rodziny XML, który pozwala na definiowanie nowych typów deskryptorów, a także na rozszerzanie i przedefiniowanie typów już istniejących.

MPEG-7 obejmuje standardem [5]:

1. Pewne zbiory deskryptorów - (ang. *descriptor*) (D), tj. zbiory obiektów, które mogą reprezentować cechę obiektu multimedialnych (MM) zarówno w sensie składniowym jak i znaczeniowym (np. kolory, kształty, częstotliwości dźwięków), jak i cech na poziomie semantycznym (np. dane o zawodach sportowych, które zostały zarejestrowane przez kamery).

2. Pewne zbiory schematów deskryptorów (ang. *description scheme*) (DS), które stanowią zapis i znaczenie relacji między swoimi składowymi, tj. wybranymi deskryptorami, albo też mogą być rekurencyjnie schematami deskryptorów.

3. Jezyk definiowania deskryptorów i schematów deskryptorów (ang. *description definition language*)(DDL). Język DDL, będący rozszerzeniem języka XML Schema, umożliwia też rozszerzenia i modyfikacje istniejących schematów deskryptorów. Tym samym schematy deskryptorów i same deskryptory są dokumentami języka XML. Jest to część systemowa standardu, która udostępnia narzędzia do opakowania powyższych opisów (tj. deskryptorów i schematów deskryptorów) i do tworzenia postaci binarnej (por. BiM) dla przesyłanego strumienia MPEG-7. Język ten pozwala na definiowanie nowych typów deskryptorów, a także na rozszerzanie i przedefiniowanie typów już istniejących.

4. Metodę binarnego kodowania deskryptorów – BiM ( skrót pochodzi od ang. *Binary Format for MPEG-7 data*). Jest to metoda kodowania strumieniowego, spełniająca takie wymagania jak: efektywność kompresji, odporność na błędy oraz bezpośredni dostęp do swoich składowych.

Podstawowym celem standardu MPEG-7 jest stworzenie metod opisu, indeksacji i klasyfikacji obiektów (danych) multimedialnych. Dla tego celu konieczne jest stworzenie efektywnych metod wyszukiwania poszczególnych elementów czy cech danych multimedialnych. Stąd dochodzi sie do: a) metadanych: danych o danych, takich jak autor, producent, copyright; b) semantyki: obiekty, zdarzenia, ludzie; c) podziałów: regiony, segmenty; d) cech: tekstury, kolory, kontury.

Schemat opisu danego deskryptora składa się z trzech zasadniczych kroków - opisu schematu deskryptora w języku DDL, opisu struktury deskryptora w kodzie binarnym (BiM) oraz opisu semantyki deskryptora. Jeśli z powyższych trzech punktów nie wynika jednoznacznie sposób dekodowania deskryptora, to zwykle dodaje się jeszcze krok definiujący ściśle kod dekodera.

MPEG-7 wyróżnia pięć następujących kategorii schematów deskryptorów:

1. Elementy podstawowe (ang. *Basic elements*), a wśród nich:

   (a) Narzędzia tworzenia schematów (ang. *Schema tools*);
   (b) Podstawowe typy danych i struktury (ang. *Datatype & structures*);
   (c) Referencje i lokalizacje mediów (ang *Link & media localization*);
   (d) Bazowe schematy deskryptorów (ang. *Basic description schemes*) – używane przez inne schematy deskryptorów.

2. Elementy opisu i zarządzania zawartością (ang. *Content description & management*), a wśród nich:

   (a) Elementy opisu zawartości (ang *Content description*): Struktura zawartości (ang. *Structural aspects*) – cechy oparte na analizie sygnałów; znaczenie zawartości (ang. *Semantic aspects*) – informacje o zdarzeniach i obiektach.

3. Elementy zarządzania zawartością (ang. *Content management*) – mogą być dołączone również do komponentów danego obiektu MM: tworzenie i produkcja treści medialnej (ang. *Creation & production*) – twórcy, klasyfikatory, określenie odbiorców treści medialnej; typy mediów (ang. *Media*) – formaty kodowania MM, identyfikacja mediów; użycie mediów (ang. *Usage*) – prawa dostępu, jednostki uprawnione, informacje finansowe i publikacyjne.

## Podstawy standardu MPEG-7 Audio

Przechodząc do części audio warto zwrócić uwagę na znaczenie niniejszego standardu w przeszukiwaniu multimedialnych baz danych. MPEG-7 zawiera warstwę nośną niskiego poziomu narzędzi, które stosowane są ogólnie we wszelkich dźwiękach. Mechanizmy te ustalają podstawowy poziom kompatybilności wśród opisów audio i pozwalają tworzyć nowe aplikacje. Warstwa ta również integruje MPEG-7 z innymi częściami standardu. Na warstwie nośnej zbudowana jest też seria narzędzi wysokiego poziomu, które są dostosowywane do poszczególnych grup aplikacji.

Zadaniem tych grup jest przeszukiwanie i wyszukiwanie cech sygnałów. Usługa audio realizowana jest często z wykorzystaniem bazy danych skompresowanego MPEG-4 znajdującego się na jednym (lub więcej) serwerze medialnym oraz skojarzoną z nim asocjacyjną bazą metadanych MPEG-7 na serwerze kwerendowym. Dla każdego indeksowanego skompresowanego sygnału audio MPEG-4, baza metadanych MPEG-7 przechowuje pełną reprezentację melodii oraz mechanizm łączący metadane ze skojarzonym nośnikiem, na przykład adres URL. Baza danych może również przechowywać inne deskryptory opisujące, np. styl i gatunek muzyki oraz bazę dźwięków instrumentów w charakterystycznym pasażu. Wykorzystując opis sygnału za pośrednictwem deskryptorów czy schematów deskryptorów istnieje możliwość wysłania kwerendy w formie fonicznej. Kształt fali sygnału próbnego jest przenoszony na serwer zapytań (kwerend), gdzie przeprowadzany jest proces, w którym uzyskiwane są jego metadane, które są celem kwerendy. Serwer MPEG-7 wyszukuje zatem odpowiedników melodii z baz danych. Kilka najlepszych odpowiedników przenoszonych jest z powrotem do urządzenia.

W części audio tego standardu najogólniej deskryptory można podzielić na dwa podtypy:

1. deskryptory funkcji widma,
2. deskryptory funkcji czasu.

Podstawowy podział deskryptorów i schematów deskryptorów audio jest na dwie klasy:

- ogólne narzędzia niskiego poziomu (ang.*generic low–level tools*),
- narzędzia nakierowane na szczególną aplikację (ang. *application–specific tools*).

W konsekwencji tej klasyfikacji pliki dźwiękowe w standardzie MPEG-7 są opisywane za pomocą deskryptorów niskiego poziomu (*low level descriptors*)(LLD), które odnoszą się do właściwości dźwięku (czasowe, widmowe) oraz wysokiego poziomu (*high level descriptors*). Deskryptory wysokiego poziomu stanowią opis zawartości multimedialnej sporządzony pod kątem określonego zastosowania, np. opisujące barwę

dźwięku, melodię itp. Nie są to parametry uzyskiwane na podstawie analizy pliku dźwiękowego, ale np. sposób wykorzystania deskryptorów niskiego poziomu w celu umożliwienia wyszukiwania nagrań w multimedialnej bazie danych [15].

MPEG-7 dostarcza 17 podstawowych deskryptorów (niskiego poziomu)sklasyfikowanych w poszczególnych klasach:

- postawowe (ang. *Basic*),
- podstawowe widmowe (ang. *Basic Spectral*),
- parametry sygnałowe (ang. *Signal Parameters*),
- czasowy barwy dźwięku (ang. *Timbral Temporal*),
- widmowy barwy dźwięku (ang. *Timbral Spectral*),
- baza widmowa (ang. *Spectral Basis*),
- deskryptor ciszy (ang. *Silence Descriptor*).

Scharakteryzujmy pokrótce poszczególne klasy przez wyszczególnione ich składników [15, 57, 58].

- Basic – deskryptory podstawowe zawierają:
  - przebieg czasowy (*AudioWaveform*) Jest to obwiednia (górna i dolna) przebiegu czasowego spróbkowanego sygnału audio. Deskryptor ten jest używany głównie do wyświetlania przebiegu czasowego.
  - moc sygnału (*AudioPower*) Jest to wygładzony przebieg chwilowej mocy sygnału w funkcji czasu i jest używany razem z deskryptorami widmowymi.
- Basic spectral – deskryptory widmowe podstawowe zawierają:
  - obwiednię widma (*AudioSpectrumEnvelope*) Jest to obwiednia krótkoterminowego widma mocy sygnału.
  - logarytmiczną skalę częstotliwości
  - środek ciężkości widma (*AudioSpectrumCentroid*) Środek ciężkości widma mocy sygnału opisuje rozkład częstotliwościowy widma; odpowiada na pytanie czy w widmie sygnału dominują składowe nisko- czy wysokoczęstotliwościowe.
  - rozkład widma (*AudioSpectrumSpread*) Drugi moment widma mocy sygnału. Deskryptor opisuje, czy widmo sygnału jest skupione wokół środka ciężkości widma, czy też pokrywa szerszy zakres częstotliwości. Pomaga to np. odróżnić dźwięki szumowe od tonów.
  - płaskość widma (*AudioSpectrumFlatness*) Deskryptor ten opisuje płaskość widma w każdym z pasm częstotliwości. Płaskie widmo może oznaczać sygnał szumowy lub impulsowy, mała wartość parametru może wskazywać na sygnał harmoniczny.
- Signal parameters – parametry sygnału odnoszą się do sygnałów okresowych bądź quasi-okresowych i zawierają:

- częstotliwość podstawową (*AudioFundamentalFrequency*) Jest to częstotliwość podstawowa sygnału, uzyskiwana za pomocą algorytmu pitch-tracking.
- harmoniczność sygnału (*AudioHarmonicity*) Pozwala rozróżnić sygnały harmoniczne, nieharmoniczne oraz nie posiadające widma prążkowego (np. szum).

➤ TIMBRAL TEMPORAL – PARAMETRY CZASOWE BARWY DŹWIĘKU

- czas narastania dźwięku (*LogAttackTime*) Wartość skalarna opisująca logarytm (przy podstawie 10) czasu mierznego w sekundach narastania amplitudy dźwięku od ciszy do wartości maksymalnej. Pozwala opisać np. transjenty dźwięków muzycznych.
- środek ciężkości przebiegu czasowego (*TemporalCentroid*) Wartość skalarna opisująca punkt na osi czasu, w którym skupia się energia sygnału (środek ciężkości obwiedni czasowej sygnału). Pozwala rozróżnić np. dźwięki o takim samym czasie narastania, ale różnym czasie wybrzmiewania.

➤ TIMBRAL SPECTRAL – PARAMETRY WIDMOWE BARWY DŹWIĘKUzawierają:

- środek ciężkości widma (*SpectralCentroid*) Liniowa skala częstotliwości jest dzielona na zakresy, liczona jest średnia ważona częstotliwości środkowych wszystkich zakresów, ważona według mocy sygnału w każdym paśmie. Parametr ten jest skorelowany z subiektywnym wrażeniem ostrości dźwięku (*sharpness*).
- środek ciężkości harmonicznych (*HarmonicSpectralCentroid*) Średnia ważona częstotliwości prążków harmonicznych w widmie, ważona według amplitudy prążków. Dotyczy tylko składowych harmonicznych widma.
- odchyłkę harmonicznych (*HarmonicSpectralDeviation*) Różnica między amplitudami prążków harmonicznych a obwiednią widma .
- rozkład harmonicznych (*HarmonicSpectralSpread*) Rozkład harmonicznych w widmie względem środka ciężkości.
- zmienność harmonicznych (*HarmonicSpectralVariation*) Znormalizowana korelacja między amplitudami prążków harmonicznych w dwóch kolejnych ramkach sygnału. Opisuje zmienność widma harmonicznego w czasie.

➤ SPECTRAL BASIS – BAZA WIDMOWA to deskryptory służące do opisu dynamicznego 3D (trójwymiarowego) widma sygnału jako trójka: czas – częstotliwość – poziom i zawiera:

- funkcje bazowe widma (*AudioSpectrumBasis*) Zbiór funkcji bazowych, otrzymanych przez dekompozycję znormalizowanego widma mocy.
- funkcje przekształcające (*AudioSpectrumProjection*) Zbiór funkcji, które wraz z funkcjami bazowymi dostarczają informacji o dynamicznym widmie sygnału.

➤ SILENCE – CISZA Deskryptor opisujący fragment nie zawierający żadnych istotnych dźwięków. Jest użyteczny przy segmentacji dźwięku oraz jako znacznik: „nie przetwarzaj tego fragmentu".

Deskryptory niskiego poziomu (LLD) mogą być wyekstraktowane (próbkowane) z pewnego ciągu jednakowych, regularnych, przedziałów bądź z dowolnego segmentu audio. Stąd pojawia się następny podział: na deskryptory wektorowe (*AudioLLDVectorTYpe*)(np.widmo) bądź skalarne (*AudioLLDScalarTYpe*)(np. moc, częstotliwość podstawowa ).

W dalszej kolejności deskryptory powstałe na skutek próbkowania mogą, po zastosowaniu odpowiednich narzędzi, być doprowadzone do postaci skalowalnych ciągów (*scalable series*). Wtedy często mogą być reprezentowane jako wektory przechowujące różne dane zsumowane (scałkowane), takie jak: minimalne, maksymalne średnie, rozrzut wartości (ciągu deskryptorów).

Druga klasa deskryptorów powstaje po zastosowaniu operacji ekstrakcji do segmentu; przy czym przez segment rozumie się w MPEG-7, w sensie ogólnym, dowolny przedział czasowy, do którego mogą być zastosowane wymienione techniki. Ma on strukturę rekursywną w tym sensie, że sygnał audio może być w spoób hierarchiczny dekomponowany na mniejsze elementy (segmenty).

Przy wyliczaniu wartości obu typów deskryptorów przyjmuje się pewien standardowy czas próbkowania, zwany w MPEG-7 wielkością skoku (*hopSize*), który został przyjęty jako 10 $\mu s$. Autorzy standardu sugerują, aby tę wielkość (przedział) zachować przy definiowaniu nowych deskryptprów bądź jego wielkrotność całkowitą lub część powstałą z podzielenia 10 $\mu s$ przez liczbę całkowitą.

Przy tworzeniu deskryptorów widmowych zaleca się korzystać ze skali logarytmicznej [57] dla uzyskania zwartego opisu przy jednoczesnym bliskiemu człowiekowi zdolności i czułości reakcji jego ucha właśnie w skali logarytmicznej.

## Propozycja nowych deskryptorów widmowych

Celem, jaki przyświecał autorowi niniejszej rozprawy, było dążenie do utworzenia mechanizmów pozwalających na rozpoznanie źródła dźwięku ze szczególnym uwzględnienie instrumentów strunowych, które nie są wyczerpująco opisane w pracach naukowych. Postacie czasowe przebiegów instrumentów strunowych (chordofonów) charakteryzują się brakiem stanu quasi-ustalonego oraz bardzo krótkim transjentem początkowym. Oznacza to, że proces parametryzacji może odbywać się tylko z wykorzystaniem transjentu końcowego.

Bardzo istotnym powodem, dla których wszczęto poszukiwania nowych deskryptorów treści multimedialnych, jest aspekt praktyczny. Algorytmy przeszukiwania zasobów danych multimedialnych nie powinnie mieć dużej złożoności obliczeniowej. Jednoczesnie artykulacja piccicato - znacznie różniąca się od typowgo pociągnięcia smyczkiem - wymusza poszukiwanie nowych, nietypowych rozwiązań.

Przedstawione w rozdziale 9 zestawy deskryptorów skalarnych był oganiczany pod kątem ich liczności i zdolności klasyfikacji wybranych 8 instrumentów strunowych. W końcowym etapie zmniejszono ich liczność by osiągnąć zbiory kilkunastoelementowe.

Najlepszy wynik klasyfikacyjny uzyskano przy liczbie 12, por. str. 129. Wektor cech składał się z jedengo deskryptora grupy podstawowej *LLD*, a mianowicie tristimulusa Tr1 oraz 11 - w pewnym sensie - nowych deskryptorów bazowych, które są wyliczoną energią zgromadzoną bądź w poszczególnych kolumnach, warstwach bądź komórkach siatki powstałej z odpowiednich podziałów wycinka płaszczyzny widma (tj. płaszczyzny wykresu amplitudy w funkcji częstotliwości). Sposób podziału, na odpowiednie kolumny, warstwy i w konsekewencji na komórki (wycinki pewnych

warstw i kolumn, został przedstawiony w rozdziałach 8 i 9. Przepis na wyliczenie tej energii umieszczono jawnie w Dodatku zależnością (3.1).

Podsumowując, wspomniany 12-elementowy wektor cech, składający się z 12 atrybutów czysto widmowych, zawiera (por. str.129 rozprawy):

1. 1 atrybut z grupy deskryptorów opisanych w rozdziale 7 oraz 4 (równanie (4.17)) : Tr1 ;

2. 2 atrybuty z 8-kolumnowego podział częstotliwościowy: energię $W$ z kolumn 1 i 3;

3. 3 atrybuty z podziału na 40 warstw o równej szerokości: energię $W$ warstw 12, 21 i 37;

4. 6 atrybutów wynikających z rozkładu wycinka płaszczyzny widma na 4 warstwy o jednakowej[1] energii i na 8 jednakowych kolumn. Z powstałej - z wykorzystaniem tej metody - siatki wybrano 6 komórek o współrzędnych, według następujacego indeksowania [warstwa, kolumna]: [1,7], [3,1], [3,5], [3,7], [3,8], [4,1] i dla nich wyliczono skumulowana energię $W$, wszysko zgodnie z zależnością (3.1) w Dodatku.

W tekście rozprawy pomyłkowo wpisano cyfrę 7 zamiast 6 w ostatnim nawiasie, wyliczając liczbę atrybutów w punkcie 4.

[1] Równość energetyczna jest to rozumiana w sensie średnim, tj. na zbiorze trenującym klasyfikowanych próbek dźwięków.

# Literatura


[1] Shigeo Ando, Kiminori Yamaguchi, *Statistical study of spectral parameters in musical instrument tones*, J. Acoust. Soc. Am. **94** (1), 37–45, 1993.

[2] K. Blair Benson, *Audio Engineering Handbook*, McGraw Hill, 1988.

[3] Chen T., *Construction and frequency characteristics of Chinese bowed string instrument*, Proc. 15th Intern. Congress on Acoustics, Trondheim, Norway 1995, vol. III, pp. 401–404, 1995.

[4] J. Bonada, A. Loscos, P. Cano, X. Serra, *Spectral approach to the modeling of the singing voice* Presented at the 111th Convention 2001 September 21–24 New York, NY, USA, 2001.

[5] J.M. Martínez, *MPEG-7 Overview*, Klangenfurt, July 2002, ISO/IECJTC1/SC29/WG11, **N4980**, pp. 1–96.

[6] X. Serra, X. Amatriain, J. Bonada, A. Loscos, *Spectral Modeling for Higher-level Sound Transformations*, Music Technology Group, Pompeu Fabra University 2001, także w: Proceedings of the first MOSART Workshop on Current Research Directions in Computer Music, November 15-16-17, 2001, Barcelona, Spain, 2001, CD ROM.

[7] A. Horner, J. Beauchamp, *Synthesis of trumpet tones using a wavetable and a dynamic filter*, Audio Eng. Soc., **43** (10), 799–812, 1995.

[8] B. Kostek, A. Wieczorkowska, *Study of parameter relations in musical instrument patterns*, 100th AES Convention, Copenhagen, 1996, preprint 4173, J. Audio Eng. Soc. (Abstracts), **44**, No 7/8, p.634, 1996.

[9] B. Kostek, A. Wieczorkowska, *Parametric representation of musical sounds*, Archives of Acoustic, **22** (1), 3–26, 1997.

[10] J. Krimphoff, S. McAdams, S. Winsberg *Caracteristation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique*, Journal de physique IV, Colloque C5, J. de Physique III, 4, 3eme Congres Francais d'Acoustique, I, pp. 625–628, 1994.

[11] K.D. Martin, Y.E. Kim, *2pMU9. Musical instrument identyfication: A pattern-recognition approach*, Internet: `ftp://sound.media.mit.edu/pub/Papers/kdm-asa98.pdf`, presented at the 136th Meeting of the Acoustical Society of America, Norfolk, VA ,October 13, 1998.

[12] M. Paraskevas, J. Mourjopoulos, *A statistical study of the variability and features of audio signals: Some preliminary results*, 100th AES Convention, preprint 4256, Copenhagen 1996.

[13] P. Toiviainen, *Optimizing self-organizing timbre maps: Two approaches*, Joint International Conference 1996, College of Europe at Brugge, Belgium, 8–11 September 1996, II Int. Conf on Cognitive Musicology, pp. 264–271.

[14] Z. Żyszkowski, *Podstawy akustyki*, Wydawnictwo Naukowo-Techniczne, Warszawa 1987.

[15] B. S.Manjunath, P. Salembier, T. Sikora, (Eds.) *Introduction to MPEG–7. Multimedia Content Description Interface*, John Wiley & Sons, Chichester, 2002.

[16] A. Wieczorkowska, *Skuteczność rozpoznawania dźwięków instrumentów muzycznych w zależności od sposobu parametryzacji i rodzaju klasyfikatora*, Praca doktorska, Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, 1999.

[17] A. Januszajtis, *Fizyka dla politechnik, tom III, Fale.*, Wydawnictwo Naukowe PWN, Warszawa 1991, ISBN 83-09708-6.

[18] H.F. Pollard, E.V. Jansson, *A tristimulus method for the specification of musical timbre*, Acustica, **51**, 162–171, 1982.

[19] *Popularna Encyklopedia Powszechna*, tom 7, Fogra Oficyna Wydawnicza, Kraków 1995.

[20] K. Tyburek, *Rozpoznawanie zależności dźwięku instrumentów szarpanych*, IV Krajowa Konferencja „Metody i systemy komputerowe w badaniach naukowych i projektowaniu inżynierskim", Materiały konferencyjne, 285–289, Oprogramowanie Naukowo-Techniczne, ISBN 83-916420-1-1. Kraków 26–28 listopad 2003.

[21] T. Zieliński, *Od teorii do cyfrowego przetwarzania sygnałów*, Wydział EAIiE AGH Kraków 2002, ISBN 83-88309-55-2.

[22] C. Marven, G. Ewers, *Zarys cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 1999, ISBN 83-206-1306-X.

[23] R.G. Lyons, *Wprowadzenie do cyfrowego przetwarzania sygnałów*, WKŁ Warszawa 2003, ISBN 83-206-1318-3.

[24] A. Czyżewski, *Dźwięk cyfrowy. Wybrane zagadnienia teoretyczne, technologia, zastosowania*, EXIT Warszawa 1998, ISBN 83-87674-08-7.

[25] C.J. Date, *Wprowadzenie do systemów baz danych*, WNT, Warszawa 2000 wyd. II.

[26] J.D. Ullman, J. Widom *Podstawowy wykład z systemów baz danych*, WNT Warszawa 1999 wyd. I.

[27] P. Beynon-Davies, *Systemy baz danych*, WNT, Warszawa 2000 wyd. II.

[28] M. Lentner, *Oracle 9i. Kompletny podręcznik użytkownika*, Wydawnictwo PJWSTK, Warszawa 2003.

[29] Data Base Systems, *Courant Computer Science Symposia Series 6*, Prentice-Hall, N.J., 33–64, 1972.

[30] *Further Normalization of the Data Base Relational Model in Data Base systems, courant computer science symposia series 6*, Englewood Cliffs, N.J. Prentice-Hall, 1972.

[31] K.A.Ross, C.R.B. Wright, *Matematyka dyskretna*, Wydawnictwo Naukowe PWN, Warszawa 1999.

[32] R.Elmasri, S.B. Navathe *Wprowadzenie do systemów baz danych*, Helion, Warszawa, 2005.

[33] Peter Pin-Shan Chen, *The Entity-Relationship Model — Toward a United View of Data*, ACM TODS 1, No.1, March 1976.

[34] A. Jaszkiewicz, *Inżynieria oprogramowania*, Helion, Warszawa 1997.

[35] J.L. Harrington, *Obiektowe bazy danych dla każdego*, Mikom, Warszawa 2001.

[36] W. Kim, *Wprowadzenie do obiektowych baz danych*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.

[37] T.W. Leung, G. Mitchell, B. Subramanian, B. Vence, S.L. Vandenberg, S.B. Zdonik, *The Aqua Data Model and Algebra*, Technical Report No. CS-93-09, March 1993.

[38] K. Subieta, *Słownik terminów z zakresu obiektowości*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[39] K. Subieta, J. Leszczyłowski, *A Critique of Object Algebras*,Institute of Computer Science, Polish Acad. Sci., Warszawa, Poland, 1995 (także: `http://www.ipipan.waw.pl/~subieta/artykuly/CritiqObjAlg.html`, wrzesień 2005)

[40] P. Józwik, M. Mazur, *Obiektowe bazy danych — przegląd i analiza rozwiązań*, Praca dyplomowa AGH, Kraków 2002.

[41] K. Stąpor, *Automatyczna klasyfikacja obiektów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005.

[42] W. Greblicki, *Asymptotycznie optymalne algorytmy rozpoznawania i identyfikacji w warunkach probabilistycznych*, prace ICT Politechniki Wrocławskiej, Nr 18, seria Monografie, Nr 3, Wrocław 1974.

[43] M. Kurzyński, *Rozpoznawanie obiektów. Metody statystyczne*, Oficyna Wydawnicza Politechniki Wrocławskiej. Wrocław 1997.

[44] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification and Scene Analysis*, John Wiley&Sons, New York 2000.

[45] P. Cichosz, *Systemy uczące się*, WNT, Warszawa 2000.

[46] A. Dominik *Analiza danych z zastosowaniem teorii zbiorów przybliżonych*, Praca dyplomowa magisterska, Politechnika Warszawska 2004.

[47] W. Siedlecki, J. Sklansky, *On automatic feature selection*, Int. J Pattern Recognition and Artificial Intelligence, **2** (2), 197–220, 1988.

[48] D. Goldberg, *Algorytmy genetyczne i ich zastosowania*, Wydawnictwa Naukowo-Techniczne, Warszawa 1995.

[49] T. Strąkowski, *Analiza danych medycznych z zastosowaniem metod zbiorów przybliżonych*, Praca magisterska, Politechnika Warszawska — Wydział Elektroniki i Technik Informacyjnych, Instytut Informatyki. Warszawa 2003.

[50] A.Mrózek, L. Płonka, *Analiza danych metodą zbiorów przybliżonych. Zastosowania w ekonomii, medycynie i sterowaniu*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1999.

[51] Z. Pawlak, *Rough Set.Teoretical Aspects of Reasoning About Data*, Wydawnictwo Politechniki Warszawskiej, Warszawa 1990.

[52] Z. Pawlak, *Systemy informacyjne — Podstawy teoretyczne*, Wydawnictwo Naukowo-Techniczne, Warszawa 1983.

[53] K. Tyburek, W. Cudny, W. Kosiński, *Analiza rozkładu częstotliwościowego dźwięków pizzicato*, referat na INTERPOR Conference, Lubostroń k. Bydgoszczy 2006.

[54] K. Tyburek, W. Cudny, W. Kosiński, *Pizzicato sound analysis of selected instruments in the frequency domain*, Image Processing & Communications, **11**(1), 53–57, 2006.

[55] J. Swacha, M. Bandosz, Ł. Radliński, *Zaawansowane multimedia na stronach www*, III Krajowa Konferencja „Multimedialne i Sieciowe Systemy Informacyjne" MISSI, Kliczków, 2002.

[56] M. Wojciechowski, Ł. Matuszczak, *Oracle interMedia na tle standardu SQL/MM i prototypowych systemów multimedialnych baz danych*, IX Konferencja PLOUG Kościelisko, Październik 2003.

[57] Adam T. Lindsay, Ian Burnett, Schuyler Quackenbush, Melanie Jackson, *Fundamentals of audio descriptions*, in [15], pp. 283–298.

[58] Michael A. Casey, *Sound classification and similarity*, in [15], pp. 317–331.

[59] J. Massalski, M. Massalska, *Fizyka dla inżynierów*, Tom 1, Fizyka klasyczna, Wydawnictwo Naukowo-Techniczne, Warszawa 1980.

[60] Andrzej Chodkowski(Red.), *Encyklopedia Muzyki*, Wydawnictwo Naukowe PWN, Warszawa 2001.

[61] Xavier Amatriain, Jordi Bonada, Alex Loscos, Xavier Serra, *Spectral Modeling for Higher-level Sond Transformations Music Technology Group*, Pompeu Fabra University xavier.amatriain, jordi.bonada, alex.loscos, xavier.serra@iua.upf.es, `http://www.iua.upf.es/mtg`.

[62] I. Kaminskiyj, *Automatic Recognition of Musical Instruments Using Isolated Monophonic Sounds*. Ph. D Thesis. Department of Electrical and Computer Systems Engineering. Monash University. February 2004.

[63] Kostek, B. and Czyżewski, A., *Representing Musical Instrument Sounds for their Automatic Classification*. Journal Audio Engineering Society, 49(9), 768–785, 2001.

[64] Agostini, G., Longari, M., et al., *Musical Instrument Timbres Classification with Spectral Features*, Eurasip J Appl Signal Process, 2003(1), 5–14, 2003.

[65] A. Wieczorkowska, J. Wróblewski, D. Ślęzak and P. Synak, *Problems with Automatic Classification of Musical Sounds*. Proc. of the Intelligent Information Processing and Web Mining Conference IIS: IIPWM'2003, Zakopane, Poland, Advances in Soft Computing, Springer, Berlin, Heidelberg , New York, pp. 423 - 430, 2003.

[66] R. Tadeusiewicz, *Sygnał mowy*. WKŁ, Warszawa 1988.

[67] B. Kostek and A. Wieczorkowska, *Parametric Representation of Musical Sounds*. Archives of Acoustics, **22**, 1, 1997, pp. 3-26.

[68] Christian Simmermacher, Da Deng and Stephen Cranefield, *Feature Analysis and Classification of Classical Musical Instruments*. An Empirical Study Department of Information Science, University of Otago, New Zealand (ddeng,scranefield)@infoscience.otago.ac.nz, luty 2007.

[69] M. Szczerba and A Czyżewski, *Pitch Detection Enhancement Employing Music Prediction Journal of Intelligent Information Systems*, 24:2/3, 223–251, 2005. Springer Science + Business Media, Inc. Manufactured in The Netherlands.

[70] B. Kostek, *"Computing with words" Concept Applied to Musical Information Retrieval*. Electronic Notes in Theoretical Computer Science, vol. 82, No. 4, 2003.

[71] E. Łukasik, *Multimedialna baza dźwięków skrzypiec – AMATI*. Dostępne http://www.zsi.pwr.wroc.pl/zsi/missi2002/pdf/s206.pdf , luty 2007.