Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Wit Jakuczun
Nr albumu: 178839

Rozpoznawanie mówców z wykorzystaniem liniowej predykcji i analizy falkowej

Praca magisterska
na kierunku MATEMATYKA

Praca wykonana pod kierunkiem
dr hab. Jerzego Cytowskiego
Instytut Informatyki

Maj 2002
Pracę przedkładam do oceny

Data

Podpis autora pracy:

Praca jest gotowa do oceny przez recenzenta

Data

Podpis kierującego pracą:
Streszczenie

Liniowa predykcja jest narzędziem znanym i stosowanym od dawna w przetwarzaniu sygnału mowy. Powstało już wiele systemów rozpoznawania mówców, opartych na tej metodzie. Falki jako narzędzie przetwarzania sygnałów pojawiły się dość niedawno. W ninieszej pracy autor próbuje zastosować nowe metody wyboru cech oparte na falkach do rozpoznawania mówców. Uzyskane wyniki są porównane z wynikami opartymi na współczynnikach liniowej predykcji.

Słowa kluczowe

cyfrowe przetwarzanie sygnałów, rozpoznawanie mówców, liniowa predykcja, falki, analiza frekwencyjna, systemy decyzyjne

Klasyfikacja tematyczna

68T10, 65T60, 94A12
Spis treści

1. Pojęcia podstawowe ........................................ 7
   1.1. Dyskretne systemy liniowe .......................... 7
      1.1.1. Model matematyczny sygnału dyskretnego i dyskretnego systemu liniowego .......................... 7
      1.1.2. Przyczynowość i stabilność ...................... 9
   1.2. Uczenie się maszyn .................................. 10
      1.2.1. Drzewa decyzyjne ............................... 10
      1.2.2. Wektorowa kwantyzacja ......................... 12

2. Liniowa predykcja ...................................... 13
   2.1. Metoda autokorelacji ............................... 14
   2.2. Algorytm Levinsona ................................ 14
   2.3. Zastosowanie w przetwarzaniu sygnału mowy ...... 16

3. Analiza falkowa ........................................ 17
   3.1. Wstęp .............................................. 17
   3.2. Analiza wieloskalowa ............................... 17
      3.2.1. Funkcja skalująca ................................ 18
      3.2.2. Falki ........................................... 19
      3.2.3. Regularność baz falkowych .................... 20
   3.3. Falki o zwartym nośniku ............................ 20
   3.4. Szybki algorytm liczenia współczynników falkowych . 23
   3.5. Zastosowania w cyfrowym przetwarzaniu sygnałów .... 24
      3.5.1. Filtry związane z falkami .................... 24
      3.5.2. Filtracja pasmowa ............................. 26

4. Rozpoznawanie mówców ................................ 29
   4.1. Wprowadzenie ....................................... 29
   4.2. Zaproponowany system klasyfikacji ................ 30
      4.2.1. Wybór cech - analiza frekwencyjna .......... 30
      4.2.2. Klasyfikatory .................................. 31
   4.3. Nagrania ........................................... 32
      4.3.1. Baza własna .................................... 32
      4.3.2. Baza hVd ...................................... 32
   4.4. Opis eksperymentów ................................ 32
      4.4.1. Przetwarzanie nagrań ......................... 32
      4.4.2. Redukcja wymiaru danych ..................... 33
      4.4.3. Wyniki dla własnej bazy nagrań ............. 33
4.4.4. Wyniki dla bazy hVd - indywidualnie ........................................... 34
4.4.5. Wyniki dla bazy hVd - płeć ......................................................... 36
4.5. Wnioski i propozycje rozszerzeń .................................................... 36

Literatura .................................................................................................. 39
Wstęp

Wprowadzenie

Przedmiotem niniejszej pracy jest opis zastosowania falek w przetwarzaniu sygnału mowy. Fale są "świeże" w przetwarzaniu sygnałów i dotychczas powstało niewiele prac dotyczących ich zastosowania w przetwarzaniu mowy ludzkiej.

Dziedzina przetwarzania sygnału mowy jest dość obszerna. W pracy skupiam się na zagadnieniu rozpoznawania mówców. Zagadnieniem tym zajmowano się i nadal zajmuje wielu ludzi na całym świecie. Powstało bardzo dużo systemów i algorytmów. Większość autorów skupia się na tworzeniu nowych systemów klasyfikacyjnych dla których wejściem są współczynniki uzyskane w wyniku liniowej predykcji (lub jej modyfikacji).


Warto podkreślić, że opisana metoda wyboru współczynników falkowych jest metodą, którą można zastosować do innego rodzaju sygnałów niż sygnał mowy.  

Układ pracy


\footnote{Adres ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/data/hVdBaSe.tar.Z}

\footnote{Pozycja [9]}
Rozdział 1

Pojęcia podstawowe

Niniejszy rozdział zawiera pewne podstawowe pojęcia, które są wykorzystywane w dalszej części pracy. Rozdział ten można traktować jako swego rodzaju wprowadzenie do teorii cyfrowego przetwarzania sygnałów oraz teorii uczenia się maszyn.

1.1. Dyskretne systemy liniowe

Sygnał jest to funkcja dostarczająca informacje o genezie lub zmianie określonych zjawisk fizycznych. Może to być na przykład sygnał mowy, który poprzez mikrofon jest przekształcany ze swojej pierwotnej postaci, fali akustycznej, na funkcje wyrażającą zmiany wielkości elektrycznych w czasie.

W przyjmowanych modelach sygnału abstrahuje się od fizycznej jego natury i przez to pojęcie może być wyrażona dowolna funkcja, która zawiera informacje związane ze zjawiskiem fizycznym, ekonomicznym, psychologicznym.

Wyróżnia się następujące typy sygnałów:

- **sygnał ciągły** - sygnał określony dla „ciągłej wartości czasu”;
- **sygnał analogowy** - sygnał ciągły mający „ciągły zbiór wartości”;
- **sygnał dyskretny** - sygnał określony dla dyskretnych wartości czasu;
- **sygnał cyfrowy** - sygnał dyskretny mający dyskretny zbiór wartości.

Każdy sygnał ciągły, który chcemy analizować techniką cyfrową musimy najpierw *zdyskretyzować*. W wyniku tej operacji otrzymamy ciąg liczb, który będzie dyskretnym odpowiednikiem wejściowego sygnału. Ciąg taki możemy następnie przekształcić za pomocą pewnego operatora matematycznego zwanego **systemem dyskretnym**.

1.1.1. Model matematyczny sygnału dyskretnego i dyskretnego systemu liniowego

**Definicja 1.1.1** (sygnał dyskretny) Sygnałem dyskretnym nazwizmy dowolny ciąg liczb rzeczywistych lub zespolonych

\[ x \equiv \{x(n)\}, -\infty < n < \infty \]

Symbol \( x(n) \) będzie oznaczał \( n \)-ty wyraz ciągu \( x \).

**Przykład 1.1.1** Ważne ciągi
• Impuls jednostkowy jest to ciąg określony wzorem

\[ \delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \quad (1.1) \]

• Ciągiem przesuniętym o k wyrazów nazwiemy ciąg określony następująco

\[ x_k(n) = x(n-k) \quad \forall n \quad (1.2) \]

gdzie \( x \) oznacza pewien dany ciąg.

Dalszy opis przestrzeni ciągów i operatorów na tych przestrzeniach przeprowadzę na podstawie książki [1]. Zaproponowany opis wyróżnia się tym, w przeciwieństwie do opisów spotykanych w literaturze, że jest kompletny z punktu widzenia matematyki.

Definicja 1.1.2 Będę rozpatrywał następujące rodzaje przestrzeni ciągów

- \( S \) - przestrzeń wszystkich ciągów
- \( S_F = \{ x : \exists N > 0 \ x(n) = 0 \ \forall n > N \} \) - przestrzeń ciągów skończonych,
- \( S_{\infty} = \{ x : \exists N > 0 \ x(n) = 0 \ \forall n < -N \} \) - przestrzeń ciągów mających gdzieś początek,
- \( S_i = \{ x : x(n) = 0 \ \forall n < i \} \) - przestrzeń ciągów mających początek w momencie \( i \).

Wszystkie powyższe przestrzenie są przestrzeniami liniowymi ze standardowymi działaniami „po współrzędnych”.

Definicja 1.1.3 Liniowy operator \( L : S_{\infty} \to S \) nazwiemy dyskretnym systemem liniowym wtedy i tylko wtedy, gdy operator ten ograniczony do przestrzeni \( S_0 \) jest ciągły względem metryki

\[ \rho(x, y) = \sum_{k=-\infty}^{\infty} 2^{-|k|} \frac{|x(k) - y(k)|}{1 + |x(k) - y(k)|} \quad (1.3) \]

oraz spełnia warunek niezmienności względem przesunięcia

\[ \forall x, y \in S_{\infty} L[x] = y \Rightarrow \forall k \ L[x_k] = y_k \quad (1.4) \]

Przykład 1.1.2 Wykazuje się, że operator \( L : S_{\infty} \to S \) określony równaniem

\[ y(n) = \begin{cases} 0 & \text{jeżeli } x(m) = 0 \ \forall m \leq n \\ \sum_{k=1}^{N} a_k y(n-k) + \sum_{k=0}^{M} b_k x(n-k) & \text{w przeciwnym przypadku} \end{cases} \quad (1.5) \]

gdzie \( y = L[x] \) i \( a_k, b_k \neq 0 \) jest systemem liniowym w sensie przyjętej definicji. Operator ten opisuje bardzo istotną pod względem praktycznym klasę układów liniowych i w niektórych pracach jest on przyjmowany jako określenie dyskretnego systemu liniowego.

Definicja 1.1.4 Dla wygody późniejszych rozważań, wprowadzę oznaczenie na ciąg będący odpowiedzią na impuls jednostkowy

\[ L[\delta] = h \quad (1.6) \]
Uwaga 1.1.1  Będę także chciał aby, dyskretny system liniowy spełniał następującą własność

\[ y = L[x] = L\left[ \sum_{k=-\infty}^{\infty} x(k)\delta_k \right] = \sum_{k=-\infty}^{\infty} x(k)L[\delta_k] \]  

(1.7)

Warto zauważyć, że powyższa własność dla dyskretnych systemów liniowych nie zmieni wzięch względem przesunięcia można zapisać w następujący sposób:

\[ y = \sum_{k=-\infty}^{\infty} x(k)h_k \]  

(1.8)

Z powyższego równania wynika, że dyskretny system liniowy jest jednoznacznie określony przez ciąg będący odpowiedzią na impuls jednostkowy. Opisuje ono splot dwóch ciągów i jest oznaczane przez

\[ y = x \ast h \]  

(1.9)

Definicja 1.1.5  Wyróżnia się dwa rodzaje dyskretnych systemów liniowych niezmiennych względem przesunięcia

- system o Skończonej Odpowiedzi Impulsowej \((h \in S_F\) w równaniu (1.9)). Oznacza się przez FIR (ang. finite impulse response) lub SOI.
- system o Nieskończonnej Odpowiedzi Impulsowej \((h \notin S_F\) w równaniu (1.9)). Oznacza się przez IIR (ang. infinite impulse response) lub NOI.

1.1.2. Przyczynowość i stabilność

Przyczynowość jest podstawowym warunkiem realizowalności, w czasie rzeczywistym, systemu liniowego. Dla systemu przyczynowego każda zmiana dowolnego sygnału wyjściowego musi być poprzedzona zmianą sygnału wejściowego.

Definicja 1.1.6  System \( y = L[x] \) będzie nazywał przyczynowym jeśli

\[ \forall x \in D \quad x \in S_i \Rightarrow y \in S_i \]  

(1.10)

gdzie \( D \) jest dziedziną rozpatrywanych sygnałów.

Ciąg \( h \) będący odpowiedzią na impuls jednostkowy jednoznacznie określa, czy system jest przyczynowy, czy nie. Zachodzi mianowicie następujące twierdzenie

Twierdzenie 1.1.1  System liniowy jest przyczynowy wtedy i tylko wtedy, gdy \( h \in S_0 \) (tzn. \( \forall n < 0 h(n) = 0 \)), gdzie \( h \) jest odpowiedzią systemu na impuls jednostkowy.

Inną ważną własnością jest stabilność systemu. Własność ta oznacza, iż na każdy ograniczony sygnał wejściowy powinniśmy otrzymać ograniczony sygnał wyjściowy.

Definicja 1.1.7  System stabilny

System \( L: D \rightarrow S \) będzie nazywał stabilnym, jeżeli

\[ \forall x \in D \quad \| x \|_{\infty} < M \Rightarrow \exists M_1 > 0 \quad \| y \|_{\infty} < M_1 \]  

gdzie \( y = L[x] \).
Zachodzi następujące twierdzenie

**Twierdzenie 1.1.2** System liniowy jest stabilny wtedy i tylko wtedy, gdy

\[ \sum_{k=-\infty}^{\infty} h(n) < \infty \]

1.2. Uczenie się maszyn

Problem uczenia się maszyn jest problemem bardzo rozległym. Pojawił się praktycznie wraz z pierwszymi maszynami liczącymi. W wyniku wytyżonej pracy naukowców związanych z takimi dziedzinami jak statystyka, czy logika powstało wiele metod i algorytmów.

W skrócie problem uczenia się można opisać następująco. Załóżmy, że \( X \) jest pewną (nieznaną) przestrzenią, tzw. uniwersum. Przestrzeń ta składa się z obiektów opisanych pewną skończoną liczbą atrybutów. Na tak zdefiniowanej przestrzeni obiektów można zadać tzw. pojęcia, poprzez wskazanie podzbioru obiektów spełniających pewne warunki. Dodatkowo zakładá się, że system (maszyna-uczeń) zna jedynie kilka przykładowych obiektów, zwanych obiektami treningowymi. Można postawić następujący problem

Na podstawie przykładów treningowych należy znaleźć jak najlepszą aproksymację funkcji \( d \) (tzw. hipotezę)

Na rysunku (1.1) przedstawiona jest opisana sytuacja.

![Rysunek 1.1: Uniwersum z 4 zadanymi pojęciami (d = 1, 2, 3, 4) oraz obiektami treningowymi (czarne) i testowymi (białe)](image)

1.2.1. Drzewa decyzyjne

Jedną z metod generowania hipotez są drzewa decyzyjne. Drzewo decyzyjne jest formalnie zwykłym drzewem, znany z informatyki teoretycznej, tzn. jest to spójny skierowany acykliczny graf. Węzły, gałęzie i liście drzewa decyzyjnego mają pewną dodatkową interpretację.
Liść drzewa decyzyjnego zawiera etykietę pojęcia opisywanego przez dane drzewo

 Każdy węzeł jest etykietowany tzw. testem, czyli pewną funkcją określoną na podzbiorze atrybutów obiektów z uniwersum. Testy rozróżnia się na binarne (przyjmujące tylko dwie wartości np. 0,1) oraz wielowartościowe. Najczęściej stosowane są testy oparte na jednym atrybutucie (np. aura=?). W przypadku atrybutów rzeczywistych stosuje się testy postaci \( a_{10} \leq 0.34 \).

 Ścieżka łącząca korzeń drzewa z liściem wyznacza tzw. regułę decyzyjną, tzn. każdy obiekt, którego atrybuty spełniają testy na tej ścieżce będzie zaliczony do obiektów należących do pojęcia reprezentowanego przez liść drzewa.

Rysunek (1.2) przedstawia przykładowe drzewo decyzyjne [2].

![Rysunek 1.2: Drzewo decyzyjne dla stanów pogody](image)

Drzewa decyzyjne są ciekawe z wielu powodów. Po pierwsze za ich pomocą można opisać praktycznie dowolnie złożone pojęcia. Są dużo bardziej czytelne od innych sposobów opisu pojęć (np. sieci neuronowych). Istnieją w miarę tanie algorytmy (ze względu na czas i pamięć) tworzenia drzew decyzyjnych¹.

Z dotychczasowego opisu wynika, że budowanie drzewa w pełni zależy od doboru testów w węzłach. Bardzo często stosowaną miarą jakości testu jest entropia podziału zbioru obiektów treningowych w wyniku zastosowania testu. Inną równie często stosowaną miarą jest liczba obiektów, należących do różnych pojęć, rozróżnianych przez dany test.


¹ Pozycja [2]
1.2.2. Wektorowa kwantyzacja


![Rysunek 1.3: Dane zgrupowane w trzech klastrach. Kółka reprezentują obiekty, trójkąty reprezentują środki klastrów.](image)


Algorytm klasyfikacyjny oparty na *wektorowej kwantyzacji* przedstawia się następująco

1. Dla obiektów treningowych należących do tego samego pojęcia przeprowadzana jest wektorowa kwantyzacja. Uzyskuje się w ten sposób dla każdego pojęcia, zbiór centroidów (tzw. słownik) opisujących to pojęcie.

2. Chcąc sklasifykować nowy obiekt, należy w każdym słowniku znaleźć centroid leżący najbliżej nowego obiektu. Wynikiem klasyfikacji jest pojęcie do którego należy najbliżej leżący centroid.

3. W przypadku grupy nowych obiektów, dla każdego słownika oddzielnie sumowane są odległości najbliższych centroidów od poszczególnych obiektów. Wynikiem klasyfikacji jest pojęcie, którego słownik uzyskał najmniejszą sumę.
Rozdział 2

Liniowa predykcja

Problem liniowej predykcji można w skrócie opisać następująco

Na podstawie wartości próbek sygnału 
\{x(k-1), \ldots, x(k-n)\} należy wyznaczyć wartość x(k)

W dziedzinie przetwarzania dyskretnych sygnałów stochastycznych metoda należy do pewnej ogólniej klasy metod tzw. filtracji ortogonalnej\(^1\). W dalszej części rozdziału będę zakładał, że mam do czynienia z sygnałem stochastycznym stacjonarnym drugiego rzędu, zgodnie z konwencją przyjętą w [8].

Niech \(y\) będzie pewnym sygnałem losowym. Będę zakładał, że znam pewną przeszłość tego sygnału, tzn. próbki \(\{y(n-M), \ldots, y(n-1)\}\). Na podstawie tych próbek zbuduję pewien predyktor postaci

\[
\hat{y}(n) = \sum_{k=1}^{M} \alpha_k y(n-k)
\]

Współczynniki \(\alpha_k\) dobierę tak, aby zminimalizować błąd średniokwadratowy między sygnałem \(y\) a \(\hat{y}\). Niech \(e(n) = y(n) - \hat{y}(n)\) oznacza błąd predykcji. Przyjmując \(a_k = -\alpha_k\) dla \(k = 1, \ldots, M\) oraz \(a_0 = 1\) mogę \(e(n)\) wyrazić następująco

\[
e(n) = \sum_{k=0}^{M} a_k y(n-k)
\]

Aby zminimalizować \(e(n)\) w sensie średniokwadratowym muszę rozpatrzyć wyrażenie postaci

\[
\alpha^M = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{M} \sum_{l=0}^{M} a_k y(n-k) y(n-l) a_l
\]

W praktyce mamy do czynienia z sygnałami o skończonym czasie trwania, dlatego wyrażenie na \(\alpha\) można zapisać następująco

\[
\alpha = \sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} \sum_{k=0}^{M} \sum_{l=0}^{M} a_k y(n-k) y(n-l) a_l
\]

w gdzie \(N\) jest długością sygnału \(y\). Widać, że minimum (jedyne) dla \(\alpha\) jest osiągane wtedy i tylko wtedy, gdy

\[
\frac{\partial \alpha}{\partial a_k} = 0 \quad \text{dla } k = 1, \ldots, M
\]

\(^{1}\text{Pozycja [11]}\)
Otrzymałem zatem układ $M$ równań na $M$ niewiadomych. W dalszej części opiszę jedną z metod rozwiązywania powyższego układu.

### 2.1. Metoda autokorelacji

Będę zakładal, że sygnał $y$ jest sygnałem skończonym o długości $N$.

**Definicja 2.1.1** Macierzą autokorelacji rzędu $M$ dla sygnału $y$ jest macierz $C^M = (c^M_{ij})_{i,j=1,\ldots,M}$ dana wzorem

$$c^M_{i,j} = \sum_{n=0}^{N-1} y(n-i)y(n-j)$$

Ze stacjonarności sygnału $y$ wynika, że wyrazy macierzy $C^M$ można zapisać wzorem

$$c^M_{i,j} = \sum_{n=0}^{N-1-|i-j|} y(n)y(n+|i-j|) = r(|i-j|)$$

Wynika stąd, że macierz $C^M$ jest symetryczna oraz dodatkowo elementy leżące na jej kolejnych przekątnych są sobie równe. Macierz o takich własnościach nazywana jest macierzą Toeplitza. Korzystając z macierzy $C^M$ wyrażenie na błąd predykcji rzędu $M$ zapisuje się następująco

$$\alpha = \sum_{k=0}^{M} \sum_{l=0}^{M} a_k c^M_{k,l} a_l$$

Zatem, zgodnie z wcześniejszymi rozważaniami muszę rozwiązać następujący układ równań liniowych

$$2 \sum_{k=0}^{M} a_k c^M_{k,l} = 0 \quad l = 1,\ldots,M \quad (2.3)$$

**Uwaga 2.1.1** Z numerycznego punktu widzenia powyższy układ nie stanowi wielkiego problemu, ponieważ w praktyce $M$ jest rzędu kilkunastu. Problem stanowi jednak to, że często trzeba rozwiązywać takie układy kilkadziesiąt razy na sekundę. Standardowe metody, takie jak eliminacja Gaussa okazują się zbyt kosztowne.

### 2.2. Algorytm Levinsoña

Przedstawię teraz jedną z metod rozwiązywania układu (2.3). Została ona zaproponowana przez Levinsona\(^2\). Metoda ta polega na stopniowym zwiększaniu stopnia predyktora (liczby współczynników).

Różniczkując wyrażenie na błąd (2.1) względem $a_k$ otrzymuję, że

$$\frac{\partial \alpha}{\partial a_k} = 2 \sum_{n=0}^{N-1} e(n) \frac{\partial e(n)}{\partial a_k} = 2 \sum_{n=0}^{N-1} e(n)y(n-k)$$

\(^{2}\)Pozycja [11]
zatem, uwzględniając (2.2), otrzymuję
\[
\alpha = \sum_{n=0}^{N-1} e(n)e(n) = \sum_{n=0}^{N-1} c(n) \left[ \sum_{k=0}^{M} a_k y(n-k) \right] \\
= \sum_{n=0}^{N-1} e(n)y(n) + \sum_{k=0}^{M} a_k \sum_{n=0}^{N-1} y(n-k) \\
= \sum_{n=0}^{N-1} \left[ \sum_{k=0}^{M} a_k y(n-k) \right] y(n) \\
= \sum_{k=0}^{M} a_k r(k) \tag{2.4}
\]

Dla wygody notacyjnej wprowadzę następujące oznaczenia

- \(\{a_k^M\}_{k=0,...,M}\) - szukane współczynniki dla predyktora rzędu \(M\)
- \(\alpha^M\) - błąd predyktora rzędu \(M\)

W każdym kroku algorytmu Levinsona rozwiązywany jest układ równań

\[
\begin{pmatrix}
1 & a_1^{M+1} & \cdots & a_M^{M+1} \\
a_1 & a_1^M & \cdots & a_M^M \\
\vdots & \vdots & \ddots & \vdots \\
a_M & a_M^M & \cdots & a_M^M \\
\end{pmatrix}
\begin{pmatrix}
1 \\
a_1^{M+1} \\
\vdots \\
a_M^{M+1} \\
\end{pmatrix}
= \begin{pmatrix}
\alpha^{M+1} \\
0 \\
\vdots \\
0 \\
\end{pmatrix}
\]

przy wykorzystaniu rozwiązania układu dla predyktora rzędu \(M\)

\[
\begin{pmatrix}
1 & a_1^M & \cdots & a_M^M \\
\vdots & \vdots & \ddots & \vdots \\
0 & a_M^M & \cdots & a_M^M \\
\end{pmatrix}
\begin{pmatrix}
1 \\
a_1^M \\
\vdots \\
a_M^M \\
\end{pmatrix}
= \begin{pmatrix}
\alpha^M \\
0 \\
\vdots \\
0 \\
\end{pmatrix}
\]

Następujące twierdzenie opisuje krok iteracyjny omawianego algorytmu.

**Twierdzenie 2.2.1 (Algorytm Levinsona)** 
Niech będą dane

- Współczynniki autokorelacji \(r(0), \ldots, r(M+1)\) dla stacjonarnego sygnału losowego \(y\)
- Współczynniki \(a_1^M, \ldots, a_M^M\) dla predyktora rzędu \(M\)
- Błąd średniokwadratowy \(\alpha^M\)

Wówczas prawdziwe są następujące zależności rekurencyjne

\[
\begin{pmatrix}
1 \\
a_1^{M+1} \\
\vdots \\
a_M^{M+1} \\
\end{pmatrix}
= \begin{pmatrix}
1 \\
a_1^M \\
\vdots \\
a_M^M \\
\end{pmatrix}
+ \rho_{M+1} \begin{pmatrix}
0 \\
a_1^M \\
\vdots \\
a_M^M \\
\end{pmatrix}
\]

\(^3\text{Pozycja [11]}\)
gdzie
\[ \rho_{M+1} = -\frac{r(M+1) + r(M)a^M_M + \ldots + r(1)a^M_1}{\alpha^M} \] (przy czym \( |\rho_{M+1}| \leq 1 \))

oraz
\[ \alpha^{M+1} = \alpha^M(1 - \rho_{M+1}^2) \] (czyli \( \alpha^{M+1} \leq \alpha^M \))

### 2.3. Zastosowanie w przetwarzaniu sygnału mowy


Model ten jest wygodnie opisywać wykorzystując z-transformatę. Mianowicie

\[ S(z) = E(z)G(z)V(z)L(z) \]

gdzie kolejno

- \( S(z) \) jest z-transformatą próbek sygnału mowy
- \( E(z) \) jest z-transformatą sygnału pobudzenia (ciąg impulsów okresowych lub o amplitudzie zmieniającej się losowo).
- \( G(z) \) jest z-transformatą transmitancji filtru modelującego głośnię. Najczęściej \( G(z) \) jest dwubiegunowym filtrem dolnoprzepustowym o częstotliwości odcięcia ok. 100Hz
  \[ G(z) = \frac{1}{(1 - z^{-1})^2} \]
- \( V(z) \) jest z-transformatą transmitancji filtru modelującego tor głosowy. Jest to filtr biegunowy będący iloczynem \( K \) współczynników
  \[ V(z) = \prod_{i=1}^{K} \frac{1}{1 - 2e^{-c_iT}\cos(b_iT)z^{-1} + e^{-2c_iT}z^{-2}} \]
  Stałe \( b_i \) oraz \( c_i \) wyznaczają tzw. formanty
- \( L(z) \) jest z-transformatą transmitancji filtru modelującego emisję warg. Filtr ten ma dość prostą strukturę (jest to filtr preemfazy)
  \[ L(z) = 1 - z^{-1} \]

Niech \( A(z) = \frac{1}{V(z)G(z)L(z)} \). Jest to tzw. filtr inwersyjny. Model generowania sygnału mowy można zatem zapisać następująco
\[ S(z) = E(z)\left(\frac{1}{A(z)}\right) \]

Okazuje się, że współczynniki filtru inwersyjnego \( A(z) \) są identyczne ze współczynnikami liniowej predykcji zastosowanej do sygnału mowy poddanego operacji preemfazy. Fakt ten spowodował, że liniowa predykcja stała się jedną z podstawowych metod analizy sygnału mowy.

---

\(^4\) Pozycja [3]
\(^5\) Pozycja [6]
Rozdział 3

Analiza falkowa

3.1. Wstęp

Falki są całkiem nową teorią w matematyce stosowanej. Pomimo tego jest to dziedzina dość szybko rozwijana. Spowodowane jest to tym, że falki posiadają własności, które czynią je przydatnymi w takich dziedzinach jak równania różniczkowe, teoria aproksymacji, przetwarzanie sygnałów. W poniższym rozdziale spróbuję przedstawić teorię związaną z falkami, a szczególnie z falkami o zwartych nośnikach. Opiszę także zastosowanie falek w cyfrowym przetwarzaniu sygnałów.

3.2. Analiza wieloskalowa

Definicja 3.2.1 Analizą wieloskalową (wielorozdzielczą) nazywamy układ domkniętych pod-przestrzeni \( \{V_j\} \) spełniających następujące warunki

\[
\ldots \subset \subseteq \subseteq \Ltwo
\]

(3.1)

\[
\bigcup_{j \in \mathbb{Z}} V_j = \Ltwo
\]

(3.2)

\[
\bigcap_{j \in \mathbb{Z}} V_j = \{0\}
\]

(3.3)

\[
f \in V_j \iff f(2^{-j} \cdot) \in V_0
\]

(3.4)

\[
f \in V_0 \iff f(\cdot - n) \in V_0 \quad \forall n \in \mathbb{Z}
\]

(3.5)

\[
\exists \Phi \in V_0: \{\Phi(\cdot - n)\}_{n \in \mathbb{Z}} \text{ jest baząortonormalną w } V_0
\]

(3.6)

Dla wygody wprowadzę pewne oznaczenia

\[
f_{j,k}(x) = 2^{j/2} f(2^j x - k)
\]

\[
P_j(f) = \sum_{n \in \mathbb{Z}} \langle f, \Phi_{j,n} \rangle \Phi_{j,n}
\]

Podstawową cechą tak zdefiniowanego układu podprzestrzeni jest istnienie bazyortonormalnej falek \( \{\Psi_{j,k}\} \) przestrzeni \( \Ltwo \), takiej że

\[
P_j(f) = P_{j-1}(f) + \sum_{k \in \mathbb{Z}} \langle f, \Psi_{j,k} \rangle \Psi_{j,k}
\]

(3.7)
W dalszej części pokażę jak można konstruować taką właśnie bazę w przypadku ogólnym. Dla każdego \( j \in \mathbb{Z} \) zdefiniuję podprzestrzeni \( W_{j-1} \) będącą ortogonalnym dopełnieniem podprzestrzeni \( V_{j-1} \) w podprzestrzeni \( V_j \). Otrzymam zatem następujący rozkład przestrzeni

\[
V_j = V_{j-1} \oplus W_{j-1}
\]

oraz

\[
W_j \perp W_j' \quad \text{jeśli} \quad j \neq j'
\]

Z tego, że dla \( j < j' \) zachodzi \( W_j \subset V_{j'} \perp W_{j'} \), wynika

\[
V_j = V_j \oplus \bigoplus_{k=0}^{j-J-1} W_{j+k} \quad \forall j > J
\]

Z własności (3.2) oraz (3.3) otrzymujemy

\[
\mathbf{L}^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j
\]

(3.8)

Dodatkowo własność (3.4) jest też spełniona dla przestrzeni \( W_j \)

\[
f \in W_j \iff f(2^{j'}\cdot) \in W_0
\]

Równanie (3.7) jest równoważne temu, że układ \( \{\Phi_{j,k}\}_{k \in \mathbb{Z}} \) dla ustalonego \( j \) stanowi bazęortonormalną dla \( W_j \). Z rozkładu (3.8) oraz z (3.2) wynika, że cały układ \( \{\Phi_{j,k}\}_{j,k \in \mathbb{Z}} \) stanowi bazę ortonormalną dla \( \mathbf{L}^2(\mathbb{R}) \). Nierdowno zauważyć, że mając takie \( \Phi \in W_0 \), że \( \Phi(\cdot - k) \) stanowi bazę ortonormalną dla \( W_0 \), układ \( \{\Phi_{j,k}\}_{j,k \in \mathbb{Z}} \) będzie stanowić szukaną bazę falek.

### 3.2.1. Funkcja skalująca

Przyjrzę się teraz bliżej funkcji (3.6), którą często nazywa się \( \text{funkcja skalująca} \). Ponieważ funkcja skalująca \( \Phi \) należy do \( V_0 \subset V_1 \) oraz układ \( \{\Phi_{1,n}\}_{n \in \mathbb{Z}} \) stanowi bazę ortonormalną w \( V_1 \), istnieje ciąg liczbowy \( (h_n)_{n \in \mathbb{Z}} \) taki, że

\[
\Phi(x) = \sum_{n \in \mathbb{Z}} h_n \Phi_{1,n}(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \Phi(2x - n)
\]

(3.9)

gdzie \( h_n = \langle \Phi, \Phi_{1,n} \rangle \) i \( \sum_{n \in \mathbb{Z}} |h_n|^2 = 1 \). Powyższe równanie możemy przepisać, używając transformaty Fouriera, w następujący sposób

\[
\hat{\Phi}(\xi) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \exp\{-in\xi/2\} \hat{\Phi}(\xi/2)
\]

(3.10)

Zbieżność obu sum zachodzi w przestrzeni \( \mathbf{L}^2(\mathbb{R}) \). Otrzymałem zatem

\[
\hat{\Phi}(\xi) = m_0(\xi/2) \hat{\Phi}(\xi/2)
\]

(3.11)

gdzie

\[
m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n \exp\{-in\xi\}
\]

(3.12)

Równania (3.9)-(3.11) zwane są \( \text{równaniami skalującymi} \). Następujące lematy opisują ważne, z punktu dalszych rozważań, własności funkcji skalujących.
Lemat 3.2.1 Załóżmy, że $\Phi(x)$ jest funkcją skalującą analizy wieloskalowej i załóżmy dodatkowo, że $\hat{\Phi}(\xi)$ jest ciągłe w 0. Wtedy $|\hat{\Phi}(0)| = \frac{1}{2\pi}$.

Lemat 3.2.2 Ortogonalność układu $\Phi(-k)$ jest równoważna temu, że

$$\sum_{n \in \mathbb{Z}} |\hat{\Phi}(\xi + 2\pi l)|^2 = \frac{1}{2\pi}$$

p.w. (3.13)

Wnioskiem z powyższego lematu jest pewna ważna własność funkcji $m_0$. Własność ta odgrywa dużą rolę w zastosowaniach fal ek w cyfrowym przetwarzaniu sygnałów.

Lemat 3.2.3 Funkcja $m_0$ spełnia następujące równanie

$$|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1$$

p.w. (3.14)

Wnioskiem z powyższego lematu jest pewna ważna własność funkcji $m_0$. Własność ta odgrywa dużą rolę w zastosowaniach fal w cyfrowym przetwarzaniu sygnałów.

Lemat 3.2.4 Funkcja $f$ należy do $W_0$ wtedy i tylko wtedy gdy

$$\hat{f}(\xi) = e^{i\xi/2}m_0(\xi/2 + \pi)v(\xi)\hat{\Phi}(\xi/2)$$

gdzie $v$ jest funkcją $2\pi$-okresową. Zachodzi dodatkowo

$$\|f\|^2 = \frac{1}{2\pi} \int_{0}^{2\pi} |v(\xi)|^2 d\xi$$

Następnym krokiem będzie opisanie bazyortonormalnej w $W_0$.

Twierdzenie 3.2.2 Niech funkcja $\Psi$ będzie dana wzorem (3.14). Układ $\{\Psi(\cdot - k)\}_{k \in \mathbb{Z}}$ jest bazą ortonormalną podprzestrzeni $W_0$ wtedy i tylko wtedy, gdy $|v(\xi)| = 1$, tzn.

$$\hat{\Psi}(\xi) = m_1(\xi/2)\hat{\Phi}(\xi/2)$$

lub

$$\Psi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \Phi(2x - n)$$

gdzie $m_1(\xi) = e^{i\xi/2}m_0(\xi + \pi)v(\xi) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} g_ne^{-im\xi}.}$

19
Konsekwencją ortogonalności funkcji \( \Psi \) do przestrzeni \( V_0 \) jest następująca zależność między funkcjami \( m_0 \) i \( m_1 \).

**Lemat 3.2.5** Funkcje \( m_0 \) i \( m_1 \) związane są następującą zależnością

\[
m_1(\xi)\overline{m_0(\xi)} + m_1(\xi + \pi)\overline{m_0(\xi + \pi)} = 0
\]  
(3.18)

Ciekawym i ważnym spostrzeżeniem jest to, że mając daną funkcję skalującą możemy zbudować falkę z nią związaną korzystając tylko ze współczynników \( \{h_n\}_{n \in \mathbb{Z}} \) (3.9). Naturalnym wydaje się pytanie o to dla jakich ciągów liczb \( \{x_n\}_{n \in \mathbb{Z}} \) możemy znaleźć funkcję skalującą \( \Phi \) spełniającą równanie skalujące

\[
\Phi(x) = \sum_{n \in \mathbb{Z}} x_n \Phi(2x - n)
\]

oraz czy funkcja ta zadaje analizę wieloskalową. W części poświęconej falkom o nośnikach zwartych opisz jak szukać takich „dobrych” współczynników.

### 3.2.3. Regularność baz falkowych

Dla ortonormalnych baz falkowych istnieje ścisły związek między regularnością falki \( \Psi \) a krot- nością zm. \( \hat{\Psi}(\xi) \) dla \( \xi = 0 \).

**Twierdzenie 3.2.3** Przypuśćmy, że \( \Psi \) jest funkcją na \( \mathbb{R} \) taką, że układ \( \{2^{j/2}\overline{\Psi(2^j \cdot -k)}\}_{j \in \mathbb{Z}, k \in \mathbb{Z}} \) jest ortonormalny w \( L^2(\mathbb{R}) \). Załóżmy, że dla pewnego \( l = 0, 1, 2, \ldots \) mamy

\[
\Psi \text{ jest klasy } C^l
\]

wszystkie pochodne \( \Psi^{(s)}(x) \) dla \( s = 0, 1, 2, \ldots, l \) są ograniczone na \( \mathbb{R} \)

\[
|\Psi(x)| \leq \frac{C}{(1 + |x|^\alpha)} \quad \text{dla pewnego } \alpha > l + 1
\]

Wtedy

\[
\int_{-\infty}^{\infty} x^s \Psi(x) = 0 \quad \text{dla } s = 0, 1, \ldots, l
\]

**Wniosek 3.2.1** Przy założeniach poprzedniego twierdzenia funkcja \( m_0 \) zdefiniowana poprzez (3.12) może zostać zapisana następująco

\[
m_0(\xi) = \left(1 + e^{-i\xi}\right)^{l+1} \frac{L(\xi)}{2}
\]  
(3.19)

gdzie \( L \) jest \( 2\pi \)-okresowa oraz należy do \( C^l \).

### 3.3. Falki o zwartym nośniku

Fakt istnienia regularnych falek o zwartych nośnikach był do nie dawna nieznany. Przez wiele lat jedyną znaną bazą przestrzeni \( L^2(\mathbb{R}) \) złożoną z funkcji o nośnikach zwartych była baza Haara. Funkcje z tej bazy nie są jednakże nawet ciągłe co czyni je mało przydatnymi w zastosowaniach praktycznych. Dopiero w swoich pracach pani Ingrid Daubechies podała „przepis” na tworzenie falek dowolnie regularnych mających zwarte nośniki.
Konstrukcja przedstawiona przez panią Daubechies opiera się na dwóch spostrzeżeniach. Po pierwsze, aby zapewnić zwartość nośnika funkcji skalującej, funkcja \( m_0 \) musi być jakimś wielomianem trygonometrycznym. Po drugie chcąc zapewnić odpowiednią regularność uzyskanie bazy wielomian ten powinien spełniać warunek (3.19).

Następujące twierdzenie [10] podaje w jaki sposób mając pewien wielomian trygonometryczny (spełniający odpowiednie warunki) skonstruować związaną z nim funkcję skalującą zadającą pewną analizę wieloskalową.

**Twierdzenie 3.3.1** Niech \( m_0(\xi) = \sum_{k=0}^{\infty} a_k e^{-ik\xi} \) będzie wielomianem trygonometrycznym spełniającym następujące warunki

\[
|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1 \quad \text{dla wszystkich} \quad \xi \in \mathbb{R} \\
m_0(0) = 1 \\
m_0(\xi) \neq 0 \quad \text{dla} \quad \xi \in [-\frac{\pi}{2}, \frac{\pi}{2}] 
\]

Wtedy iloczyn nieskończony

\[
\Theta(\xi) = \prod_{j=1}^{\infty} m_0(2^j \xi)
\]

zbiera niemal jednostajnie, a więc funkcja \( \Theta(\xi) \) jest ciągła. Ponadto należy ona do \( L^2(\mathbb{R}) \). Funkcja \( \Phi \) dana przez \( \hat{\Phi} = \frac{1}{\sqrt{2\pi}} \Theta \) ma nośnik zawarty w odcinku \([T,S]\) oraz jest funkcją skalującą pewnej analizy wieloskalowej. W szczególności jej przesunięcia są ortogonalne.

Wyjaśnienia wymaga pojawienie się nieskończonego iloczynu w wyrażeniu na \( \hat{\Phi} \). Okazuje się, że biorąc pod uwagę równanie skalujące (3.11) otrzymamy, że

\[
\hat{\Phi}(\xi/2) = m_0(\xi/2)\hat{\Phi}(\xi/2) = \ldots = \prod_{j=1}^{\infty} m_0(2^j \xi)
\]

Pozostaje jeszcze problem znajdowania takich wielomianów trygonometrycznych, które będą spełniały jego założenia. Od razu widać, że najwięcej kłopotów będzie z warunkiem

\[
|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1
\]

**Twierdzenie 3.3.2** Wielomian trygonometryczny postaci

\[
m_0(\xi) = \left( \frac{1 + e^{-i\xi}}{2} \right)^N \mathcal{L}(\xi)
\]

spełnia warunek

\[
|m_0(\xi)|^2 + |m_0(\xi + \pi)|^2 = 1
\]

wtedy i tylko wtedy, gdy \( L(\xi) = |\mathcal{L}(\xi)|^2 \) może być zapisana jako

\[
L(\xi) = P(\sin^2 \xi/2)
\]

przy

\[
P(y) = P_N(y) + y^N R(\frac{1}{2} - y)
\]

gdzie

\[
P_N(y) = \sum_{k=0}^{N-1} \binom{N-1 + k}{k} y^k
\]

a \( R \) jest wielomianem wybranym, tak aby \( P(y) > 0 \) dla \( y \in [0,1] \).
Powyższe twierdzenie charakteryzuje nam jedynie \( |m_0|^2 \). Korzystając z lematu Riesza możemy „wyciągnąć pierwiastek” aby otrzymać \( m_0 \).

**Lemat 3.3.1 (Riesz)** Niech \( A \) będzie dodatnim wielomianem trygonometrycznym niezmiennym przy podstawieniu \( \xi \rightarrow -\xi \). Istnieje wtedy wielomian trygonometryczny \( B \) stopnia takiego samego jak \( A \) taki, że

\[
|B(\xi)|^2 = A(\xi)
\]

Przykładem zastosowania powyższych twierdzeń są tzw. falki Daubechies. Zgodnie z konwencją przyjętą w książce [4] falki te indeksowane są parametrem \( N \). Podczas ich konstrukcji przyjęto \( R \equiv 0 \) (twierdzenie 3.3.2). Falki te mają \( 2N \) znikających momentów, należą do klasy \( C^{\mu N} \) dla \( \mu \approx 0.2 \) a ich nośnik ma długość \( 2N - 1 \). Na rysunkach (3.2) oraz (3.1) przedstawione są wykresy falki i funkcji skalującej dla \( N = 5 \).

![Rysunek 3.1: Funkcja skalująca Daubechies 5](image1.png)

![Rysunek 3.2: Falka Daubechies 5](image2.png)
3.4. Szybki algorytm liczenia współczynników falkowych

Z tego co już powiedziałem na temat falek nie do końca wynika jak można by efektywnie wykorzystać je w praktyce. Okazuje się, że istnieje bardzo efektywny algorytm wyznaczania współczynników rozwinięcia funkcji w bazie falkowej. Algorytm ten wykorzystuje analizę wieloskalową i należy do grupy tzw. algorytmów hierarchicznych, tzn. takich w których ostateczny wynik osiągany jest w kilku etapach.

Na początku przytoczę wzory na pewne iloczyny skalare

\[ \langle \Phi_{j-1,k}, \Phi_{j,n} \rangle = \frac{h_{k-2n}}{n} \]
\[ \langle \Psi_{j-1,k}, \Phi_{j,n} \rangle = \frac{g_{k-2n}}{n} \]
\[ \langle \Phi_{j,n}, \Phi_{j-1,k} \rangle = \frac{h_{k-2n}}{n} \]
\[ \langle \Phi_{j,n}, \Psi_{j-1,k} \rangle = \frac{g_{k-2n}}{n} \]

gде \( h_n \) oraz \( g_n \) są współczynnikami w równaniach (3.9) oraz (3.17). Wyposażony w powyższe wzory, których wyprowadzenie jest bardzo proste i opiera się na prostej zamianie zmiennej pod całkę, mogę opisać algorytm.

Niech funkcja \( f \in L^2(\mathbb{R}) \). Wprowadzę teraz oznaczenia na współczynniki rzutu tej funkcji na podprzestrzenie \( V_j \) i \( W_j \).

\[ P_{V_j}(f) = \sum_{n \in \mathbb{Z}} \lambda_{j,n} \Phi_{j,n} \]
\[ P_{W_j}(f) = \sum_{n \in \mathbb{Z}} \beta_{j,n} \Psi_{j,n} \]

Ponieważ \( V_j = V_{j-1} \oplus W_{j-1} \) więc

\[ P_{V_j}(f) = \sum_{n \in \mathbb{Z}} \lambda_{j,n} \Phi_{j,n} = \sum_{k \in \mathbb{Z}} \lambda_{j-1,k} \Phi_{j-1,k} + \sum_{l \in \mathbb{Z}} \lambda_{j-1,l} \Psi_{j-1,l} \]

Mnożąc powyższe równanie skalarnie przez \( \Phi_{j-1,k} \) oraz \( \Phi_{j-1,l} \), mając na uwadze wzory (3.21) otrzymuję

\[ \lambda_{j-1,k} = \sum_{n \in \mathbb{Z}} h_{n-2k} \lambda_{j,n} \]
\[ \beta_{j-1,k} = \sum_{n \in \mathbb{Z}} g_{n-2k} \lambda_{j,n} \]

Wzory (3.23) i (3.24) są tak naprawdę wzorami wyrażającymi co drugi wyraz splotu ciągu \( \lambda_{j,n} \) odpowiednio z ciągami \( h_{n} \) oraz \( g_{n} \). Mam już zatem wzory dzięki, którym mogę dokonać analizy dowolnej funkcji z pewnej podprzestrzeni \( V_j \).

Wzory realizujące operację odwrotną otrzymamy się mnożąc równanie (3.22) przez \( \Phi_{j,n} \) oraz uwzględniając równania (3.21). Dokładniej

\[ \lambda_{j,n} = \sum_{k \in \mathbb{Z}} \lambda_{j-1,n} h_{n-2k} + \sum_{l \in \mathbb{Z}} \beta_{j-1,l} g_{n-2l} \]

Podobnie jak w poprzednim przypadku, tak i w tym, wzór wyrażający \( \lambda_{j,n} \) można interpretować jako pewien splot. Mianowicie jeśli ciągi \( \lambda_{j-1,k} \) oraz \( \lambda_{j-1,l} \) zastąpiemy ciągami odpowiednio

\[ \tilde{\lambda}_{j-1,n} = (\ldots, 0, \lambda_{j-1,1}, \ldots) \]
oraz
\[ \tilde{\beta}_{j-1,n} = (\ldots, 0, \beta_{j-1,0}, 0, \beta_{j-1,1}, 0, \ldots) \]
to wzór na \( \lambda_{j,n} \) mogę zapisać następująco
\[ \lambda_{j,n} = \sum_{k \in \mathbb{Z}} \tilde{\lambda}_{j-1,n} h_{n-k} + \sum_{l \in \mathbb{Z}} \tilde{\beta}_{j-1,l} g_{n-l} \]
Rysunek (3.3) przedstawia schemat działania powyższego algorytmu.

Rysunek 3.3: Schemat działania algorytmu obliczania współczynników falkowych

Kończąc opis algorytmu muszę wspomnieć o pewnym problemie technicznym z nim związanym. Chodzi mianowicie o przypadek w którym ciąg \( \lambda_{j,n} \) jest skończony, tzn. jedynie dla skończonej liczby indeksów \( n \) wyrazy tego ciągu są różne od zera. Dla funkcji \( f \) oznacza to ograniczenie jej nośnika. Problem polega na tym, że współczynniki falkowe dla falek, których nośnik tylko zahacza o nośnik funkcji \( f \), zostaną zafałszowane. Dlatego, aby zminimalizować utratę informacji, będę zakładal okresowość funkcji \( f \).

3.5. Zastosowania w cyfrowym przetwarzaniu sygnałów

W niniejszym podrozdziale opiszę zastosowanie falek w przetwarzaniu sygnałów. Opiszę również pewien układ filtrów, który realizuje identyczną operację co algorytm wyznaczania współczynników falkowych.

3.5.1. Filtry związane z falkami

Na początku przyjrzymy się wzorom na transformatę Fouriera funkcji skalującej i falki z nią związanej.
\[
\tilde{\Phi}(\xi) = m_0(\xi/2)\tilde{\Phi}(\xi/2) \\
\tilde{\Psi}(\xi) = m_1(\xi/2)\tilde{\Phi}(\xi/2)
\]
gdzie
\[
m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n e^{-im\xi} \\
m_1(\xi) = e^{i\xi} m_0(\xi + \pi) v(\xi) \quad |v(\xi)| = 1
\]
Z lematu (3.2.1) wynika, że funkcja \( m_0(\xi) \) jest ciągła w 0 oraz \( m_0(0) = 1 \), zatem z lematu (3.2.3) otrzymuję
\[
\begin{align*}
\begin{cases}
m_0(0) &= 1 \\
m_0(\pi) &= 0
\end{cases}
\end{align*}
\] (3.26)

Funkcje spełniające powyższą własność będę nazywał uogólnionymi filtrami dolnoprzepustowymi. Korzystając z powyższych własności funkcji \( m_0 \) oraz ze wzoru na \( m_1 \) można pokazać, że
\[
\begin{align*}
\begin{cases}
m_0(0) &= 0 \\
m_0(\pi) &= 1
\end{cases}
\end{align*}
\] (3.27)

Podobnie jak poprzednio, funkcje spełniające powyższy warunek będę nazywał uogólnionymi filtrami górnoprzepustowymi. Rysunek (3.4) przedstawia przykładowe wykresu modułu funkcji \( m_0 \) i \( m_1 \) dla falki Daubechies o numerze 5.

Rysunek 3.4: Wykres \(|m_0|\) oraz \(|m_1|\) dla falki Daubechies przy \( N = 5 \)
3.5.2. Filtracja pasmowa

W teorii przetwarzania sygnałów znana jest operacja filtracji pasmowej. Polega ona na podziale dziedziny częstotliwościowej na kilka niezależnych pasm, w których sygnał jest następnie niezależnie przetwarzany.

![Diagram of bandpass filter](image.png)

Rysunek 3.5: Układ realizujący filtrację dwupasmową

Z powodów, które wyjaśnię wkrótce chciałbym opisać przykład układu realizującego filtrację dwupasmową. Schemat takiego układu przedstawia rysunek (3.5). Na rysunku tym \( H_i \) oraz \( F_i \) oznaczają pewne filtry typu FIR. Symbol ↓ 2 oznacza tzw. decymator, którego działanie polega na usunięciu co drugiego elementu z ciągu wejściowego. Z kolei symbol ↑ 2 oznacza tzw. ekspander, a jego działanie polega na wstawieniu zera pomiędzy każde dwie próbki ciągu wejściowego.

Na początku opiszę co się dzieje z sygnałem pomiędzy filtrami (rysunek 3.6).

![Diagram of decimator-expander](image.png)

Rysunek 3.6: Układ decymator-ekspander

Jeśli \( y = (\ldots, y_0, y_1, \ldots) \) to po przejściu przez decymator otrzymamy \( (\ldots, y_0, y_2, y_4, \ldots) \). Dalej po przejściu przez ekspander otrzymamy, że \( u = (\ldots, y_0, 0, y_2, 0, y_4, \ldots) \). Sygnał ten możemy zapisać następująco

\[
u = \frac{1}{2}(\ldots, y_0 + y_0, y_1 - y_1, y_2 + y_2, y_3 - y_3, \ldots) =
\]

\[
= \frac{1}{2}[\ldots, y_n, \ldots] + [\ldots, (-1)^n y_n, \ldots]
\]

co po zastosowaniu transformaty Fouriera prowadzi do

\[
U(\xi) = \frac{1}{2}[Y(\xi) + Y(\xi + \pi)]
\]

Zatem transformata Fouriera sygnału \( \hat{x} \) na wyjściu układu będzie dana wzorem

\[
\hat{X}(\xi) = \frac{1}{2}[H_0(\xi)F_0(\xi) + H_1(\xi)F_1(\xi)]X(\xi)
\]

\[
+ \frac{1}{2}[H_0(\xi + \pi)F_0(\xi) + H_1(\xi + \pi)F_1(\xi)]X(\xi + \pi)
\]

26
Z powyższego równania wynika, że gdyby transformaty filtrów spełniały układ równań

\[
\begin{align*}
H_0(\xi)F_0(\xi) + H_1(\xi)F_1(\xi) &= 2 \\
H_0(\xi + \pi)F_0(\xi) + H_1(\xi + \pi)F_1(\xi) &= 0
\end{align*}
\]

(3.28)
to sygnał wyjściowy \( \hat{x} \) byłby identyczny z sygnałem oryginalnym \( x \). Własność ta nazywana jest \textit{własnością doskonałej rekonstrukcji sygnału} (ang. \textit{Perfect Reconstruction Property}).

Opierając się na lemach (3.2.3) i (3.2.5) otrzymuję, że jeśli przyjmę

\[
\begin{align*}
H_0(\xi) &= 2^{1/2}m_0(\xi) \\
H_1(\xi) &= 2^{1/2}m_1(\xi) \\
F_0(\xi) &= 2^{1/2}m_0(\xi) \\
F_1(\xi) &= 2^{1/2}m_1(\xi)
\end{align*}
\]

to otrzymam układ mający własność doskonałej rekonstrukcji. Co więcej działanie powyższego układu jest identyczne z opisany wcześniej algorytmem wyznaczania współczynników falkowych.

Rysunek (3.7) ujawnia bardzo ważną własność układu filtrów falkowych. Mianowicie współczynniki falkowe zawierają jednocześnie informację o charakterystyce częstotliwościowej (kołowe podprzestrzenie \( W_j \)) oraz informację czasową (falki w kolejnych podprzestrzeniach). Każda podprzestrzeń \( W_j, \ j \leq 0 \) odpowiada pewnemu fragmentowi dziedziny częstotliwości, co w sumie pozwala scharakteryzować własności częstotliwościowe sygnału. Dodatkowo, przeciwnie do dyskretniej transformaty Fouriera, współczynniki falkowe, z tych podprzestrzeni, zawierają także informację czasową. W literaturze często spotyka się określenie tzw. \textit{dziedziny czasowo-częstotliwościowej}.
Rysunek 3.7: Podział osi częstotliwości w wyniku iterowania układem filtrów falkowych
Rozdział 4

Rozpoznawanie mówców

4.1. Wprowadzenie

Problem rozpoznawania mówców jest jednym z problemów związanych z przetwarzaniem mowy ludzkiej. Można go podzielić na dwa podproblemy

- **Identyfikacja** Na podstawie nagrania należy ustalić tożsamość osoby
- **Weryfikacja** Na podstawie nagrania należy zweryfikować tożsamość osoby

Każdy z powyższych podproblemów można dalej uszczegółowić następująco

- Rozpoznawanie niezależne/zależne od treści nagrania
- System zamknięty/otwarty, gdzie przez system zamknięty rozumieim, taki, który nie potrafi reagować na nieznanych mówców a przez system otwarty taki, który potrafi reagować w takim przypadku.

Systemy rozpoznawania mówców mogą mieć wiele zastosowań. Niektóre z nich to

- systemy zabezpieczeń (np. banki)
- telekomunikacja (np. usługi na telefon, inteligentne automatyczne sekretarki)
- kryminalistyka (np. identyfikacja przestępców)

Budowa typowego systemu rozpoznawania mówców przedstawiona jest na rysunku (4.1). Widać, że systemy te składają się z dwóch części. Pierwsza część odpowiedzialna jest za kodowanie sygnału mowy. Wynikiem takiego kodowania jest przeważnie jakiś zbiór cech. Dzisiaj najbardziej powszechnymi metodami kodowania mowy w celach rozpoznawania mówców są

- Liniowa Predykcja (LPC)
- Cepstrum
- Współczynniki MEL
- Połączenie powyższych metod

Wynik etapu kodowania służy za wejście do następnego etapu, czyli tzw. klasifikatora. W skrócie można powiedzieć, że zadaniem klasifikatora jest znalezienie jakichś reguł, dzięki którym będzie możliwe rozróżnienie między poszczególnymi mówcami. W literaturze najczęściej wymienia się następujące rodzaje klasifikatorów
• Kwantyzacja wektorowa (VQ)
• Sieci neuronowe wielowarstwowe (MLP)
• Gauss Mixture Models (GMM)
• Połączenie powyższych metod

4.2. Zaproponowany system klasyfikacji

W niniejszym podrozdziale opiszę jak można wykorzystać falki do problemu rozpoznawania mówców. Powszechnie wiadomo, że falki sprawdziły się np. w przetwarzaniu obrazów. Moim celem jest pokazanie, że także przetwarzanie sygnału mowy za pomocą falek może dawać całkiem ciekawe wyniki.

4.2.1. Wybór cech - analiza frekwencyjna

Ciekawą własnością falek jest to, że dla rzeczywistych sygnałów liczba istotnych współczynników falkowych jest niewielka w stosunku do długości sygnału. Jest to tak zwana własność *kompresji*. Postanowilem wykorzystać powyższą własność do kodowania sygnału mowy.

Niech \( \{f_k\}_{k=1,...,K} \) będzie zbiorem sygnałów o długości \( N \). Pojawia się pytanie, czy można wydobyć jakieś wspólne cechy dla tych funkcji. Piotr Wojdyło w swojej pracy [9] zaproponował następującą metodę

1. Ustal stałe \( \theta > 0 \) i \( 0 < M < K \), \( \theta \in \mathbb{R} \), \( M \in \mathbb{Z} \)
2. Dla każdego sygnału \( f_k \) znajdź indeksy tych współczynników falkowych, które co do modułu są większe od \( \theta \)
3. Mając tak indeksy wybierz, te które występowały dla co najmniej \( M \) sygnałów
4. Otrzymany zbiór indeksów jest szukanym zbiorem cech
Można powiedzieć, że metoda ta pozwala wyznaczyć falki, których współczynniki były naj-
większe dla jak największej liczby sygnałów.

W przypadku pracy Wojdyłły zaproponowana metoda dała bardzo dobre rezultaty. Po-
stanowieniem zatem zastosować tę metodę do kodowania sygnału mowy. Jednakże okazało się,
że jest to metoda zbyt mało elastyczna. Chodziło mianowicie o arbitralnie wybieraną próg
\( \theta \). Bardzo szybko okazało się, że dla różnych mówców trzeba było wybierać całkiem różne \( \theta \),
przy tej samej liczbie \( M \). Rozwiązanie tego problemu okazało się bardzo proste

1. Ustal \( \theta \in (0,1] \) oraz \( 0 < M < K, M \in \mathbb{Z} \)

2. Dla każdego sygnału wybierz tyle największych, co do modułu, współczynników falko-

3. Zlicz ile razy był wybierany współczynnik o danym indeksie dla całego zbioru sygnałów

4. Wybierz \( M \) najczęstszych

Zaletą tak zmodyfikowanej metody jest to, że stała \( \theta \) oznacza jakość aproksymacji jaką chem

4.2.2. Klasyfikatory

W celu lepszego zbadania nowej metody kodowania sygnału mowy postanowiłem zastosowa-
wać dwa rodzaje klasyfikatorów. Pierwszym był klasyczny klasyfikator oparty na wektorowej
kwantyzacji (VQ), drugim zaś był klasyfikator regulowy PART\(^1\). Zastosowanie dwóch róż-
nych klasyfikatorów pozwoliło mi wyciągnąć dużo więcej wniosków na temat współczynników
falkowych. W przypadku klasyfikatora VQ, głównie brałem pod uwagę stosunek jakości kla-

1\(^1\) Jest to klasyfikator budujący reguły na podstawie słynnego drzewa decyzyjnego C4.5 stworzo-

Dodatkowo w przypadku algorytmu PART zastosowałem następujące schematy klasyfikacji:

- Wszyscy razem Algorytm PART generował reguły ze zbioru wszystkich danych tre-

- Pary Dla każdej pary mówców zbudowałem oddzielny klasyfikator rozróżniający między

\( \theta \) procent energii całego sygnału

Dla każdego sygnału wybierz tyle największych, co do modułu, współczynników falko-

\( \theta \) procent energii całego sygnału

Zaletą tak zmodyfikowanej metody jest to, że stała \( \theta \) oznacza jakość aproksymacji jaką chem zapewnić. Intuicyjnie w wyniku zastosowania tej metody wyznaczane są falki, które najczęściej brały udział w kompresji sygnałów. W wyniku eksperymentów ustaliłem, że wybranie \( \theta \) z przedziału \( (0.9, 1) \) dawało najciekawsze rezultaty.
4.3. Nagrania

4.3.1. Baza własna

W nagraniach wzięły udział cztery osoby (dwie kobiety i dwóch mężczyzn). Za tekst treningowy posłużył fragment opowiadania E.A. Poe „Opowieść Artura Gordona Pyma z Nantucket”. Każde nagranie trwało ok. 4min. Nagrania były przeprowadzane przy pomocy karty dźwiękowej podłączonej do komputera typu PC oraz prostego mikrofonu. Częstotliwość próbowania wynosiła 44100Hz a dokładność 16 bitów na próbkę. Z powodu niskiej jakości mikrofonu nagrania zostały sztucznie podgłośnione. Dodatkowo z powodów praktycznych zmniejszono szybkość próbowania do 8000Hz (po uprzednim przefiltrowaniu przez filtr dolnoprzepustowy o częstotliwością odcięcia 4000Hz). Oprócz nagrań treningowych stworzono też nagrania testowe. Nagranie testowe składały się z następujących wypowiedzi:

- _Nazywam się ..._
- _Mój numer to ..._
- _Cześć to ja. Mam bardzo piłą sprawę, bardzo cię proszę o telefon_
- _Lista 64 słów_

Czas trwania tych nagrań wahał się od 2-50 sekund. W sumie przygotowałem 19 nagrań testowych.

4.3.2. Baza hVd

Baza zawiera kilkanaście izolowanych słów wypowiedzianych w języku angielskim przez kilkunastu mówców. Nagrania były wykonane w cichym pomieszczeniu a następnie spróbkowane z częstotliwością 10kHz oraz dokładnością 12 bitów na próbkę. Do celów mojego eksperymentu wybrałem następujących mówców:

**Mężczyźni**: Andrew, Andy, Bill, Tim, David, Mike, Nick

**Kobiety**: Sarah, Sue, Wendy, Rose, Kate, Jo, Penny

**Dzieci**: Alex

Każdy z wybranych mówców nagrał 18 izolowanych słów w języku angielskim.


4.4. Opis eksperymentów

4.4.1. Przetwarzanie nagrań

Ponieważ nagrania były zbyt długie aby je przetwarzać w całości trzeba było podzielić je na kawałki. Wybrano kawałki o długości 256 próbek co odpowiadało około 25-32ms. Był to kompromis pomiędzy wygodą implementacyjną a długością trwania fonemu. Każdy fragment
nagrania był przemnażany przez okienko Hanninga. W związku z tym, że w wyniku okienko-
wania traci się dane przy końcach fragmentów zastosowano zakładki, tzn. następnny fragment „zahaczał” o poprzedni. Wybrano długość zakładki 10% lub 50% długości wycinka. Następnie w przypadku liniowej predykcji stosowano preemfazę (oprócz jednego eksperymentu). Dopiero tak przetworzone wycinki były zamieniane na współczynniki falkowe oraz liniowej predykcji.

4.4.2. Redukcja wymiaru danych

W przypadku współczynników falkowych zastosowałem zmodyfikowaną analizę frekwencyjną w celu wybrania istotnych współczynników. Dzięki temu uzyskałem redukcję wymiaru danych z 256 do 4,11 lub 10 w zależności od eksperymentu. Liczba współczynników uzyskiwanych z liniowej predykcji w każdym przypadku odpowiadała liczbie współczynników falkowych.

4.4.3. Wyniki dla własnej bazy nagrań

<table>
<thead>
<tr>
<th>Mówca</th>
<th>Współczynniki</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>15 8 5 6 14 9 20 19 21</td>
</tr>
<tr>
<td>M2</td>
<td>8 15 5 6 9 14 19 20 18</td>
</tr>
<tr>
<td>K1</td>
<td>8 15 9 14 20 19 5 21 18</td>
</tr>
<tr>
<td>K2</td>
<td>15 8 14 9 20 19 21 18 22</td>
</tr>
<tr>
<td>Ostatecznie</td>
<td>5 6 8 9 14 15 18 19 20 21 22</td>
</tr>
</tbody>
</table>

Tablica 4.1: Wybrane współczynniki falkowe dla bazy własnej, po zastosowaniu zmodyfikowanej analizy frekwencyjnej przy $\theta = 0.9$ i $M = 9$

Klasyfikator regułowy

Do eksperymentu wybrałem trzech mówców z bazy: M1, K1, K2. Zarówno w przypadku falek jak i liniowej predykcji uzyskiwany zbiór reguł był nieficznym a same reguły były dość krótkie (kilku składnikowe). Współczynniki falkowe dały dość satysfakcjonujące rezultaty chociaż okazały się gorsze (pod względem jakości klasyfikacji) od liniowej predykcji. Poniższe tabele zawierają wyniki jakie uzyskałem stosując trzy schematy klasyfikacji

- Wszyscy razem

<table>
<thead>
<tr>
<th>Mówca</th>
<th>Falki</th>
<th>Liniowa predykcja</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>40%</td>
<td>100%</td>
</tr>
<tr>
<td>K1</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>K2</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td>58.3%</td>
<td>100%</td>
</tr>
</tbody>
</table>

- Pary

<table>
<thead>
<tr>
<th>Mówca</th>
<th>Falki</th>
<th>Liniowa predykcja</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>K1</td>
<td>66.6%</td>
<td>100%</td>
</tr>
<tr>
<td>K2</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td>91.6%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Klasyfikator VQ

Do eksperymentu z klasyfikatorem VQ użyłem wszystkich mówców. Wielkość słownika wahała się od 16 do 1024 centroidów dla każdego mówcy. Dodatkowo zamieściłem wyniki jakie uzyskałem stosując liniową predykcję bez preemfazy. Znowu falki okazały się gorsze od liniowej predykcji (z preemfazą), jednak dużo lepsze od liniowej predykcji bez preemfazy.

- **Falki**

<table>
<thead>
<tr>
<th>Mówca</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>M2</td>
<td>66%</td>
<td>66%</td>
<td>50%</td>
<td>33%</td>
<td>33%</td>
<td>50%</td>
<td>33%</td>
</tr>
<tr>
<td>K1</td>
<td>50%</td>
<td>50%</td>
<td>25%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>K2</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td><strong>78.94%</strong></td>
<td><strong>78.94%</strong></td>
<td>68.42%</td>
<td>68.42%</td>
<td>68.42%</td>
<td>73.68%</td>
<td>68.42%</td>
</tr>
</tbody>
</table>

- **Liniowa predykcja z preemfazą**

<table>
<thead>
<tr>
<th>Mówca</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>83.33%</td>
<td>100%</td>
<td>100%</td>
<td>66.66%</td>
<td>83.33%</td>
<td>50%</td>
<td>66.66%</td>
</tr>
<tr>
<td>M2</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>K1</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>K2</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td><strong>94.73%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td>89.47%</td>
<td><strong>94.73%</strong></td>
<td><strong>84.21%</strong></td>
<td><strong>89.47%</strong></td>
</tr>
</tbody>
</table>

- **Liniowa predykcja bez preemfazy**

<table>
<thead>
<tr>
<th>Mówca</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>60%</td>
<td>80%</td>
<td>100%</td>
<td>100%</td>
<td>80%</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>M2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>K1</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>K2</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td><strong>55.55%</strong></td>
<td><strong>61.11%</strong></td>
<td><strong>66.66%</strong></td>
<td><strong>66.66%</strong></td>
<td>61.11%</td>
<td><strong>55.55%</strong></td>
<td><strong>55.55%</strong></td>
</tr>
</tbody>
</table>

4.4.4. Wyniki dla bazy hVd - indywidualnie

Do tego eksperymentu wybrałem następujących mówców

**Mężczyźni** Andrew, Andy, Bill, Tim

**Kobiety** Sarah, Sue, Rose, Wendy

**Dzieci** Alex

Jako część treningową wybrałem 9 spośród 18 słów (w sposób losowy). Testy przeprowadzałem na pozostałych 9 słowach. W sumie czas nagrania treningowego dla każdego mówcy wynosił od 5–9s a czas nagrań testowych 0.5–1s. W wyniku analizy frekwencyjnej wybrałem następujące współczynniki falkowe: 15, 14, 13, 8, 19, 5, 4, 18, 187, 20.
**Klasyfikator regułowy**

W przypadku klasyfikatora regułowego zastosowałem schemat *Pary*. Zastosowanie falek zauwaczało dość małą liczbą reguł potrzebnych do rozróżnienia większości par (3-99). Niestety w przypadku niektórych par algorytm PART miał duże kłopoty ze zbudowaniem dobrego zbioru reguł. Oto uzyskane wyniki

<table>
<thead>
<tr>
<th>Mówca</th>
<th>Falki</th>
<th>Liniowa predykcja</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Andrew</td>
<td>33.33%</td>
<td>88.88%</td>
</tr>
<tr>
<td>Andy</td>
<td>55.55%</td>
<td>88.88%</td>
</tr>
<tr>
<td>Bill</td>
<td>88.88%</td>
<td>100%</td>
</tr>
<tr>
<td>Tim</td>
<td>88.88%</td>
<td>100%</td>
</tr>
<tr>
<td>Rose</td>
<td>100%</td>
<td>77.77%</td>
</tr>
<tr>
<td>Sarah</td>
<td>0.0%</td>
<td>88.88%</td>
</tr>
<tr>
<td>Sue</td>
<td>33.33%</td>
<td>33.33%</td>
</tr>
<tr>
<td>Wendy</td>
<td>100%</td>
<td>77.77%</td>
</tr>
<tr>
<td>Łącznie</td>
<td>66.66%</td>
<td>83.95%</td>
</tr>
</tbody>
</table>

**Klasyfikator VQ**

- **Falki**

<table>
<thead>
<tr>
<th>Mówca</th>
<th>16</th>
<th>32</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alex</td>
<td>88.88%</td>
<td>88.88%</td>
<td>100%</td>
</tr>
<tr>
<td>Andrew</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Andy</td>
<td>33.33%</td>
<td>77.77%</td>
<td>77.77%</td>
</tr>
<tr>
<td>Bill</td>
<td>88.88%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Tim</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Rose</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Sarah</td>
<td>55.55%</td>
<td>88.88%</td>
<td>77.77%</td>
</tr>
<tr>
<td>Sue</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Wendy</td>
<td>77.77%</td>
<td>88.88%</td>
<td>100%</td>
</tr>
<tr>
<td>Łącznie</td>
<td>82.71%</td>
<td>93.82%</td>
<td><strong>95.06%</strong></td>
</tr>
</tbody>
</table>
4.4.5. Wyniki dla bazy hVd - płeć

Eksperyment miał na celu zbadać możliwość identyfikacji płci na podstawie nagrania. W eksperymencie wzięli udział wszyscy mówcy wymienieni w opisie bazy. Do celów treningowych wybrałem po 9 słów od następujących mówców:

**Mężczyźni**: Andrew, Andy, Bill, Tim

**Kobiety**: Sarah, Sue, Rose, Wendy

Testy przeprowadziłem na nagraniach (nie biorących udział w treningu) pochodzących od mówców z bazy oraz na grupie zupełnie nowych mówców (3 kobiety, 3 mężczyźni). Analiza frekwencyjna zaowocowała wybraniem 4 współczynników falowych: 4, 5, 14, 15. Jako klasyfikator użyłem algorytmu opartego na wektorowej kwantyzacji. Tym razem współczynniki falowe okazały się dużo lepsze od liniowej predykcji, zarówno w przypadku mówców biorących udział w treningu jak i w przypadku zupełnie nowych mówców.

**Falki**

<table>
<thead>
<tr>
<th>Grupa</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wszyscy</td>
<td>76.11%</td>
<td>76.11%</td>
<td>80.55%</td>
<td><strong>81.66%</strong></td>
</tr>
<tr>
<td>Mówcy z treningu</td>
<td>95.83%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Nowi mówcy</td>
<td>62.96%</td>
<td>62.03%</td>
<td>67.59%</td>
<td>86.11%</td>
</tr>
</tbody>
</table>

**Liniowa predykcja**

<table>
<thead>
<tr>
<th>Grupa</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wszyscy</td>
<td><strong>74.44%</strong></td>
<td>66.66%</td>
<td>72.77%</td>
<td>71.11%</td>
</tr>
<tr>
<td>Mówcy z treningu</td>
<td>91.66%</td>
<td>83.33%</td>
<td>84.72%</td>
<td>83.33%</td>
</tr>
<tr>
<td>Nowi mówcy</td>
<td>62.96%</td>
<td>55.55%</td>
<td>64.81%</td>
<td>62.96%</td>
</tr>
</tbody>
</table>

4.5. Wnioski i propozycje rozszerzeń

Pierwszym nasuwającym się wnioskem jest to, że zaproponowana metoda wymaga wielu ulepszeń. Uważam, że uzyskiwane wyniki można polepszyć dobierając innego rodzaju falki, np.
falki na odcinku czy pakiety falkowe. Dodatkowo bardziej staranny dobór systemu klasyfikacyjnego (np. opartego na zbiorach przybliżonych) mógłby dać lepsze wyniki, jak to było w przypadku pracy [9]. Lepsza jakość uzyskiwana w przypadku liniowej predykcji wynikała prawdopodobnie z tego, że liniowa predykcja jest silnie związana z modelem generowania sygnału mowy a falki nie. Zaproponowana metoda nie wykorzystuje żadnych szczególnych cech sygnału mowy. Jest przez to metodą dużo bardziej ogólną. Eksperyment z rozpoznawaniem płci pokazuje, że liniowa predykcja jest bezradna gdy jej rząd jest niski. Daje to nadzieję, że staranniejszy dobór bazy falkowej mógłby zaowocować mniejszą liczbą wybranych współczynników a przez to mogłoby się okazać, że falki są lepsze od liniowej predykcji. Warto podkreślić, że w większości opisywanych w literaturze systemów liczba atrybutów jest niemniejsza niż kilkanaście. Własności kompresyjne falek mogą więc okazać się cenne, szczególnie w przypadku baz składających się z setek mówców. Podsumowując, mogę wskazać następujące kierunki rozszerzeń:

- Pakiety falkowe, które dają bardziej staranny podział dziedziny częstotliwości
- Falek drugiej generacji³, bardziej dopasowane do sygnału mowy
- Innego rodzaju klasyfikatory: sieci neuronowe, GMM, zbiory przybliżone.

³Pozycja [7]
Bibliografia