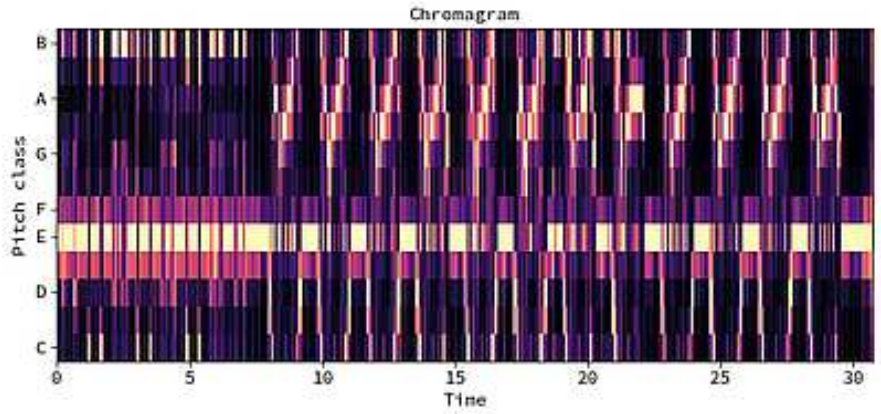
ANALIZA DŹWIĘKÓW MUZYCZNYCH



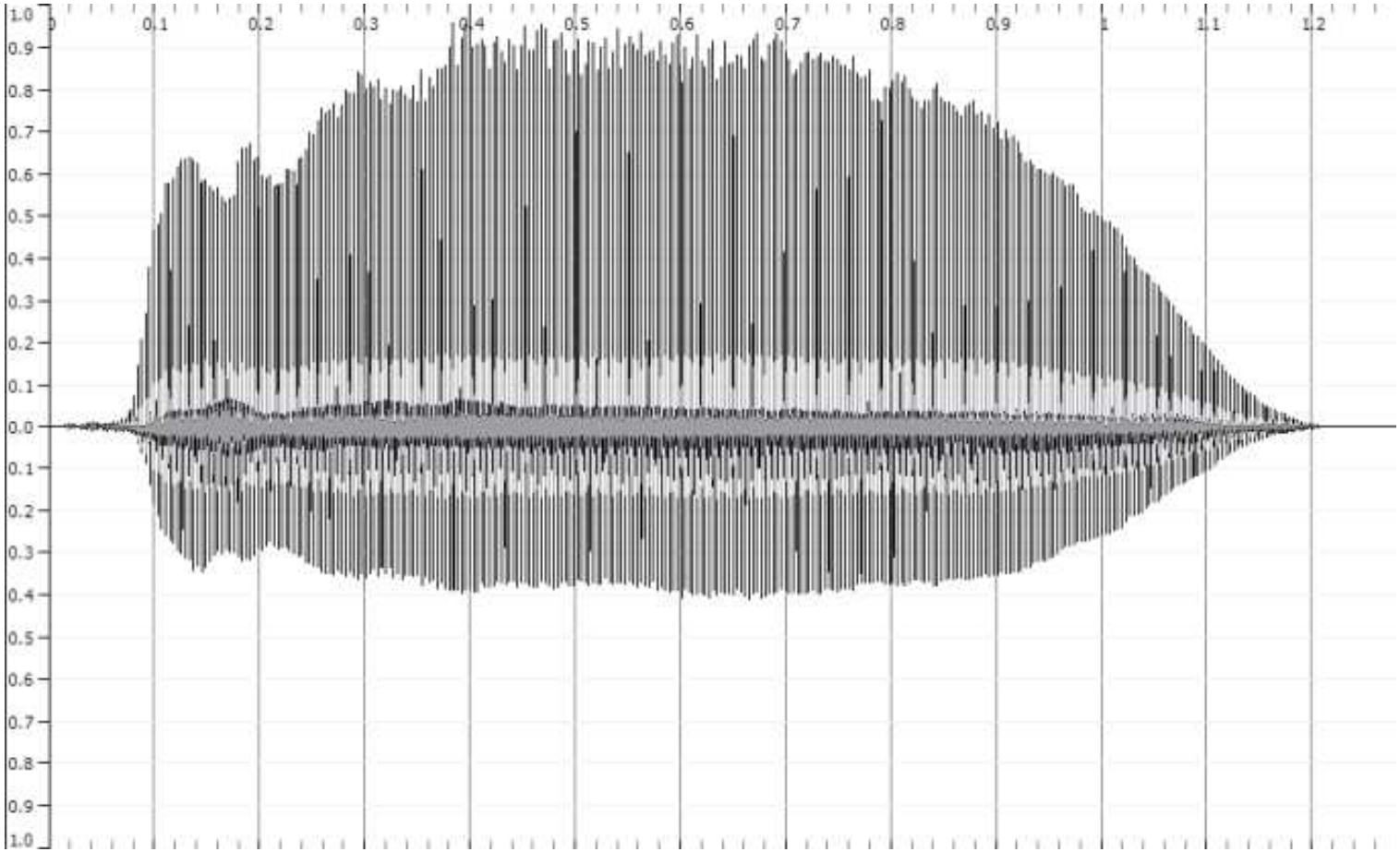
Opracowanie: Grzegorz Szwoch

Politechnika Gdańska, Katedra Systemów Multimedialnych

Analiza czasowa dźwięków muzycznych

Postać czasowa dźwięku trąbki:

Amplituda



Czas

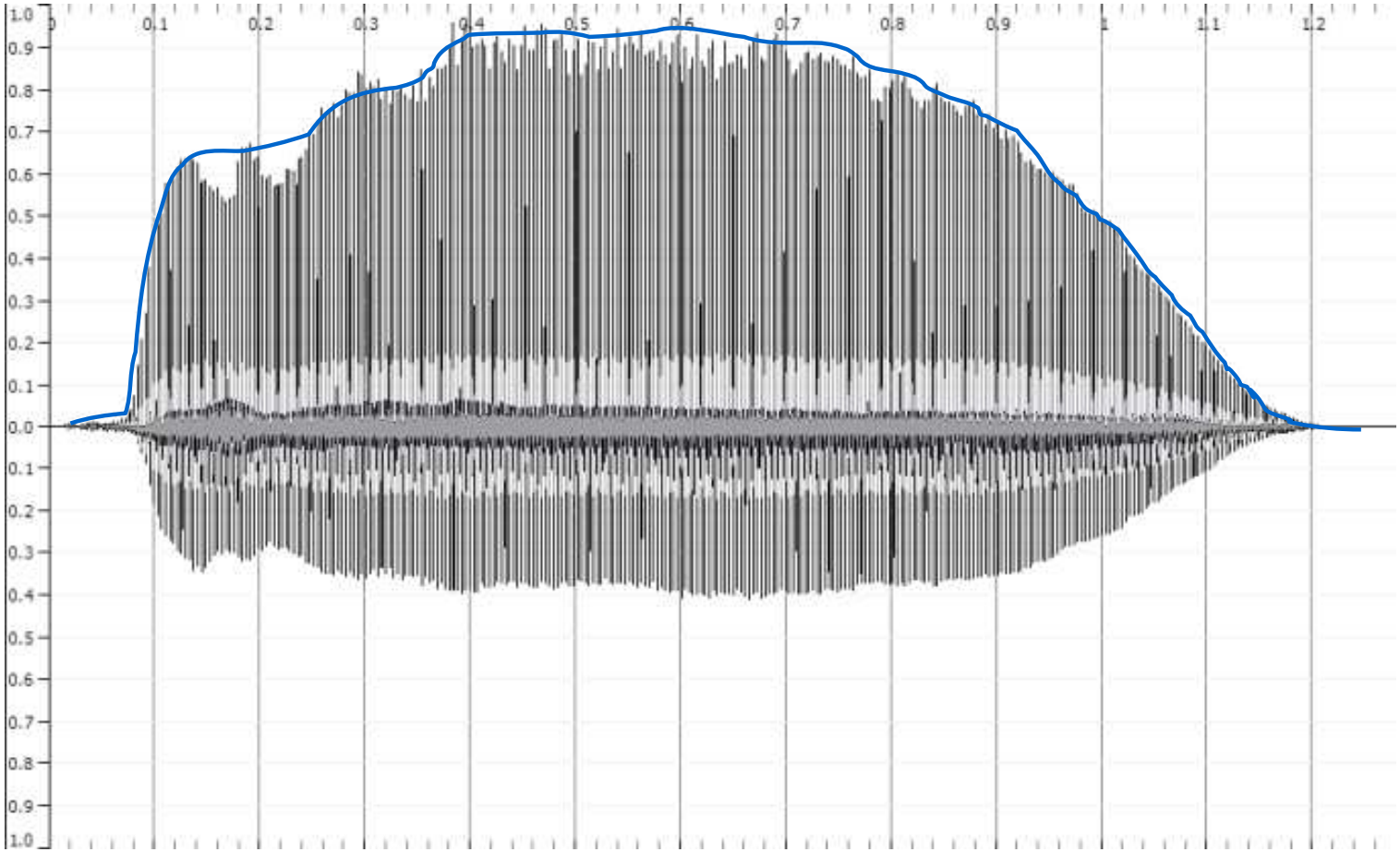
Obwiednia dźwięku

- Obwiednia (*envelope*): linia „podążająca” za szczytami wykresu czasowego.
- Informuje o zmianach głośności sygnału.
- Jest przydatna do wyznaczania:
  - początków nut (*onset detection*),
  - faz dźwięku (transjent, stan ustalony, wybrzmiewanie).
- Układ nadążania za obwiednią (*envelope follower*): wyznacza obwiednię dźwięku.
- Najprostsza realizacja: podniesienie sygnału do kwadratu i jego uśrednianie w krótkich fragmentach (np. średnia ruchoma).

Analiza czasowa - obwiednia

Obwiednia dźwięku trąbki:

Amplituda



Czas

Fazy dźwięku

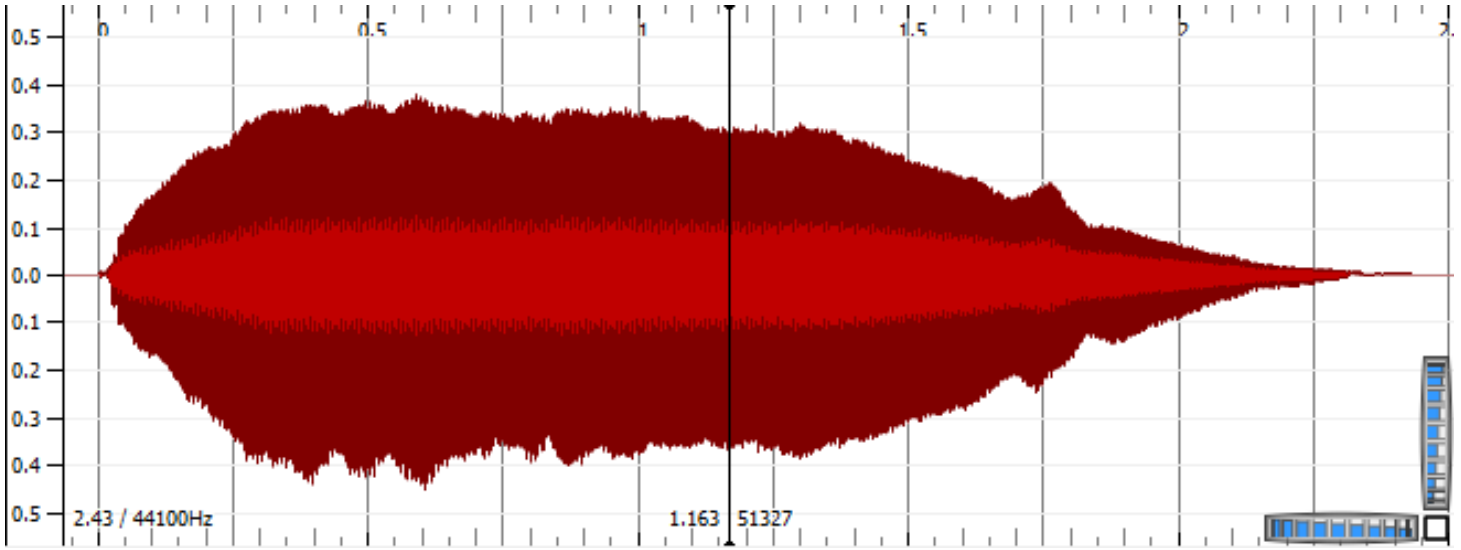
W przebiegu czasowym dźwięków muzycznych możemy wyróżnić fazy:

- ataku (*attack*) – transjent początkowy, budowanie dźwięku,
- podtrzymania (*sustain*) – stan ustalony, stabilny dźwięk,
- wybrzmiewania (*release*) – wygaszanie dźwięku.

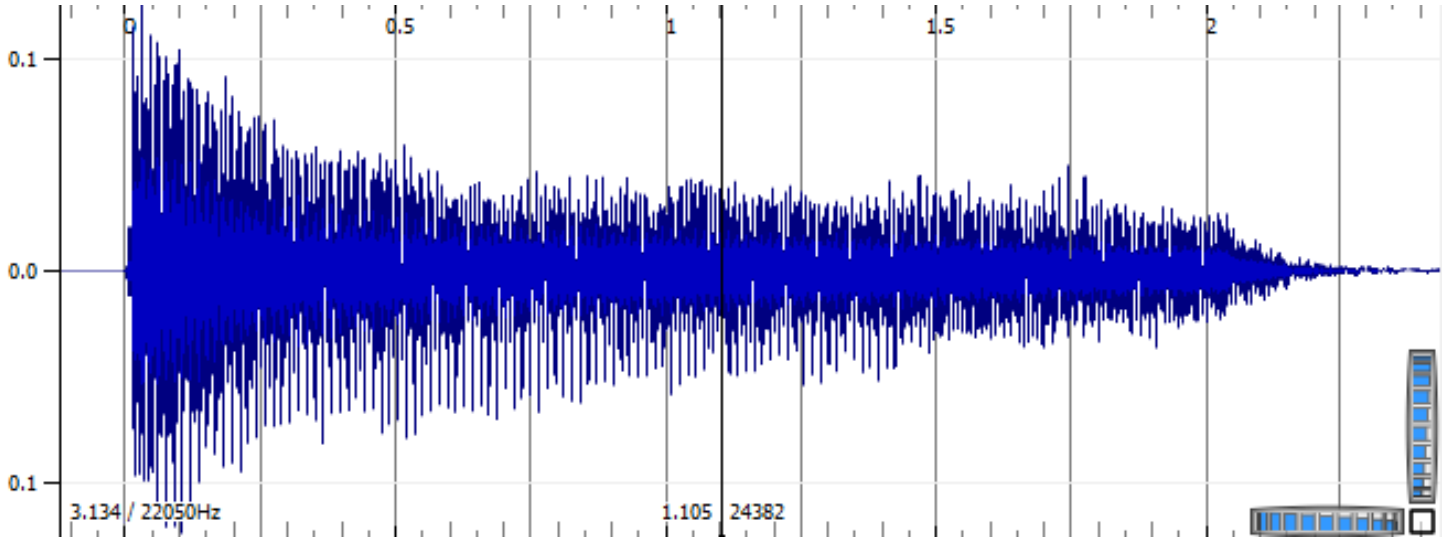
Czas trwania faz dźwięku zależy od:

- charakteru instrumentu (np. dęty/strunowy),
- sposobu gry (artykulacji).

Przykłady obwiedni dźwięków muzycznych



Trąbka



Fortepian

Analiza częstotliwościowa

- Widmo dźwięku – rozkład natężenia składowych dźwięku w zależności od częstotliwości tych składowych.
- Analiza częstotliwościowa – obliczenie widma dla wybranego fragmentu dźwięku (okna analizy).
- Jest to widmo statyczne (chwilowe)
  – dla określonego fragmentu dźwięku.
- Analiza częstotliwościowa dostarcza informacji o strukturze widmowej sygnału.
- Typowe dźwięki muzyczne mają najczęściej strukturę harmoniczną (wieloton).
- Widmo dźwięku decyduje o jego barwie. Pozwala rozróżniać instrumenty muzyczne.

Analiza częstotliwościowa

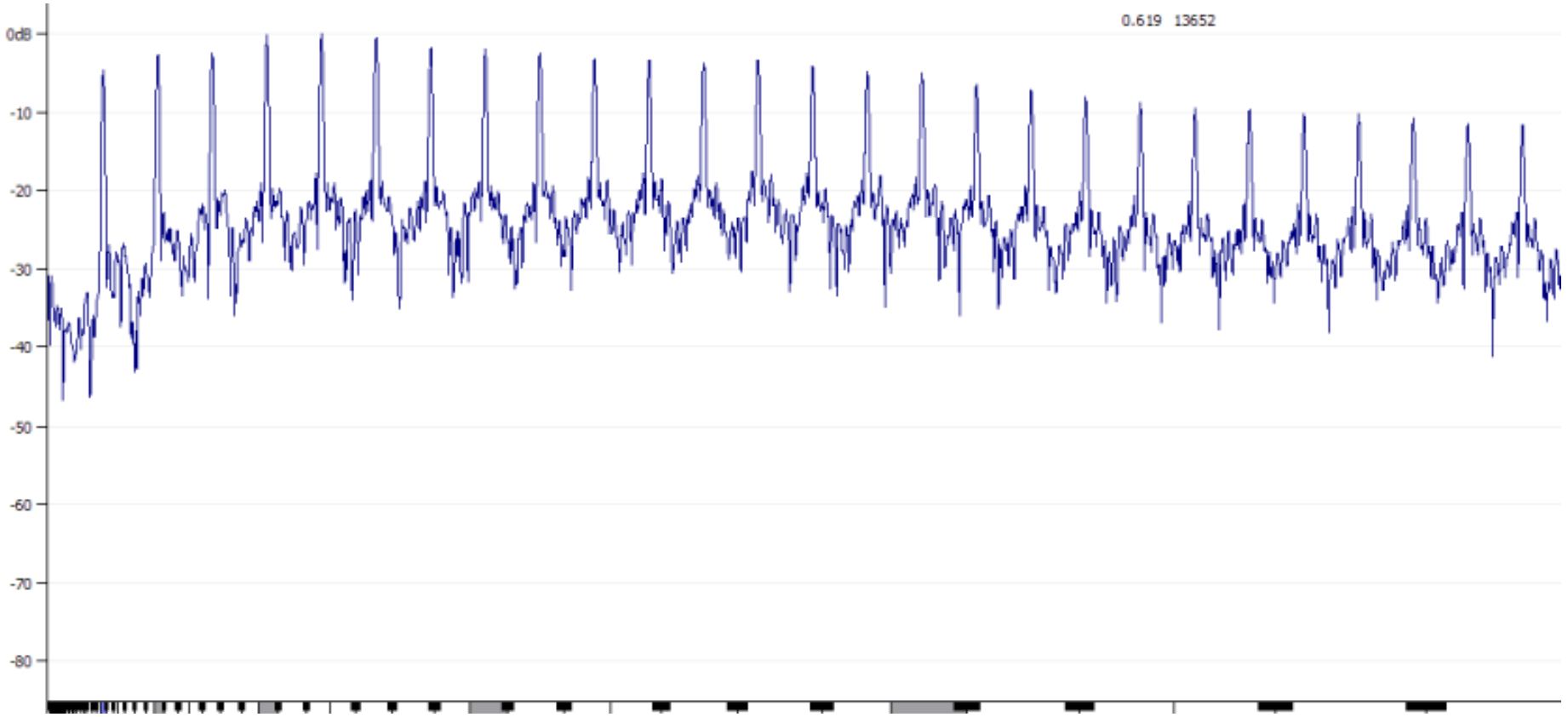
Jak obliczamy widmo:

- „wycinamy” z sygnału blok próbek,
- mnożymy go przez funkcję okna (np. Hamminga),
- obliczamy transformatę Fouriera (FFT)
  - przechodzimy do dziedziny częstotliwości,
- moduł FFT daje nam widmo amplitudowe w funkcji częstotliwości, kąt fazowy FFT – widmo fazowe.

Widmo amplitudowe sygnałów muzycznych przedstawiamy zazwyczaj w skali decybelowej ($20 \log_{10}$).

Analiza częstotliwościowa

Widmo statyczne dźwięku trąbki (w stanie ustalonym)
poziom widma [dB] w funkcji częstotliwości



Dźwięki perkusyjne

Dźwięki typu perkusyjnego mają widmo o charakterze szumowym, nie ma tutaj „prążków”.

Analiza czasowo-częstotliwościowa

- Widmo statyczne jest obserwowane w wybranym odcinku skali czasu.
- Widmo dźwięków muzycznych jest dynamiczne - zmienne w czasie („ewoluuje”).
- Potrzebujemy zbadać widmo w różnych punktach osi czasu, dla całego dźwięku.
- Analiza czasowo – częstotliwościowa: połączenie wyników analizy częstotliwościowej dla kolejnych odcinków dźwięku.
- STFT: *Short-term Fourier Transform*

Analiza czasowo-częstotliwościowa

Analiza czasowo-częstotliwościowa STFT:

- dzielimy sygnał na okna, często z zakładką,
- dla każdego okna obliczamy widmo statyczne (FFT),
- łączymy wyniki z poszczególnych okien.

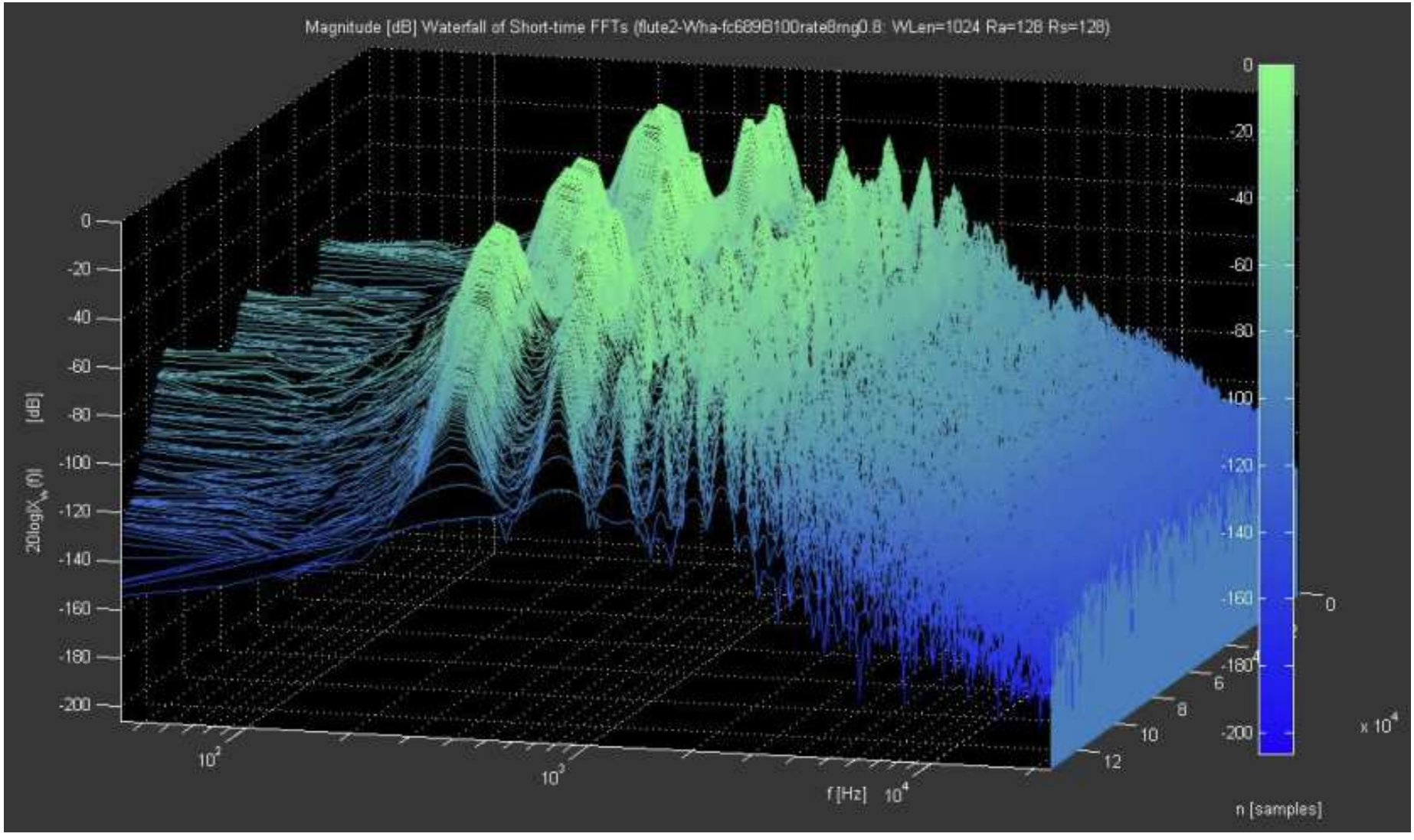
Powstaje trójwymiarowy opis sygnału:
czas / częstotliwość / amplituda widmowa.

Sposoby prezentacji wyniku STFT:

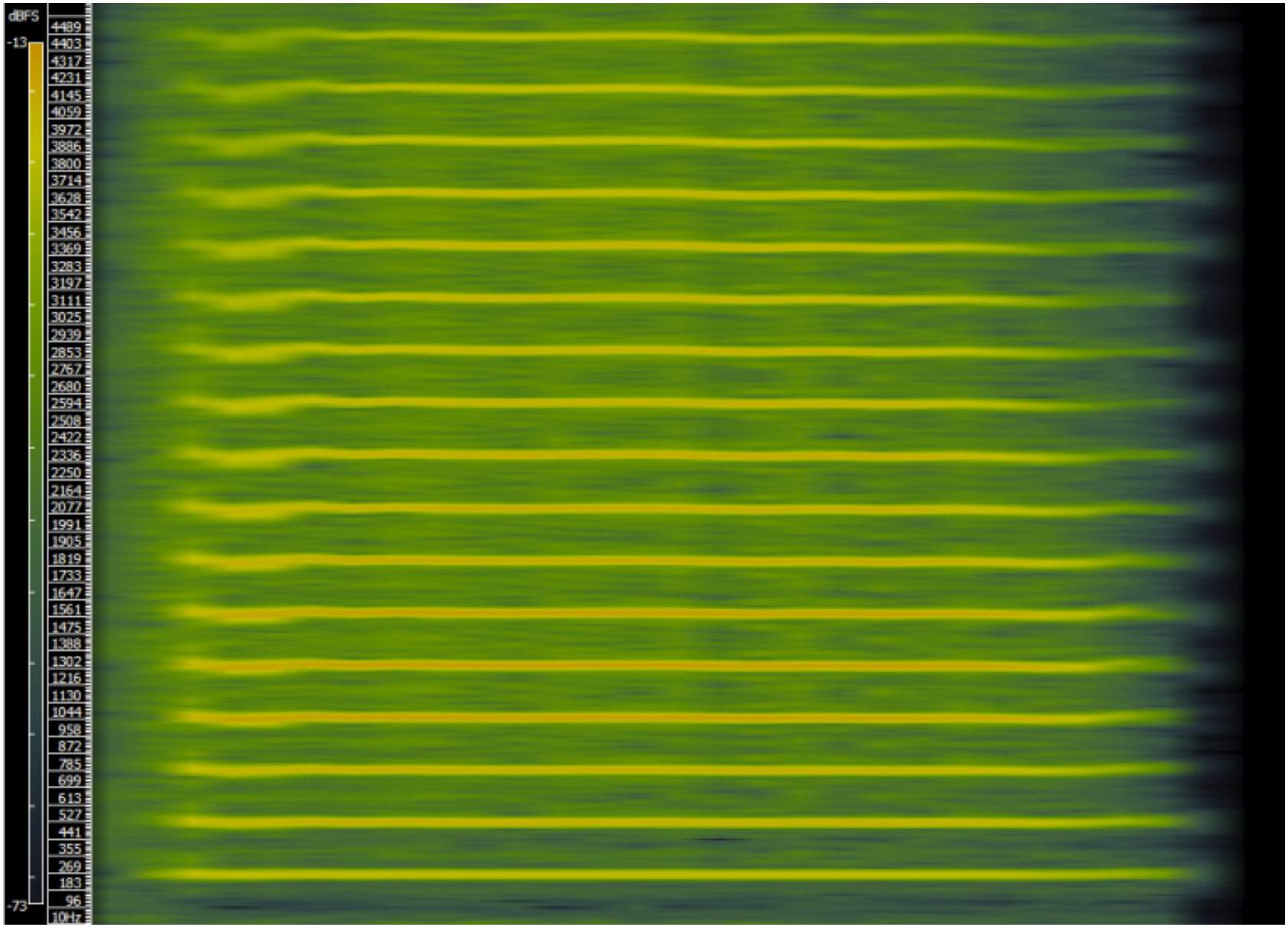
- wykres 3D typu *waterfall*,
- spektrogram 2D: czas/częstotliwość, amplituda widmowa przedstawiana za pomocą barwy lub intensywności.

Wykres typu waterfall

Ilustracyjny, dobrze widoczna zmienność widma, mało czytelne szczegóły



Spektrogram

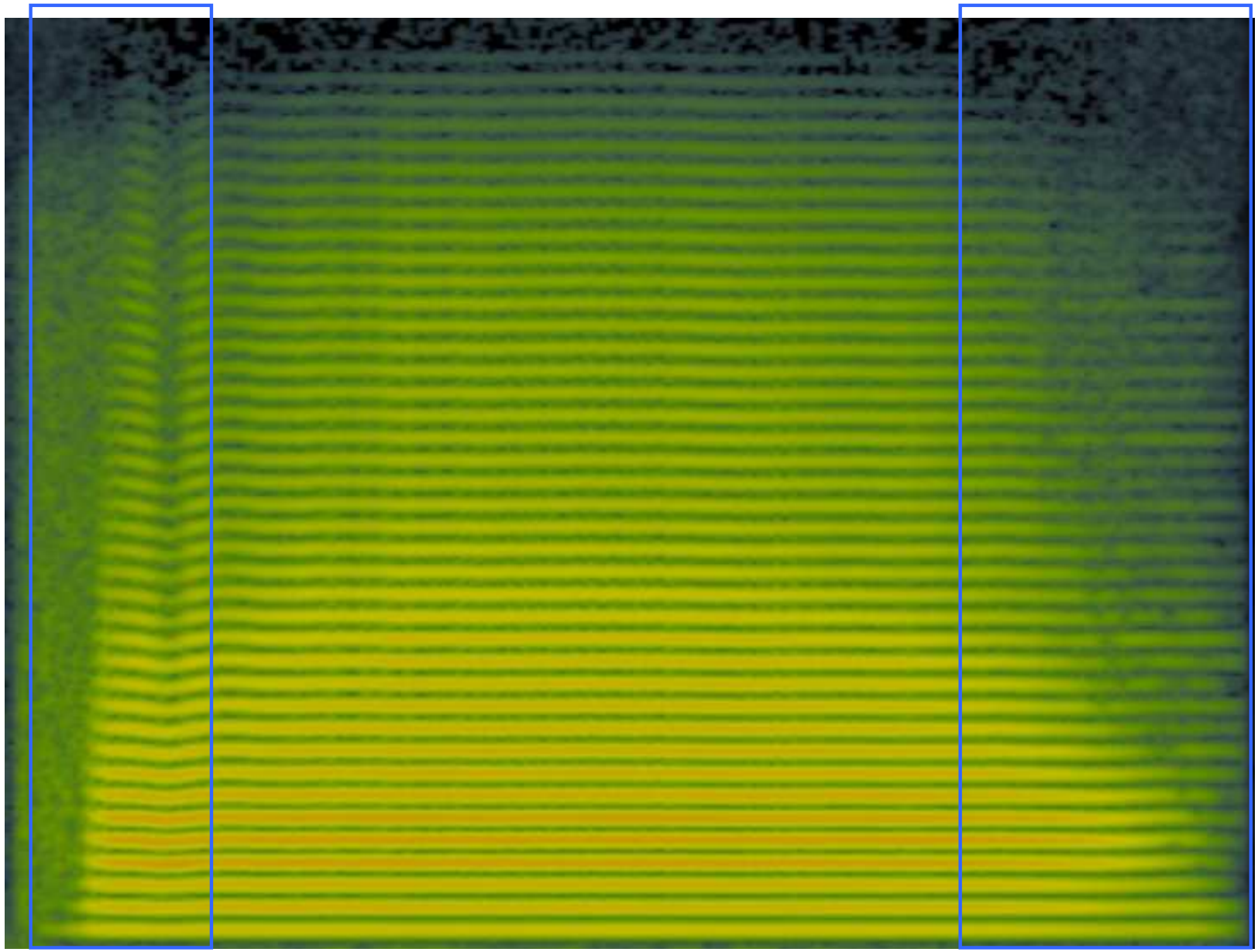


Spektrogram a fazy dźwięku

Obserwując spektrogram można wyznaczyć czasy trwania faz dźwięku:

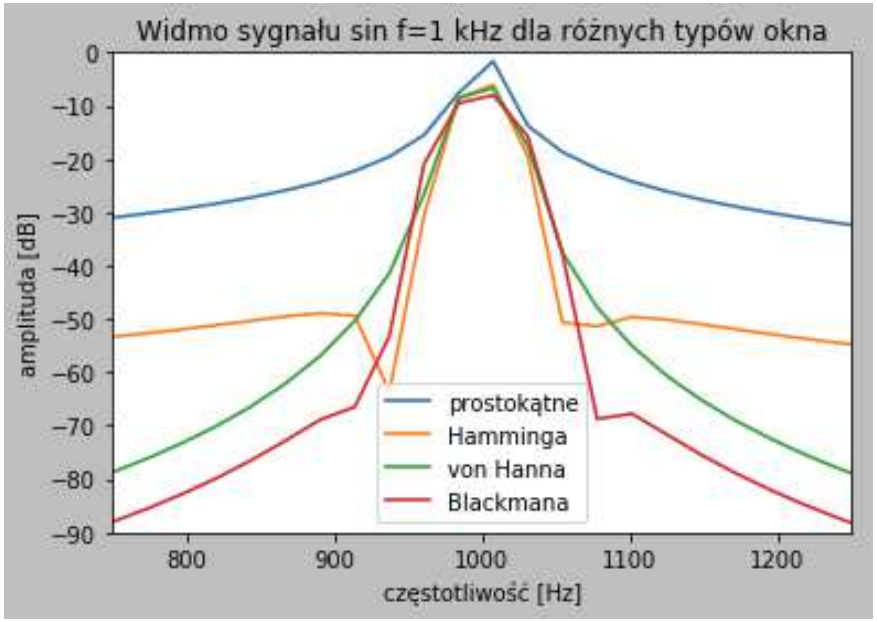
- faza ataku: stopniowe budowanie widma od niższych do wyższych składowych,
- stan ustalony – stabilne widmo, może się zmieniać na skutek artykulacji (np. wibrato),
- faza wybrzmiewania: stopniowe zanikanie widma od wyższych do niższych składowych.

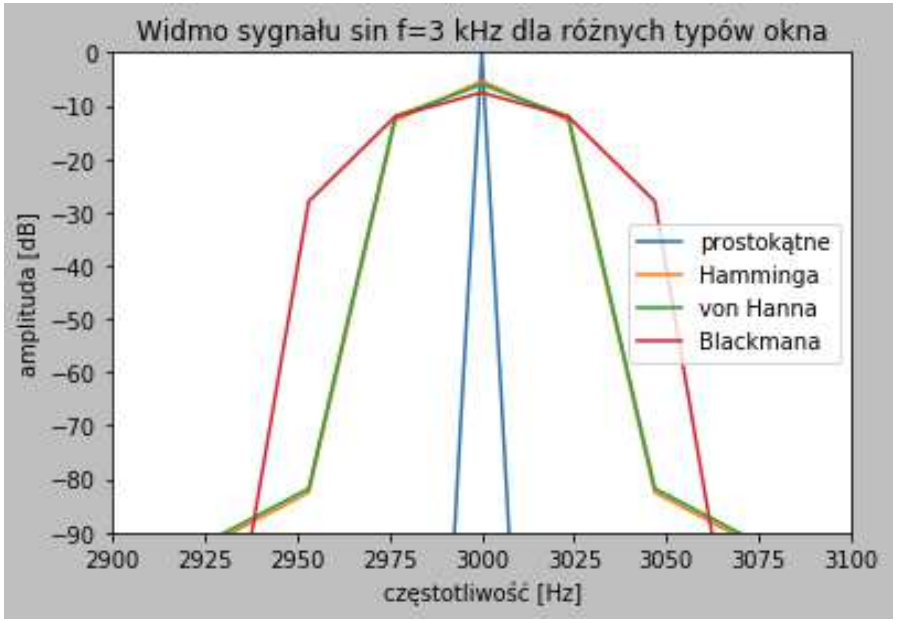
Spektrogram a fazy dźwięku



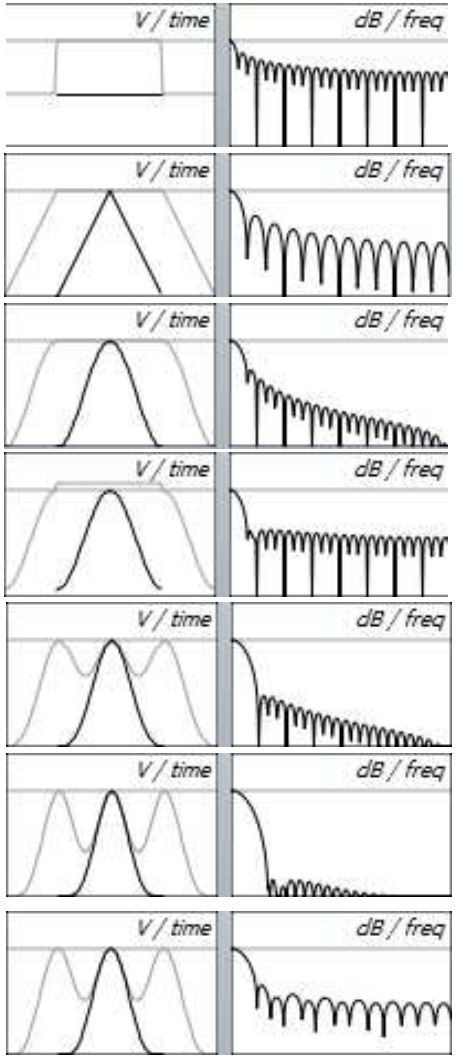
Funkcje okna

- Jeżeli wycinamy z sygnału blok próbek (stosujemy okno prostokątne), powstaje zjawisko przecieków widma.
- Aby temu zapobiec, stosujemy funkcje okna.
- Dobór okna jest kompromisem między tłumieniem listków bocznych a szerokością prążka.
- Nie ma okna „najlepszego” (uniwersalnego).





Funkcje okna



Prostokątne

Trójkątne (Bartletta)

Hanna („hanning”)

Hamminga

Blackmana

Blackmana-Harrisa

Parzena

Rozdzielczość częstotliwościowa

Jaki rozmiar okna analizy?

- Dla FFT zaleca się rozmiar 2^N.
- Weźmy więc 512 próbek.
- W wyniku FFT dostajemy tyle samo wartości, pokrywających zakres od $-F_S/2$ do $F_S/2$.
- Każda wartość FFT wpada do jednego z 512 przedziałów o szerokości $F_S/512$.
- Dla F_S = 48 kHz daje to 93,75 Hz
- Tyle wynosi rozdzielczość częstotliwościowa analizy – nie rozróżnimy dwóch składowych widma odległych o mniej niż 93,75 Hz.

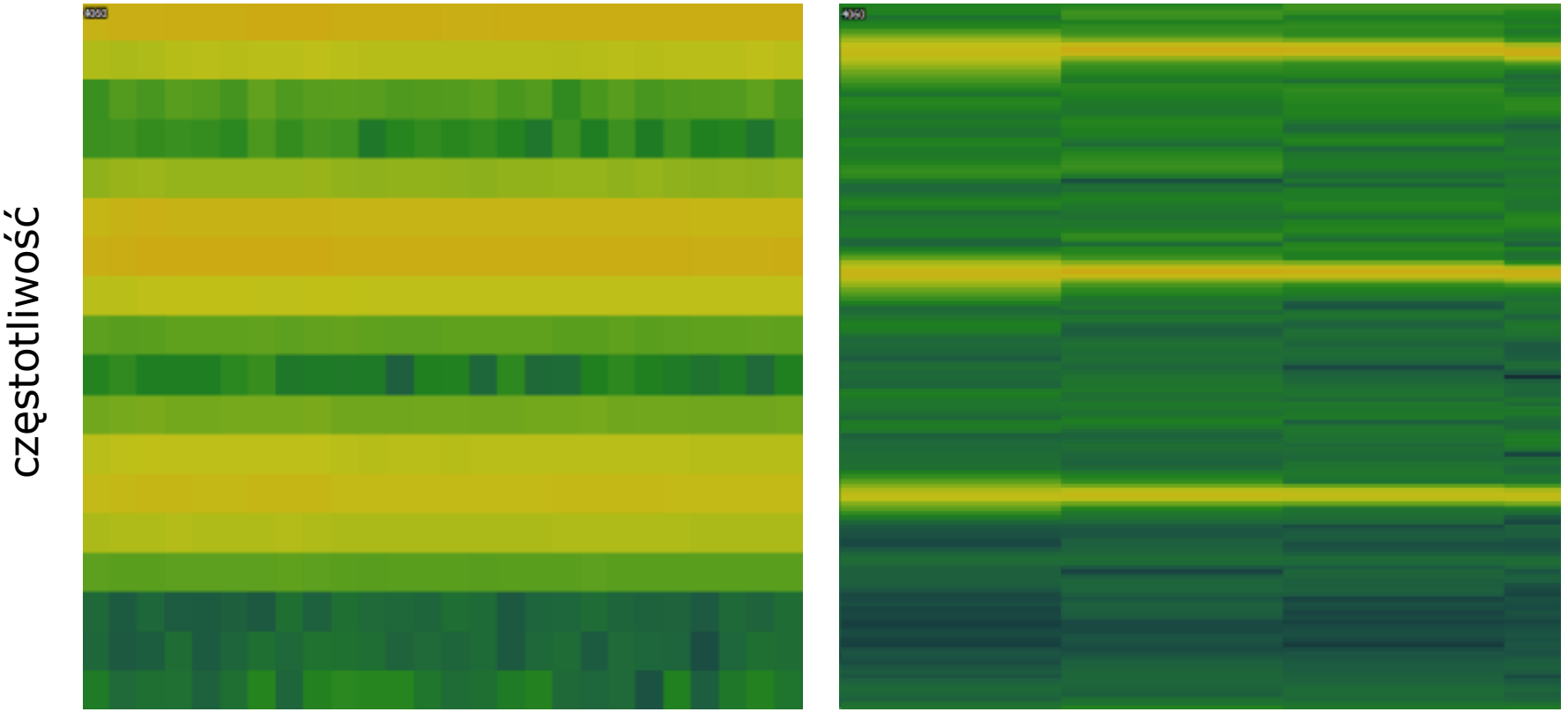
Rozdzielczość czasowa

Zatem czy większe okno jest lepsze?

- Weźmy 4096 próbek (8x więcej).
- Rozdzielczość częstotliwościowa (dla 48 kHz):
  11,72 Hz (8x większa).
- Ale teraz okno pokrywa zakres 85,33 ms
  (8x większy, wcześniej 10,67 ms).
- Tyle wynosi rozdzielczość czasowa analizy.
- Nie rozróżnimy zdarzeń czasowych
  w odstępie mniejszym niż 85,33 ms.
- Poprawiliśmy rozdzielczość częstotliwościową, ale pogorszyliśmy czasową.

Rozdzielczość czasowo-częstotliwościowa

Spektrogram dla okna 512 i 4096 próbek, wyłączone zakładkowanie



Zakładkowanie

- Zakładkowanie (*overlapping*) – sąsiednie przedziały czasowe analizy widmowej pokrywają się (przesuwamy okno o mniej niż jego długość).
- Najczęściej stosuje się zakładkę: 25%, 50%, 75% dł. okna.
- Korzyści:
  - zwiększenie rozdzielczości czasowej,
  - zmniejszenie zniekształceń sygnału spowodowanych mnożeniem przez funkcję okna.
- Wada: zwiększenie czasu analizy (więcej przedziałów).

Uzupełnianie zerami

Zero-padding: często stosowany „trik” przy STFT:

- bierzemy krótsze okno (np. 1024 próbki),
- uzupełniamy zerami, najczęściej do długości x2,
- obliczamy FFT z takiego bloku.

Nie uzyskujemy więcej danych w stosunku do FFT 1024, ale:

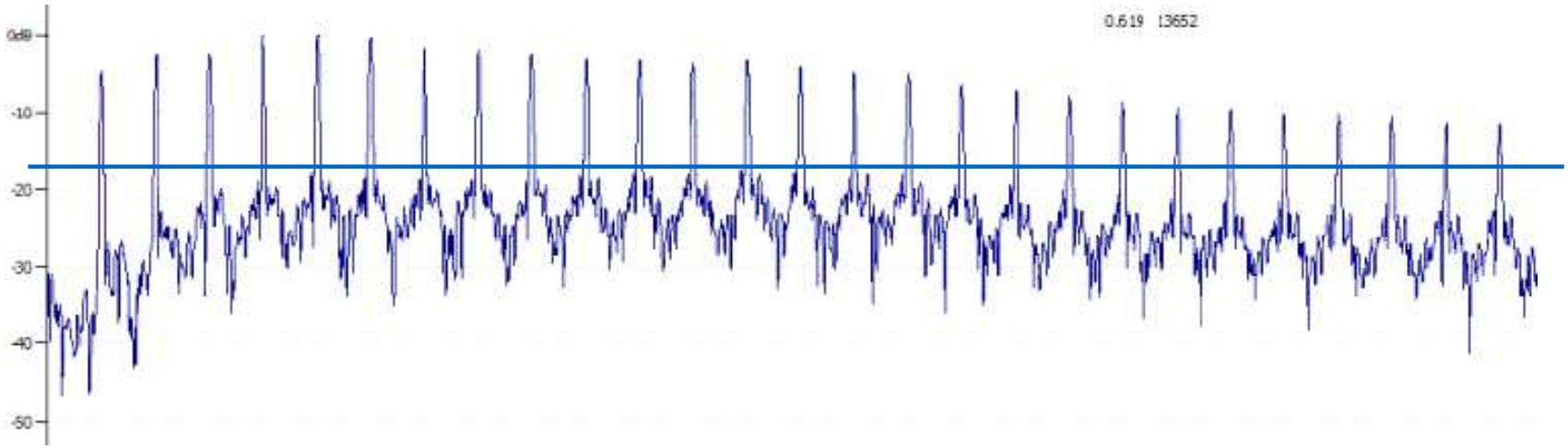
- zachowujemy (w przybliżeniu) r. częstotliwościową,
- poprawiamy r. czasową.

STFT – praktyczne rady

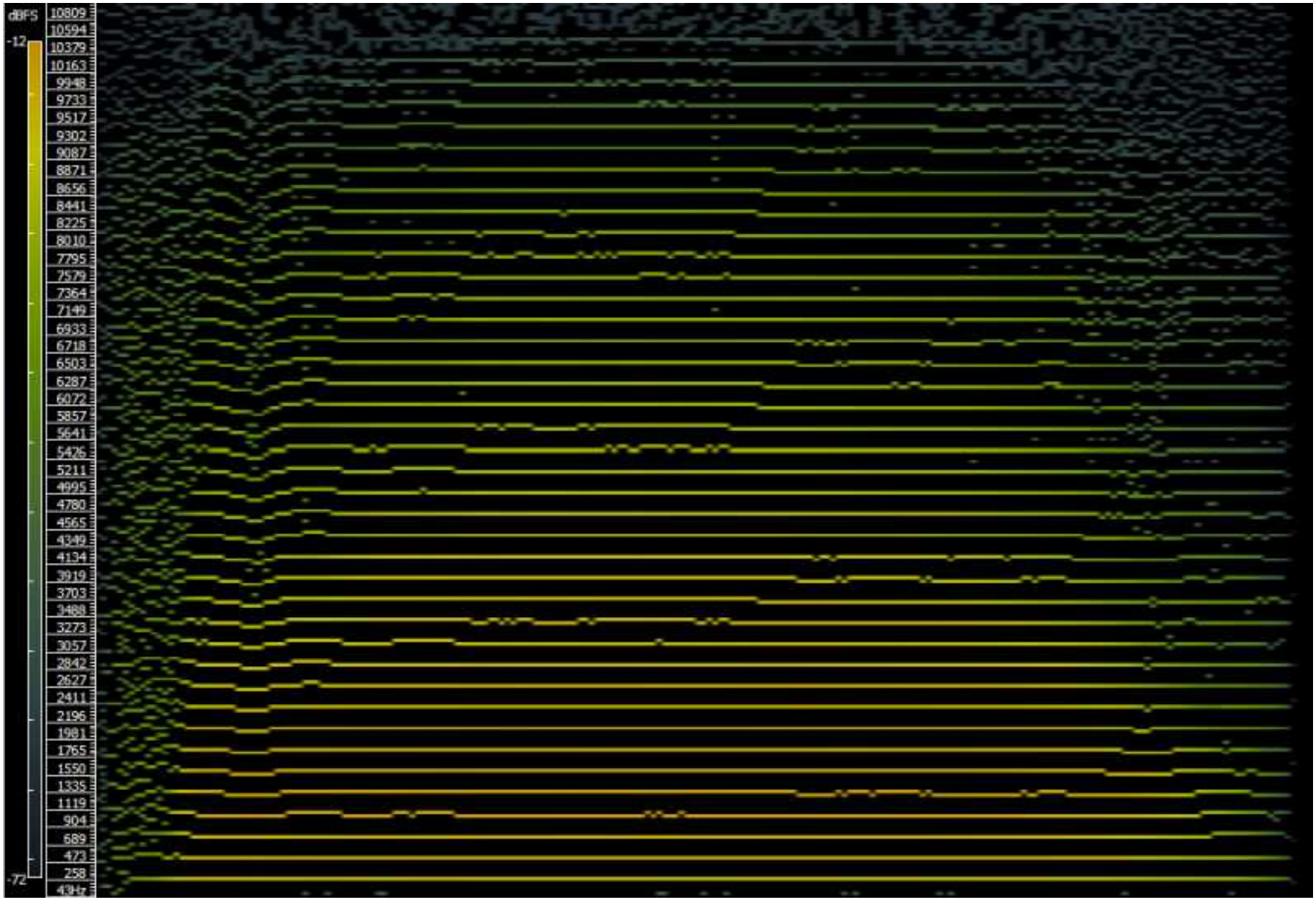
- Rozmiar okna: 2048 na start (1024-4096).
- Zakładkowanie: 50% lub 75%.
- Okno: Hamminga, Blackmana lub von Hanna.
- Większe okno (np. 4096) gdy widmo jest mało zmienne (np. stan ustalony) lub gdy częstotliwość sygnału jest mała.
- Mniejsze okno (np. 512), większe zakładkowanie i uzupełnianie zerami gdy widmo jest silnie zmienne (np. w fazie ataku).
- Zbyt duże okno (np. 16384): nie ma istotnego zysku dokładności, a bardzo wydłużamy analizę.

Analiza widma progowanego

- Możemy uwzględnić w analizie tylko wartości widma o poziomie większym niż pewien próg.
- Pomijamy szumy i małe składowe, lepiej uwidaczniamy maksima widmowe.
- Szczególnie przydatne dla dźwięków harmonicznych.



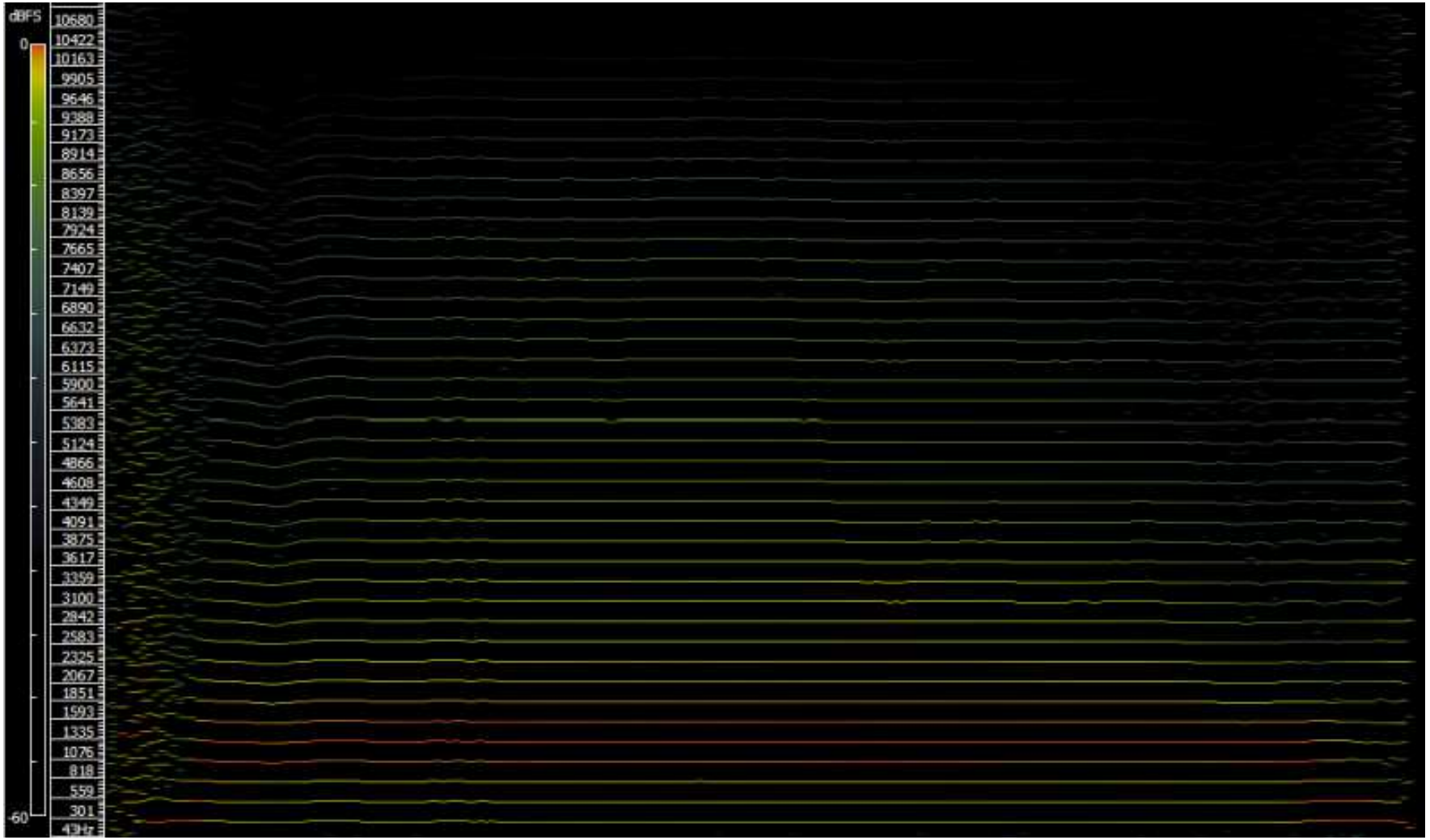
Analiza widma progowanego



Analiza lokalnych maksimów widma

- Wyszukiwane są lokalne maksima widma.
- Na wykresie są pokazywane tylko znalezione maksima.
- Bardzo użyteczna dla dźwięków instrumentów muzycznych mających charakter harmoniczny.
- Analiza pozwala wyraźnie pokazać zmienność częstotliwości i amplitud prążków harmonicznych.
- Przydatna np. do:
  - śledzenia częstotliwości podstawowej,
  - parametryzacji barwy dźwięku,
  - uzyskiwania parametrów do resyntezy dźwięku.

Analiza lokalnych maksimów widma

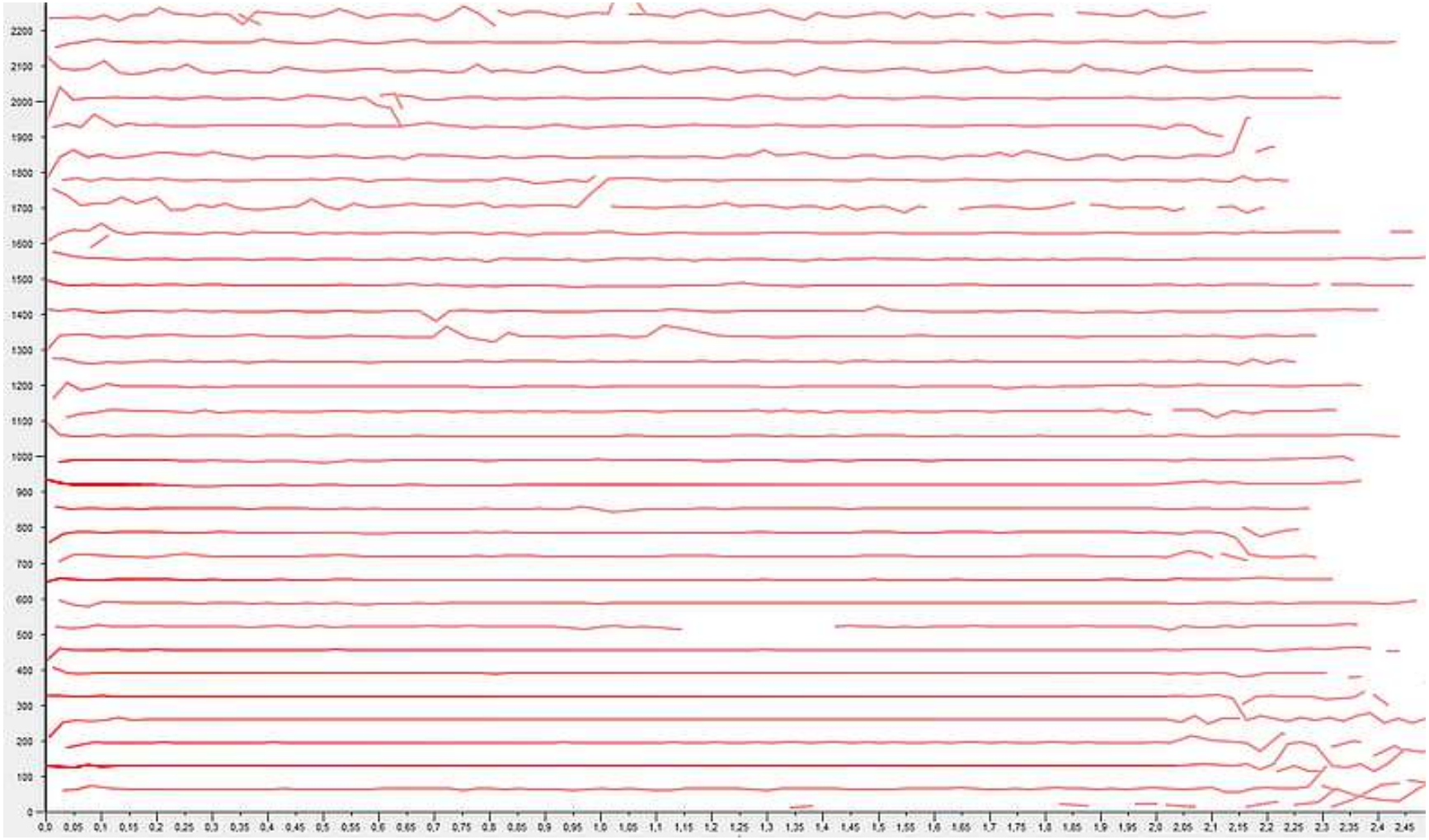


Analiza McAulay-Quatieri (MQ)

Analiza MQ stanowi rozwinięcie poprzednich analiz:

- wyszukiwanie maksimów widmowych w STFT,
- odrzucanie maksimów mniejszych niż próg,
- łączenie wyników z kolejnych ramek – tworzenie ciągłych ścieżek składowych widmowych (*tracking*),
- odrzucanie krótkotrwałych ścieżek,
- wynikiem analizy MQ jest zestaw ścieżek reprezentujących istotne składowe widmowe, np. harmoniczne.

Analiza McAulay-Quatieri (MQ)



Analiza autokorelacyjna

- Autokorelacja – miara zgodności sygnału z przesuniętą w czasie kopią tego samego sygnału.
- Wynik wyznaczany dla różnych wartości przesunięcia.
- W przypadku sygnałów harmonicznych, autokorelacja jest największa gdy przesunięcie jest równe okresowi lub jego wielokrotności.
- Jest to jedna z metod wyznaczania częstotliwości podstawowej sygnału (wysokości dźwięku).

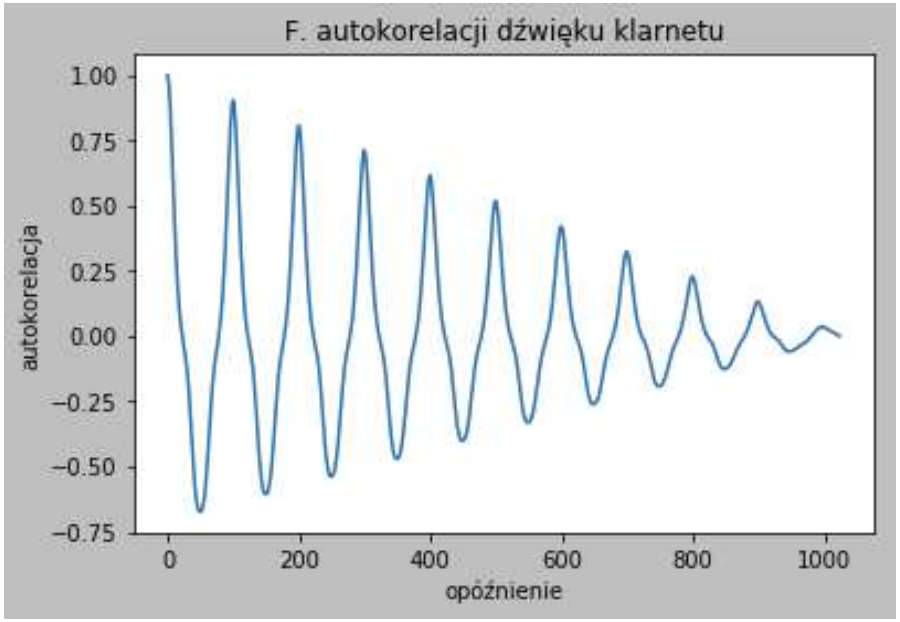
Analiza autokorelacyjna

Sposób obliczania autokorelacji:

- pobranie bloku próbek (musi być dłuższy niż okres),
- obliczenie FFT,
- obliczenie modułu widma,
- obliczenie IFFT z modułu widma, obcięcie części urojonej,
- bierzemy połowę wyniku (usuwamy symetryczną kopię),
- normujemy dzieląc przez wartość dla zerowego przesunięcia.

Analiza autokorelacyjna

Funkcja autokorelacji dla dźwięku klarnetu:



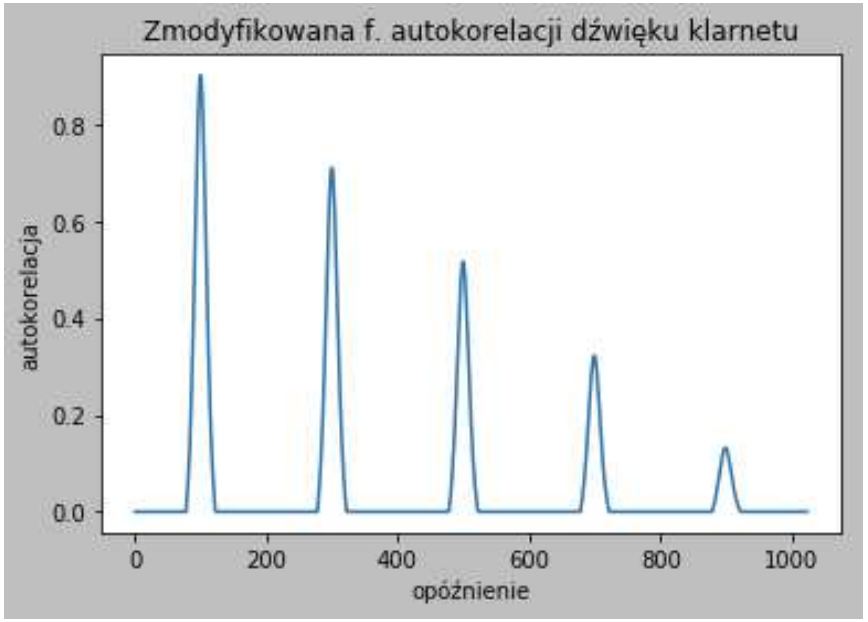
Pierwsze maksimum: $x = 100$

Częstotliwość: $f = f_S / x = 44100 / 100 = 441$ Hz

Analiza autokorelacyjna

Sposób modyfikacji funkcji autokorelacji dla łatwego obliczenia częstotliwości sygnału:

- po obliczeniu FFT, moduł widma podnosimy do drugiej lub trzeciej potęgi,
- zerujemy ujemne wartości autokorelacji,
- od funkcji autokorelacji odejmujemy jej pierwszą połowę „rozciągniętą” na całość – eliminujemy maksimum dla zerowego przesunięcia,
- zerujemy ujemne wartości,
- szukamy maksimum w wyniku.



Autokorelacja vs. FFT

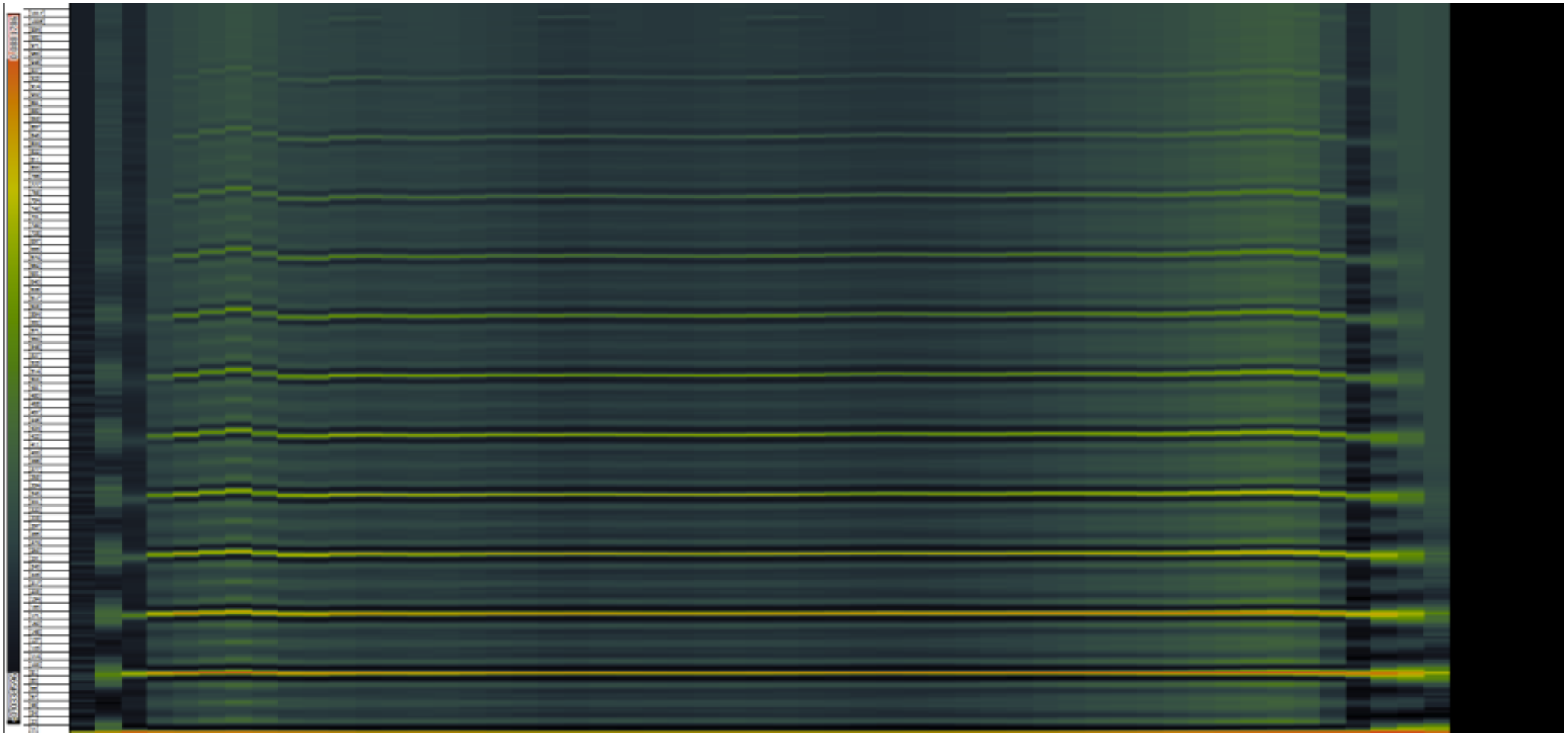
Porównanie dokładności estymacji częstotliwości podstawowej.

- Muzyk miał zagrać dźwięk a^1: 440 Hz.
- FFT z oknem 2048: 430,664 Hz
  – zbyt mała rozdzielczość częstotliwościowa.
- Dopasowanie paraboli do prążka FFT: 441,335 Hz.
- Autokorelacja: 441 Hz.

Wniosek: nie należy stosować FFT do estymacji częstotliwości podstawowej dźwięku. Jest to zbyt niedokładna metoda. Autokorelacja zapewnia lepszą dokładność.

Analiza autokorelacyjna

Autokorelogram – podobny do spektrogramu, uwidacznia zmienność harmonicznych

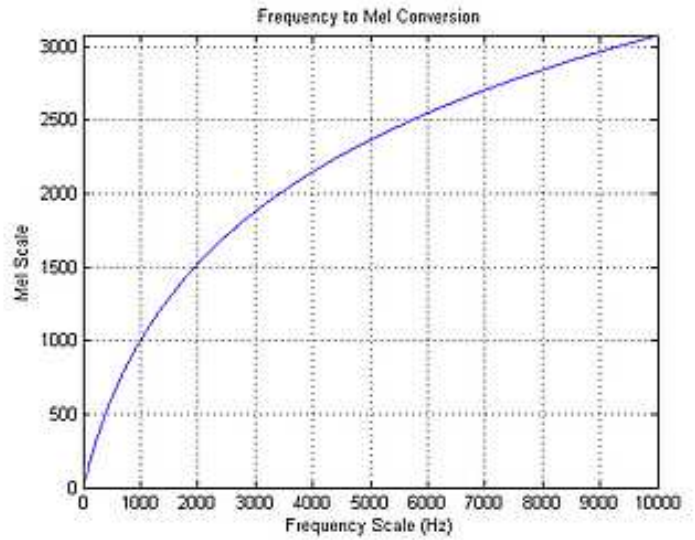


Analiza w skali melowej

- Standardowe analizy są przeprowadzane na skali częstotliwości.
- Dla dźwięków muzycznych istotna jest ich wysokość.
- Skala melowa wiąże częstotliwość z wysokością dla tonów prostych (sinusów).

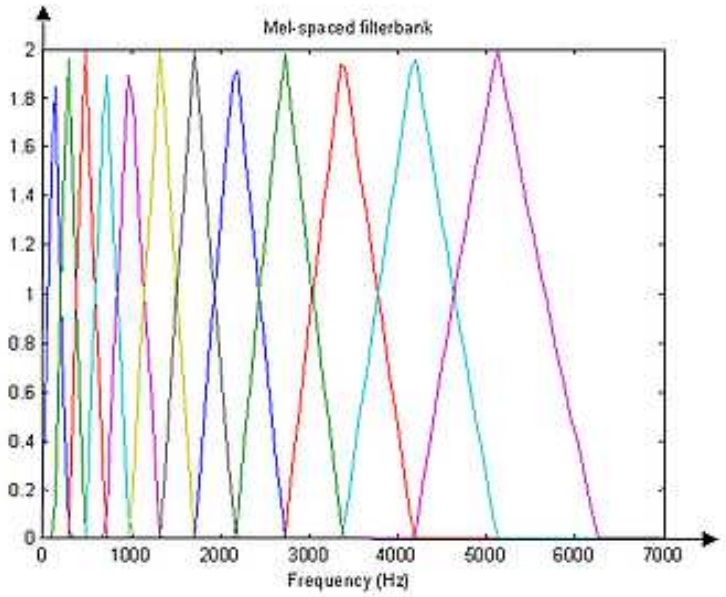
$$M = 1125 \ln(1 + f / 700)$$

$$M = 2595 \log_{10}(1 + f / 700)$$



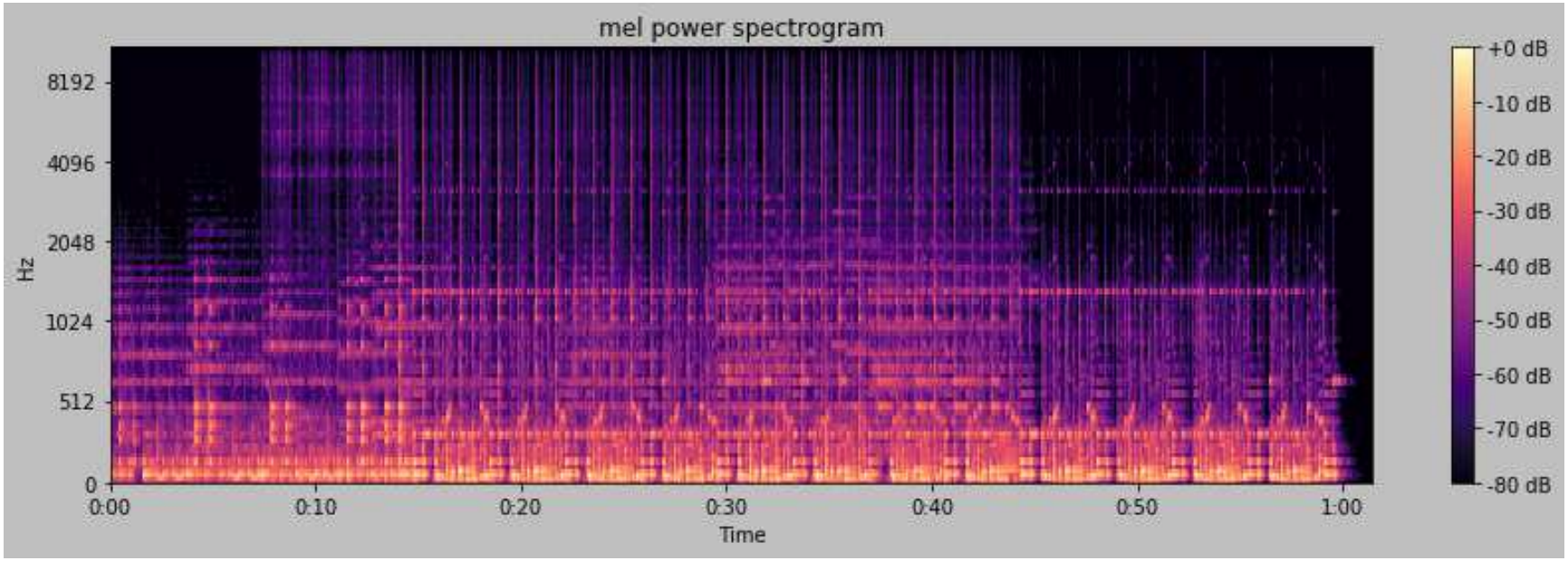
Bank filtrów melowych

- Zbiór filtrów w postaci trójkątnych funkcji wagowych.
- Zakres częstotliwości transformowany do skali melowej i dzielony na *N* równych pasm (np. *N* = 128).
- Każdy filtr pokrywa jedno pasmo melowe.
- Wynik filtracji: suma wartości widma sygnału przemnożonego przez daną funkcję wagową.



Spektrogram w skali melowej

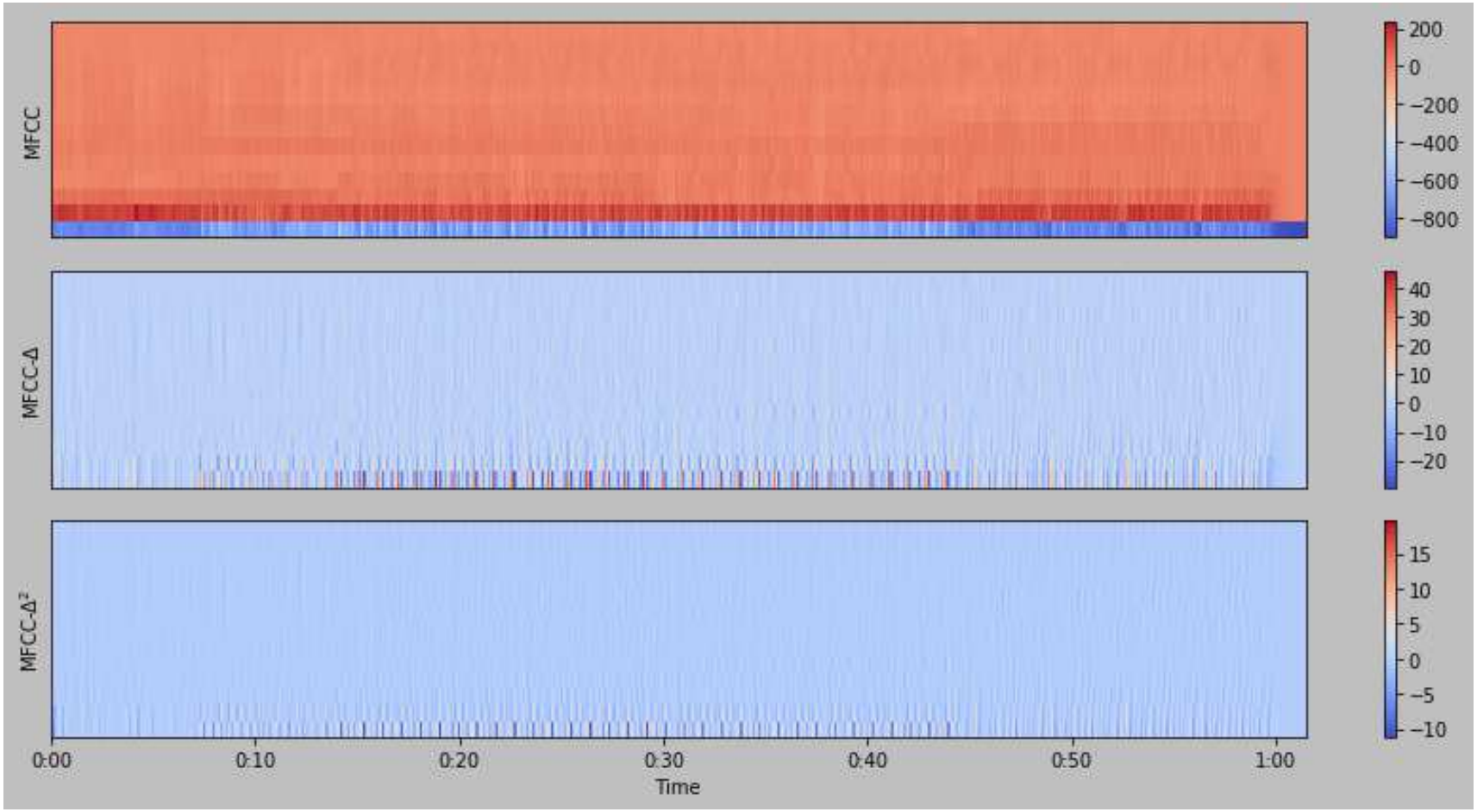
Obliczenie widmowej gęstości mocy (kwadrat modułu FFT) i podział na 128 pasm melowych.



Współczynniki mel-cepstralne

- Podział sygnału na ramki (funkcja okna).
- Obliczenie FFT i kwadratu modułu widma.
- Zastosowanie banku filtrów melowych, sumowanie pasm.
- Logarytmowanie wyników (cepstrum).
- Obliczenie dyskretnej transformaty kosinusowej (DCT) dla wartości mel-cepstralnych – parametry MFCC.
- MFCC-Δ: pochodna MFCC, czyli różnica MFCC między obecną a poprzednią ramką.
- MFCC-Δ^2: druga pochodna MFCC, czyli różnica pochodnych MFCC dla obecnej i poprzedniej ramki.
- Do parametryzacji często pomija się skrajne pasma.

Wykres MFCC i pochodnych



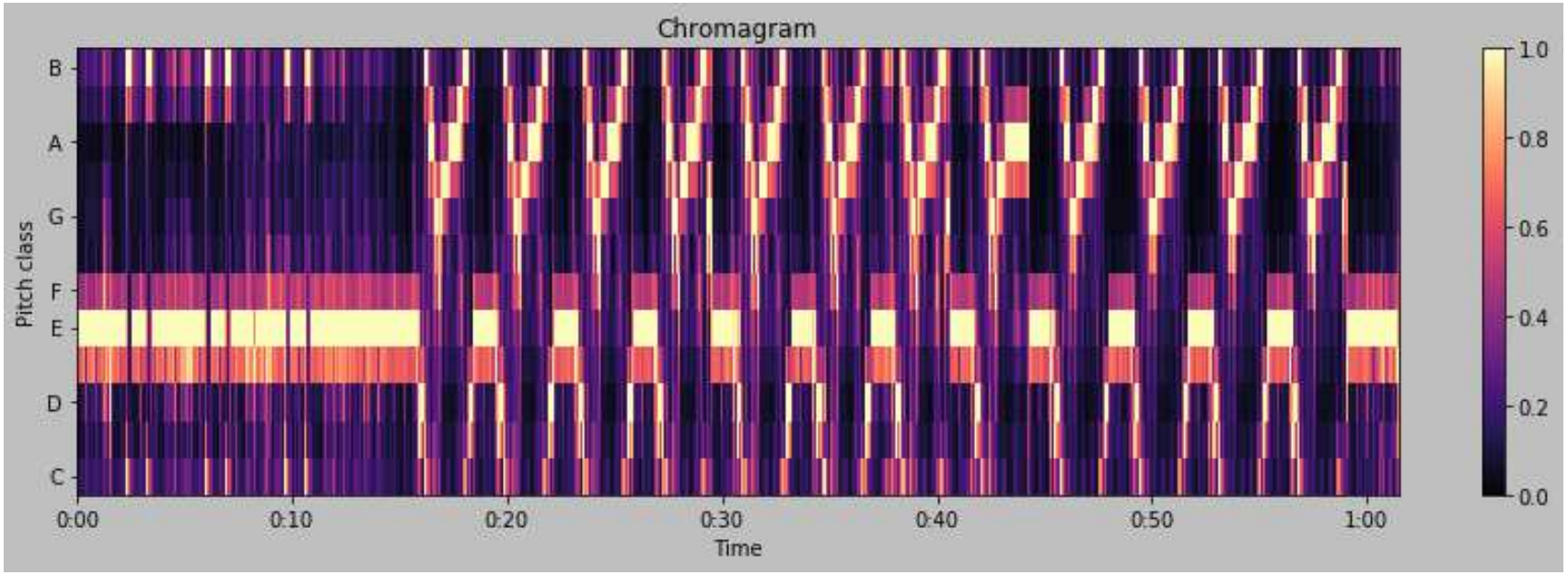
Współczynniki MFCC są używane m.in. do parametryzacji barwy dźwięków muzycznych.

Klasy wysokości

- Klasa wysokości dźwięku muzycznego (*pitch class*): zbiór dźwięków o danej wysokości półtonu (np. C) ze wszystkich oktaw.
- Klasy wysokości: C, C#, D, D#, E, F, F#, G, G#, A, A#, H.
- Dźwięki z różnych oktaw (np. C, c, c^1, c^2, …) należą do tej samej klasy wysokości.
- Składowe widma są przypisywane do klas wysokości.
- Oblicza się sumę składowych widmowych należących do danej klasy wysokości.
- Jest to jedna z metod analizy i parametryzacji barwy dźwięku, używana np. do rozpoznawania utworów.

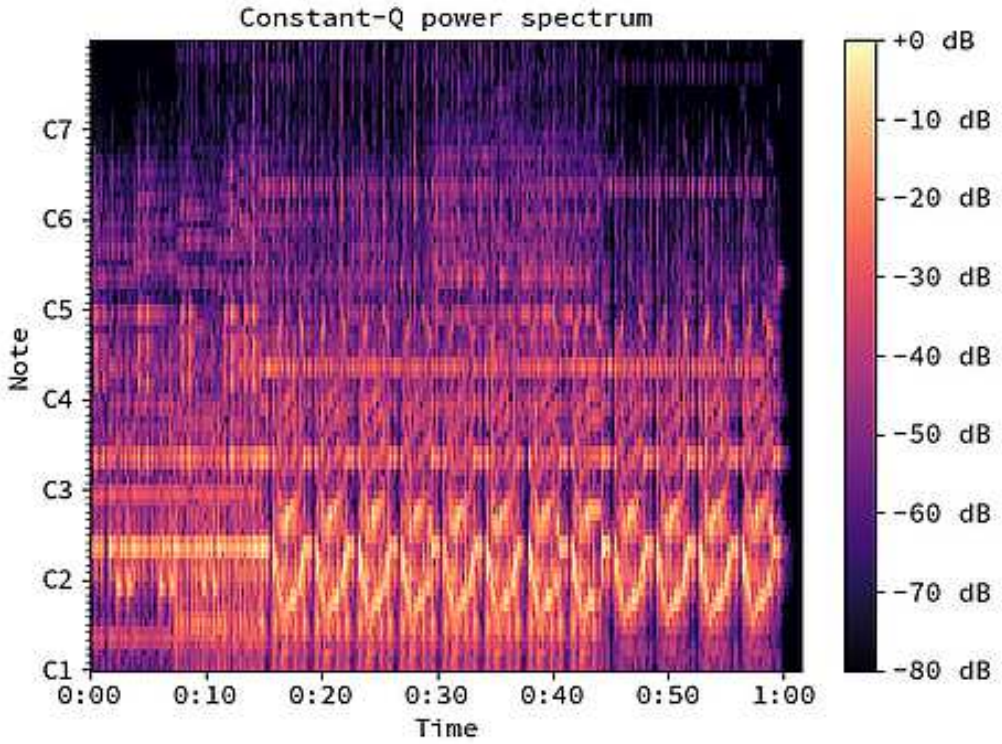
Chromagram

Chromagram – wykres obrazujący wartości klas wysokości w funkcji czasu.



Transformacja constant-Q

- Przekształcenie *Constant-Q* odpowiada analizie sygnału przez bank filtrów o stałej dobroci, czyli o paśmie zwiększającym się z częstotliwością.
- Uzyskujemy logarytmiczną skalę częstotliwości, podobną do skali melowej.
- Wysokość składowych przedstawia się w skali muzycznej.



Detekcja początków nut

Detekcja początków nut (*note onset detection*):

- analiza fraz muzycznych,
- wyznaczanie miejsc, w których zaczynają się dźwięki,
- zastosowania:
  - segmentacja nagrania na pojedyncze „nuty”, np. rozpoznawanie melodii, automatyczny zapis nutowy,
  - estymacja rytmu (tempa utworu).

Detekcja początków nut

Typowy schemat analizy:

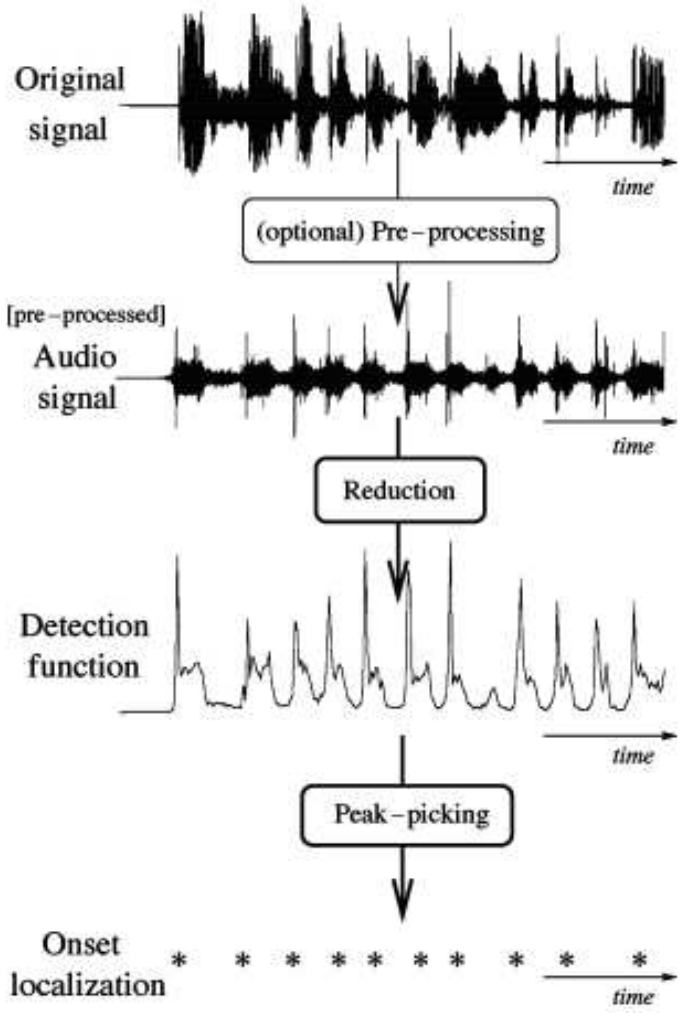


Fig. 2. Flowchart of a standard onset detection algorithm.

Bello et al.: A tutorial on onset detection in music signals

Wstępne przetwarzanie (filtracja)

Obliczenie funkcji detekcji

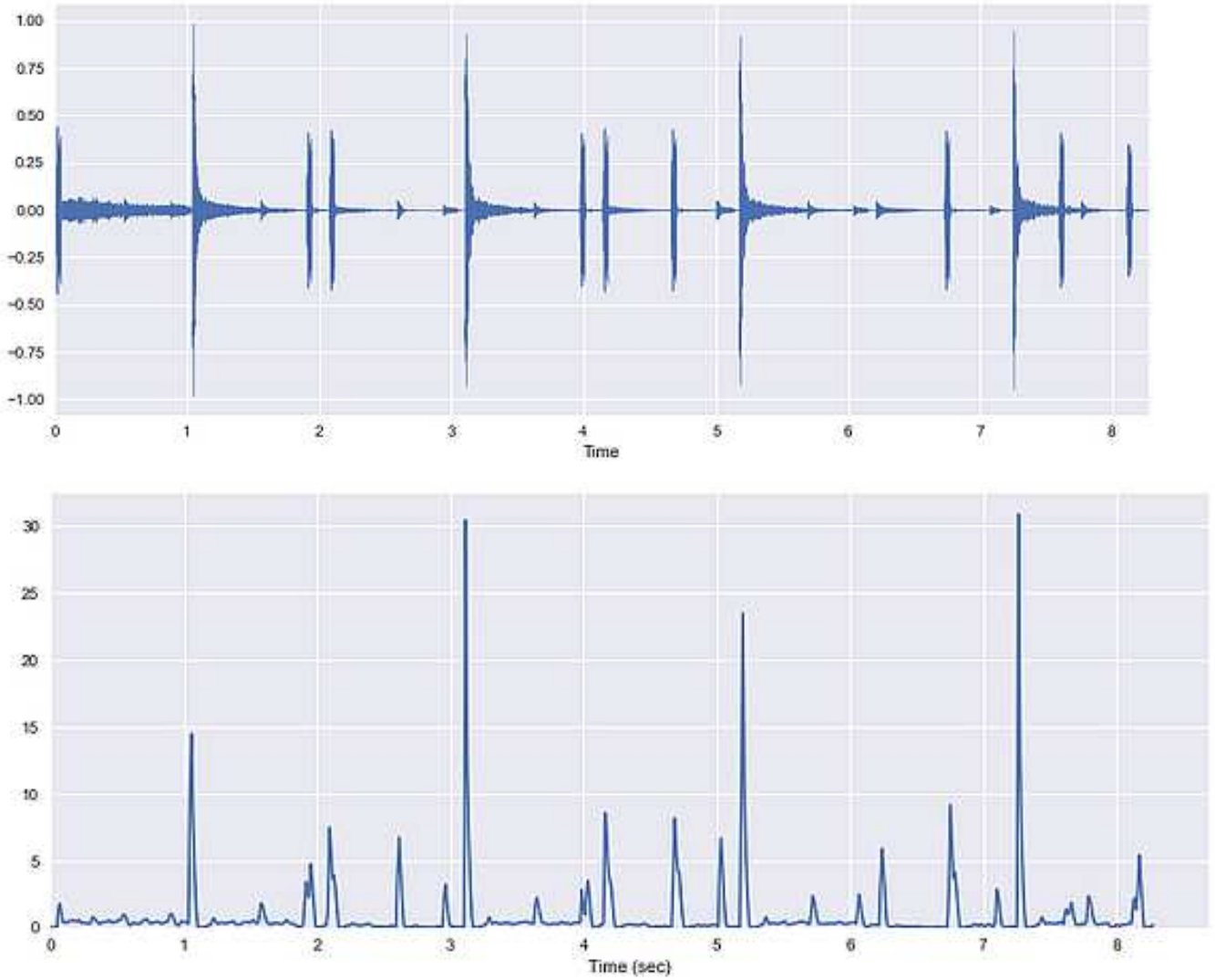
Znalezienie maksimów funkcji

Detekcja początków nut

- Wstępne przetwarzanie: filtracja, kompresja dynamiki, itp. (różne podejścia).
- Obliczanie funkcji detekcji – istnieje bardzo wiele metod.
- Metoda *spectral flux*:
  - oblicza się reprezentację widmową (spektrogram, mel-spektrogram, itp.),
  - oblicza się różnicę widma bieżącej i poprzedniej ramki,
  - sumuje się te różnice (zwykle pomija się ujemne wartości).
- Znajdowanie początków nut: wyszukanie maksimów w funkcji detekcji i identyfikacja istotnych maksimów.

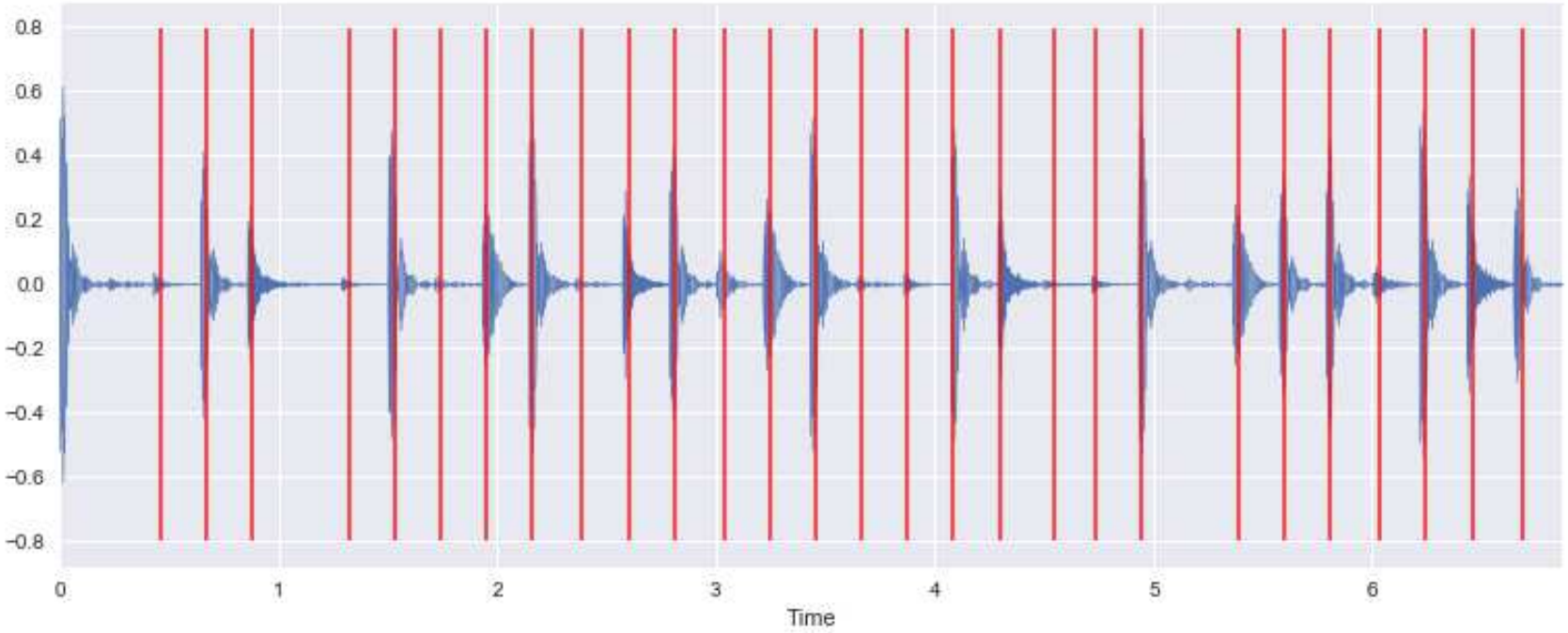
Detekcja początków nut

Wyznaczanie funkcji detekcji (spectral flux)



Detekcja początków nut

Wynik detekcji:



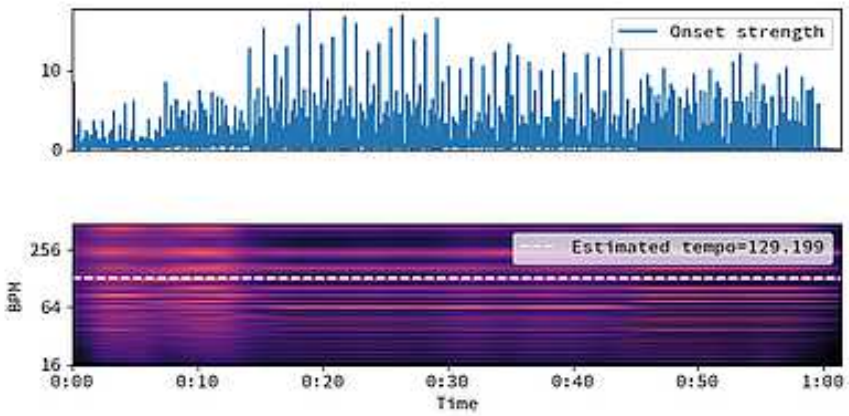
Na podstawie pomiaru odstępów czasowych między początkami nut można oszacować tempo utworu.

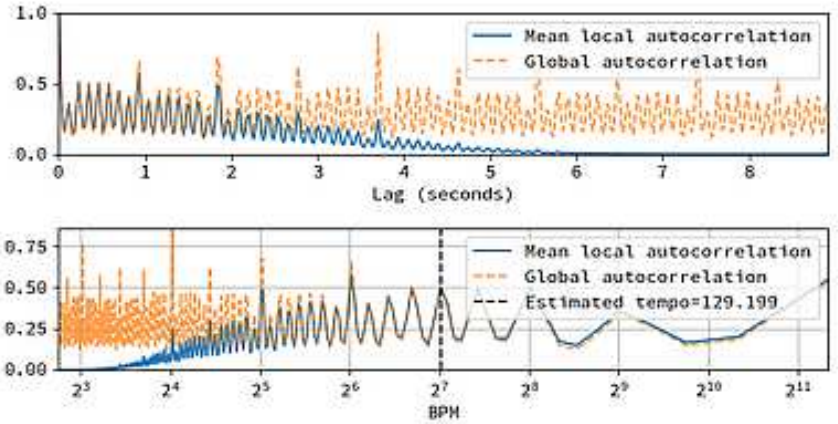
https://musicinformationretrieval.com/onset_detection.html

Estymacja tempa

Dwie metody oszacowania tempa utworu na podstawie funkcji *spectral flux*:

- obliczenie spektrogramu funkcji *spectral flux* (tempogram) i znalezienie maksimum,
- obliczenie funkcji autokorelacji dla funkcji *spectral flux* i znalezienie maksimum.





http://librosa.github.io/librosa/generated/librosa.feature.tempogram.html

Analiza falkowa (wavelet)

Analiza falkowa (*wavelet analysis*)

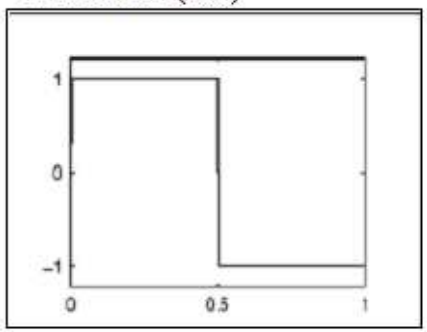
- Sygnał jest reprezentowany jako suma ważona falek (*wavelet*) o różnym przesunięciu i skali.
- Przesunięcie odpowiada osi czasu w FFT.
- Skala („rozciągnięcie” falki) odpowiada osi częstotliwości w FFT (duża skala – mała częstotliwość).
- Transformacje falkowe:
  - CWT – ciągła transformacja falkowa, stałe odstępy między wartościami skali (1, 2, 3, 4, …)
  - DWT – dyskretna transformacja falkowa, kolejne skale są potęgami dwójki (1, 2, 4, 8, …)

Analiza falkowa (wavelet)

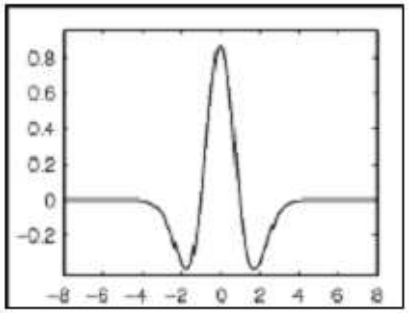
Falka (*wavelet*) jest sygnałem o skończonej długości i o zerowej wartości średniej.

Stosuje się wiele różnych kształtów falek.

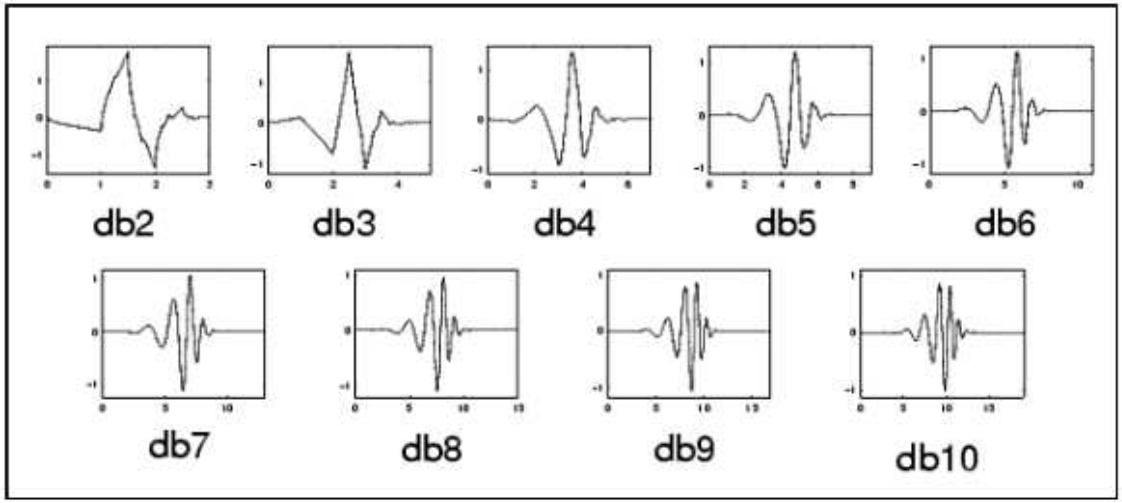
Falka Haara(db1)



Kapelusz Meksykański



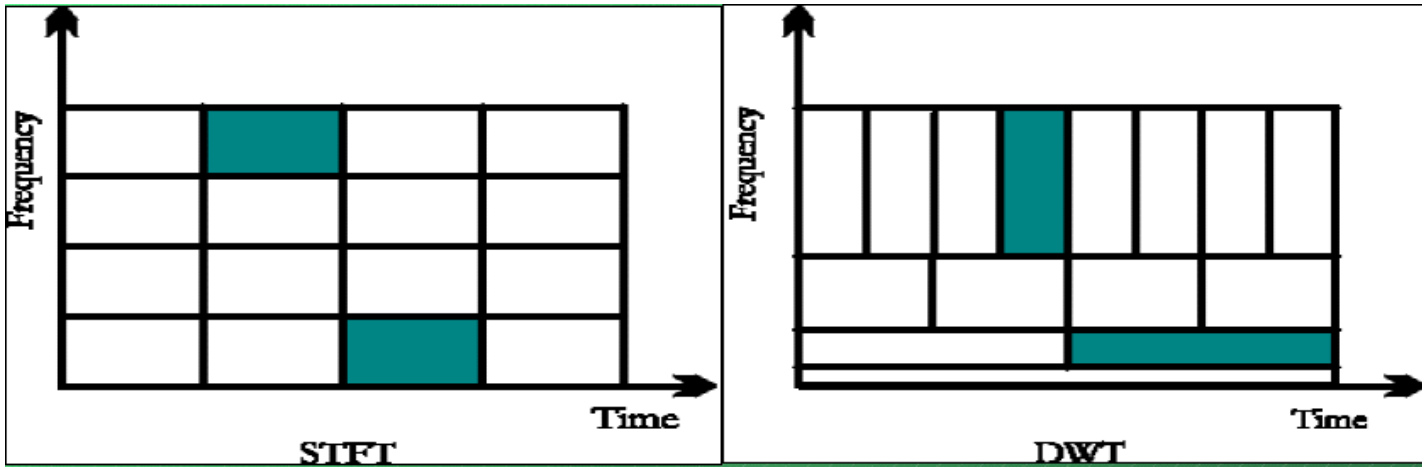
Falki Daubechies



Analiza falkowa vs. FFT

W porównaniu z analizą FFT:

- zmienna rozdzielczość czasowo-częstotliwościowa,
- niskie częstotliwości: lepsza rozdzielczość częstotliwościowa (długi okres, małe odległości między prążkami widma),
- wysokie częstotliwości: lepsza rozdzielczość czasowa (krótki okres, duże odległości między prążkami).



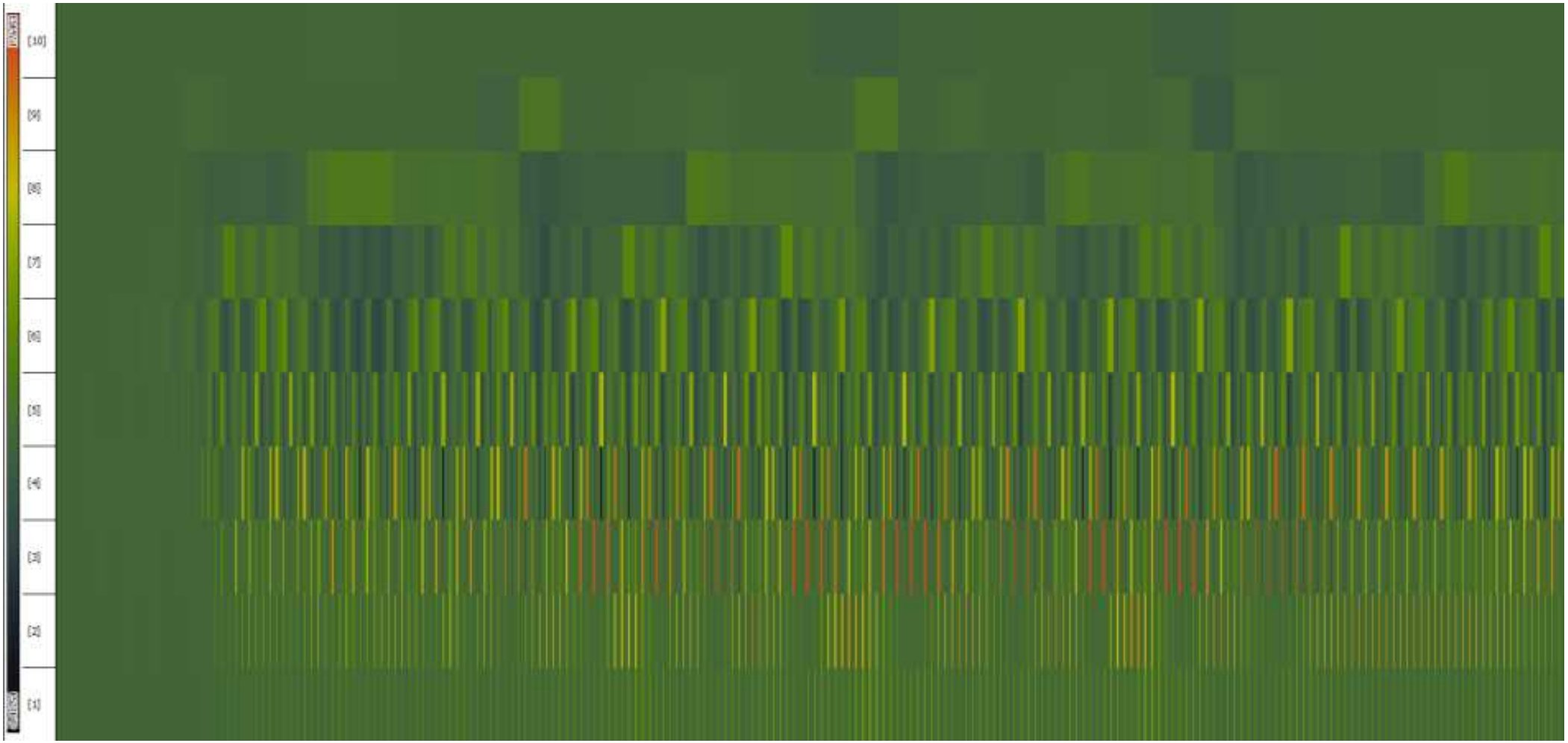
Skalogram

Wynik analizy falkowej można przedstawić w formie skalogramu:

- oś pozioma – czas (przesunięcie falki),
- oś pionowa – skala falki (większa skala odpowiada mniejszej częstotliwości),
- kolor – wartość amplitudy składowej falki.

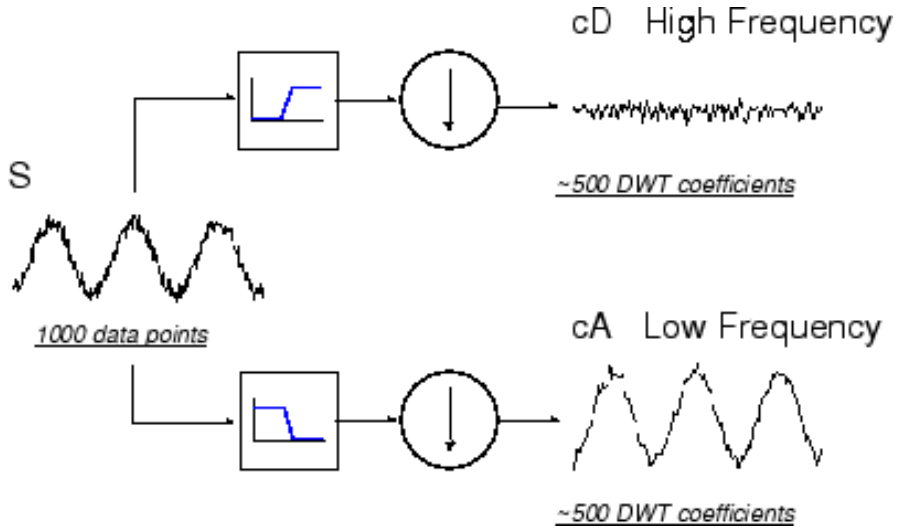
Skalogram jest mniej czytelny niż FFT pod względem analizy częstotliwościowej sygnału. Pozwala on uwidocznić zmiany przebiegu składowych o różnych skalach.

Analiza falkowa – skalogram dźwięku trąbki



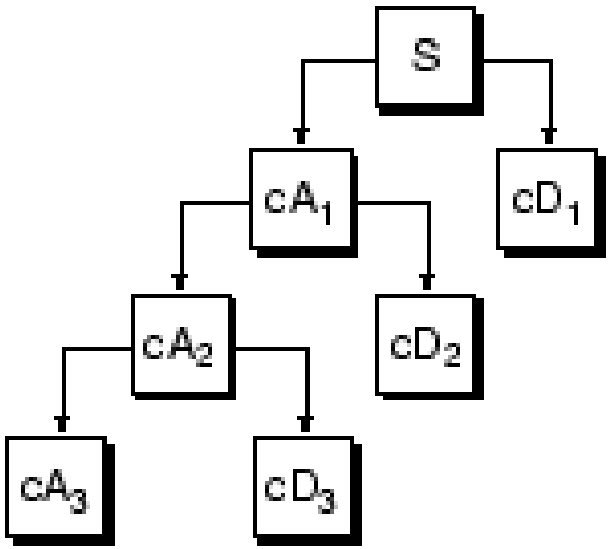
Dekompozycja falkowa

- Sygnał jest filtrowany przez dwa komplementarne filtry: dolnoprzepustowy (DP) i górnoprzepustowy (GP).
- Współczynniki tych filtrów odpowiadają kształtowi falki.
- Sygnał aproksymujący (A) – wynik filtracji DP.
- Sygnał szczegółowy (D) – wynik filtracji GP.
- Sygnały A i D są poddawane decymacji – dwukrotne zmniejszenie liczby próbek.



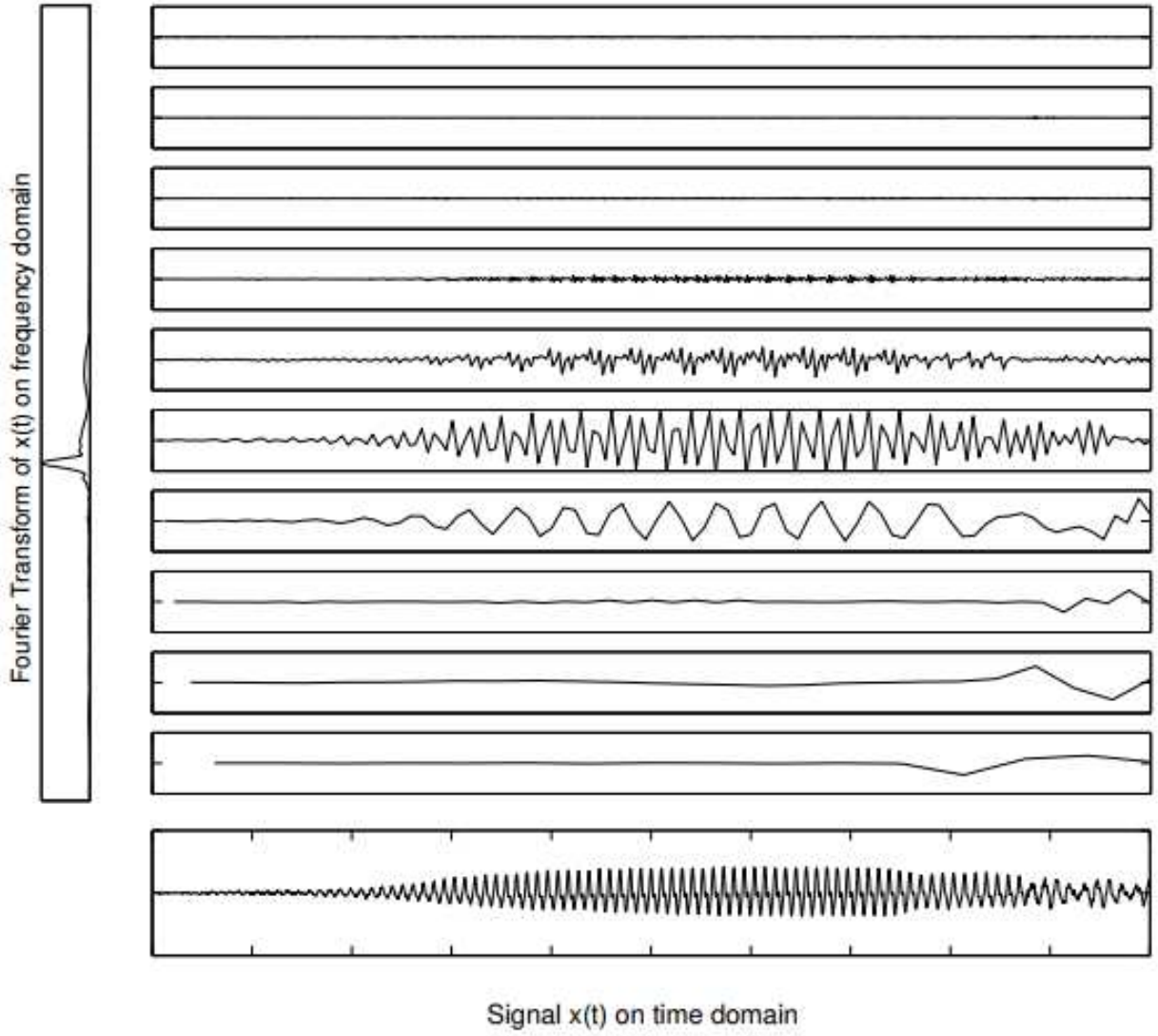
Dekompozycja falkowa

- Sygnał A po decymacji jest ponownie poddawany dekompozycji.
- Proces jest powtarzany – powstaje określona liczba poziomów dekompozycji.
- Uzyskujemy szereg poziomów: D_1, D_2, …, D_N, A_N.
- Możliwe jest złożenie sygnału z powrotem.



Dekompozycja falkowa

Przykład dekompozycji falkowej dźwięku fletu



Analiza falkowa - zastosowania

Zastosowania analizy falkowej dla dźwięków muzycznych:

- parametryzacja do celów klasyfikacji, np. rozpoznawania instrumentów muzycznych,
- wykrywanie zmian w przebiegu czasowym dla różnych zakresów częstotliwości (np. początków nut),
- usuwanie szumów i zakłóceń w nagraniach

Oprogramowanie

Wybrane darmowe oprogramowanie do analizy dźw. muz.:

- *Audacity* – analiza czasowa, widmowa, autokorelacja, wykrywanie początków nut.
- *Sonic Visualiser* – analiza czasowo-częstotliwościowa dźwięków muzycznych, liczne wtyczki *VAMP*.
- *LibROSA* – moduł języka Python do analizy i parametryzacji dźwięków muzycznych.

Dużo informacji i przykładów na stronie:
Notes on Music Information Retrieval
https://musicinformationretrieval.com/