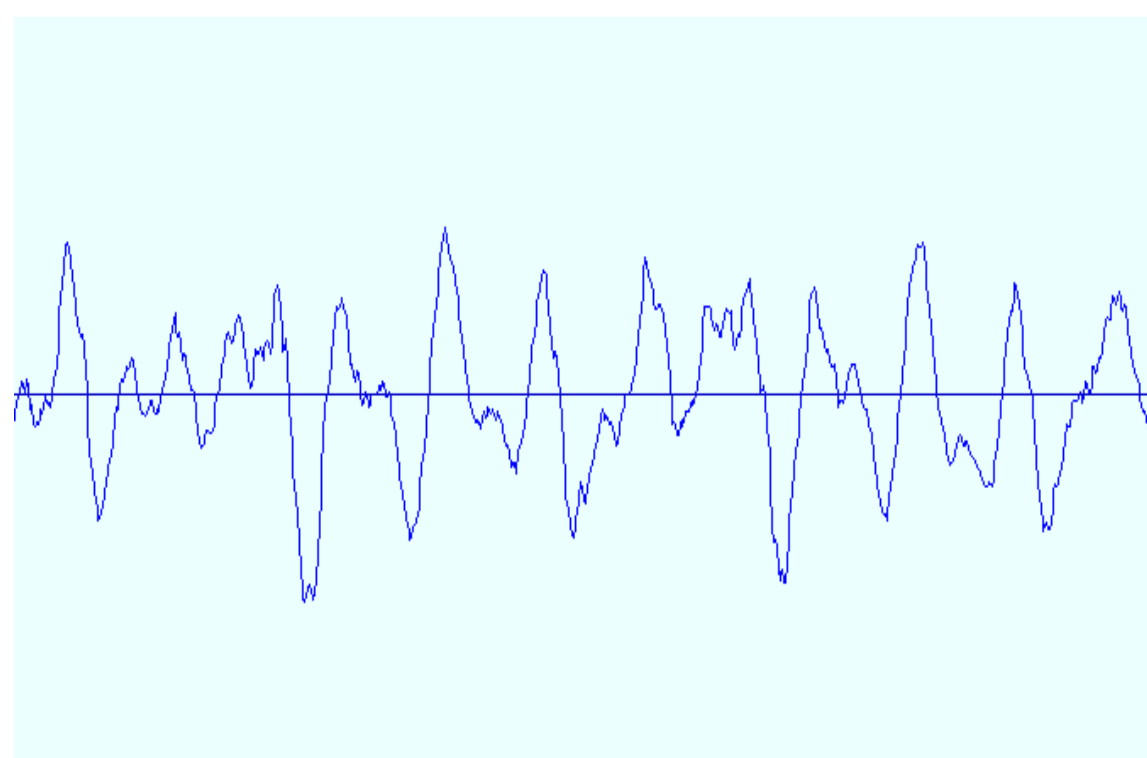


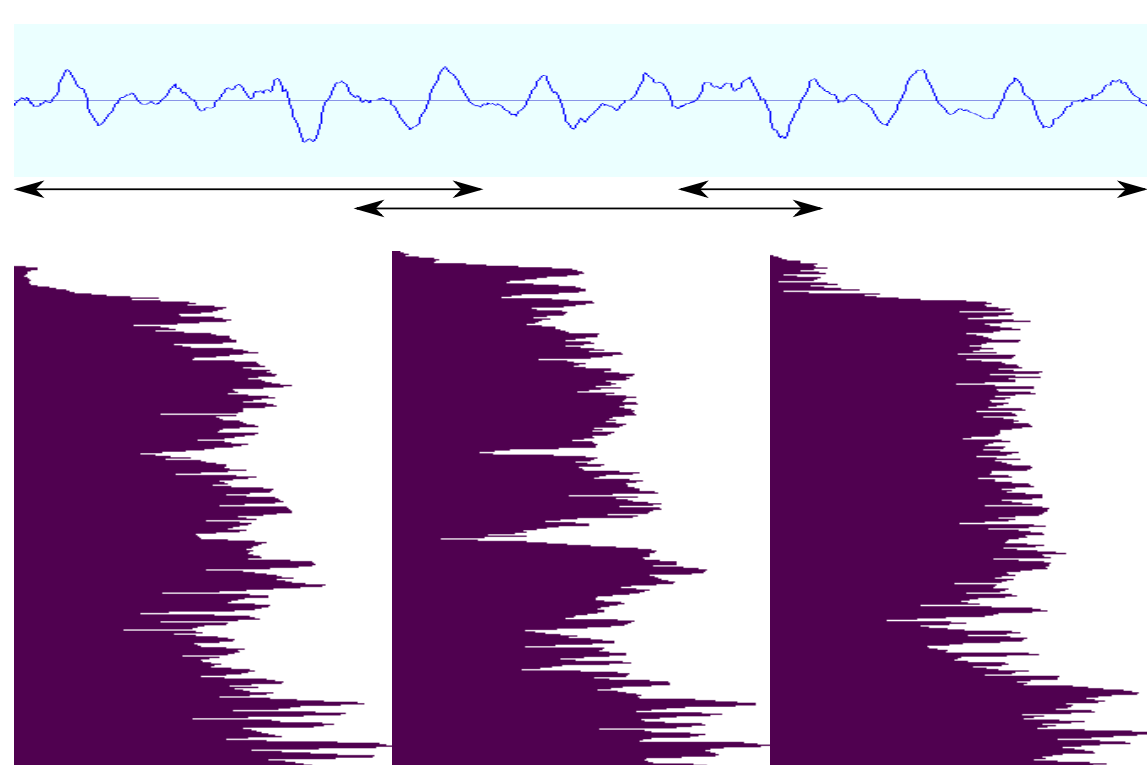
# Efektywny algorytm transkrypcji muzyki polifonicznej

Piotr Róžański, Marcin Radoszewski, Mariusz Tycz, Maciej Włoch, Bartosz Zasada  
Wydział Matematyki i Informatyki  
Uniwersytet Mikołaja Kopernika w Toruniu



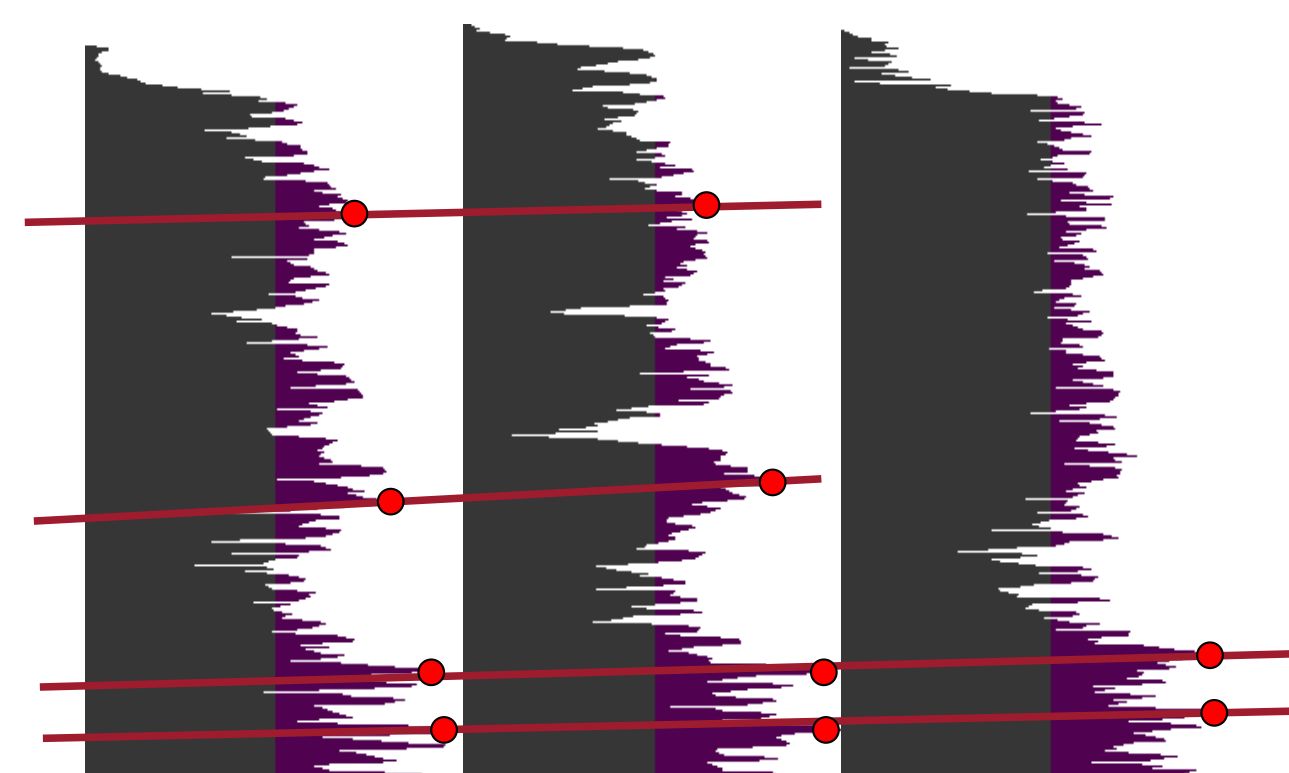
Muzyka w postaci cyfrowej ma postać ciągu liczb (próbek), przedstawiających wartości chwilowe sygnału w równomiernie oddalonych od siebie chwilach czasu. Ustalenie odstępu czasowego pomiędzy kolejnymi próbkami nosi nazwę *próbkiwania*, zaś przedstawienie wartości sygnału w postaci liczb całkowitych o określonej precyzji, nosi nazwę *kwantyzacji*.

Transkrypcja muzyki polega na analizie cyfrowego sygnału dźwiękowego, wyodrębnianiu z niego słyszalnych struktur muzycznych i przedstawieniu ich w postaci zapisu nutowego.



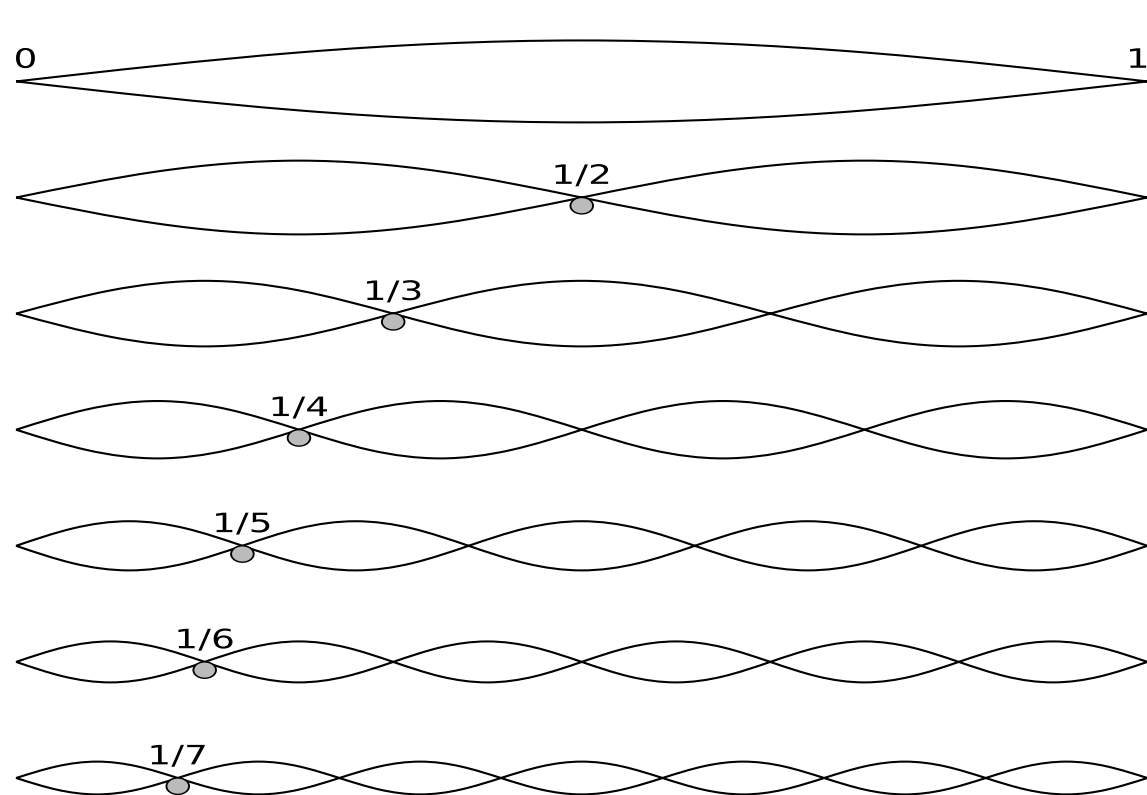
Przy pomocy transformaty Constant Q można obliczyć, na podstawie przebiegu sygnału w zadanym przedziale czasowym, widmo dźwięku odpowiadające temu przedziałowi. Jak się okazuje, jest ono w pośredni sposób powiązane ze słyszonym w danej chwili układem dźwięków.

Aby analizować przebieg dźwięku w czasie, należy wyznaczyć widmo dźwięku dla każdego z wielu, niekoniecznie rozłącznych, przesuniętych względem siebie przedziałów czasowych.



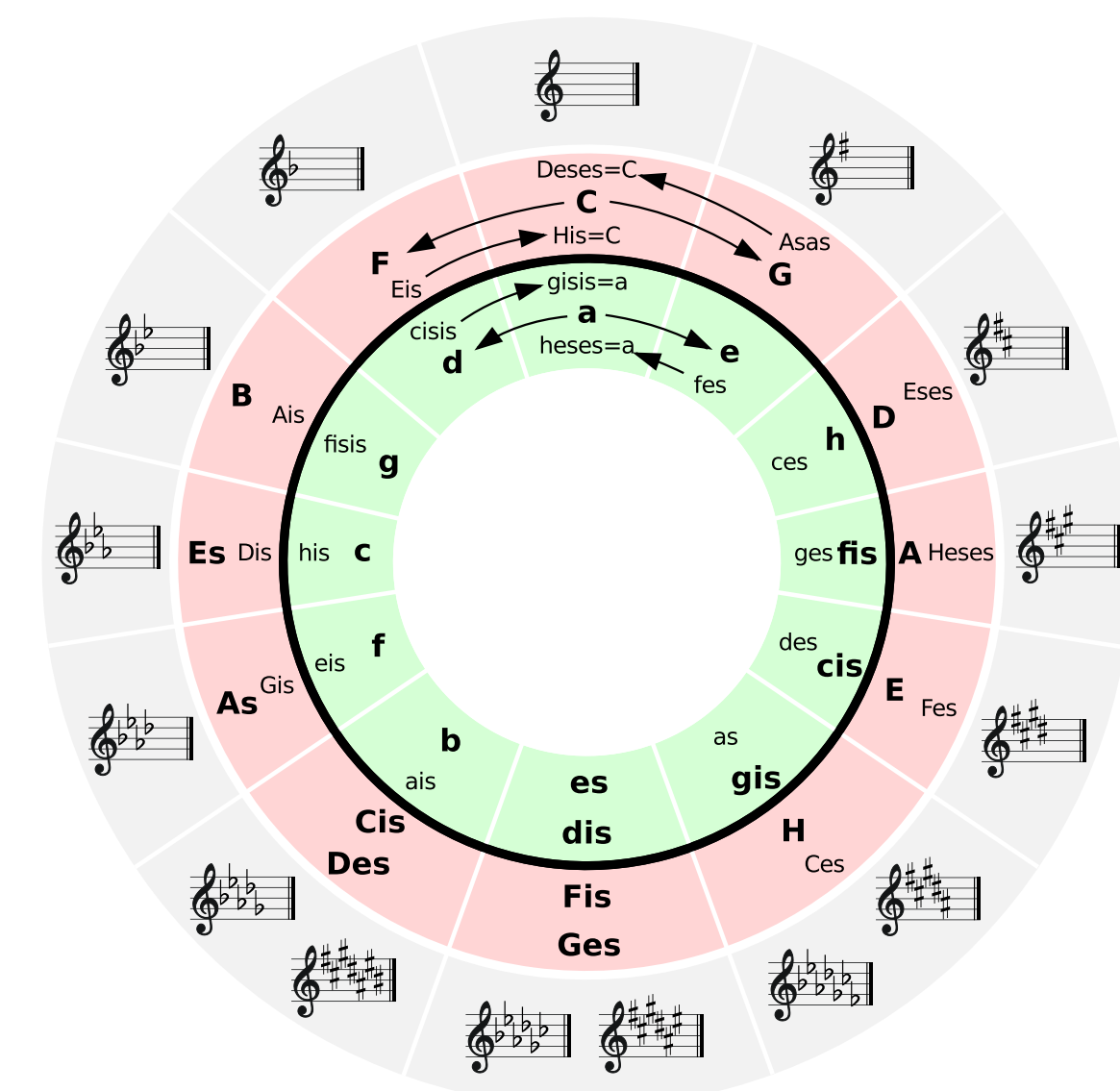
Na podstawie uzyskanych widm wykonywana jest estymacja poziomu szumu, wyznaczany jest próg detekcji, a następnie wyszukiwane są maksima lokalne widma przekraczające próg detekcji. Przy pomocy techniki "phase vocoder" poprawiana jest znajomość częstotliwości poszczególnych maksimów, eliminowane są także artefakty numeryczne.

W miarę przetwarzania kolejnych obrazów widma, znalezione w nich maksima łączone są w tony, zawierające powiązane czasowo dźwięki o zbliżonej częstotliwości. Przy dodawaniu kolejnego maksimum do przebiegu tonu, algorytm ocenia, czy stanowi on jego konfigurację, czy raczej repetycję, czyli powtórzenie dźwięku na tej samej wysokości.



Jak się okazuje, widma rzeczywistych instrumentów, nawet pochodzące od pojedynczych dźwięków, zawierają cały szereg częstotliwości stanowiących całkowite wielokrotności częstotliwości podstawowej.

W związku z powyższym, należy wyznaczyć powiązania harmoniczne pomiędzy wykrytymi tonami, a następnie do dalszej analizy zachować jedynie te, które z największym prawdopodobieństwem stanowią tony podstawowe.



Dla każdej z możliwych tonacji muzycznych wyznaczana jest ilość dźwięków gamowłaściwych spośród wszystkich dźwięków znalezionych w utworze. Pod uwagę brane jest 12 tonacji durowych oraz 12 tonacji molowych (moll harmoniczny). Wybierana jest tonacja, dla której ilość dźwięków gamowłaściwych jest największa. W przypadku remisu, wybierana jest tonacja o mniejszej liczbie znaków przykluczowych.

Wyznaczony przebieg dźwięków można przedstawić w postaci zapisu nutowego, przybliżając długości poszczególnych dźwięków w postaci wartości rytmicznych.



Opisany algorytm został zaimplementowany w programie IMARE, będącym projektem studenckim z Programowania Zespołowego na Uniwersytecie Mikołaja Kopernika. Program, który powstał w ramach pracy zaliczeniowej, jest dostępny bez opłat do niekomercyjnego użytku.

- implementacja w języku Java
- wykorzystanie obliczeń równoległych
- obsługa plików WAVE oraz MP3
- możliwość ustawienia parametrów transkrypcji
- możliwość transkrypcji w czasie rzeczywistym