# Temat

**„Projekt i implementacja adaptacyjnego systemu korekcji parametrów nagrań dźwiękowych”**

Celem pracy jest zaprojektowanie i implementacja adaptacyjnego systemu korekcji parametrów nagrań dźwiękowych z wykorzystaniem metod uczenia maszynowego.

## Zadania:

### Przygotowanie szkicu teoretycznego i raportu literaturowego

Literatura:

* Everest F. – Podręcznik akustyki

### Projekt aplikacji

Aplikacja ma wyznaczać:

* Barwę dźwięku – ulepszona spectral centroid
* Klarowność – energia pierwszych 50ms vs reszta
* Czas pogłosu – T60/T30/Early Decay time
* Pierwsze odbicie

### Przygotowanie środowiska pracy (instalacja oprogramowania i dodatkowych bibliotek programistycznych)

Python

### Implementacja kodu

Pycharm

### Testy aplikacji

Testy subiektywne oprogramowania – ocena działania programu w porównaniu z nagraniami wzorcowymi – ocena wierności odtworzenia charakteru nagrania, ocena naturalności brzmienia (osobno dla algorytmu machine learning).

# Wstęp

Celem pracy dyplomowej jest projekt oraz implementacja systemu analizy akustyki pomieszczenia, jako próbki nagrania audio w celu ekstrakcji parametrów akustycznych charakterystycznych dla tego nagrania, takich jak wpływ pomieszczenia na barwę dźwięku, czas pogłosu. Parametry te nadają nagraniu charakterystyczne brzmienie. System ma na celu nałożenie zmierzonych parametrów na czyste nagranie studyjne.

Nagranie poddawane analizie w początkowej fazie projektowania algorytmu jest odpowiedzią impulsową pomieszczenia, co dokładniej opisane zostanie w dalszej części pracy.

Analiza próbek nagrań dźwiękowych w następnych rozdziałach odbywa się w kontekście ludzkiego postrzegania zachodzących zjawisk, aniżeli ich opisu w zakresie fizyki. Dobór parametrów charakteryzujących badany sygnał ma więc na celu odzwierciedlenie elementów, które dobrze reprezentują postrzeganie zmiany zawartości sygnału po jego obróbce.

Parametryzacja nagrań w dalszym etapie pracy dyplomowej zostanie rozszerzona o uczenie maszynowe będące elementem eksperymentalnym oraz mające na celu usprawnienie obliczeń. Wpływ uczenia maszynowego na rozpoznanie charakteru nagrania (barwy itp.) zostanie zadany poprzez poddanie go ocenie subiektywnej przez słuchaczy.

# Teoria

## Czas pogłosu

Czas pogłosu definiuje się jako zanik energii o 60dB. Jest to tysiąckrotny zanik poziomu ciśnienia akustycznego, co zapisuje się:

.

Parametr T30 jest to czas, po jakim poziom energii spadnie o 60dB względem stanu początkowego, wyznaczany na podstawie nachylenia krzywej zaniku poziomu energii w zakresie 30dB, co umożliwia wyznaczenie czasu pogłosu nawet w przypadku, gdy mamy do czynienia z szumami.

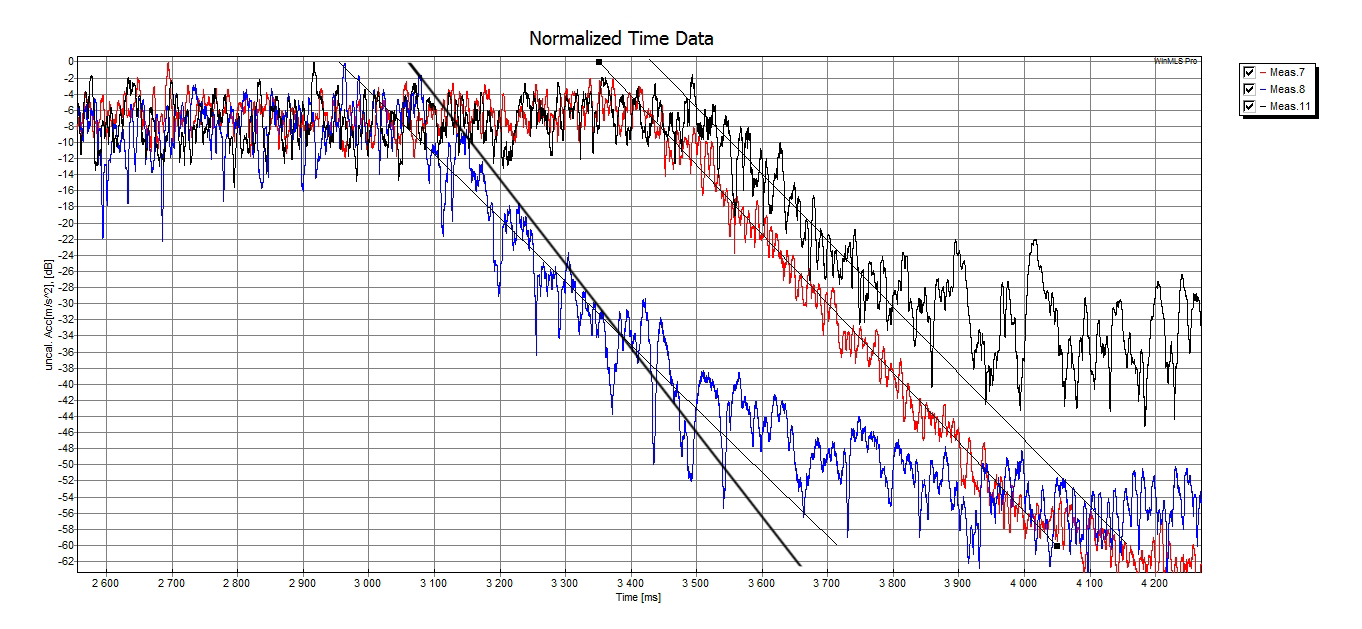
### Przykładowy pomiar zaniku poziomu energii w pomieszczeniu.

Tabela 1 Pomiar wymiarów pomieszczenia.

|  |  |
| --- | --- |
| wymiary pomieszczenia [m] | |
| długość | 5,92 |
| szerokość | 5,40 |
| wysokość | 3,40 |

Tabela 2 Współczynniki pochłaniania wybranych elementów pomieszczenia.

|  |  |  |
| --- | --- | --- |
|  | wsp. pochłaniania | powierzchnia [m2] |
| drzwi | 0,18 | 2,03 |
| meble | 0,17 | 7,00 |
| okno | 0,18 | 7,16 |
| ściany | 0,02 | 59,04 |
| drzwi | 0,17 | 1,76 |
| podłoga | 0,31 | 31,97 |
| sufit | 0,05 | 31,97 |



Wykres 1. Krzywe zaniku poziomu energii [czerwona - 2kHz, niebieska - 500Hz, czarna – bez filtra]

## Charakterystyka częstotliwościowa w zależności od długości impulsu.

Do pomiaru czasu pogłosu metodą impulsową należy uwzględnić rodzaj źródła tegoż sygnału. W poniższych przykładach zaprezentowano sygnał impulsowy z generatora odtwarzany na głośnikach. Źródłem takiego sygnału mogą także być: pękający balon, wystrzał z pistoletu.

W zależności od charakteru źródła, możemy mieć do czynienia z bardzo wąskim pasmem użytecznych częstotliwości, bądź też niskim poziomem sygnału, gdy sygnał ma bardzo krótki czas trwania (co zauważyć można na poniższych przykładach)



* Rysunek 2 Charakterystyka częstotliwościowa dla impulsu 50 μs

![](data:None;base64,)

Rysunek 3 Charakterystyka częstotliwościowa dla impulsu 75 μs



Rysunek 4 Charakterystyka częstotliwościowa dla impulsu 100 μs



Rysunek 5 Charakterystyka częstotliwościowa dla impulsu 200 μs



Rysunek 6 Charakterystyka częstotliwościowa dla impulsu 250 μs



Rysunek 7 Charakterystyka częstotliwościowa dla impulsu 500 μs



Rysunek 8 Charakterystyka częstotliwościowa dla impulsu 750 μs



Rysunek 9 Charakterystyka częstotliwościowa dla impulsu 1000 μs

## Barwa dźwięku

### **Spectal centroid**

Spectral centroid, czyli środek ciężkości charakterystyki częstotliwościowej sygnału jest parametrem, który został użyty w algorytmie w celu aproksymacji barwy dźwięku rozumianej jako częstotliwości dominujące w danym nagraniu w określonym odcinku czasu.

Do algorytmu wyznaczającego parametr centroidy częstotliwościowej dodano ważenie amplitudy zgodnie z założeniami prawa Webera Fechnera, ze względu na to, iż rozdzielczość FFT na wykresie wynosi 10Hz, co każdy punkt. Stąd też występuje zagęszczenie pomiarów dla wyższych częstotliwości dla skali logarytmicznej (która reprezentuje sposób postrzegana wrażeń dźwiękowych przez człowieka). Stąd, aby uzyskać bardziej naturalny dla ludzkiego ucha wynik działania tego algorytmu dodano ważenie wartości amplitudy FFT w odniesieniu do częstotliwości.

Przykłady

Przykład na wykresie [sinus 440Hz + szum biały] (kolejno FFT sygnału, watość SC ważone, wartość SC nieważone):



Kolejny przykład – utwór muzyczny

Kolejno:

* FFT w skali liniowej (X lin Y lin) [wartości Y nieważone]
* FFT w skali logarytmicznej (X log Y lin) [wartości Y nieważone]
* FFT X log i Y log [wartości Y nieważone]
* FFT X log i Y log [wartości Y ważone]

Legenda:

Poprawka ta widocznie udoskonala znajdowanie środka wagi widma.

# Projekt aplikacji – omówienie

Do implementacji projektu wybrano język Python w wersji 3,7, który charakteryzuje się szerokim zakresem zastosowań ze względu na wysokopoziomową składnię pozwalającą na szybką implementację złożonych czy rozbudowanych aplikacji. Wsparcie w postaci różnorodnych i specjalistycznych bibliotek dla tego języka umożliwia znacząco szybsze rozwiązywanie zaawansowanych problemów, które na przykład w przypadku C++ zajęłyby wielokrotnie więcej nakładu pracy (co mogłoby być niewspółmierne do zalet płynących z zastosowania C++). Wersja 3,7 umożliwia wsparcie w możliwym późniejszym rozwoju aplikacji.

## Schemat działania aplikacji

## Opis działania

W pierwszym kroku algorytm dokonuje odczytu pliku w formacie wave, który umożliwia zapis nagrania bez kompresji. Wstępnie określany jest zapis:

* ilość kanałów,
* częstotliwość próbkowania,
* ilość próbek.

Do wykonania tego etapu utworzono klasę WaveFile:

class **WaveFile**:  
 PATH = str()  
  
 def \_\_init\_\_(self, path):  
 self.PATH = path  
 self.FILE = wave.open(self.PATH)  
 [self.CHANNELS, self.SAMP\_WIDTH, self.FRAMERATE, self.NFRAMES, \*self.rest] = self.FILE.getparams()  
 self.RAW = np.fromstring(self.FILE.readframes(-1), dtype='int32')  
 self.RAW = self.RAW/max(self.RAW)

Instancja klasy inicjowana jest za pomocą adresu do analizowanego pliku w poniższy sposób:

orig\_file = WaveFile('./PathToFile/orig.wav')

Opis wykorzystanych zmiennych:

* PATH - ścieżka do pliku,
* FILE - obiekt modułu wave,
* CHANNELS - liczba kanałów,
* SAMP\_WIDTH - ilość bajtów dla jednej próbki (przeważnie jest to 2),
* FRAMERATE - częstotliwość próbkowania,
* NFRAMES - liczba próbek,
* RAW - neprzetworzony zapis pliku wave.

Odczyt wartości próbek zapisany do zmiennej RAW (będącej jednowymiarową tabelą typu int32[[1]](#footnote-1)), która w następnym kroku algorytmu zostaje normalizowana (w wyniku operacji dzielenia typ zmiennej ulega przekształceniu na float).

Na tej podstawie podejmowane są decyzje dotyczące kolejnych etapów działania algorytmu.

### Pogłos - Badanie odpowiedzi impulsowej

Najpierw obliczany jest kwadrat dla każdej wartości próbki:

for i, val in enumerate(frames):  
 f\_squared[i] = do\_square(val)

Nastepnie odwrócono kolejność próbek:

f\_squared = np.flip(f\_squared)

Po czym dokonano całkowania i utworzono oś czasu X:

integral = integrate(f\_squared)  
time = np.linspace(0, len(integral)/FP, num=len(integral))

OPISAĆ:

db30o = x[y.searchsorted(-30, 'left')]  
db30 = round(db30o\*1000\*CHOP)  
db0o = x[y.searchsorted(-0.2, 'right')]  
db0 = round(db0o\*1000\*CHOP)  
print("0dB---> at: " + str(db0) + "ms")  
print("-30dB-> at: " + str(db30) + "ms")  
print("-60dB-> at: " + str(db30+(db30-db0)) + "ms [extrapolated]")  
print("T30------>: " + str((db30-db0)\*2) + "ms")  
  
x1 = int((db30o)\*FP) # spadek o 30dB  
x2 = int((db0o)\*FP) # odniesienie  
x3 = int((db0o+((db30o-db0o)\*2))\*FP) # spadek o 60dB szacowany

Pzenieść z opracowanego rozdziału pogłos

Na pliku reprezentującym odpowiedź impulsową pomieszczenia przeprowadzana zostaje operacja całkowania.

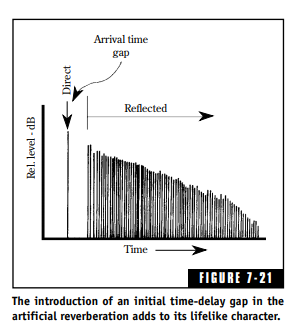
Równanie 1 PN-EN ISO 3382

Z wykresu krzywej zaniku poziomu energii szacowany jest spadek o 60dB na podstawie zakresu 30dB [parametr T30].

## Pogłos

Oddanie naturalności brzmienia, w kontekście zjawiska pogłosu wymaga między innymi rozpatrzenia jednego z ważnych elementów, jakim jest opóźnienie pomiędzy źródłem dźwięku, a pierwszym odbiciem, które to nadaje wrażenie wielkości pomieszczenia.

Fala dźwiękowa pokonując bezpośrednią drogę dociera najszybciej do słuchacza, następnie docierają wczesne odbicia. Opóźnienie to mierzone jest przez algorytm [fragment kodu z omówieniem].



Kolejnym elementem jest gęstość ech występujących w ogonie pogłosowym, która jest wymagana, aby dźwięk wydawał się naturalny i aby nie wystapił efekt flutter. Zgodnie z badaniami Schroedera szacuje się, że gęstość ta wynosi około 1000 ech na sekundę.

### Pomiar czasu pogłosu

Algorytm wyliczający czas pogłosu bazuje na metodzie impulsowej. Do wykonania nagrań odpowiedzi impulsowej jako użyto różnego źródła dźwięku różnego rodzaju, między innymi – pękający balon, trzaśnięcie, klaśnięcie, aby oszacować wpływ jakości nagrania odpowiedzi impulsowej na późniejszą operację splotu [całe badanie ma charakter subiektywnego postrzegania działania algorytmu, aniżeli odwzorowywania pomiarów zgodnie ze sztuką].

Do zminimalizowania poziomu tła w nagraniach dokonano analizy w pasmach częstotliwościowych [jakich? – pasma +spadki] przy użyciu filtrów cyfrowych, co pozwoliło na uzyskanie większego zakresu (około 30-50dB) zaniku poziomu dźwięku pozwalającego oszacować jego spadek o 60dB.

Nagrania wykonano w trzech miejscach położenia mikrofonu oraz trzech miejscach położenia źródła dźwięku (co daje razem 9 nagrań na jedno pomieszczenie).

Z założeń projektowych algorytmu, przy analizie nagrania pominięto częstotliwości mniejsze, jak 120Hz ze względu na dodatkowe trudności występujące w analizie tychże częstotliwości, takie jak rezonanse osiowe, styczne i skośne, a także wpływ jakości mikrofonu na pomiar częstotliwości w tym zakresie. Zjawiska te mogą wpływać na dodatkowe błędy, które utrudnią analizę działania projektowanego algorytmu, z tego powodu zostaną one pominięte, a analiza zostanie przeprowadzana w przypadkach „idealnych”, czyli niestwarzających dodatkowych problemów.

# Uczenie maszynowe

## Wstęp

W celach eksperymentalnych założono implementację algorytmów uczenia maszynowego do wcześniej omówionego programu.

Uczenie maszynowe pozwala na optymalizację oraz znajdywanie rozwiązań dla złożonych problemów w mniej konwencjonalny sposób, niż wierne sztuce metody opisywane w normach.

1. 32 bitowa liczba całkowita z zakresu od do . [↑](#footnote-ref-1)