


**TTS Studio**  
Instrukcja obsługi

## Spis treści

<b>1. Logowanie</b>	<b>3</b>
<b>2. Widok główny</b>	<b>3</b>
2.1. Pole tekstowe . . . . .	4
2.2. Pole parametrów syntezy . . . . .	6
2.3. Pole parametrów fonemów . . . . .	7
2.3.1. Wysokość fonemu . . . . .	7
2.3.2. Czas trwania fonemu . . . . .	8
2.3.3. Głośność fonemu . . . . .	8
2.3.4. Modyfikacja wartości . . . . .	8
2.3.5. Tryby nawigacji . . . . .	8
2.3.6. Presety . . . . .	9
2.3.7. Resetowanie . . . . .	9
2.4. Pole odtwarzacza . . . . .	9
2.5. Pole eksportu do pliku . . . . .	9
2.5.1. Częstotliwość próbkowania . . . . .	10
2.5.2. Kodek audio . . . . .	10
2.6. Pole zarządzania projektami i komunikatami . . . . .	10
<b>3. Widok projektów</b>	<b>11</b>
<b>4. Inne</b>	<b>13</b>
4.1. Powiadomienia . . . . .	13
4.2. Skróty klawiszowe . . . . .	13

## 1. Logowanie

W celu uzyskania dostępu do TTS Studio, należy wpisać w polach formularza **nazwę użytkownika** oraz **hasło**, a następnie wcisnąć  lub kliknąć przycisk **Zaloguj**. W przypadku wpisania nieprawidłowych danych, aplikacja wyświetli komunikat o błędzie.

## 2. Widok główny

Jest to główne miejsce pracy nad komunikatem głosowym. Na (Rys. 1) przedstawiono interfejs graficzny aplikacji, z ponumerowanymi i zaznaczonymi na kolorowo polami, odpowiedzialnymi za poszczególne funkcjonalności.



Rys. 1: Widok główny aplikacji z zaznaczonymi na kolorowo poszczególnymi polami

Interfejs graficzny składa się z następujących pól:

1. Pole tekstowe
2. Pole parametrów syntezy
3. Pole parametrów fonemów
4. Pole odtwarzacza
5. Pole eksportu do pliku
6. Pole zarządzania projektami i komunikatami

Będą one szczegółowo omówione w odpowiadającym ich sekcjach.

## 2.1. Pole tekstowe

Jest ono przeznaczone do wpisania **komunikatu** przygotowanego do syntezy. Jego treść może być zwykłym tekstem, natomiast aplikacja obsługuje również **tagi SSML**, czyli specjalny system znaczników, które pozwalają na **szczegółowe dopracowanie** brzmienia komunikatu do indywidualnych potrzeb użytkownika. Aby ułatwić korzystanie z tagów SSML, na dole pola tekstowego znajdują się przyciski do ich dodawania.

Istnieją **dwa typy tagów**: takie, które są **umieszczane w dowolnym miejscu** tekstu, czyli:

### Przykład

```
To było <break time="100ms"/> niesamowite!
```

oraz takie, które **obejmują dowolny fragment** tekstu, czyli:

### Przykład

```
Banany kosztują aż <say-as interpret-as="currency"> 4.99 zł </say-as>!
```

Przed wyborem tagów pierwszego typu trzeba jedynie zaznaczyć kursorem **miejsce w tekście**, w które chcemy, aby tag został wstawiony. Dla tagów drugiego typu należy zaznaczyć **fragment tekstu**, który ma być przez nie objęty. W przeciwnym razie, w miejscu kursora zostanie wygenerowany tag z pustym miejscem na tekst.

W przypadku wstawiania tagów **wymagających jednostek** należy pamiętać, żeby **uzupełnić** je o pożądaną wartość, przed symbolem jednostki, bez spacji.

### Przykład

```
<prosody rate="150%"> Przed użyciem zapoznaj się z ulotką. </prosody>
```

TTS Studio umożliwia wykorzystanie następujących tagów:

1. `break` umożliwia dodanie **ciszy**. Można ją zdefiniować poprzez podanie czasu w **sekundach** lub w **milisekundach** lub wybrać jedną z **predefiniowanych długości** przerw.

### Przykład

```
Hej, Ty! <break strength="weak"/> Jak mogłeś!
```

2. `emphasis` służy do położenia lub zredukowania **nacisku** na konkretny fragment komunikatu. Dostępne jest kilka **predefiniowanych poziomów intensywności**.

### Przykład

```
To <emphasis level="moderate"> było </emphasis> niesamowite!
```

3. `phoneme` umożliwia dodanie fragmentu zapisanego **międzynarodowym alfabetem fonetycznym** (IPA). Zaleca się jego wykorzystywanie w ostateczności, gdy nie można już wpłynąć na syntezę, modyfikując zapis ortograficzny. Symbole fonetyczne można wprowadzać przy użyciu klawiatury IPA, którą włącza się przetączykiem nad polem tekstowym. Jest ona w formie **pływającego** widgetu, więc można ją umieścić w dowolnym miejscu okna.

**Przykład**

```
Firma <phoneme ph="g'ugl"/> zanotowała znaczący wzrost cen akcji
```

4. `pitch` powoduje ustawienie **średniej wysokości głosu** na wybraną część komunikatu. Wysokość można ustawić, podając częstotliwość bezwzględnie (w **hercach**), względnie (w **procentach** lub **półtonach**) od częstotliwości bazowej. Dostępne są również **predefiniowane** wartości.

**Przykład**

```
<prosody pitch="-7st"> Proszę uregulować należności. </prosody>
```

5. `range` odpowiada za sterowanie **zróżnicowaniem wysokości głosu**. Parametr ten można ustawić, podając częstotliwość bezwzględnie (w **hercach**) lub względnie (w **procentach** lub **półtonach**). Dostępne są również **predefiniowane** wartości.

**Przykład**

```
<prosody range="0Hz"> Jestem tylko robotem. </prosody>
```

6. `rate` służy do ustawiania **prędkości mówienia**. Parametr ten można zdefiniować względnie (w **procentach**) lub wybrać którąś z **predefiniowanych** wartości.

**Przykład**

```
<prosody rate="250%"> Przed użyciem zapoznaj się... </prosody>
```

7. `say as` to **bardzo użyteczny tag**, który wykorzystywany jest do czytania: dat, skrótowców, numerów telefonów, cen, jednostek, adresów mailowych lub adresów stron internetowych, odpowiednio je rozwijając. Używanie tych tagów może być pomocne **komunikatach generowanych automatycznie**, przyjmujących zmienne wartości jako argumenty.

**Przykład**

```
Potrzebujemy <say-as interpret-as="unit"> 2.5 kg </say-as> ziemniaków.
```

8. `sound icon` pozwala na dodanie do komunikatu **predefiniowanych nagrań**, zawierających **podstawowe zwroty** jak na przykład: *Dzień dobry!* lub *Do usłyszenia!* oraz **reakcje** takie

jak: *Aha!*, *Hmm...* lub *Mhm*. Warto je wykorzystywać, w przypadku gdy chce się nadać komunikatowi bardziej **naturalny** charakter. Należy zaznaczyć, że nie jest on tagiem SSML, jak pozostałe. Zapisuje się go jako nazwę predefiniowanego nagrania, otoczoną z dwóch stron **dwukropkami**.

#### Przykład

```
:hmm2: To interesująca propozycja!
```

9. `voice` pozwala na to, aby wybrany fragment komunikatu był czytany przez konkretnego mówcę.


#### Przykład

```
<voice name="Masza1"> Hej! Jestem Masza! </voice>
```

10. `volume` odpowiada za sterowanie głośnością wybranych partii komunikatu.

#### Przykład

```
<prosody volume="x-loud"> Hej! </prosody> Wybacz, że krzyknęłam.
```

Domyślnie audio jest generowane automatycznie, w miarę wpisywania treści komunikatu. Automatyczną aktualizację stanu można wyłączyć przełącznikiem nad polem tekstowym. Wtedy przed odtworzeniem syntezy wymagana jest aktualizacja stanu przez użytkownika, która następuje poprzez wciśnięcie , gdy w polu tekstowym jest migający kursor.

## 2.2. Pole parametrów syntezy

Ustawione w tym panelu parametry są zastosowywane do **całej** generowanej **wypowiedzi**. Są one realizowane przy pomocy tagów SSML, widocznych na górze i na dole pola tekstowego. Dostępnych jest kilka opcji:

1. `Opcje głosu` umożliwiają wybranie **języka** (pole z flagą) oraz dostępnego w jego ramach **mówcy** (pole z symbolem płci) w wybranym **wariancie** (pole z kratką).
2. `Ekspresja` odpowiada za **zróżnicowanie wysokości głosu** w procentach względem wartości domyślnej. Wysoka wartość tego parametru powoduje, że głos będzie bardziej **melodyjny**, natomiast jego niska wartość sprawi, że głos stanie się bardziej **monotonny** (robotyczny).
3. `Wysokość` służy do ustalenia **średniej wysokości głosu** w procentach względem wartości domyślnej.
4. `Głośność` odpowiada za wzmocnienie lub osłabienie energii wypowiedzi o wybraną wartość w decybelach.

5. **Prędkość** służy do sterowania **prędkością wypowiedzi** w procentach względem wartości domyślnej.

Parametry takie jak **Ekspresja** oraz **Wysokość** mogą być ustawiane nie tylko przy pomocy suwaków, ale również poprzez wybranie jednej z predefiniowanych wartości. Dostępne są one po kliknięciu przycisku **Presety**, przy polu z nazwą parametru.

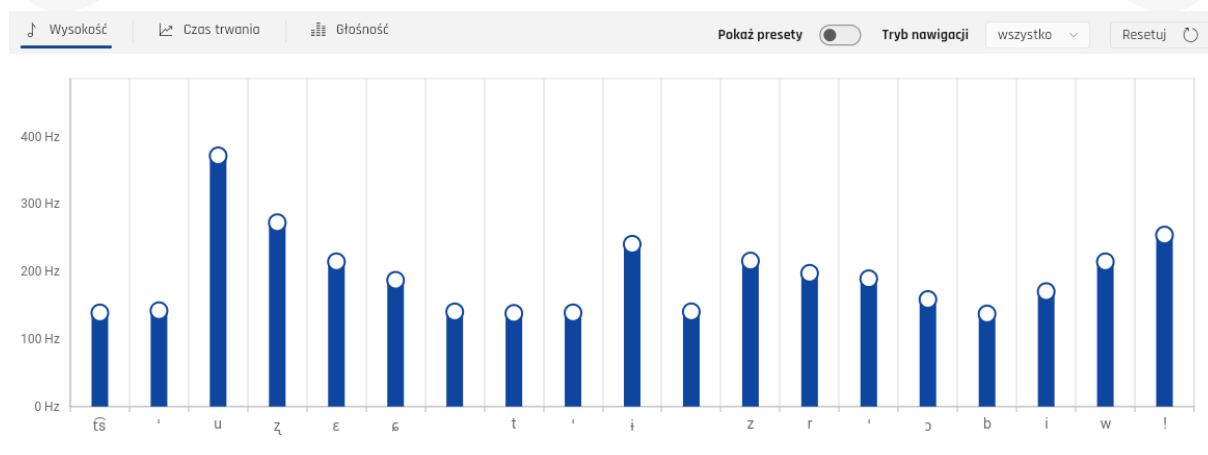
Aby szybko przywrócić stan parametrów do wartości domyślnych, można kliknąć przycisk **Resetuj** w dolnej części panelu. Należy jednak pamiętać, że resetowanie **nie dotyczy** parametru **Opcje głosu**. Wybrany język, mówca oraz wariant pozostaną niezmienione.

## 2.3. Pole parametrów fonemów

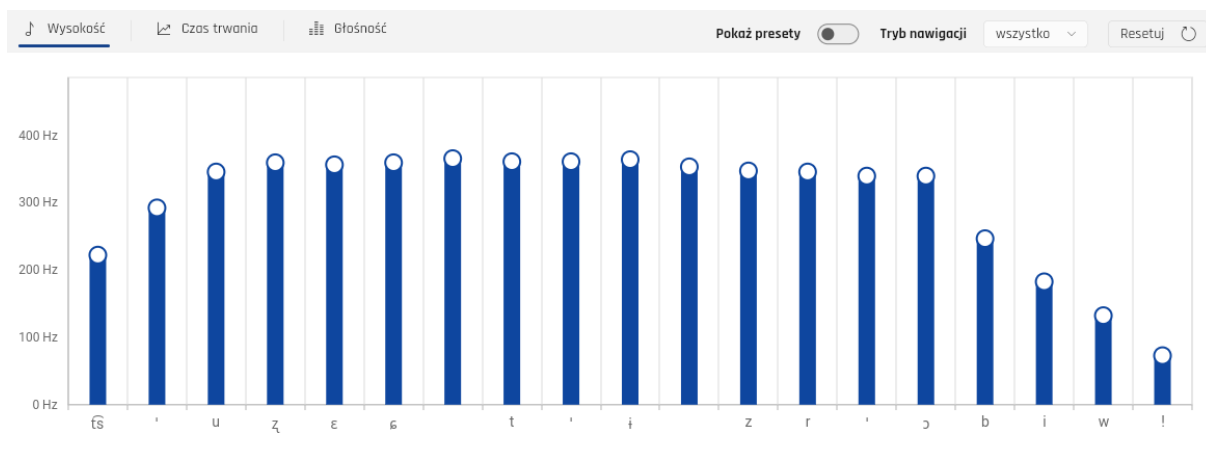
W wielu językach (również w polskim) zapis ortograficzny nie zawsze odpowiada wymowie. Na przykład *rz* czasami czytane jest jak [z] (*drzewo*), czasami jak ʂ (*krzak*), a czasami jak [rz] (*cyberzłoczyńca*). Dlatego, tekst komunikatów transkrybowany jest na **międzynarodowy alfabet fonetyczny (IPA)**. Każdy z fonemów posiada trzy sterowalne parametry: **wysokość**, **czas trwania** oraz **głośność**. *TTS Studio* umożliwi manipulację tymi parametrami dla **każdego fonemu z osobna**.

### 2.3.1. Wysokość fonemu

Sterując wysokościami poszczególnych fonemów (w hercach), można modelować **kontur melodii** wypowiedzianych zdań. Kontrola nad tym parametrem jest kluczowa, gdyż to on w dużej mierze kształtuje **charakter wypowiedzi**. Dla przykładu, podnosząc melodię na końcu zdania, można zaintonować pytanie (Rys. 2), a wysoka, monotonna melodia, opadająca na końcu zdania daje wrażenie krzyku (Rys. 3). Ponadto podwyższenie intonacji na wybrane fonemy może służyć do **akcentowania** niektórych sylab lub nawet całych wyrazów.



Rys. 2: Próba intonacji pytania



Rys. 3: Próba intonacji krzyku

### 2.3.2. Czas trwania fonemu

Czasy trwania fonemów (w milisekundach) również są użytecznym parametrem. Zwiększanie ich dla niektórych fonemów może służyć do **akcentowania**. Ponadto, zwiększając czas trwania interpunkcji, można zwiększyć przerwy między wyrazami.

### 2.3.3. Głośność fonemu

Głośność fonemów (w decybelach) w stosunku do jej wartości domyślnej może zostać wykorzystana do **zdynamizowania** wypowiedzi, różnicując głośność jej fragmentów.

### 2.3.4. Modyfikacja wartości

Wartości wysokości fonemów, czasu ich trwania oraz głośności mogą być ustawiane zarówno **pojedynczo**, jak i **zbiorczo**. Pojedynczy fonem można modyfikować poprzez **przeciąganie** jego **suwaka** lub **klikanie** na odpowiednią wartość w jego polu. Aby zastosować zbiorczą edycję, należy kliknąć suwaki fonemów, które chce się zmodyfikować, z wciśniętym przyciskiem `ctrl`. Zaznaczone suwaki zmieniają kolor na **pomarańczowy**. Po zaznaczeniu można je przesuwac podobnie jak pojedyncze fonemy.

### 2.3.5. Tryby nawigacji

Przy długich komunikatach wykres fonemów może stracić na czytelności. Dlatego w górnej części pola dostępne są trzy tryby nawigacji: `wszystko`, `zdanie` i `słowo`. Dzięki nim można podglądać cały komunikat, pojedyncze zdania lub słowa. W miarę pisania komunikatu, gdy osiągnie on pewną długość, **tryb nawigacji automatycznie przełącza się** najpierw w tryb `zdanie`, a potem w `słowo`.



### 2.3.6. Presety

Pole z wykresami posiada w swoim górnym pasku przełącznik **Pokaż presety**. Włączając go, użytkownik dostaje dostęp do ukrytego panelu, który umożliwi edycję **wszystkich suwaków widocznych obecnie na wykresie** (gdy nie ma żadnych zaznaczonych) **lub wszystkich zaznaczonych**, które są w obecnym widoku. Dostępne są trzy tryby:

1. **zwiększ** **dodaje** do aktualnej wartości suwaków zadaną wartość.
2. **zmniejsz** **odejmuje** od aktualnej wartości suwaków zadaną wartość.
3. **spłaszcz** **zmniejsza różnicowanie** parametrów fonemów poprzez zbliżanie ich do wartości średniej. Szczególnie przydatne w przypadku spłaszczania melodii.

Po ustawieniu trybu i współczynnika należy wcisnąć przycisk **Zastosuj**. Zaleca się ustawianie małych współczynników i wielokrotne zastosowywanie zmian, aż do osiągnięcia pożądanego efektu.

### 2.3.7. Resetowanie

Aby przywrócić suwaki do wartości domyślnych, należy kliknąć przycisk **Resetuj**. Należy jednak pamiętać, że resetowane są suwaki **tylko w aktualnym widoku**. Resetując suwaki dla czasu trwania fonemów, ich wartości dla czasu trwania i głośności pozostaną niezmienione.

## 2.4. Pole odtwarzacza

W dolnej części ekranu znajduje się odtwarzacz. Posiada on możliwość **odtworzenia** zsyntetyzowanego komunikatu, **pauzowania** oraz **przesuwania** do dowolnego miejsca na nagraniu. Należy zwrócić szczególną uwagę, że **przy zmianie** treści komunikatu, parametrów syntezy lub właściwości poszczególnych fonemów, **audio jest generowane na nowo**.

## 2.5. Pole eksportu do pliku

Aby zapisać plik dźwiękowy z wygenerowanym komunikatem na dysk, należy wybrać **częstotliwość próbkowania** oraz **kodek audio**, po czym kliknąć przycisk **Pobierz audio**. Następnie, wyświetlone zostanie okno dialogowe, w którym program poprosi o podanie **nazwy pliku**. Po kliknięciu **Pobierz** zostanie wysłane do przeglądarki żądanie pobrania pliku. W zależności od ustawień przeglądarki, plik zostanie pobrany albo do **domyślnej lokalizacji**, albo system poprosi o **wybranie folderu**.

### 2.5.1. Częstotliwość próbkowania

Dostępne są trzy częstotliwości próbkowania. Najwyższa, 22050 Hz pozwala na **rzetelną ocenę jakości** syntezy. Pośrednia, 16000 Hz jest dobra na przykład do komunikatów w **aplikacjach mobilnych**. Najniższa z nich, czyli 8000 Hz, służy głównie do symulacji, jak komunikaty będą słyszane na **linii telefonicznej**.

### 2.5.2. Kodek audio

Dostępne są cztery kodeki audio. `PCM16` jest najprostszym, **bezstratnym** kodekiem. Plik taki waży więcej, natomiast ma najlepszą jakość. Zapisywany jest z rozszerzeniem `wav`. `OGG_VORBIS` oraz `OGG_OPUS` (nowszy) są natomiast kodekami kompresji **stratnej**. Oznacza to, że pliki będą miały znacznie mniejszy rozmiar, kosztem drobnej (często niezauważalnej) utraty jakości. Zapisywane są z rozszerzeniem `ogg`. Ostatnie kodeki, czyli `A_LAW` oraz `MU_LAW` również przeprowadzają kompresję stratną. Znajdują one zastosowanie w telefonii. Pliki są mniejsze ze względu na konwersję dynamiki do mniejszej liczby bitów. Zapisywane są z rozszerzeniem `wav`.

## 2.6. Pole zarządzania projektami i komunikatami

W górnej części okna, nad polem tekstowym, znajduje się pięć przycisków:

1. `Nowy projekt`
2. `Zapisz` (przy wczytanym projekcie)
3. `Zapisz jako`
4. `Kopiuj komunikat`
5. `Wygeneruj komunikat`

Poniżej, znajdują się karty otwartych projektów. Gdy w projekcie istnieją niezapisane zmiany, na jego karcie, przy nazwie projektu, pojawia się **gwiazdka**. Gdy otwarte są **więcej niż dwie karty**, każda z nich posiada ikonkę krzyżyka, umożliwiającą jej zamknięcie. W przypadku otwarcia tylko jednego projektu, nie można go zamknąć. W widoku głównym zawsze musi być otwarty **przynajmniej jeden projekt**.

Każdy projekt posiada **autora**. Projekty innych autorów można otwierać, natomiast nie można ich edytować. Zapisywać zmiany można tylko we własnych projektach. W przypadku potrzeby modyfikacji projektu innego użytkownika zaleca się zapisanie go jako inny projekt, do którego będzie się miało uprawnienia.

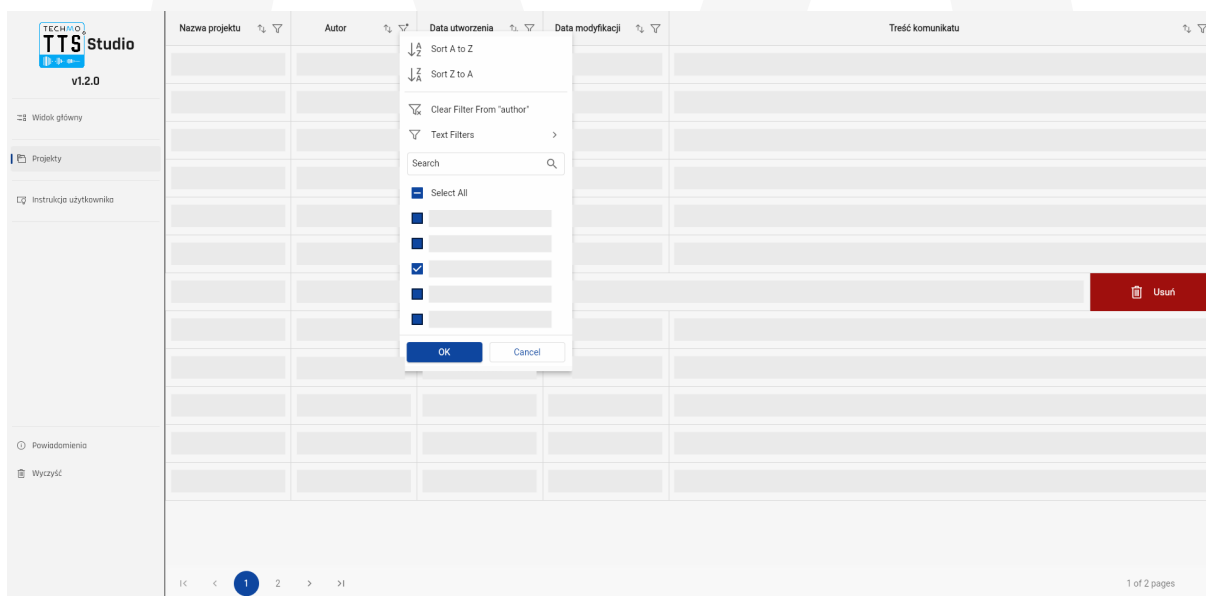
Przycisk `Kopiuj komunikat` pozwala na skopiowanie do schowka pełnej treści komunikatu, rozumianej jako treść wpisana przez użytkownika oraz tagi w górnej i dolnej części pola

tekstowego. Należy jednak podkreślić, że kopiowanie komunikatu nie uwzględnia zmian wprowadzonych w parametrach indywidualnych fonemów.

Aby komunikat zawierał w sobie wszystkie zmiany poczynione przez użytkownika (m.in. edycja wartości parametrów poszczególnych fonemów), należy kliknąć **Wygeneruj komunikat**. Spowoduje to zakodowanie komunikatu w specjalnym tagu SSML `say-as`, którego można użyć bezpośrednio w serwisie TTS. Należy pamiętać o tym, że aby zakodowany komunikat został odpowiednio zinterpretowany przez serwis, serwis musi operować na zgodnej z TTS Studio wersji zasobów oraz posiadać zgodną z aplikacją wersję. **Ręczna edycja zawartości wygenerowanego komunikatu może spowodować nieoczekiwane efekty syntezy.**

### 3. Widok projektów

W tym miejscu widoczne są wszystkie projekty osób w danej firmie (Rys. 4). Każdy projekt jest odrębnym komunikatem głosowym, z zapisanymi wszystkimi ustawieniami. Projekty innych użytkowników można wczytać i edytować, natomiast nie można zapisać wprowadzonych zmian. Aby je zachować, należy zapisać je jako odrębny projekt, którego będzie się właścicielem.



Rys. 4: Widok projektów z zasłoniętymi prywatnymi danymi

Aby **otworzyć** dany projekt, należy go **dwukrotnie kliknąć**. Aby go **usunąć**, trzeba **przesunąć linijkę w lewo**, aż pokaże się czerwony przycisk **Usuń** z ikonką kosza (Rys. 4), który należy kliknąć.

Tabela z projektami obsługuje **sortowanie** oraz **filtrowanie**, których kryteria można ustawić klikając na **ikonę lejka** w nagłówkach kolumn. Filtrowanie może być przeprowadzone na dwa sposoby. Pierwszy z nich (domyslny) polega na zaznaczaniu i odznaczaniu wartości istniejących

w danej kolumnie, natomiast drugi, dostępny po kliknięciu [Text Filters](#) lub [Date Filters](#) opiera się na podawaniu kryteriów połączonych operatorami koniunkcji (*and*) lub alternatywy (*or*).



## 4. Inne

### 4.1. Powiadomienia

W lewej części ekranu znajduje się pole powiadomień. Dostępne są ich trzy rodzaje:

1. Sukces
2. Info
3. Błąd

Powiadomienia można wyczyścić przyciskiem **Wyczyść**.

### 4.2. Skróty klawiszowe

W całym programie można używać skrótów **ctrl** + **Z** (do cofania) lub **ctrl** + **↑** + **Z** (do przywracania zmian).