

# Aplikacja przetwarzająca mowę ludzką na tekst

---

## Podział funkcjonalności aplikacji

Aplikacja składa się z trzech paneli.

**Na żywo** - panel umożliwiający przetwarzanie mowy na żywo używając mikrofonu.

**Z pliku** - panel umożliwiający przetwarzanie plików audio z nagrany głosem. Aplikacja wspiera pliki audio z rozszerzeniem WAV.

**Ustawienia** - panel umożliwiający wybranie języka aplikacji i języka przetwarzanej mowy.

## Korzystanie z paneli

### Panel na żywo

Aby rozpocząć przetwarzanie, należy nacisnąć przycisk **rozpocznij rozpoznawanie mowy**. Przetworzony tekst pojawi się po rozpoznaniu na ekranie.

### Panel z pliku

Po otwarciu tego panelu zostanie wyświetlony eksplorator plików. Należy wybrać w nim plik, który chcemy przetworzyć, a następnie nacisnąć **rozpocznij transkrypcję plików**. Po przetworzeniu na panelu pojawi się pole tekstowe z gotową transkrypcją.

### Panel Ustawienia

Na tym panelu znajdują się dwa rzędy ustawień. Pierwszy z nich służy do wybrania języka aplikacji, drugi do wybrania języka przetwarzanej mowy. Aby zmienić ustawienie, należy na nią kliknąć. Wybrana opcja zostanie pogrubiona.

---

# Konstrukcja plików aplikacji

Aplikacja składa się z czterech plików. Każdy z nich jest niezbędny do jej poprawnego działania

`main.py` główny skrypt aplikacji

`readConfig.py` skrypt odpowiedzialny za wczytywanie danych aplikacji takich jak dane językowe, czy konfiguracyjne aplikacji

`recognition.py` skrypt odpowiedzialny za przetwarzanie mowy na tekst

`app.kv` plik kivy, plik odpowiedzialny za przechowywanie interfejsu graficznego

## Wymagania do poprawnego działania

Aplikacja w celu poprawnego działania potrzebuje odpowiedniej wersji Pythona, bibliotek zewnętrznych i sterowników audio.

### **Ogólne wymagania aplikacji to:**

Najnowsza wersja sterowników audio

Python w wersji 3.8

Biblioteka kivy w wersji 2.0.0

Biblioteka SpeechRecognition w wersji 3.8.1

Biblioteka pydub w wersji 0.25.1

Biblioteka pyaudio w wersji 0.2.11

# Instalacja aplikacji

## Windows

Windows 8 | Windows 10 | Windows 11

1. Pobierz i zainstaluj w wersji 3.9 **Python** (<https://www.python.org/downloads/>).
2. Otwórz terminal i uruchom komendę `pip install Kivy==2.0.0 SpeechRecognition==3.8.1 pydub==0.25.1 numpy==1.20.3 pyaudio==0.2.11`
3. Gdy wystąpią problemy z instalacją **pyaudio** należy ręcznie pobrać odpowiedni plik .whl (<https://www.lfd.uci.edu/~Egohke/pythonlibs/#pyaudio>), a następnie zainstalować go, używając komendy `pip install <ścieżka pliku>\<nazwa pobranego pliku .whl>`
4. Uruchom aplikację komendą `python main.py` upewniając się, że ścieżka terminala to ścieżka z plikami aplikacji

## Linux - systemy z wbudowanym wsparciem PPA

Linux Mint | Ubuntu

1. Uruchom terminal
2. Wpisz komendę `sudo add-apt-repository ppa:deadsnakes/ppa`
3. Wpisz komendę `sudo apt-get update`
4. Wpisz komendę `sudo apt-get install python3.8 python3.8-dev build-essential python3.8-distutils ffmpeg xclip pulseaudio`
5. Wpisz komendę `sudo apt install python3-pip`
6. Wpisz komendę `sudo apt-get update`
7. Wpisz komendę `sudo apt-get install libasound-dev portaudio19-dev`
8. Wpisz komendę `python3.8 -m pip install kivy==2.0.0 SpeechRecognition==3.8.1 pydub==0.25.1 numpy==1.20.3 pyaudio==0.2.11`
9. Upewnij się, że ścieżka terminala prowadzi do folderu z plikami aplikacji. Uruchom aplikację używając komendy `python3.8 main.py`

## Linux

MX Linux | Debian | elementary OS | i inne

1. Uruchom terminal
2. Wpisz komendę `sudo apt-get update`
3. Wpisz komendę `sudo apt install build-essential zlib1g-dev libncurses5-dev libgdbm-dev libnss3-dev libssl-dev libsqlite3-dev libreadline-dev libffi-dev curl libbz2-dev`
4. Wpisz komendę `curl -O https://www.python.org/ftp/python/3.8.2/Python-3.8.2.tar.xz`
5. Wpisz komendę `tar -xf Python-3.8.2.tar.xz`
6. Wpisz komendę `cd Python-3.8.2`

7. Wpisz komendę `./configure --enable-optimizations`
8. Wpisz komendę `nproc`. Wyświetli ona liczbę rdzeniów maszyny. Ta liczba będzie potrzebna w następnym kroku.
9. Wpisz komendę `make -j <liczba rdzeniów>` korzystając z liczby która została wyświetlona przy ostatniej komendzie
10. Wpisz komendę `sudo make altinstall`
11. Wpisz komendę `sudo apt install python3-pip`
12. Wpisz komendę `sudo apt-get update`
13. Wpisz komendę `sudo apt-get install libasound-dev portaudio19-dev`
14. Wpisz komendę `sudo python3.8 -m pip install kivy==2.0.0 SpeechRecognition==3.8.1 pydub==0.25.1 numpy==1.20.3 pyaudio==0.2.11`
15. Upewnij się, że ścieżka terminala prowadzi do folderu z plikami aplikacji. Uruchom aplikację używając komendy `python3.8 main.py`

## Testowane systemy

Funkcjonalność aplikacji została sprawdzona na systemach Windows 10, Windows 11, Linux Mint 20.3, MX Linux 21.1, Ubuntu 22.04.

Na tych systemach operacyjnych aplikacja nie wykazała problemów z działaniem. Poniżej znajdują się zakładki które pomogą ci rozwiązać problemy z aplikacją, jeżeli na takie trafiłeś.

## Rozwiązywanie problemów

**Rozpoznawanie mowy na żywo nie działa, a terminal wypisuje błędy z rodziny `OSError`:**

*Błąd sterowników. Sprawdź czy zainstalowałeś poprawną wersję **pyaudio** dla swojego systemu. Zaktualizuj sterowniki systemowe. Na systemach Linux użyj komendy `sudo apt-get update && sudo apt-get upgrade`*

**Rozpoznanie mowy na żywo pokazuje, że mikrofon działa, ale nie pojawia się transkrypcja.**

*Sprawdź, czy masz połączenie z internetem.*

***Plik został przetworzony, ale nie pojawia się transkrypcja.***

1. *Sprawdź czy masz połączenie z internetem.*
2. *Sprawdź czy wybrałeś poprawny język transkrypcji.*

***Interfejs graficzny nie działa poprawnie lub aplikacja crashuje***

1. *Sprawdź czy korzystasz z **Python3.8***
2. *Sprawdź czy korzystasz z **kivy** w wersji **2.0.0**. Wersja **2.1.0 nie jest** wspierana*