


Karolina Pieniowska  <https://orcid.org/0000-0002-5023-8952>
Uniwersytet Adama Mickiewicza w Poznaniu, Wydział Neofilologii, Zakład Fonetyki
karolina.pieniowska@amu.edu.pl

Narzędzia do analizy mowy - przegląd dostępnych programów i bibliotek

Streszczenie

Artykuł przedstawia zestawienie popularnych narzędzi do analizy mowy w formie programów dostępnych do pobrania i w formie bibliotek w różnych językach programowania. W pierwszej części zestawione zostały programy służące do wizualizacji sygnału mowy, edytowania, analizy (na przykład pomiarów częstotliwości podstawowej, intensywności czy formantów) oraz anotacji (segmentacji, transkrypcji i etykietowania nagrań). W drugiej części przedstawiono wybrane biblioteki dostępne na stronie GitHub, które służą do akustycznej, fonetyczno-fonologicznej oraz prozodycznej analizy nagrań. Wszystkie narzędzia zostały opisane z uwzględnieniem ich funkcji i możliwości, źródeł, autorów, licencji, na jakich są udostępniane. W ostatnim rozdziale artykułu podjęto próbę ewaluacji opisanych programów, biorąc pod uwagę liczbę i użyteczność ich funkcjonalności.

Słowa kluczowe: analiza mowy, analiza akustyczna, prozodia, programy, biblioteki

Wstęp

Nowe technologie komputerowe umożliwiły szybki i znaczny rozwój badań językoznawczych i spowodowały powstanie nowej dziedziny łączącej technologie informatyczne z lingwistyką – językoznawstwo komputerowe. Oprogramowanie i urządzenia wykorzystywane w badaniach językoznawczych umożliwiły rozwój nowych obszarów działalności naukowej i rozwoju technologii, takich jak tłumaczenie maszynowe, systemy do rozpoznawania i syntezy mowy, automatyczna ekstrakcja terminów itd. (Łukasik, 2009).

W celu rozwoju badań powyższych dyscyplin tworzone są często korpusy językowe. Korpus mowy to zbiór nagrań mowy wraz z zestawem informacji: transkrypcjami, anotacjami i metadanymi, często również z podstawowymi analizami statystycznymi danych (Leech, 2005). Informacje te są zgodne ze standardami, co umożliwia wykorzystywanie korpusów przez różne systemy, oprogramowania (Leech, 2005).

W dobie szybkiego rozwoju technologii mowy powstało wiele programów oraz bibliotek w różnych językach programowania służących do analizy mowy. Językoznawcy komputerowi stoją przed wyzwaniem wyboru najbardziej efektywnych metod i narzędzi do badań.

W tej pracy przedstawiono zestawienia programów oraz bibliotek dostępnych na GitHub do analizy akustycznej, fonetyczno-fonologicznej i prozodycznej mowy. Narzędzia zostały opisane z uwzględnieniem ich głównych funkcji i możliwości, dostępności (licencji) i twórców. W ostatnim rozdziale artykułu podjęto próbę oceny narzędzi, biorąc pod uwagę szczególnie ich funkcjonalność.

1. Programy do analizy mowy

1.1. Praat

<http://www.fon.hum.uva.nl/praat/> [dostęp: 15.05.2019]

Praat to program do naukowej analizy mowy i zjawisk fonetycznych, który stworzyli i rozwinęli dwaj profesorowie fonetyki Paul Boersma i David Weenink z Instytutu Nauk Fonetycznych na Uniwersytecie Amsterdamskim. Praat jest bardzo elastycznym narzędziem do analizy mowy. Oferuje szeroki zakres standardowych i niestandardowych procedur, w tym analizę spektrograficzną, syntezę artykulacyjną i sieci neuronowe. Posiada całą gamę funkcji do analizy mowy, syntezy mowy, segmentacji i tagowania nagrań.

Praat umożliwia analizę mowy, na przykład analizę spektralną (spektrogramy), analizę wysokości tonu, analizę formantów, analizę intensywności, przerwy głosowe. Za pomocą programu można również wykonać syntezę artykulacyjną i akustyczną Klatta. Wspiera on tworzenie i przeprowadzanie testów identyfikacyjnych i dyskryminacyjnych. W programie Praat można łatwo przeprowadzić segmentację, transkrypcję i etykietowanie nagrań w rozmiarze nawet do 2 GB. Praat wykorzystuje również algorytmy i sieci neuronowe. Można w nim wykonywać analizy statystyczne, takie jak skalowanie wielowymiarowe, analiza głównych składowych i analiza dyskryminacyjna.

Program działa pod systemami Windows, Linux, Macintosh, FreeBSD i Solaris i jest dostępny bezpłatnie, do pobrania ze strony. Kod źródłowy Praata jest dostępny na licencji GNU General Public License. Autorzy zachęcają do dodawania nowych funkcjonalności i ulepszania programu poprzez modyfikowanie i rozszerzanie kodu źródłowego w języku C lub C++.

1.2. AnnotationPro

<http://annotationpro.org/> [dostęp: 15.05.2019]

AnnotationPro to program do anotacji plików dźwiękowych i tekstowych, który umożliwia tworzenie wielu równoległych warstw anotacji. Narzędzie wyposażone jest w funkcję zaznaczania fragmentów nagrań i ich przybliżania do segmentu lub punktu, odtwarzania zapętlenia dźwięku. W widoku spektrogramu można łatwo oznaczyć granice, wyświetlać je i ukrywać w miarę potrzeby, co usprawnia pracę nad anotacją. Program pomaga w zarządzaniu plikami, grupowaniu ostatnio używanych plików, co umożliwia zachowanie porządku i łatwy dostęp do tych materiałów, które w danym momencie mają

być wykorzystane. AnnotationPro automatycznie otwiera zestaw plików składający się z audio i anotacji. AnnotationPro ma intuicyjny interfejs, a graficzna reprezentacja cech sygnału mowy jest innowacyjnym rozwiązaniem stosowanym w tego typu oprogramowaniu.

AnnotationPro wspiera również projektowanie i przeprowadzanie testów percepcji poprzez funkcje, które umożliwiają wybór operacji dostępnych dla uczestników testu, takich jak anonimowość testu, ukrywanie nazw plików, liczba powtórzeń sygnału, dostępność opcji edycji, identyfikacja warstw anotacji dostępnych podczas testu itp.

Każdy użytkownik może uzupełnić program AnnotationPro o dodatkowe funkcje za pomocą wtyczek. Lista proponowanych, stworzonych wtyczek jest dostępna na stronie annotationpro.org. Twórcy oprogramowania udostępnili instrukcję do tworzenia wtyczek oraz link do API programu. Na liście wtyczek gotowych do pobrania i instalacji można znaleźć:

- statystyki – AnnotacjaPro + TGA – umożliwia przetwarzanie zbiorów plików (tryb Workspace),
- średnia ruchoma segmentów (*segment rate moving average*),
- średnia ruchoma nPVI (*nPVI moving average*) – oblicza znormalizowany indeks zmienności par (*normalized pairwise variability index*),
- szybkość segmentu w warstwie – oblicza szybkość segmentów w warstwie anotacji (w segmentach na sekundę),
- kwadraty czasu trwania – oblicza czas trwania kolejnych segmentów w z-punktach w warstwie anotacji i wyniki wykresów,
- analiza przestrzeni obiektów – przypisuje etykiety do obszarów wielokątów,
- eksportowanie wybranej warstwy do pliku CSV,
- eksportowanie wybranej warstwy jako siatek tekstowych,
- eksportowanie wybranej warstwy (określone przez „layerNames”) do formatu Text-Grid (Praat),
- usuwanie segmentów (określone przez „pausePattern”) z jednego lub więcej plików (tryb przestrzeni roboczej),
- funkcja znajdź i zamień – zastępuje jeden ciąg innym lub usuwa ciąg z anotacji (tryb przestrzeni roboczej),
- usuwanie pustych segmentów – usuwa segmenty, które nie zawierają żadnej etykiety w anotacji.

AnnotationPro jest dostępne do celów badawczych i edukacyjnych i może być używane bezpłatnie pod warunkiem zachowania informacji o prawach autorskich. Prawa autorskie do AnnotationPro należą do dr hab. Katarzyny Klessy z Uniwersytetu Adama Mickiewicza w Poznaniu.

1.3. Anvil

<http://www.anvil-software.org> [dostęp: 15.05.2019]

ANVIL to darmowe narzędzie służące do anotacji, które zostało opracowane przez Michaela Kippa z Uniwersytetu Nauk Stosowanych w Augsburgu. Program umożliwia anotację wielowarstwową w schemacie wybranym przez użytkownika. Podczas pracy anotujący ma przed sobą elementy oznaczone kolorami na wielu ścieżkach skoordynowanych według

czasu nagrania. Spośród specjalnych funkcji narzędzie oferuje ścieżki czasowe, obiekty beczasowe, linki wielopoziomowe, przeglądanie danych 3D przechwytywania ruchu oraz analizę zgodności kodowania czy narzędzie służące do zarządzania całym korpusami anotowanych nagrań. Anvil jest używany na wielu polach badań, między innymi w lingwistyce, interakcji człowiek–komputer, psychoterapii, animacji komputerowej, etologii, antropologii czy oceanografii, choć oryginalnie rozwinięto go w roku 2000 w celu badań gestów.

Dodatkową cechą ANVIL przydatną przy pracy z anotacją jest możliwość importowania danych z narzędzi fonetycznych takich jak opisany wyżej Praat. Pozwala to na wygodną i dokładną transkrypcję mowy. Narzędzie może wyświetlać nagrania w formie falowej lub konturu tonu. Dane zapisywane w plikach Anvil bazują na formacie XML, a opracowanych nagrań po wyeksportowaniu można użyć w analizie przy pomocy narzędzi statystycznych, takich jak Statistica czy SPSS. Nadchodząca wersja będzie obsługiwała również pliki w formacie ELAN.

Narzędzie napisane jest w języku Java, działa na platformach Windows, Macintosh oraz Unix.

1.4. Audíamus

<https://www.nthieberger.net/audiamus.htm> [dostęp: 15.05.2019]

Narzędzie opracowane przez Nicka Thiebergera ze Szkoły Języka i Lingwistyki Uniwersytetu Melbourne. Służy do budowania korpusów połączonych transkryptów, nagrań dźwiękowych i wideo, jednocześnie nie wymagając segmentacji plików audiowizualnych. Nie ma ograniczenia liczby transkryptów czy rozmiaru obsługiwanych plików. Każda zakładka opracowywanego modelu prezentuje pojedynczy transkrypt. Do programu wprowadza się dane zsynchronizowane czasowo, które opracowuje się wcześniej przez inne narzędzia jak na przykład SoundIndex czy Transcriber.

Transkrypcje w Audiamusie zapisywane są otwartym tekstem, a więc zarówno treść, jak i kody czasowe mogą być dowolnie edytowane. Dzięki temu dane można traktować jako główną kopię transkryptu, którą ulepsza się stopniowo wraz z użyciem. By uniknąć blokowania dostępu do danych przez użycie formatu plików o ograniczonej dostępności, Audiamus wyposażony jest w opcję masowego eksportu całego połączonego tekstu i kodów czasowych do otwartego tekstu lub jakiegokolwiek wybranego przez użytkownika formatu.

1.5. Dolmen

<https://dolmen.readthedocs.io/en/latest/> [dostęp: 15.05.2019]

Dolmen to darmowy zestaw narzędzi oferowany w formie otwartego kodu źródłowego opracowany przez Julię Eychenne’a z Uniwersytetu Studiów Zagranicznych w Hankuk. Służy do analizy danych lingwistycznych, oferując interfejs przyjazny użytkownikowi umożliwiający zarządzanie, anotację i tworzenie zapytań do korpusów językowych (*query language corpora*). Program jest dobrym narzędziem do pracy z danymi wyrównanymi w czasie. Główne możliwości i funkcje Dolmena:

- można prowadzić projekty przez organizowanie plików w projekty i zarządzanie wersjami;

- metadane w plikach mogą być anotowane wraz z właściwościami z możliwością sortowania i porządkowania danych;
- współdziałanie z programem Praat: Dolmen może analizować pliki formatu TextGrid i otwierać pliki bezpośrednio w narzędziu Praat;
- rozbudowana wyszukiwarka pozwala na tworzenie i zapisywanie złożonych zapytań oraz wzorców wyszukiwania między warstwami;
- oparty jest na standardach: pliki programu Dolmen są zapisywane w formatach XML i Unicode;
- silnik skryptów: Dolmen może być rozbudowany za pomocą wtyczek napisanych w JavaScript/JSON;
- działa na platformach Windows, Mac OS X oraz GNU/Linux i jest dostępny darmo na zasadach regulowanych przez GNU General Public License (GPL).

1.6. ELAN

<https://tla.mpi.nl/tools/tla-tools/elan/> [dostęp: 15.05.2019]

ELAN (*Eudico Linguistic Annotator*) to narzędzie do anotacji stworzone przez Instytut Psycholingwistyki Maxa Plancka w Nijmegen. Pozwala na tworzenie, edytowanie, wizualizację oraz wyszukiwanie anotacji dla danych audio i wideo. ELAN stworzony został, by dostarczyć solidną bazę technologiczną dla anotacji i opracowywania nagrań multimedialnych. Choć narzędzie zaprojektowano w celu analizy języka mówionego, migowego oraz gestów, to może być używane przez każdego pracującego z korpusami mediowymi w rodzaju danych audio lub wideo, w celach anotacji, analizy i dokumentacji.

ELAN umożliwia:

- wyświetlanie sygnału mowy i audiowizualnych wraz z anotacją,
- powiązanie anotacji do ścieżek nagrań względem czasu,
- łączenie anotacji z innymi anotacjami,
- definiowanie nieograniczonej liczby warstw anotacji przez użytkownika,
- użytkowanie różnych zestawów znaków,
- eksportowanie w postaci plików zawierających tekst podzielony tabulatorami,
- import oraz eksport pomiędzy narzędziami ELAN i Shoebox,
- wyszukiwanie.

1.7. WaveSurfer

<https://sourceforge.net/projects/wavesurfer/> [dostęp: 15.05.2019]

WaveSurfer to narzędzie pozwalające na wizualizację oraz manipulację nagrań dźwiękowych, dostępne na zasadzie otwartego kodu źródłowego. Głównymi jego zastosowaniami są anotacja oraz transkrypcja dźwięku, a także analiza dźwięku i mowy. Program WaveSurfer można wzbogacić w funkcje przez użycie wtyczek, jak również przez osadzenie go w innych aplikacjach.

Cechy narzędzia WaveSurfer:

- konfigurowalność – użytkownik może tworzyć własne konfiguracje;
- wsparcie dla lokalizacji językowej;
- elastyczność – możliwość dodawania nowych funkcji przez architekturę wtyczek;

- osadzalność – program może funkcjonować jako widżet w innych aplikacjach;
 - formaty transkrypcji plików – zapis i odczyt HTK (oraz MLF), TIMIT, ESPS/Waves+, a także Phondat; wsparcie dla kodowania i Unicode.
- WaveSurfer jest udostępniany na zasadach regulowanych przez licencję BSD 3-Clause.

1.8. LaBB-CAT

<http://labbcats.sourceforge.net/> [dostęp: 15.05.2019]

LaBB-CAT to przede wszystkim repozytorium służące do przechowywania zsynchronizowanych w czasie nagrań audio/wideo. Zsynchronizowane transkrypcje mogą być tworzone przy użyciu programu Transcriber lub Praat (które można wykorzystać do utworzenia pliku synchronizującego transkrypcję z nagraniem audio/wideo). Transkrypcja jest następnie przesyłana do LaBB-CAT, co umożliwia gromadzenie informacji o mówcach i o transkrypcjach.

Transkrypcje i anotacje mogą być wyszukiwane za pomocą słów kluczowych lub wyrażeń regularnych. Wyniki wyszukiwania lub całe transkrypty mogą być zapisywane w różnych formatach, a powiązane z nimi fragmenty nagrań mogą zostać odtworzone za pomocą oprogramowania do analizy akustycznej. Cały proces przebiega za pośrednictwem przeglądarki.

LaBB-CAT jest udostępniany darmowo na zasadach regulowanych przez GNU General Public License (GPL).

2. Biblioteki do automatycznej ekstrakcji cech akustycznych, fonetycznych, fonologicznych i prozodycznych mowy

2.1. My-Voice-Analysis

<https://github.com/Shahabks/my-voice-analysis> [dostęp: 15.05.2019]

My-Voice Analysis to biblioteka Pythona do analizy głosu bez konieczności tworzenia transkrypcji. Program segmentuje wypowiedzi, wykrywa granice sylaby, częstotliwość podstawową i formanty. Wbudowane funkcje rozpoznają płeć i nastrój mówcy oraz mierzą wskaźniki artykulacji, tempo mowy, wypełnione pauzy, częstotliwość podstawową. Pełna lista miar i analiz wykonywana przez My-Voice Analysis zawiera:

- rozpoznawanie płci i typu/nastroju mowy (czytanie, mowa spontaniczna, nieujawniająca emocji, entuzjastyczna itd.),
- liczbę sylab [sylaba jest definiowana i rozpoznawana na podstawie szczytów intensywności (dB)],
- liczbę pauz,
- liczbę pauz wypełnionych,
- prędkość mowy (liczba sylab dzielona przez sekundy pierwotnego czasu trwania),
- artykulację (prędkość),
- całkowity czas mowy z uwzględnieniem pauz i wypełniaczy,
- całkowity czas mowy z wykluczeniem pauz i wypełniaczy,
- stosunek wypowiedzi do całkowitego czasu nagrania,
- średnią częstotliwość podstawową,

- odchylenie standardowe częstotliwości podstawowej,
- medianę częstotliwości podstawowej,
- minimalną wartość częstotliwości podstawowej,
- maksymalną wartość częstotliwości podstawowej,
- pomiar rozkładu częstotliwości podstawowej (25 kwantyl),
- pomiar rozkładu częstotliwości podstawowej (75 kwantyl).

Biblioteka została opracowana przez instytut MySolution Lab w Japonii i jest przeznaczona dla językoznawców, naukowców, programistów i badaczy terapii mowy i języka oraz jest udostępniona bezpłatnie. Narzędzie jest cały czas rozwijane. Prawa autorskie należą do Shahab Sabahi.

2.2. MyProsody

<https://github.com/Shahabks/myprosody> [dostęp: 15.05.2019]

MyProsody to biblioteka Pythona do pomiaru cech akustycznych mowy. Autor rozumie badanie prozodii mowy przez badanie melodii i rytmu mowy w kontekście ich wpływu na znaczenie wypowiedzi. Fonetyczne badanie prozodii jest badaniem suprasegmentalnych cech mowy. Na poziomie fonetycznym prozodię charakteryzuje wysokość głosu (częstotliwość podstawowa), intensywność akustyczna oraz rytm (czas trwania fonemu i sylaby). Algorytm segmentuje nagrane wypowiedzi i wykrywa granice sylaby, kontury częstotliwości podstawowej i formanty. Pełna lista pomiarów i analiz wykonywanych przez algorytm zawiera na przykład liczbę pauz, liczbę słów, wskaźnik intonacji, liczbę i czas trwania pauzy, liczbę słów na minutę, formanty, częstotliwość podstawową (F0) oraz jej statystyki (odchylenie standardowe, wartość minimalna i maksymalna, średnia). Biblioteka została opracowana przez instytut MySolution Lab w Japonii i jest udostępniona bezpłatnie. Narzędzie jest cały czas rozwijane.

2.3. Machine-Learning-Algorithm-for-Voice-Analysis

<https://github.com/Shahabks/Machine-Learning-Algorithm-for-Voice-Analysis> [dostęp: 15.05.2019]

Machine-Learning-Algorithm-for-Voice-Analysis to algorytm analizujący cechy akustyczne głosu i tworzący klasyfikator akustyczny stworzony w języku programowania Python. Proces uczenia maszynowego zajmuje około 12 godzin (12 000 danych i 47 funkcji początkowych). W celu uruchomienia algorytmu należy przygotować zestaw danych w postaci plików csv. Prawa autorskie należą do Free Software Foundation. Algorytm jest udostępniany bezpłatnie na zasadach zawartych w licencji GNU General Public License. Autor algorytmu udostępnia kod bezpośrednio zainteresowanym, wymaga kontaktu.

2.4. pyAudioAnalysis

<https://github.com/tyiannak/pyAudioAnalysis> [dostęp: 15.05.2019]

Oprogramowanie stworzone w ramach projektu ufundowanego przez Komisję Europejską. Narzędzie pyAudioAnalysis to biblioteka Pythona dostępna na zasadzie otwartego kodu źródłowego. Dostarcza szeroką gamę procedur analizy dźwięku, takich jak:

- ekstrakcja cech dźwięku i ich prezentacja (np. mfccs – nieliniowa analiza cepstralna z wyznaczaniem współczynników cepstralnych, spektrogram, chromagram),
- klasyfikacja nieznanymi dźwięków,
- dostrajanie parametrów oraz ewaluacja klasyfikatorów segmentów dźwięku,
- wykrywanie zdarzeń dźwiękowych i wykluczanie okresów ciszy z długich nagrań,
- wykonywanie nadzorowanej segmentacji (segmentacja łączna – klasyfikacja),
- wykonywanie nienadzorowanej segmentacji (np. diaryzacja mówców),
- wyodrębnianie miniatur audio,
- uczenie maszynowe i wykorzystanie modeli regresji dźwięku (przykładowe zastosowanie: rozpoznawanie emocji),
- stosowanie redukcji wymiarowości podczas wizualizacji danych dźwiękowych i podobieństw treści.

Oprogramowanie pyAudioAnalysis zostało napisane w języku Python w wersji 2.7 przy pomocy szeroko stosowanych otwartych bibliotek. Najważniejsze spośród tych zewnętrznych zależności pyAudioAnalysis to:

- Numpy – podstawowa biblioteka do obliczeń numerycznych przy użyciu Pythona; użyta przede wszystkim dla jej możliwości prezentacji i obsługi macierzy oraz tablic, jak również zestawu odpowiednich podstawowych funkcji tablic;
- Matplotlib – tworzenie wizualizacji danych 2D;
- SciPy – oferuje zoptymalizowane obliczenia numeryczne i metodologie naukowe; narzędzie pyAudioAnalysis wykorzystuje SciPy w procedurach podstawowego przetwarzania sygnału (jak konwolucja), obliczeń liniowych, obliczeń FFT (szybkiej transformacji Fouriera) oraz odczytu i zapisu plików w formacie WAVE;
- MLpy – pakiet Pythona używany do uczenia maszynowego, wykorzystywany głównie z powodu oferowanych możliwości implementacji k-średnich oraz SVM (maszy-na wektorów podpierających);
- sklearn – pakiet uczenia maszynowego i ekstrakcji danych, użyty dla jego implementacji HMM (ukryty model Markova) oraz LDA (rozkład Dirichleta z ukrytymi zmiennymi).

Biblioteka pyAudioAnalysis została już wykorzystana w kilku aplikacjach do analizy dźwięku, takich jak: funkcjonalność inteligentnego domu (ang. *smart-home*) poprzez wykrywanie zdarzeń dźwiękowych, rozpoznawanie emocji mowy, klasyfikacja depresji na podstawie cech audiowizualnych, segmentacja muzyki, multimodalna rekomendacja filmów oparta na treści, zastosowania związane ze zdrowiem (m.in. monitorowanie nawyków żywieniowych). Narzędzie pyAudioAnalysis jest licencjonowane na licencji Apache.

2.5. Neuro-Speech

<https://github.com/jcvasquezc/NeuroSpeech> [dostęp: 15.05.2019]

Neuro-Speech to oprogramowanie o otwartym kodzie źródłowym przeznaczone do analizy mowy u osób z zaburzeniami neurodegeneracyjnymi, szczególnie osób z chorobą Parkinsona. Oprogramowanie jest dedykowane lekarzom, zwłaszcza logopedom i neurologom, ale może być wykorzystywane również przez pacjentów oraz naukowców zajmujących się

badaniami nad mową patologiczną. Program mierzy wybrane zmienne do oceny możliwości komunikacyjnych pacjentów i obejmuje analizy fonacji, artykulacji, prozodii i zrozumiałości.

Po analizie nagrania mowy Neuro-Speech generuje raport medyczny, który przedstawia deficyty mowy u pacjentów i porównuje ich wyniki do średniej wyników zdrowej grupy kontrolnej.

Neuro-Speech jest napisane w języku C++ i korzysta ze skryptu do analizy mowy napisanego w Pythonie. W celu pełnego wykorzystania możliwości programu konieczne jest pobranie dodatkowego oprogramowania, które jest udostępniane bezpłatnie przez twórców, tj. Anaconda, Praat, ffmpeg.

Neuro-Speech jest udostępniane darmowo na zasadach regulowanych przez GNU General Public License (GPL).

2.6. DisVoice

<https://github.com/jcvasquezc/DisVoice> [dostęp: 15.05.2019]

DisVoice to biblioteka Pythona stworzona w celu ekstrakowania cech mowy patologicznej. Program oblicza artykulację fonacji i cechy prozodyczne mowy z samogłosek oraz ciągłych wypowiedzi w celu oceny umiejętności komunikacyjnych pacjentów z różnymi zaburzeniami mowy. Brane są pod uwagę zmiany mowy spowodowane rakiem krtani, guzkami czy hipernasalność związana z rozszczepem wargi i podniebienia, a także zaburzenia neurodegeneracyjne, takie jak choroba Parkinsona.

W celu pełnego wykorzystania biblioteki konieczne jest pobranie programu Praat oraz biblioteki Pysptk.

2.7. WORLD

<http://www.kisc.meiji.ac.jp/~mmorise/world/english/> [dostęp: 15.05.2019]

Kod źródłowy dostępny jest na stronie GitHub: <https://github.com/mmorise/World>.

WORLD to bezpłatne oprogramowanie do wysokiej jakości analizy mowy, edycji, manipulacji i syntezy. Biblioteka ma funkcje umożliwiające szacowanie częstotliwości podstawowej (F0), aperiodyczności i obwiednię widmową, co pozwala na generowanie mowy na podstawie szacowanych parametrów.

Kod źródłowy został napisany w języku C++ na Uniwersytecie Meiji w Tokio i wydany na zmodyfikowanej licencji BSD, a używane algorytmy nie są opatentowane. WORLD jest wykorzystywany jako syntezytor śpiewu przez japońskie UTAU i metoda analizy dla syntezytoru głosu CeVIO. Najnowsza wersja 0.2.0 jest znacznie lepsza od poprzedniej wersji oprogramowania znanej pod nazwą Tandem-Straight.

3. Inne biblioteki

Na stronie GitHub można znaleźć znacznie więcej bibliotek, algorytmów i rozwiązań z zastosowaniem w lingwistyce komputerowej. Należą do nich:

- Tensorflow CTC Speech Recognition – biblioteka do rozpoznawania mowy wykorzystująca CTC (*Connectionist Temporal Classification*), tj. rodzaj sieci neuronowych,

- magphase – MagPhase Vocoder – analiza i synteza mowy do aplikacji rozwiązań TTS (*Text-To-Speech*),
- MATLAB realtime/interactive speech tools,
- PolyglotDB – narzędzia korpusu Montreal w języku Python do przechowywania danych językowych i API zapytań lingwistycznych,
- Speechnet – biblioteka Pythona do automatycznego rozpoznawania mowy,
- language-detection-speech-using-dnn – biblioteka Pythona do rozpoznawania języka w plikach dźwiękowych.

4. Podsumowanie i ewaluacja przedstawionych rozwiązań

Narzędzi do analizy mowy jest wiele i większość umożliwia przeprowadzanie podstawowych analiz sygnału mowy, zazwyczaj nieautomatycznie. W przypadku programów, które mają funkcję automatycznej transkrypcji nagrań, funkcja obejmuje język angielski. Wszystkie wymienione wyżej programy są używane przez badaczy języka mówionego, jednak w literaturze najczęściej wymieniany jest Praat. Programy dostępne do pobrania online mają znaczną przewagę nad bibliotekami ze względu na ich graficzny interfejs, który ułatwia zrozumienie i obsługę programu. Zaprezentowane biblioteki mają bardzo dużo ciekawych i użytecznych funkcji, jednak korzystanie z nich jest możliwe tylko przez osoby ze znajomością podstaw programowania w danym języku. Biblioteki pobiera się z poziomu konsoli, tak samo wykonuje się wszelkie analizy, co może stanowić utrudnienie.

Wśród wyżej wymienionych programów najwięcej funkcji oferują Praat oraz AnnotationPro. Pozostałe narzędzia służą głównie do graficznej prezentacji sygnału mowy oraz anotacji. Praat posiada wiele przydatnych funkcji do analizy akustycznej, fonetycznej i fonologicznej oraz prozodycznej mowy. AnnotationPro ma bardziej ograniczone możliwości, jednak ze względu na możliwość łatwego tworzenia i dodawania wtyczek może okazać się również przydatny. Dodatkowym atutem jest to, że autorzy przedstawili na stronie internetowej instrukcję tworzenia dodatków. Z drugiej strony przedstawione biblioteki mają znacznie więcej funkcji automatycznych. Użytkownik wskazuje plik dźwiękowy, który ma być poddany analizie, a program wyświetla wyniki. Taki rodzaj pracy jest zdecydowanie szybki i wygodny, a jej wyniki będą powtarzalne i obiektywne.

5. Wnioski

Po przeglądzie i analizie przedstawionych narzędzi można uznać, że najwięcej funkcji i możliwości daje program Praat. Biorąc pod uwagę fakt, że wiele innych narzędzi wymaga instalacji i komunikacji z Praatem, to zdaje się on nieuniknionym, podstawowym elementem potrzebnym do analizy mowy. W przypadku bibliotek można zauważyć, że są one tworzone, wzbogacane o nowe funkcje i publikowane w nowych wersjach. Najwięcej możliwości dają My-Voice-Analysis i pyAudioAnalysis. Co ciekawe, biblioteki te również posiadają funkcje działające w połączeniu z programem Praat. Niemniej każde

z omówionych rozwiązań ma swoje zalety i unikalne możliwości, jednak ostateczny wybór narzędzia zależy głównie od tego, czego w danym momencie użytkownik potrzebuje oraz jakie posiada umiejętności z zakresu programowania.

Bibliografia

- Bachan J., Wagner A., Klessa K., Demenko G. (2015), *Consistency of Prosodic Annotation of Spontaneous Speech for Technology Needs*, Proceedings of the 7th Language & Technology Conference.
- Giannakopoulos T. (2015), *pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis*, <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0144610&-type=printable> [dostęp: 6.12.2019].
- Leech G. (2005), *Adding Linguistic Annotation*, [in:] M. Wynne, *Developing Linguistic Corpora: a Guide to Good Practice*, Oxford.
- Łukasik M. (2009), *Anotacja korpusów tekstów specjalistycznych*, „Języki Specjalistyczne 9. Wyraz – tekst – interpretacja”.
- Orozco-Arroyave J.R., Vásquez-Correa J.C., Vargas-Bonilla J.F., Arora R., Dehak N., Nidadavolu P., Nöth E. (2018), *NeuroSpeech: An open-source software for Parkinson's speech analysis*, “Digital Signal Processing”, July.
- Rykowski J. (2014), *Metody i narzędzia rozpoznawania mowy w zastosowaniach niekomercyjnych*, „Napędy i Sterowanie”, R. 16, nr 6.
- Wagner A., Bachan J., Klessa K., Demenko G. (2015), *Przegląd wybranych aspektów analizy prozodii mowy spontanicznej na potrzeby technologii mowy*, „Prace Filologiczne”, t. LXVI.

Abstract

Speech analysis tools - overview of available programs and libraries

The article presents a list of popular speech analysis tools in the form of programs available online to download and in the form of libraries in various programming languages. The first part presents programs used to visualise, to edit, to analyse the speech signal (for example, measurements of the fundamental frequency, intensity or formants) and annotation (segmentation, transcription and labelling of recordings). The second part presents selected libraries available on the GitHub website, which are used for acoustic, phonetic-phonological and prosodic analysis of speech. All tools have been described taking into account their functions and capabilities, sources, authors, licenses on which they are made available. The last chapter of the article presents the evaluation of the described programs taking into account the number and usability of their functions.

Keywords: speech analysis, acoustic analysis, prosody, programs, libraries