

W ramach projektu przeprowadzone zostaną badania mające na celu opracowanie metody automatycznej transkrypcji fonetycznej mowy (w języku angielskim), opartej na wykorzystaniu połączenia informacji pochodzących z analizy sygnałów fonicznych i wizyjnych. Spodziewany jest efekt synergii przy połączeniu danych pochodzących z różnych modalności, który pozwoli na uzyskanie dużej skuteczności opracowywanej metody. W szczególności, przeprowadzone zostaną badania podstawowe nad związkami pomiędzy zróżnicowaniem alofonicznym w mowie, tj. różnicami w charakterze tych samych głosek wynikających z różnego ułożenia narządów mowy w zależności od środowiska fonetycznego (tj. głosek siedzących lub cech prozodycznych) a obiektywnymi parametrami sygnału. Zostaną również wnikliwie zbadane parametry sygnału mowy (akustyczne i wizyjne) charakterystyczne dla Polaków uczących się języka angielskiego, w tym przy wykorzystaniu artykulografu elektromagnetycznego, pozwalając uzyskać dodatkowe dane odnoszące się do ruchów narządów mowy, umożliwiającego pogłębienie analiz różnic między sposobu wymawiania głosek.

Transkrypcja fonetyczna polega na zamianie tekstu na ciąg symboli fonetycznych opisujących dokładnie wymowę danych słów. Do opisu używa się symboli Międzynarodowego Towarzystwa Fonetycznego (IPA - International Phonetic Association), wraz z dodatkowymi znakami opisującymi warianty danych fonemów. Do chwili obecnej wykonywanie dokładnej transkrypcji wymaga zaangażowania specjalistów wyszkolonych w dziedzinie fonetyki. Dzięki metodzie automatycznej proces ten będzie znacznie ułatwiony. Założeniem jest opracowanie na tyle dokładnej metody, aby pozwalała rozpoznawać drobne różnicowania alofoniczne i akcentowe. Dzięki tej analizie sygnałów wizyjnych i fonicznych transkrypcja fonetyczna mowy może zostać przeprowadzona z większą dokładnością przy wykorzystaniu jedynie modalności akustycznej. Wniosek ten jest poparty wynikami wcześniej prowadzonych badań.

Wyniki projektu będą mogły być wykorzystane przez lingwistów oraz osoby uczące się języka angielskiego. Pogłębione badania nad zróżnicowaniem głosek w kontekście parametrów sygnałów akustycznych i wizyjnych przyczyni się również do zaawansowania stanu wiedzy w dziedzinie audiowizualnego rozpoznawania mowy, a co za tym idzie w dziedzinie interakcji człowieka z komputerem.

Opracowanie metody automatycznej transkrypcji alofonicznej ma szereg potencjalnych zastosowań i stanowi rozwój stanu wiedzy dla rozlicznych dziedzin, w tym:

- **fonetyka i fonologia** - dostępne w słownikach i bazach danych transkrypcje fonetyczne oparte są na subiektywnej ocenie specjalistów dokonujących transkrypcji. Metoda zautomatyzowana będzie narzędziem do szybkiego przetworzenia dużych korpusów językowych w obiektywny i powtarzalny sposób, umożliwiając pogłębione badania fonetyczne i fonologiczne oraz ułatwiając odkrycia naukowe w tej dziedzinie. Wskazano również istniejące różnice, które wymagają podania na wejściu informacji w formie tekstowej. Istnieją również metody transkrypcji bazujące wyłącznie na sygnale akustycznym, jednak nie działają one na poziomie alofonicznym i nie są wspomaganie wizyjnie;

- **metodyka nauczania języka** - według British Council angielskiego uczy się na świecie 750 milionów ludzi, w tym wielu Polaków. Jednym z ważniejszych aspektów nauki jest poprawa wymowy. Metoda automatycznej transkrypcji fonetycznej zwiększa znacząco możliwości w zakresie usprawnienia procesu nauczania wymowy angielskiej, w tym również nauczania na odległość;

- **automatyczne rozpoznawanie akcentu** - podczas gdy do rozpoznania treści wypowiedzi wystarcza informacja o rozpoznanych głoskach, poprawne rozpoznanie akcentu wymaga informacji o zróżnicowaniu alofonicznym głosek. Istnieją prace z zakresu automatycznego rozpoznawania akcentu opierające się na lepszym, statystycznym podejściu do problemu. Przewidziana do opracowania metoda stworzy możliwość porównania subtelnych cech wypowiedzi ze znanymi wzorcami i ułatwia rozpoznanie akcentu mówcy, co znajduje szerokie zastosowanie w dziedzinie interakcji człowieka z komputerem. Jednym z przykładowych zastosowań jest prezentowanie użytkownikowi treści dostosowanej do jego pochodzenia geograficznego lub etnicznego;

- **audiowizualne rozpoznawanie mowy** - technologia audiowizualnego rozpoznawania mowy potrzebuje przełomu, aby dorównać pod względem popularności akustycznemu rozpoznawaniu mowy. Obecny stan nauki w tej dziedzinie pozwala jedynie na poprawę skuteczności rozpoznawania mowy w warunkach szumowych, zaś nie na poprawę dokładności analizy. Do takiego przełomu może się przyczynić zidentyfikowanie parametrów sygnału wizyjnego, które w wyraźny sposób korelują z aspektami fonetycznymi, co staje się możliwym dzięki pogłębionym badaniom prowadzonym w tym projekcie.

Projekt podzielony jest na pięć faz:

1. **W fazie przygotowawczej** zostanie zebrana eksperymentalna baza alofonów i wizemów, służyć jako materiał badawczy w kolejnej fazie projektu. Zostaną nagrane reprezentacje akustyczne i wizyjne (wizemy) wszystkich alofonów języka angielskiego w kontekście spółgłoskowym i samogłoskowym. Wykorzystane zostaną zaawansowane kamery rejestrujące obraz w wysokiej rozdzielczości z szybkością rzędu 100 ramek na sekundę oraz kamery termowizyjne. Dodatkowo zarejestrowane zostaną sygnały z artykulografu elektromagnetycznego (EMA), które zawierają informacje o dokładnym ułożeniu artykulatorów. Nagranie bazy poprzedzi przygotowanie aparatury i rekrutacja mówców. Wśród mówców będą zarówno osoby posługujące się natywnie językiem angielskim, jak i Polacy uczący się języka angielskiego, będący na różnym stopniu zaawansowania nauki.

2. **W fazie eksperymentalnej** prowadzona będzie pogłębiona analiza zarejestrowanych alofonów i wizemów. Analiza będzie przeprowadzana zarówno fonetycznie, jak i numerycznie. W zakresie eksperymentów fonetycznych specjaliści z dziedziny językoznawstwa będą analizowali zgromadzone alofony i wizemy oraz sygnały z artykulografu, wskazując drogi poszukiwania parametrów artykulacyjnych w sygnałach fonicznych i wizyjnych. W zakresie eksperymentów numerycznych obliczane będą z sygnału wizyjnego i fonicznego parametry mające za zadanie dobrze odzwierciedlać zróżnicowanie alofoniczne głosek. Wynikiem fazy eksperymentalnej będzie identyfikacja parametrów i wzorców, które odzwierciedlają różnice pomiędzy poszczególnymi alofonami. Opracowane zostaną również odpowiednie algorytmy decyzyjne umożliwiające odróżnienie poszczególnych alofonów w przestrzeni parametrów.

3. **W fazie treningowej** opracowana zostanie metoda automatycznej transkrypcji mowy. Na podstawie danych treningowych stworzone będą algorytmy przetwarzania sygnałów, klasyfikatory i odpowiednie modele pozwalające na zamianę informacji w postaci fali akustycznej na ciąg symboli fonetycznych.

4. **W fazie testowej** zostanie przetestowana dokładnie stworzonej metody. Obliczone zostaną odpowiednie miary skuteczności na poziomie symboli (alofonów) i słów. Dokładność metody zostanie zbadana w zależności od akcentu natywnego

mówcy i od stopnia zaawansowania w języku angielskim (w przypadku mówców uczących się angielskiego).

5. W fazie podsumowującej sformułowane zostaną wnioski płynące z przeprowadzonych badań. Przygotowana zostanie też ilustracja audiowizualna alofonów języka angielskiego w formie obrazów wizyjnych stwarzających z dźwiękiem. Nagrane w czasie projektu bazy danych alofonów i wizemów zostaną opublikowane i udostępnione społeczności naukowej w postaci odpowiedniej bazy multimedialnej.

Wyniki badań miały globalny zasięg. Spodziewane jest zaawansowanie stanu wiedzy z dziedziny systemów informacyjnych, fonetyki i fonologii angielskiej. Istnieją również inne liczne dziedziny, w których opracowywana metoda może być potencjalnie zastosowana. Rezultaty projektu zostaną opublikowane w czasopiśmie punktowanym o światowym zasięgu, oraz zaprezentowane na renomowanych międzynarodowych konferencjach.