

Wydział: IMiR Kierunek: Inżynieria Akustyczna	Skład zespołu: Lider projektu Krzysztof Prochner (285418) Zespół Bartosz Staniewski (285426) Wojciech Wronka (285436)	Rok: III
		Grupa: 9:30-11:00
		Przedmiot: Technologia Mowy
Data wykonania: 03-24.01.2018	Temat: Projekt: Szachy cyfrowe sterowane głosowo	

Raport końcowy

Zakładany, oraz finalny efekt projektu.

Początkowo, stworzony przez nas interfejs miał za pomocą systemu *sarmata* wychwytywać komendy głosowe, podawane przez użytkownika, i przekazywać je dalej do zewnętrznego programu szachowego w stosowanym przez nią języku.

Końcowy efekt używa autorskiej gry w szachy, napisanej za pomocą biblioteki *python-chess*. Takie rozwiązanie, poza ewentualną kwestią prawną, okazało się mieć sporo zalet. Bezpośredni wgląd w kod używanej gry znacząco ułatwił jej komunikację z segmentem rozpoznającym mowę, oraz pozwolił na natychmiastowe wprowadzanie zmian, na przykład w języku gry. Dzięki temu w końcowej fazie rozwoju projektu, mogliśmy łatwo poprawiać funkcjonalność programu.

Zastosowane ustawienia parametrów.

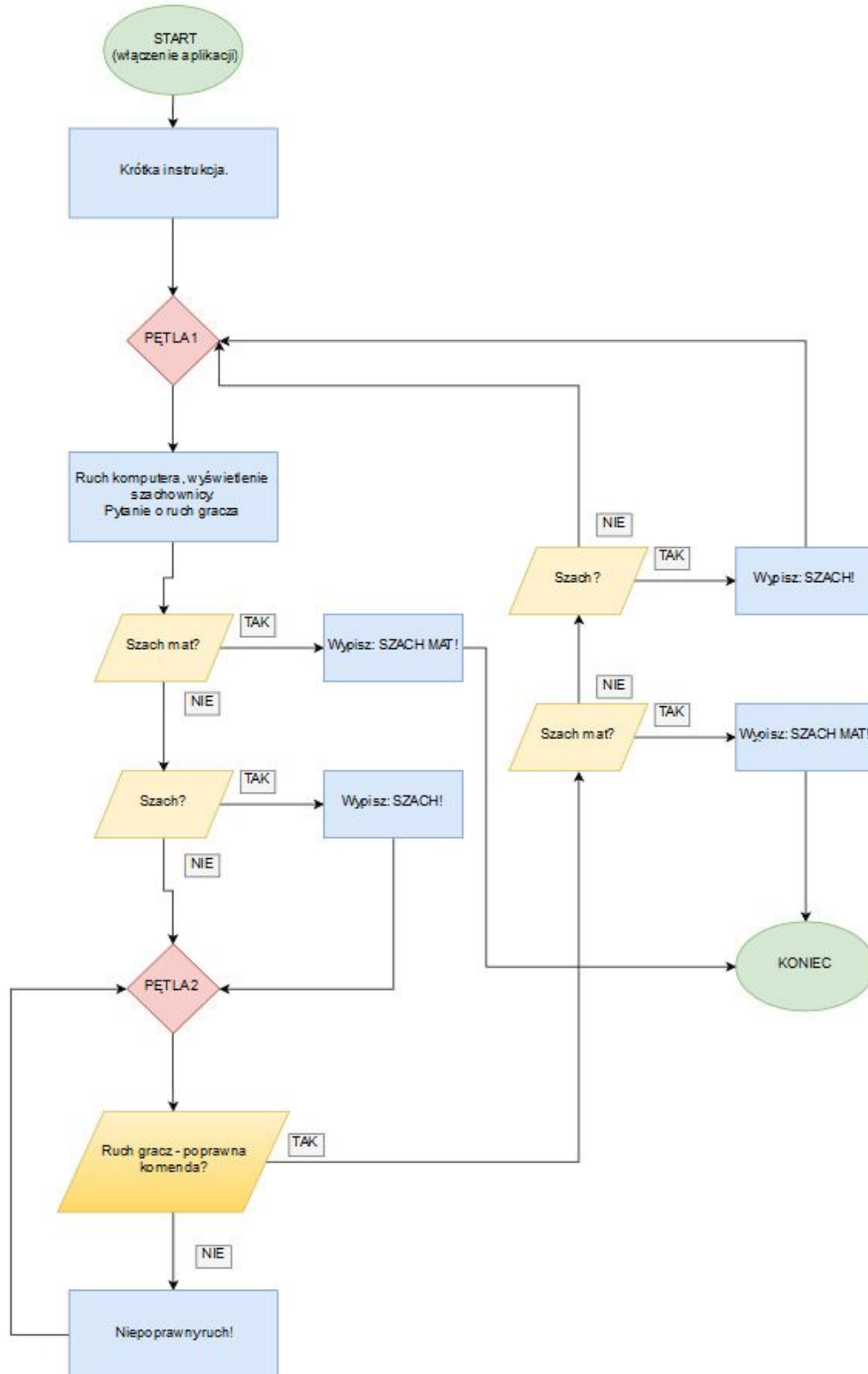
Threshold : 500

Chunk_Size : 2048

Czas oczekiwania : 30

Takie ustawienia stanowiły optymalne wartości dla wychwytywania mowy, na sprzecz z którego jest prezentowany projekt.

Struktura programu i schemat interfejsu.



rys. 1 - schemat interfejsu

Zmiany w gramatyce.

Podczas rozwoju interfejsu, podjęta została decyzja o wprowadzeniu zmian w gramatyce programu. Pozwoliło to na zmniejszenie prawdopodobieństwa błędnej interpretacji poleceń przez program, oraz całkowicie wyeliminowało kwestię ich wieloznaczności.

Innymi słowy, gramatyka została zoptymalizowana pod kątem już działającego programu.



rys.2 - Diagram przedstawiający gramatykę programu

Problemy z poprawnym rozpoznawaniem komend

Podjęta została próba rozwiązania problemu z rozpoznawaniem niektórych głosek (*a/ha*, *be/ce/e*). Pierwszą teorią było zbyt wolne uruchamianie nagrywania wypowiedzi, a w konsekwencji fonemów charakterystycznych dla danych głosek. Po sprawdzeniu nagranych, a błędnie rozpoznanych próbek teoria ta została obalona. Nagrania są zapisywane w całości. Winą obarczyć można zewnętrzny system rozpoznawania, który nie radzi sobie z rozpoznawaniem głosek rozpoczynających wypowiedź, pozbawionych poprzedzającego je kontekstu. Rozwiązaniem może być wymaganie od użytkownika poprzedzenia nazw pól figurą szachową (czyli wprowadzenie zmian do gramatyki).

Problemy z wykrywaniem odgłosów tła

Podczas ostatnich testów interfejsu zauważono, że głośniejsze odgłosy otoczenia, mogą mieć wpływ na funkcjonalność całego programu. Jest to prawdopodobnie spowodowane charakterystyką wydawanych poleceń (bardzo krótkie elementy, pojedyncze litery, cyfry). Oczwistym rozwiązaniem pozwalającym na uniknięcie takich problemów byłaby aktywacja nagrywania komend zewnętrznym przyciskiem - na przykład program odbierałby sygnał audio tylko przy naciśnięciu spacji. Coś takiego kłóciłoby się jednak z założeniami projektu.

Wnioski z projektu w kontekście technologii mowy.

Praca nad projektem pozwoliła nam poczuć na własnej skórze problemy z jakimi wciąż boryka się technologia mowy. Nawet tak stosunkowo prosty system, jak wykorzystany w naszym interfejsie, posiada wiele czynników, które wpływają na efektywność jego pracy, a które ciężko wyeliminować, ponieważ są niezależne od twórców. Tak naprawdę, aby program działał w stu procentach poprawnie, musiałby być używany w idealnych warunkach. Ogólniej mówiąc, interfejsy takie jak ten zaprezentowany przez nas, to raczej ciekawostka. Wykorzystana technologia nie nadaje się w tej chwili do użytku codziennego, głównie ze

względu na swoją wciąż dość dużą zawodność. Wiele problemów można rozwiązać zmniejszając wygodę użytkowania problemu, na przykład wymagając dłuższych, jednoznacznych komend, czy używania mikrofonu ze słuchawek (zbierającego mniej zakłóceń związanych z tłem akustycznym).

Wyrażamy zgodę na prezentację wyników projektu, którego dotyczy ten raport na stronie Zespołu Przetwarzania Sygnałów, stronie przedmiotu oraz w ramach wydarzeń promujących uczelnię, takich jak Noc Naukowców, Dzień Otwarty AGH, Festiwal Nauki.

Wyrażamy zgodę na publikację naszych imion i nazwisk podczas prezentacji wyników projektu, którego dotyczy ten raport.