1. Użytkownik wpisuje i co parę sekund wysyła do backendu część audio i poprawia speech to text przez llm

