

**Tomasz PRAUZNER**

Akademia im. Jana Długosza w Częstochowie, Polska

## **Praktyczne wykorzystanie symulacji dźwięku w kształceniu technicznym studentów**

### **Wstęp**

Stworzenie inteligentnej maszyny jest jak dotychczas sferą marzeń naukowców i wydaje się, iż długo jeszcze będzie pojęciem futurystycznym. Jednakże wszelkie próby i udoskonalenia są kolejnym krokiem w celu powstania właśnie takiej zaawansowanej technologicznie maszyny. Nasuwa się pytanie: jaki jest cel prowadzonych prac w tym zakresie? Zdecydowanie na pierwsze miejsce wysuwa się pragnienie czy dążność do opracowania takiej sztucznej inteligencji, która dorównałaby w sprawności jej twórcy, a nawet ją przewyższyła. Kolejnym i zdecydowanie pragmatycznym celem jest stworzenie aplikacji komputerowych wspomagających jej pracę i umożliwiających nadanie jej wymiaru bardziej ludzkiego. Przykładem może być wzbogacenie aplikacji informatycznych o system komunikacji dźwiękiem.

### **1. Funkcjonalność syntezy mowy informacyjnej**

Podstawowym założeniem opracowania aplikacji komputerowych, jako syntezy mowy, było po prostu „lenistwo” użytkowników aplikacji. Z założenia człowiek poszukuje jak najprostszych form rozwiązań sytuacji problemowych. Powstaje pytanie: po co czytać, skoro można słuchać, po co pisać, skoro można powiedzieć? Realizacja tych dwóch funkcji przebiega odmiennie i daje użytkownikowi różnorodne udogodnienia w pracy. Pomysł ten przeniesiony został w szybkim czasie na inne rozwiązania i zdobył sobie uznanie użytkowników systemów informatycznych.

Do głównych zastosowań syntezy mowy w aplikacjach elektronicznych zaliczyć można:

- zastosowanie w procesie dydaktycznym (wykorzystanie narządu słuchu jest mniej uciążliwe dla pracującej osoby przy monitorze niż czytanie z monitora);
- syntezy mowy jest bardzo często jedyną aplikacją w porozumiewaniu się pomiędzy użytkownikiem a komputerem, np. w grupie osób niepełnosprawnych (w wielu krajach istnieje nakaz udźwiękowienia stron internetowych dla tej grupy odbiorców);

- w zakresie urządzeń (serwery ISR) i aplikacji multimedialnych, np. do odtworzenia aplikacji internetowych, filmów itp. w nowych urządzeniach telekomunikacyjnych, np.: netbooki, iPady, telefony komórkowe nowej generacji<sup>1</sup> (technologia „text to speech” – TTS), bankofony, w różnego rodzaju aplikacjach on-line: słowniki (rys. 1);
- stanowią przedmiot dalszych badań w grupie naukowców dążących do osiągnięcia nowej jakości i zastosowań aplikacji wykorzystujących syntezę mowy.



**Rys. 1. Babylon 9 – słownik on-line**

Źródło: <http://polski.babylon.com>

Syntezytor mowy, czyli proces i system przetwarzania tekstu pisanego na dźwięk, wykorzystuje zaawansowane techniki lingwistyczne i matematyczne w celu uzyskania jak najbardziej naturalnego przekazu<sup>2</sup>. Początkowo na dużą skalę wykorzystywano techniki alofoniczne, polegające na budowaniu słów z poszczególnych elementów – najczęściej pojedynczych liter. Metoda alofoniczna wymaga ogromnej mocy obliczeniowej, a skutek też nie jest zachwycający – wypowiedź jest mechaniczna, pozbawiona krzty emocji. Ta niedogodność nie występuje w najpopularniejszej obecnie metodzie konkatencyjnej. Jakość generowanej mowy jest tak dobra, że można ją pomylić z głosem prawdziwego lektora. W przeciwieństwie do metody alofonicznej tutaj analizie są poddawa-

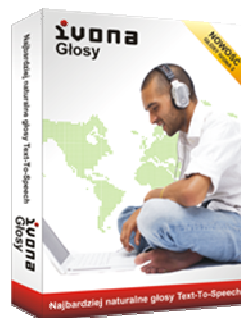
<sup>1</sup> Realizacja technologii TTS w telefonach komórkowych opartych na oprogramowaniu Symbian jest obecnie w fazie wdrażania. Przykładem może tu być chociażby oprogramowanie Nuance Talks firmy Brand & Gröber Communications. Inne oprogramowanie produkcji rodzimej: Speaker Mobile stworzone przez programistów z grupy IVOSoftware.

<sup>2</sup> Synteza mowy – (ang. TTS – Text-To-Speech) – zamiana tekstu w formie pisanej na sygnał akustyczny, którego brzmienie naśladuje brzmienie ludzkiej mowy.

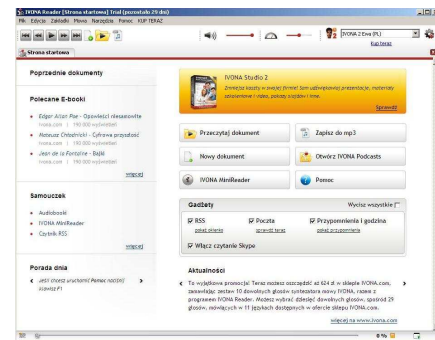
ne całe słowa, które wcześniej rzeczywiście zostały nagrane przez lektora [Wolan 2012].

## 2. Przegląd wybranych aplikacji syntezy mowy

Ogólnie ofertę tę można podzielić na cztery grupy programów: pierwszą grupą są syntezy opracowane na rynek rodzimy, opracowań pod kątem specyfiki języka rodzimego. Druga grupa to aplikacje, które są efektem prac firm o zasięgu międzynarodowym (Text-To-Speech Synthesis), które oferują już bogatszą ofertę w zakresie językoznawstwa. Trzecia grupa programów to ta, która nie tylko potrafi „czytać tekst” drukowany, ale i daje możliwość rozpoznawania mowy za pośrednictwem urządzeń rejestrujących (Automatic Speech Recognition). Czwarta grupa to typowe aplikacje on-line (Interactive TTS Demo). Do dalszych rozważań wybrałem dwa przykładowe produkty oferowane na rodzimym rynku. Są to programy: IVONA 2 (rys. 2, 3) oraz syntezy on-line firmy Interactive Voice News. Z pewnością każdy z tych produktów mógłby być oceniony bardziej dokładnie, gdyby istniała możliwość uzyskania aplikacji pełnej, bez ograniczeń. Z drugiej strony wersje z narzuconymi ograniczeniami pozwalają na bezpłatny hosting utworzonych w ten sposób plików. Aplikacje w wersji *demo*, jakie można uzyskać w licencji *freeware*, dają ograniczone możliwości. Tak jest w wersji oprogramowania firmy Interactive Voice News. Aplikacja IVONA 2 jest oferowana jako *trial*, a więc trzydziestodniowa, ale daje pełne możliwości jej wykorzystania.



Rys. 2. IVONA Głosy  
Źródło: [www.daden.co.uk](http://www.daden.co.uk)

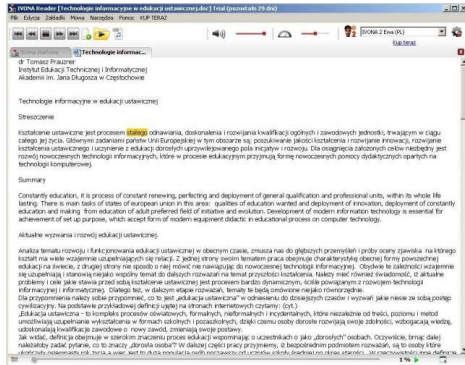


Rys. 3. Strona startowa programu IVONA  
Źródło: [www.daden.co.uk](http://www.daden.co.uk)

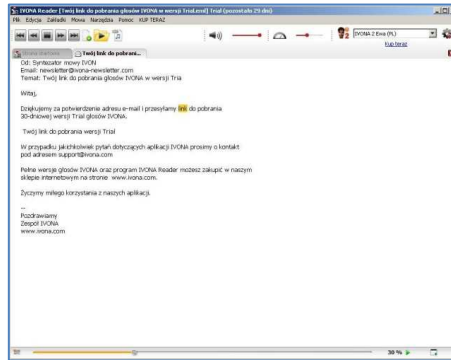
Program IVONA to zbiór aplikacji skierowany do użytkownika w zależności od przeznaczenia jej wykorzystania. Najistotniejsze wydają się możliwości doboru wielu wersji językowych oraz (jako udogodnienie) wersje głosów, zarówno męskiego, jak i żeńskiego w kilku modulacjach<sup>3</sup>. Powstał też syntezy

<sup>3</sup> Informacje udostępnione za zgodą firmy IVONA; [www.ivona.com/pl](http://www.ivona.com/pl)

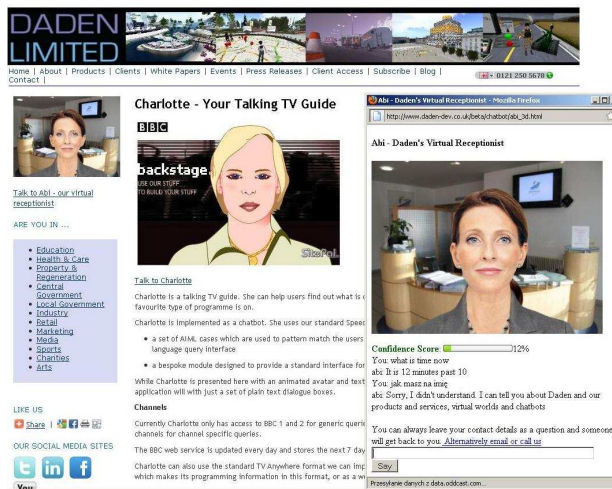
mowy na urządzenia mobilne z systemem Android (iPod, tablet), w urządzeniach firm Apple, IBM PC, w systemach Windows, Linux. Obydwa programy oferują bogate możliwości w wyborze instalacji konkretnej aplikacji w zależności od potrzeby użytkownika. Poniżej przedstawiono zdjęcia z ekranu aplikacji, które zostały zaimplementowane do odczytu audio. Są nimi: dokument tekstowy opracowany w edytorze tekstu MSWord (rys. 4), interpretacja wiadomości z poczty elektronicznej (rys. 5) oraz odczyt informacji ze strony internetowej w przeglądarce Mozilla Firefox.



Rys. 4. IVONA w edytorze tekstu  
Źródło: opracowanie własne.



Rys. 5. IVONA w programie pocztowym  
Źródło: opracowanie własne.



Rys. 6. Aplikacja DADEN LIMITED  
Źródło: www.daden.co.uk

Wykorzystanie syntezy mowy to również rozwiązania o szerszym przeznaczeniu. Przykładem może być chociażby strona internetowa o nazwie Daden Limited, będąca próbą stworzenia świata wirtualnego oraz tworzenia interaktywnych wirtualnych osobowości. Obecnie możemy na tej stronie odnaleźć trzy modele wirtualne o nazwach: Built Environment Visualisation, Training and Education, Chatbots. Aplikacje te mają za zadanie symulację różnych aspektów rzeczywistości, a więc ukazanie środowiska zewnętrznego, szkolenia i edukacji w sytuacjach problemowych oraz ukazanie wirtualnej postaci. Szczególnie interesująca jest aplikacja postaci, z którą możemy prowadzić interaktywną rozmowę. Co ciekawe, w wyniku tej konwersacji zauważyć można specyficzny – osobowościowy wizerunek wirtualnej osoby, która w inteligentny sposób prowadzi z nami dialog na przeróżne tematy, w tym możemy uzyskać od tej osoby interesujące nas odpowiedzi na postawione pytania. Aplikacja działa poprawnie wykorzystując słownik pojęć w języku angielskim. Zadane pytanie w języku polskim są dla systemu niezrozumiałe, co widać na liście odpowiedzi na zadawane próbne pytania (rys. 6).

### 3. Wnioski końcowe

Świat wirtualny pomimo iż jest jeszcze niedoskonały, stanowić może dla nas nieograniczone możliwości w symulacji rzeczywistości. To nie tylko nowinka techniczna, to przede wszystkim możliwość opracowania coraz lepszych materiałów dydaktycznych. W obrębie dydaktyki jesteśmy świadkami powstania nowej subdyscypliny, która już w najbliższym czasie zadziwi nas swoją ofertą. Otwiera ona zupełnie nowe możliwości dla pedagogów, wykładowców, nauczycieli i uczniów. Po całkowicie plastycznym wykreowaniu cyfrowego świata otrzymujemy możliwość tworzenia unikalnych doświadczeń edukacyjnych. Wirtualne kreowanie rzeczywistości posiada w sobie pewien potencjał, chociażby w postaci jej autorów i z pewnością jeszcze bardziej zrewolucjonizuje proces dydaktyczny. To również szansa na wzrost popularności nauczania e-Learning. Symulacja audio już wkrótce wzbogaci na tyle materiał szkoleń, iż stanie się on bardziej przyjazny ich uczestnikom.

### Literatura

- Darell R. (2012), *VODER: World's First Synthetic Speech Synthesizer* [Retro], <http://www.bitrebels.com>
- Portal internetowy, (2011), *Isle of Man Newspapers, Talking TV guide*, <http://www.iomtoday.co.uk>
- Prauzner T. (2010), *Applications of multimedia devices as teaching aids* [w:] *Annales UMCS Informatica AI X*, red. R. Szczygieł, Lublin.
- Prauzner T. (2011), *Media education – today and tomorrow* [w:] *Science For Education – Education For Science*, red. H. Orsolya, I. Psenakova, Nitre.

- Prauzner T., Ptak P. (2012), *The role of standardization in the development of e-learning* [w:] *Annales Universitatis Mariae Curie-Skłodowskiej Informaticales*.
- Prauzner T. (2006), *Zastosowanie programów symulacyjnych w nauczaniu przedmiotów technicznych* [w:] *Prace Naukowe Akademii im. Jana Długosza w Częstochowie*, red. J. Wilsz, Częstochowa.
- Wolan J. (2012), *Synteza mowy dla Androida*, <http://down01.ivona.com>

### **Streszczenie**

Głównym celem niniejszej dysertacji jest zapoznanie czytelnika z szeroko rozumianą problematyką syntezy ludzkiego głosu w procesie komunikacji na drodze człowiek–maszyna. Realizacja aplikacji multimedialnych wzbogaconych w generowany system mowy jest szczególnie interesującą propozycją wykorzystywaną w edukacji zdalnej na odległość.

**Słowa kluczowe:** symulacja, synteza mowy.

### **Practical using of sound simulation in the selected computer applications**

#### **Abstract**

The communication process occurring between the user and the computer is still quite limited by considerable difficulties in interpreting the written commands. The main aim of the present article is then to describe various aspects of human speech synthesis in the human–computer communication. Multimedia applications with synthetic speech systems are very a promising idea in distance learning.

**Key words:** simulation, speech synthesizer.