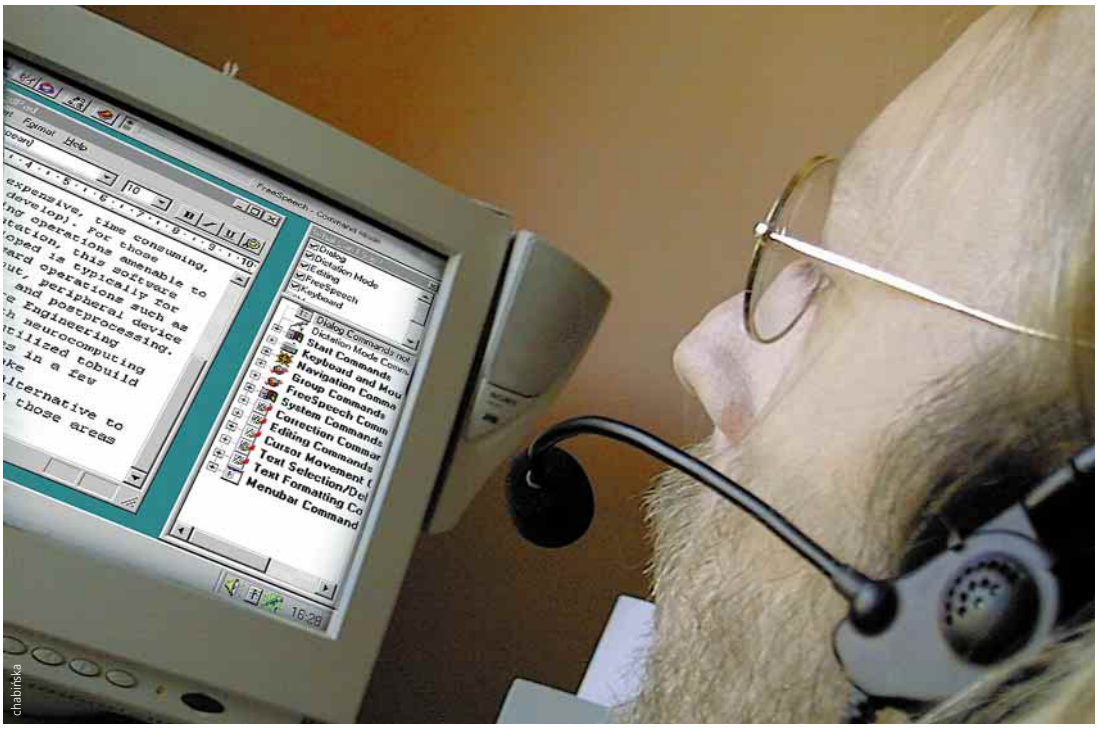


chabińska

Śmierć klawiaturom!

Oprogramowanie służące do rozpoznawania mowy nie jest na rynku żadną nowinką. Jednak dopiero mniej więcej rok temu pojawiły się w sprzedaży aplikacje umożliwiające rozpoznawanie mowy ciągłej.

Tegoroczne targi CeBIT stały się dla kilku niezależnych producentów okazją do przedstawienia szerokiej publiczności nowej technologii. Prezentacje nowego oprogramowania cieszyły się dużą popularnością, o czym świadczyły tłumy zbierające się wokół stoisk, na których odbywały się seminaria. Trudno się dziwić takiemu zainteresowaniu, zważywszy fakt, że oferowane do tej pory aplikacje kazały użytkownikom stosować pauzy pomiędzy wszystkimi wypowiadanymi słowami. Aby się przekonać, jak nienaturalny jest taki sposób mówienia, wystarczy spróbować wypowiedzieć dowolne zdanie, pamiętając o konieczności dokonywania przerw pomiędzy poszczególnymi wyrazami. Niewygodne i męczące, prawda?

Nie jest żadną tajemnicą, że badania nad technologią komputerowego rozpoznawania mowy prowadzone były od kilku dziesięcioleci; wszak pierwsze prace w tej dziedzinie pojawiły się na długo przed skonstruowaniem pecetów! Powstaje zatem pytanie: jakiego rodzaju przełom dokonał się w ciągu ostatnich dwóch lat, że możliwe stało się stworzenie oprogramowania pozwalającego na rozpoznawanie mowy ciągłej...?

Po pierwsze: wydajność

Odpowiedź jest szokująca. Okazuje się, że w ostatnich latach nie został dokonany – jak można by przypuszczać – żaden przełom! Najistotniejszym czynnikiem, który miał największy wpływ na powstanie aplikacji rozpoznających mowę ciągłą, był bowiem... wzrost wydajności komputerów klasy PC. Pojawienie się na rynku procesorów Pentium taktowanych z częstotliwością ponad 100 MHz pozwoliło producentom stworzyć przeznaczone dla masowego odbiorcy narzędzia dopiero w ubiegłym roku. Drugim, równie istotnym czynnikiem był spadek cen pamięci operacyjnych, dzięki czemu możliwe stało się wyposażenie komputera osobistego w odpowiednią dla potrzeb algorytmów rozpoznających mowę ciągłą wielkość pamięci RAM.

Czterech muszkieterów

Najważniejszymi i najbardziej znanymi na rynku aplikacjami do rozpoznawania mowy ciągłej są: *Dragon Naturally Speaking*, rodzina produktów z serii *ViaVoice* IBM-a oraz *Voice Xpress Plus* belgijskiej firmy Lernout&Hauspie. W trakcie targów PC Expo w czerwcu 1998 w Nowym Jorku dwa pierwsze programy zostały zaprezentowane w swych najnowszych edycjach. Od niedawna na rynku dostępny jest jeszcze pakiet *FreeSpeech 98* firmowany przez Phillipsa.

Wymienione aplikacje wymagają (minimum) procesora Pentium taktowanego z częstotliwością 133–166 MHz. W skład każdego pakietu wchodzi specjalny mikrofon, dający najlepsze efekty w pracy z tego typu aplikacjami. Wymagania odnośnie do pamięci operacyjnej są zmienne i zależą przede wszystkim od środowiska, w którym program ma pracować (32 MB RAM dla systemu Windows 95 lub 48 MB RAM dla Windows NT) oraz od wersji językowej (podane informacje dotyczą edycji angielskojęzycznych). W przypadku pakietów rozpoznających mowę np. niemiecką roszczenia pamięciowe rosną o kolejne 16–32 MB. Dzieje się tak ze względu na specyfikę języka naszych zachodnich sąsiadów; główny kłopot sprawiają tu wyrazy złożo-

słowniczek

fonem – najmniejsza, niepodzielna jednostka mowy (głoska)
klasyfikator neuronowy – przekształcenie matematyczne, które dla zadanego wektora cech odnajduje odpowiadający mu fonem, wykorzystując do tego sieć neuronową
transformata Fouriera – przekształcenie matematyczne, które zamienia sygnał mowy na złożenie szeregu sinusoidalnych przebiegów o różnych częstotliwościach i fazach; współczynniki transformaty określają amplitudy tychże przebiegów i tworzą tzw. widmo częstotliwościowe, wykorzystywane w dalszej analizie
ukryty łańcuch Markowa – metoda statystyczna pozwalająca na wybór najbardziej prawdopodobnej sekwencji fonemów w przetwarzanym sygnale mowy
wektor cech – specjalnie dobrany podzbiór współczynników transformaty Fouriera najlepiej nadających się do rozpoznawania fonemów
zjawiska prozodyczne – zjawiska zachodzące w mowie, takie jak intonacja, tempo wypowiedzi oraz akcenty

ne, takie jak np. Regenwald (las zwrotnikowy) czy Großplattenbauweise (budownictwo z tzw. „wielkiej płyty"). Rozpoznanie, czy chodzi tutaj o dwa lub więcej następujących po sobie oddzielnych słów, czy też o jeden będący ich językowym „zlepkiem", jest problemem trudnym i dopiero analiza kontekstowa pozwala programom na definitywne rozstrzygnięcie, o jakie słowa chodzi.

Programy dostępne są w kilku wersjach językowych, przede wszystkim zachodnioeuropejskich. Plany poszczególnych firm przewidują oczywiście tworzenie kolejnych edycji narodowych, wśród nich nie ma jednak – z dość oczywistych powodów – języków Europy Centralnej i Wschodniej (jedynym wyjątkiem bywa język rosyjski).

Produkty pozwalają w mniejszym lub większym stopniu zarówno dyktować tekst, jak i sterować za pomocą głosu programem, z którym współpracują. Dyktowane zdania pojawiają się na ekranie stopniowo w miarę ich wypowiadania przez użytkownika; jeśli jakieś słowo lub fraza zostanie rozpoznana niepoprawnie, program umożliwia wydanie komend głosowych cofających kursor do źle rozpoznanego fragmentu i następnie jego modyfikację.

Dla tych, którzy często przemieszczają się z miejsca na miejsce przewidziano możliwość użycia specjalnych cyfrowych dyktafonów. Za ich pomocą można nagrać dowolny tekst, a następnie – po powrocie do biura lub domu – „zamienić" na tekst.

Trening czyni mistrza

Pomimo że w dostępnych na rynku produktach wykorzystywane są tzw. algorytmy „niezależne od mówcy" (ang. speaker independent), istnieje wiele możliwości „poprawiania" skuteczności ich rozpoznawania. Po pierwsze algorytm może dopasować się do właściwości akustycznych (wyznaczyć charakterystykę częstotliwościową) użytego do rozpoznawania mikrofonu. Po drugie stopniowo usprawnia procedury rozpoznawania poszczególnych fonemów, tak aby uwzględnić specyfikę brzmienia głosu określonej osoby – w rezultacie pozwala to skuteczniej rozróżniać podobne do siebie głoski, np. „k" od „t", „ś" od „sz" itp. Po trzecie może brać pod uwagę specyfikę wymowy pewnych grup liter w określonych wyrazach (np. słowo „wziąłem" przez niektórych wymawiane jest jako „wzięłem", a słowo „prezydent" jako „prezydęt"). Wreszcie adaptacji poddawany jest także słownik, do którego użytkownik dodaje nowe wyrazy, zadając ich wymowę (np. nazwiska lub nazwy własne).

Wyznaczanie właściwości akustycznych mikrofonu oraz modyfikacja rozpoznawania fonemów to procesy, które odbywają się najczęściej zanim użytkownik zacznie korzystać z programu. W tym celu wiele systemów do dyktowania wymaga od mówcy, aby przeczytał on zadany fragment specjalnie przygotowanego tekstu (ok. 15–30 min czytania). Niemal regułą jest możliwość wielokrotnego powtarzania takiego treningu – za każdym razem z innym, wybranym przez producenta tekstem – co może istotnie wpłynąć na poprawę jakości rozpoznawania (nawet do ok. 98%).

Gadający Windows

W opcję dyktowania wyposażone już zostały pakiety biurowe firm Lotus i Corel: *SmartSuite 98* i *SmartSuite Millennium Edition* oraz *WordPerfect Suite 8*, znajdujące się w sprzedaży od kilku miesięcy, wyprzedzając tym samym w tej dziedzinie dominujący obecnie na rynku *Microsoft Office*, z którego edytorem *Microsoft Word* w wersji 97 współpracują jednak wszystkie opisane powyżej aplikacje.

Microsoft od dawna myśli o włączeniu funkcji rozpoznawania mowy, jednak nie do Office'a, lecz w skład systemu Windows. Rozumowanie giganta z Redmond jest proste: ponieważ pewnego dnia technologia ta stanie się zasadniczym sposobem ▶ 192

porównanie

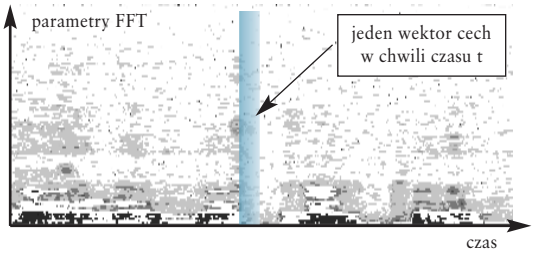
Krótki test

Aby określić przydatność omawianych w artykule pakietów dla polskich użytkowników, chcących je wykorzystać do dyktowania tekstu w języku angielskim, przeprowadziliśmy prosty test. Polegał on na przepisaniu tekstu w tym języku za pomocą klawiatury, a następnie podjęciu próby jego rozpoznania przez programy rozpoznające mowę ciągłą. Osoba przeprowadzająca test pisała na klawiaturze z przeciętną prędkością, natomiast jej angielska wymowa daleka była od doskonałości. Czas trenowania programów przed ich użyciem wynosił ok. 30 minut.

Na poprawne przepisanie 161 słów testujący potrzebował ok. 10 minut (daje to prędkość 16,1 słów/minutę), na przedyktowanie tekstu i poprawienie wszystkich błędnie rozpoznanych słów – ok. 13 minut (czyli z prędkością 12,4 słów/minutę) przy użyciu *Dragon NaturallySpeaking*, ok. 15 min przy wykorzystaniu *IBM ViaVoice* (10,7 słów/min) oraz ok. 17 min (9,5 słowa/min) w przypadku programu *FreeSpeech 98*. Zważywszy fakt, że eksperymentator mówił po amerykańsku znacznie gorzej od przeciętnego Amerykanina, jakość wbudowanych algorytmów rozpoznawania mowy należy uznać za bardzo dobrą (zwłaszcza w przypadku programu firmy Dragon). Innymi słowy, osoby lepiej mówiące po angielsku z łatwością osiągną daleko lepsze efekty.

technologie

Rozpoznawanie mowy ciągłej

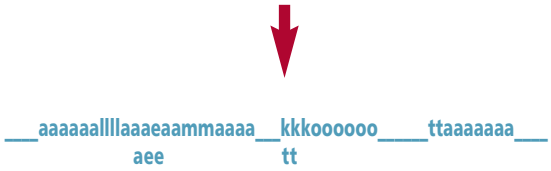


1 Wczytywanie wypowiedzi w postaci cyfrowej za pośrednictwem mikrofonu i karty dźwiękowej. Operacja ta realizowana jest jednocześnie z kolejnymi etapami procesu; kolejne fragmenty wypowiedzi pobierane są przez podproces „odcinający je po kawałku" od nieprzerwanego strumienia danych wejściowych, a następnie poddawane są dalszej obróbce.

2 Zamiana „odciętego" fragmentu do postaci parametrycznej, niezbędnej dla przeprowadzenia rozpoznawania fonemów oraz wyznaczenia tzw. wektorów cech (np. wyliczenie widma transformaty Fouriera – FFT, czyli Fast Fourier Transform).

Jeden zestaw parametrów (wektor cech) wyliczany jest dla tzw. „okien" – jednakowo długich fragmentów o długości kilkudziesięciu milisekund. Okno przesuwane jest wzdłuż osi czasu o kilka do kilkunastu milisekund; kolejne okna częściowo na siebie „zachodzą".

Przykładowy wektor cech na rysunku reprezentowany jest przez pionową kreskę.



3 Rozpoznawanie każdego wektora cech i wyznaczenie odpowiadającego mu fonemu. Liczba wektorów cech przypadających na jeden fonem jest zazwyczaj większa niż 1 (na rysunku reprezentowane jest to poprzez powielone ciągi głosek). Niekiedy wyniki nie są jednoznaczne; w takiej sytuacji moduł odpowiedzialny za rozpoznanie fonemów podaje kilka wariantów (tu zaprezentowane jedne pod drugimi).

Do rozpoznawania fonemów na podstawie wektorów cech stosuje się najczęściej techniki analizy statystycznej lub tzw. klasyfikatory neuronowe.

4 Analiza zjawisk prozodycznych:
- wyznaczenie intonacji;
- wyznaczenie mocy akustycznej fonemów (potrzebne do odnajdywania akcentów);
- wyznaczenie czasu trwania poszczególnych fonemów (potrzebne do odnajdywania akcentów).

5 Przekształcenie ciągu rozpoznanych fonemów (jak w punkcie 3) na tekst wynikowy z jednoczesnym usunięciem zbędnych powtórzeń fonemów przy użyciu narzędzia statystycznego, tzw. ukrytych łańcuchów Markowa (ang. Hidden Markov Model – HMM). Rezultatem tego kroku jest stworzenie grafu wszystkich wariantów rozpoznania poszczególnych fonemów danego fragmentu wypowiedzi.

6 Przeprowadzenie na stworzonym w poprzednim kroku grafie analizy słownikowej (odszukanie wszystkich słów w słowniku, które mogły zostać wypowiedziane przez użytkownika) i gramatycznej na podstawie związków frazeologicznych (tj. statystyki występowania słów obok siebie) oraz wyników analizy zjawisk prozodycznych z punktu 4. W efekcie otrzymywany jest najbardziej prawdopodobny tekst wypowiedzi („Ala ma kota").

komunikowania się człowieka z maszyną, zatem w chwili, gdy rynek i przemysł będą gotowe na ich przyjęcie, należy funkcję rozpoznawania mowy włączyć do systemu operacyjnego (dodać tutaj należy, że chodzi tu nie tylko o możliwość dyktowania tekstu aplikacjom, lecz i sterowania całym systemem za pomocą głosu). Dzięki temu każdy program potencjalnie mógłby wykorzystywać mowę tak samo jak obecnie korzysta z „dobrodziejstw" okienkowego interfejsu użytkownika czy też ze standardu obsługi drukarek. Nawigacja głosem i możliwość dyktowania tekstów niewątpliwie ułatwiłaby i usprawniłaby obsługę systemu nie tylko początkującym.

Microsoft opracował standard Microsoft Speech Application Programming Interface (w skrócie: SpeechAPI lub SAPI), definiujący sposób realizacji aplikacji do przetwarzania mowy w środowisku Windows. Standard ten dokładnie opisuje także użycie technik rozpoznawania mowy w aplikacjach użytkownika. W maju br. ukazała się najnowsza wersja pakietu do programowania w standardzie SAPI – *Microsoft SpeechAPI SDK 4.0*.

Mimo to większość sprzedawanych w chwili obecnej programów do rozpoznawania mowy nie wykorzystuje SAPI ▶ 195

(wyjątkiem są produkty IBM-a). Dlatego też 10 września 1997 roku Microsoft podpisał ze wspomnianą już kilkakrotnie belgijską firmą Lernout&Hauspie umowę, w ramach której Microsoft zainwestował 45 mln dolarów w L&H, Belgowie natomiast zobowiązali się w zamian zaprojektować i zrealizować aplikacje wykorzystujące SAPI.

Zamiary Microsoftu szybko wzbudziły protesty konkurentów L&H. Szczególnie ostro protestuje firma Dragon Software, dla której technologia rozpoznawania mowy jest w zasadzie jedynym polem działania; innymi słowy, włączenie funkcji głosowych do jądra systemu operacyjnego błyskawicznie zredukowałoby dochody producenta Naturally Speaking praktycznie do zera. Nie ma zatem co się dziwić, że w sytuacji, gdy systemy operacyjne Microsoftu zdominowały rynek komputerów PC, Dragon obawia się o swoją przyszłość.

Należy tutaj przypomnieć o toczącym się obecnie procesie w kwestii stosowania (lub nie) przez Microsoft praktyk monopolistycznych (m.in. poprzez włączenie do systemu Windows 98 funkcji internetowych; patrz CHIP 7/98, s. 26–28). Nie ulega najmniejszej wątpliwości, że rezultat sądowego starcia będzie miał ogromny wpływ na dalszy rozwój środowiska Windows, a więc i na plany włączenia do jądra systemu np. funkcji rozpoznawania mowy.

Do you speak... Polish?

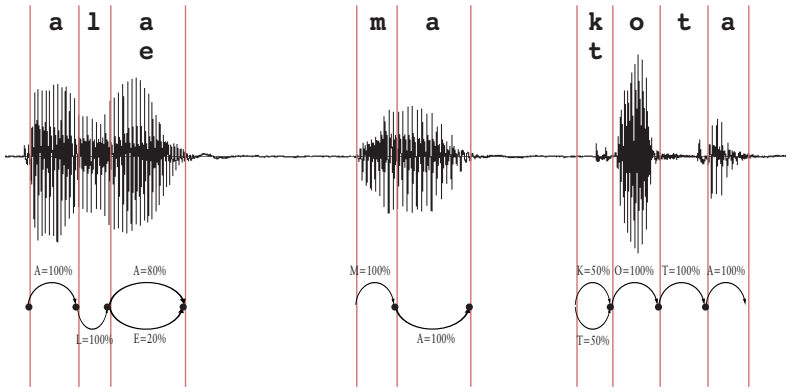
Jak już wspomnieliśmy, w najbliższej przyszłości żadna z firm trudniących się tworzeniem aplikacji do dyktowania nie zamierza realizować polskojęzycznych wersji swoich programów. Nie oznacza to jednak, że w naszym kraju nie są prowadzone żadne prace przez rodzimych specjalistów.

Na razie najbardziej zaawansowaną technologicznie aplikacją do rozpoznawania mowy dostępną na polskim rynku jest *Lektor 4.0* sopockiej firmy Drive (test pakietu opublikujemy w jednym z następnych numerów CHIP-a). Jest to de facto pakiet służący przede wszystkim do syntezy mowy, jednak wersja 4.0 wyposażona została dodatkowo przez producenta w funkcje rozpoznawania wydawanych głosem poleceń. Nie jest to zatem system przeznaczony do dyktowania tekstu, tak jak w przypadku omówionych powyżej produktów. W chwili obecnej kilkuosobowy zespół firmy pracuje nad nową ▶ 196

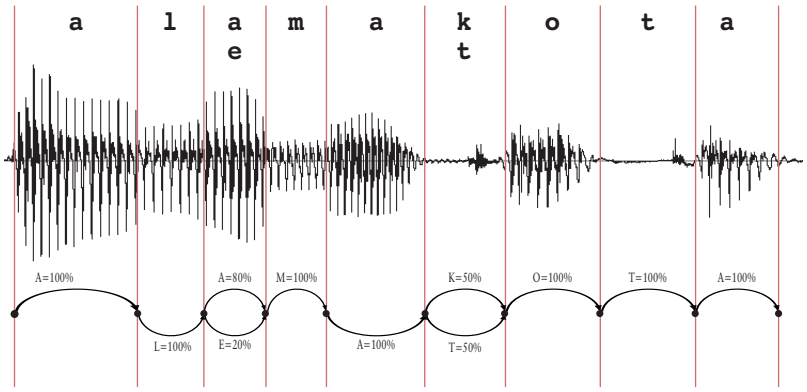
technologie

Mowa dyskretna i mowa ciągła

Technika rozpoznawania mowy ciągłej jest podobna do techniki rozpoznawania mowy dyskretnej operującej na dużym zestawie słów. W obu przypadkach każde słowo analizowane jest poprzez rozbicie na fonemy. Różnica polega na wykrywaniu końców wyrazów – w mowie dyskretnej wyraz kończy się odpowiednio długo trwającą ciszą, natomiast w mowie ciągłej przerwy takiej zazwyczaj nie ma i o podziale wypowiedzi na wyrazy w dużej mierze decyduje informacja kontekstowa (wynik analizy słownikowej, gramatycznej, zjawisk prozodycznych itp.), która pozwala wybrać ciąg słów będący najlepszym wariantem dla danego fragmentu mowy. Poniżej przedstawiono przykład wypowiedzi „Ala ma kota" dla mowy dyskretnej.



Potencjalne procentowe prawdopodobieństwa rozpoznania tekstu są następujące:
„Ala ma kota" – 80%
„Ale ma kota" – 20%
„Ala ma tota" – 0% (brak w języku polskim słowa „tota")
„Ale ma tota" – 0% (brak w języku polskim słowa „tota")
W słowie „Ala" system ma wątpliwości co do drugiej litery „a", która rozpoznana jest z prawdopodobieństwem 80% – w 20% fonem ten przypomina literę „e". Ze względu na to, iż w języku polskim występuje zarówno słowo „Ala", jak i „ale", o wyborze pierwszego z tych słów zadecyduje prawdopodobieństwo rozpoznania. W słowie „kota" fonem „k" jest rozpoznany jako „k" lub „t" (z takim samym prawdopodobieństwem), jednak słowo „tota" nie występuje w języku polskim, dlatego też wariant „t" zostanie odrzucony na etapie analizy słownikowej.
Kolejny rysunek przedstawia zdanie „Ala ma kota" wypowiedziane w sposób ciągły.



Prawdopodobieństwa rozpoznania tekstu są następujące (podkreślenie oznacza akcent):
„Ala ma kota" – 60% (akcent pada wyraźnie na pierwszą sylabę)
„A lama to ta?" – 40%
Brak wyraźnych przerw między słowami powoduje, że rośnie liczba potencjalnych rozwiązań. W rezultacie analizator językowy zasugerował dwa, poprawne z punktu widzenia języka polskiego, zdania: „Ala ma kota" oraz „A lama to ta?". Dopiero analiza prozodyczna – w tym wypadku czasu trwania fonemów – pozwoliła na wybór pierwszego rozwiązania, gdyż akcent w wypowiedzi wyraźnie pada na pierwszą sylabę słowa „Ala".

Przykład ten ilustruje większą złożoność rozpoznawania mowy ciągłej w porównaniu z rozpoznawaniem mowy dyskretnej. Mimo sporej komplikacji algorytmów przetwarzanie mowy ciągłej ma kilka ważnych zalet. Po pierwsze jest wygodniejsze dla użytkownika. Po drugie wyrazy wypowiadane są w naturalny sposób. Wreszcie w mowie ciągłej występuje bogatszy repertuar zjawisk prozodycznych (akcenty, intonacja, tempo), których analiza wspomaga rozpoznawanie.

wywiad

Dopiero w 2010 roku

Ryszard Tadeusiewicz, profesor nauk technicznych, autor licznych prac z zakresu komputerowej analizy mowy polskiej i pierwszej w Polsce książki o rozpoznawaniu mowy, rektor AGH

CHIP: Kiedy będzie można „pogadać" z komputerem jak z człowiekiem?
R. T.: Dyskutując o możliwości rozmowy z komputerem, nie możemy zapominać, że – wbrew pozorom – nie jest to problem z zakresu akustyki, lecz informatyczny. Dlatego w celu ustalenia, kiedy możliwa będzie swobodna rozmowa z komputerem, nie należy ograniczać się do oceny stopnia aktualnego rozwoju systemów rozpoznawania mowy. O sukcesie w tej materii zadecyduje postęp w zakresie specjalistycznego oprogramowania z dziedziny tak zwanej sztucznej inteligencji oraz postęp w badaniach naukowych dotyczących głębokiej, semantycznej struktury samego języka.

Interesujących zagadnień w problematyce komputerowego przetwarzania języka jest naprawdę bez liku. Wynika to z niesłychanej, trudnej do wyobrażenia dla niefachowców, głębi i złożoności intelektualnej tego problemu. Przy bliższym poznaniu okazuje się, że w całej informatyce trudno znaleźć problem o większym stopniu złożoności. Na pozór nie ma nic prostszego i bardziej naturalnego niż komunikacja za pomocą mowy. Wydaje się to proste i takie rzeczywiście jest – tyle że dla człowieka.

CHIP: Dlaczego nie dla komputera?
By zrozumieć polecenie sformułowane w języku naturalnym, trzeba najpierw zidentyfikować słownikowe znaczenie poszczególnych słów (jest to tak zwana analiza leksykalna), a następnie ustalić wzajemne relacje tych słów w zdaniu, czyli dokonać rozbioru gramatycznego (tzw. analizy syntaktycznej). By z systemu dialogowego mieć jakiś pożytek, należy umieć wydobyć sens, prawdziwe znaczenie wypowiedzi. W tym celu trzeba zwykle także poprawnie zinterpretować intencję osoby mówiącej. Jest to z reguły najtrudniejsza część, tak zwana analiza semantyczna. Problemu pełnej analizy semantycznej nie rozwiązano jeszcze zadowalająco dla żadnego języka.

CHIP: Kiedy możemy się zatem spodziewać stworzenia odpowiednio zaawansowanych systemów dialogowych?
R. T.: Chyba nie wcześniej niż w 2010 roku. Istotnym osiągnięciem będzie moim zdaniem jednak dopiero stworzenie systemów dialogowych, które będą mogły poprawnie interpretować polecenia człowieka wydawane głosem bez konieczności starannego formułowania wypowiedzi.

CHIP: A kiedy doczekamy się systemów rozpoznających mowę polską?
R. T.: Jeśli idzie o rozpoznawanie pojedynczych izolowanych słów języka polskiego – to nie ma żadnych przeszkód już dzisiaj. Dostępne są już prototypowe systemy tego typu, oferowane głównie niewidomym. Oczywiście nie chodzi w tym przypadku o pełną i nie ograniczoną komunikację za pomocą głosu, ale o zastąpienie klawiatury i myszki – po prostu mikrofonem. Gorzej jest z mową ciągłą...

CHIP: Czy nie można po prostu zaadoptować programów zachodnich?
R. T.: Jest ogromna różnica pomiędzy programem, który wyświetla na ekranie polskie menu, a aplikacją, która byłaby w stanie poprawnie zrozumieć głosowe polecenie użytkownika sformułowane w języku polskim. By rozumieć polecenia wydawane głosem, trzeba najpierw uporać się z różnicami fonetyki. Polska wymowa jest odmienna od angielskiej, co ma swoje bezpośrednie odbicie na płaszczyźnie akustycznej struktury sygnału mowy. Innymi słowy, istnieje trudna do wyobrażenia różnica stopnia złożoności zadań przy rozpoznawaniu mowy polskiej w stosunku do bardziej zaawansowanych osiągnięć uzyskanych dla języka angielskiego. Zupełnie inna jest gramatyka języka polskiego, z licznymi odmiennymi formami części mowy oraz z ogromną liczbą skomplikowanych reguł i jeszcze bardziej skomplikowanych wyjątków...

wersją Lektora, która wyposażona będzie m.in. w funkcje rozpoznawania mowy ciągłej. Gotowy produkt ma być przedstawiony szerokiej publiczności najprawdopodobniej na początku przyszłego roku; zaawansowanie prac oceniane jest na ok. 50%. Wedle zapowiedzi pakiet będzie współpracował z popularnymi edytorami tekstu, takimi jak np. Microsoft Word.

Od 1993 r. prace z zakresu analizy i syntezy mowy prowadzi również wrocławski Neurosoft. Ich rezultatem jest generator mowy syntetycznej *SynTalk* (patrz CHIP 6/96, s. 77). Firma pracuje obecnie nad systemem do automatycznego rozpoznawania mowy ciągłej dla języka polskiego (projekt *NeuroEar*), jednak w chwili obecnej trudno określić termin wprowadzenia gotowego produktu na rynek.

Najgorszy jest więc, jak widać, fakt, że obecnie o pakietach tego typu możemy w zasadzie przede wszystkim... poczytać. Miejmy nadzieję, że w chwili, gdy światło dzienne ujrzy nowa wersja środowiska Windows wyposażona w funkcje rozpoznawania mowy, na rynku pojawi się także polskojęzyczna edycja systemu, pozwalająca na pracę w naszym języku ojczystym.

Cezary Dołęga,
Piotr Kubiszewski

info

Internet

Speech Recognition – How it works:
http://ourworld.compuserve.com/homepages/Grant_Cari_Fairley/spechdes.htm
Microsoft Speech Technology:
http://www.research.microsoft.com/research/srg/
Dragon Systems:
http://www.dragonsys.com/
http://www.naturalspeech.com/
IBM:
http://www.software.ibm.com/viavoice/
Lernout&Houspie:
http://www.lhs.com/
Phillips:
http://www.freespeech98.com/
Voicetronics:
http://www.dateko.cz/voice.htm
Neurosoft:
http://www.neurosoft.com.pl/

Grupa dyskusyjna

Pytania, uwagi i komentarze do artykułu można umieścić na liście dyskusyjnej *news://news.vogel.pl/chip.software*.