

Kamila Pawłowicz
nr indeksu: 211563
Wydział Nauk Ekonomicznych

Systemy rozpoznawania mowy

Systemy rozpoznawania mowy (*speech recognition* – SR), pozwalają odpowiednio zaprogramowanym komputerom przetwarzać odbierane sygnały dźwiękowe na tekst lub powodują wykonanie określonych funkcji. Współcześnie ta technologia jest dość mało precyzyjna, a także wymaga od użytkownika dość dużego wkładu (szczególnie czasowego) w dostosowanie systemu do swoich potrzeb. Ponadto systemy takie sprawdzają się bardzo dobrze jedynie w kontrolowanym środowisku, ich efektywność znacznie spada w przypadku „zwykłego” stosowania.

Przewiduje się¹, że w ciągu kilku lat technologia rozpoznawania mowy zostanie rozwinięta do poziomu, który umożliwi jej wykorzystanie w codziennym życiu. Będzie to niezwykle przełom, dający wiele nowych możliwości.

Wysoko rozwinięte systemy rozpoznawania mowy pozwolą na zmianę sposobu komunikacji człowieka z komputerem. Możliwości wydają się nieograniczone – od zwykłego wydawania maszynie poleceń (tzw. *hands-free computing*), przez dyktowanie tekstu oraz automatyczne wykonywanie stenopisów, aż po możliwość głosowego sterowania urządzeniami zainstalowanymi w naszych domach.

Kolejnym bardzo ważnym zastosowaniem będzie zapewne rozbudowanie systemów bezpieczeństwa opartych o rozpoznawanie mowy użytkownika. Ludzki głos jest równie niepowtarzalny jak odciski palców i dzięki temu może okazać się bardzo praktycznym narzędziem np. kontroli dostępu do strzeżonych obiektów.

Dalszym zastosowaniem wydaje się możliwość tłumaczenia komputerowego „na żywo”. Połączenie rozpoznawania mowy z mechanizmem tłumaczącym rozpoznaną treść na inne języki całkowicie zrewolucjonizowałoby komunikację międzykulturową. Nie chodzi jedynie o ułatwienia w negocjacjach biznesowych czy politycznych. Ogólnodostępny system tłumaczenia real-time zastosowany np. w technologii Skype, dałby przeciętnym użytkownikom nieograniczone możliwości komunikacji i doprowadził do powstania prawdziwej globalnej wioski. Pokonana zostałaby podstawowa bariera – językowa. Równocześnie nie byłaby konieczna znajomość języków obcych np. przy poszukiwaniu pracy za granicą czy w czasie wycieczek. System przyczyniłby się najpewniej również do niepostykanego dotąd rozwoju wszelkiego rodzaju portali społecznościowych służących wymianie poglądów oraz dał ludziom dostęp do niezliczonej liczby źródeł informacji (obcojęzycznej stacje radiowe, telewizyjne itd.).

Bardzo precyzyjne mechanizmy rozpoznawania mowy umożliwiłyby również wprowadzenie tzw. audio miningu, czyli wyszukiwania treści w plikach dźwiękowych. Pewne próby wprowadzenia takiego rozwiązania były podejmowane przez m.in. Google, jednak jak na razie nie przyniosły oczekiwanych rezultatów. Audio mining byłby przydatny np. przy poszukiwaniu plików muzycznych i wideo w Internecie (gdy zapomnimy z jakiego filmu pochodzi słynne „run Forest, run!” wyszukiwarka natychmiast zaprezentuje nam odpowiedni fragment dzieła Roberta Zemeckisa).

Systemy SR mogłyby być wykorzystywane również w sposób czysto komercyjny. Ich rozwój przyczyniłby się do powstawania w pełni komputerowych i zautomatyzowanych centr obsługi klienta oraz infolinii. Obniży to zdecydowanie koszty prowadzenia tego typu usług oraz przyczyni się do ich popularyzacji wśród przedsiębiorców.

¹ Por. wywiad z futurologiem Rayem Kurzweilem (blogs.computerworld.com)