

Magdalena Dukielska
Wydział Matematyki, Informatyki i Mechaniki UW

Interface człowiek-maszyna za kilkanaście lat Jak komórka pomoże niewidomemu odnaleźć drogę

Interface człowiek-maszyna przez ostatnie kilkadziesiąt lat przeszedł ogromne zmiany. Bardzo ważnym aspektem tych zmian był fakt, że szerokie grono ludzi po raz pierwszy w historii otrzymało dostęp do komputerów. Sam sposób komunikacji z komputerem również szybko się zmieniał, np. w 1968r. wynaleziono myszkę komputerową, która po kilku dekadach zrewolucjonizowała interfejs człowiek-komputer. Mniej więcej w tym samym czasie zaczęto też prace nad siecią znaną pod nazwą Internet. Dziś dynamicznie rozwijające się firmy, takie jak Google, również mają ciekawe pomysły na wykorzystanie możliwości drzemiących w nowych technologiach. Warto wspomnieć chociażby o projekcie zeskanowania książek z najbardziej znaczących bibliotek świata i udostępnieniu ich za darmo w Internecie. Ta i wiele podobnych inicjatyw dają nam coraz łatwiejszy dostęp do wiedzy i pozwalają swobodnie rozwijać nasze talenty.

Warto zastanowić się czy takie same możliwości mają niepełnosprawni. Kiedyś zamknięci w swoich domach mieli bardzo mały kontakt z zewnętrznym światem. Dziś dzięki staraniom wielu ludzi oraz inicjatywom w ramach „dostępności” (*accessibility*), które m.in. dostosowują interfejsy komputerów do ich potrzeb, jest im coraz łatwiej w pełni uczestniczyć w życiu społecznym i co ważniejsze, robić to samodzielnie.

Czy osoba niewidoma może studiować informatykę i zostać programistą? Niewidomy Peter pracujący dla Google jest dowodem, że jak najbardziej tak. Okazuje się, że osoba niewidoma, dzięki zaawansowanym technologiom i wysiłkowi włożonemu w dostosowywanie interfejsów do potrzeb ludzi, jest w stanie korzystać z komputera tak samo jak każdy z nas. A wszystko dzięki syntezie mowy. Pierwsze badania nad syntezą mowy prowadzono już w XVIII w., ale przed pojawieniem się komputerów urządzenia symulujące ludzką mowę miały zastosowanie co najwyżej w rozrywce.

W dzisiejszych czasach niewidomi komunikują się z komputerem, korzystając z zaawansowanych technologicznie i drogich programów nazywanych po polsku czytnikami ekranowymi (*screen reader*). Są to programy identyfikujące i interpretujące to, co wyświetla się na ekranie komputera. W systemie Windows (niestety większość używanych obecnie czytników obsługuje tylko ten system operacyjny, choć trwają prace nad wsparciem dla Linux'a) polega to na przechwytywaniu zdarzeń wywołanych, np. naciśnięciem klawisza na klawiaturze czy przesunięciem myszy. Swoją drogą niewidomi zwykle nie korzystają z myszek, bo nie są one dla nich zbyt wygodnymi urządzeniami – niedowiarkom proponuję spróbować obsługiwać myszkę z zamkniętymi oczami. To, co powinno wyświetlić się na ekranie komputera po danym zdarzeniu jest następnie opisywane z wykorzystaniem syntezy mowy albo języka Braille'a, do czego służą do tego specjalne urządzenia, m.in. takie jak to na zdjęciu.



Czytniki ekranowe są tylko jedną z wielu technologii wspomagających osoby niepełnosprawne (tzw. *assistive technologies*). Czytnik wymaga dość dużych zasobów i niestety spowalnia działanie komputera, bo musi wykonać sporo pracy przy każdym ruchu użytkownika i dodatkowo opowiedzieć mu, co się stało w systemie. Ale mimo to, osoba niewidoma jest w stanie pracować równie sprawnie jak każdy inny doświadczony użytkownik komputera, a często nawet sprawniej. Niewidomi zwykle stosują mało popularne skróty klawiaturowe, które są

znacznie efektywniejsze niż myszki. Z kolei prędkość czytania jest zwykle ustawiona na tak dużą wartość, że dla osoby, która pierwszy raz ma kontakt z syntezą mowy, głos jest całkiem niezrozumiały, a osoba niewidoma i tak świetnie się w tym wszystkim orientuje.

W tej chwili na rynku jest sporo czytników ekranowych, w tym najbardziej znany JAWS. Warto zauważyć, że od kilku lat w najbardziej znaczących konkursach wygrywa syntezator mowy zaprogramowany w Polsce przez firmę Ivo Software. Jakość mowy jest tam wyjątkowo dobra. Włączając go, nie słyszymy sztucznego metalicznego głosu, jaki znamy z filmów si-fi, ale całkiem naturalny głos albo kobiety (Iwony) albo mężczyzny (Jacka) – we współczesnych syntezatorach mowy bez problemu da się odróżnić płeć a nawet wiek „osoby mówiącej”. Co więcej, czytanie jest równie zrozumiałe po angielsku, jak i po polsku, a chyba nie trzeba podkreślać, że dźwięki w języku polskim są dość trudne do poprawnego odtworzenia.



Technologia syntezy mowy, rozwijana w tej chwili przede wszystkim na dość wąski rynek niepełnosprawnych ma duże szanse stać się bardziej popularna. Już teraz widać, że na tym polu da się sporo osiągnąć. Wprowadzić ciągle mamy problemy z rozpoznawaniem ludzkiej mowy, ale za to jej synteza idzie nam całkiem nieźle.

Już dziś bardzo popularne są urządzenia mobilne na pograniczu telefonów i komputerów. Na swojej prezentacji latem tego roku jeden z twórców Internetu Vint Cerf właśnie w tych urządzeniach widział przyszłość interfejsu człowiek-maszyna. Coraz częściej korzysta się też z możliwości głosowej komunikacji maszyny z człowiekiem. Przykładem jest tzw. bankofon, czyli urządzenie umożliwiające klientowi uzyskiwanie informacji o stanie konta za pośrednictwem telefonu. Być może za kilka lat tego typu usługi będą dostępne też w polskich bankach.

Synteza mowy z pomocą komputerów będą mogły w przyszłości udzielać ludziom również innych rodzajów informacji: o repertuarze kin, godzinach otwarcia, kursach akcji, nowych wiadomościach email, itp. W połączeniu z ogólnodostępną lokalizacją danych telefon z wbudowanym syntezatorem mowy byłby bardzo przydatnym narzędziem usamodzielniającym ludzi niepełnosprawnych. Mógłby on m.in. ułatwiać im przemieszczanie się, dzięki wbudowanemu GPS'owi określającemu naszą pozycję, ogólnodostępnej sieci bezprzewodowej i dobrze zorganizowanemu systemowi map.

Być może za kilkanaście lat normalnym zjawiskiem będzie niewidoma osoba, która jest sama w nieznanym mieście i musi załatwić coś w urzędzie. Nie ma pojęcia, gdzie powinna się udać, ale dzięki wszechobecnej sieci może wprowadzić zapytanie, korzystając ze średniej klasy „komórki” i otrzymać adres odpowiedniego miejsca (może w nowej wersji wyszukiwarki Google sterowanej głosem i przeznaczonej specjalnie dla urządzeń mobilnych?). Budynek urzędu z pewnością będzie labiryntem wielu pokoi i korytarzy. Dla ludzi widzących będą wskazówki, jak dotrzeć w dane miejsce, ale dla niewidomego mogą one być niedostępne. Za to z pomocą komórki przy wejściu do budynku będziemy mogli bez żadnego problemu ściągnąć z darmowej sieci jego szczegółowy elektroniczny plan.

Dane geograficzne – podobne do tych, które dziś możemy oglądać w Google Earth, będą z upływem czasu znaczyć coraz więcej. Będą dostarczać szczegółowych informacji o danym miejscu w danym czasie. Być może będą zawierały również informacje historyczne – będziemy mogli obejrzeć, jak kiedyś wyglądało dane miejsce i jak będzie wyglądało w przyszłości, np. poprzez plany budowy nowych budynków. Wybierając dany obiekt, będzie można ściągnąć o nim dodatkowe informacje, w tym komentarze innych ludzi, którzy teraz są

w tym miejscu albo byli tam kiedyś. W ten sposób zawsze będziemy otrzymywać najświeższe informacje i pozostawać w kontakcie z innymi ludźmi.

Plan budynku będzie zawierał poza opisem jego topologii wiele metadanych, które pozwolą dokładnie określić przeznaczenie każdego pomieszczenia. Dzięki temu będziemy mogli z góry zdecydować, gdzie chcemy dotrzeć. Program zainstalowany na komórce z pomocą GPS'u odnajdzie, w którym miejscu budynku aktualnie się znajdujemy. Będziemy mogli wpisać (powiedzieć?), co chcemy załatwić i na tej podstawie zostanie wybrane pomieszczenie, do którego powinniśmy się udać. Następnie automatycznie zostanie znaleziona optymalna droga do tego miejsca, w szczególności omijająca różne przeszkody, takie jak np. zbyt wąskie przejścia lub podjazdy zbyt strome dla osoby na wózku inwalidzkim, o ile takie pozostaną jeszcze z naszych czasów. System będzie umiał dać konkretne wskazówki, jak dotrzeć do miejsca, które nas interesuje („idź prosto, teraz skręć o 90 stopni w lewo, idź prosto ...” albo „10 kroków prosto, potem skręć w lewo o 90 stopni, później znów 30 kroków prosto...” – mierzymy wszystko w krokach, bo czym dla człowieka przyszłości będzie metr – tylko jakąś abstrakcyjną jednostką wymyśloną dawno temu). Informacje te będą zsynchronizowane z GPS'em, więc jeśli pomylimy się i pójdziemy złą drogą, to system da nam o tym znać i powie, co najlepiej zrobić, żeby wrócić we właściwe miejsce. Oczywiście podanych wskazówek nie będziemy musieli czytać – za pomocą syntezy mowy zostaną one nam odczytane i do ich odsłuchania wystarczą nawet staromodne słuchawki.

Bibliografia

1. Technologie wspomagające – http://en.wikipedia.org/wiki/Assistive_technology
2. Synteza mowy – http://pl.wikipedia.org/wiki/Synteza_mowy
3. Czytniki ekranowe – http://en.wikipedia.org/wiki/Screen_reader
4. JAWS – http://www.freedomscientific.com/fs_products/software_jaws.asp
5. IVONA – http://www.ivo.pl/?page=ivona_on_line