

Aplicații informatice și dispozitive cu interfață vocală

Paul Fogarassy-Neszly

BAUM Engineering

Arad, str. Traian Moșoiu nr.8

pf@baum.ro

REZUMAT

În acest articol sunt descrise aplicații software și hardware cu interfață vocală. Scopul acestora este pe de o parte de a asigura accesul persoanelor cu deficiențe de citire la tehnica de calcul și comunicare, iar pe de altă parte de a îmbunătăți uzabilitatea și ergonomia acestora. Sunt prezentate atât aplicații informatice și dispozitive pentru persoane cu deficiențe de vedere, cât și de uz general.

Cuvinte cheie

Accesibilitate, sinteză vocală, recunoaștere vocală, aplicații cu interfață vocală, tehnologii asistive, deficienți de vedere.

Clasificare ACM

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6).

INTRODUCERE

Aplicațiile cu interfață vocală (*self-voicing application*) sunt aplicații care pot fi folosite de către utilizatorii cu deficiențe de citire, fără ajutorul unui cititor de ecran. Categoriile de deficienți avute în vedere sunt: nevăzătorii, slab-văzătorii (ambliopi sau deficienți de vedere), dislexicii și analfabeții. Scopul principal al acestui tip de aplicații îl constituie asigurarea accesibilității.

În ultimii ani sunt dezvoltate dispozitive cu interfață vocală pentru automatizarea serviciilor de asistență a clienților (*call centers*), precum și pentru creșterea utilizabilității aparatelor de bord, atât în industria aerospațială, cât și în industria constructoare de autovehicule. De asemenea, există echipamentele cu interfață vocală pentru uzul nevăzătorilor

La ora actuală, interacțiunea utilizatorului cu o interfață vocală este (în general) într-un singur sens și se bazează pe redarea prin voce sintetică a elementelor de interacțiune cu aplicația. Mai rar, interacțiunea are loc în ambele sensuri, ceea ce presupune că aplicația este capabilă de recunoaștere vocală. Recunoașterea vocală pentru limba română este încă în stadiu experimental; nu există încă nici o aplicație capabilă de recunoaștere vocală pentru limba română, care să prezinte un nivel decent de utilizabilitate.

APLICAȚII CU INTERFAȚĂ VOCALĂ PENTRU ASIGURAREA ACCESIBILITĂȚII

Cea mai utilizată tehnologie care asigură accesul persoanelor cu deficiențe de vedere la tehnica de calcul și comunicare sunt cititoarele de ecran și magnificatoarele de ecran. Prin natura lor, aceste aplicații au interfață vocală și/sau Braille.

Cititoarele de ecran

Cititoarele de ecran sunt aplicații informatice care au rolul de a identifica și interpreta informațiile care sunt trimise pentru afișare pe ecranul calculatorului. Tehnologia care stă la baza acestui tip de aplicații permite interceptarea și prelucrarea corespunzătoare a ieșirii grafice a sistemului; aceste informații sunt apoi redată prin sinteză vocală și/sau afișaj Braille. Cititoarele de ecran fac parte din categoria tehnologiilor asistive și se adresează în principal nevăzătorilor sau persoanelor cu deficiențe severe de vedere, dar și dislexicilor, persoanelor cu deficiențe de învățare sau persoanelor care nu știu să citească. [4]

Odată cu utilizarea pe scară largă a interfețelor grafice s-au dezvoltat și cititoare de ecran mult mai sofisticate, care permit recunoașterea informației textuale precum și redarea prin informație textuală a unor elemente grafice. Această informație este redată fie prin sinteză vocală, fie cu ajutorul unui afișaj electronic prin caractere Braille, fie simultan.

Magnificatoarele de ecran cu redare vocală

Magnificatoarele de ecran sunt aplicații informatice prin care este posibilă prezentarea mărită a conținutului ecranului. Tehnologia care stă la baza acestui tip de aplicații permite interceptarea și prelucrarea corespunzătoare a ieșirii grafice a sistemului.

Pe lângă funcțiile specifice de magnificare a informației afișate, de creștere a contrastului și de modificare a paletelor de culori, magnificatoarele de ecran profesionale permit redarea prin sinteză vocală a informației textuale afișate, sau chiar a elementelor interfeței grafice.

Redarea prin sinteză vocală a conținutului paginilor Web

Pentru un nevăzător, cel mai eficient instrument software pentru redarea conținutului ecranului calculatorului prin sinteză vocală și/sau prin afișare Braille este cititorul de ecran. Există însă multe categorii de persoane cu deficiențe de citire, care fie nu au acces la un cititor de ecran, fie (la un moment dat) nu pot utiliza o asemenea aplicație. [5]

WebVOX nu este un cititor de ecran, deși îndeplinește unele dintre funcțiile acestuia. La redarea prin sinteză vocală, aplicația permite parcurgerea paragrafelor și elementelor grafice, cu redarea vocală a conținutului sau numelui acestora. [1][2]

Textul conținut în pagina Web curentă este în general parcurs paragraf cu paragraf; în acest caz, paragraful curent este marcat prin culori cu contrast maxim (de exemplu galben sau alb pe fond negru – Figura 1).



Figura 1. Sistem client-server pentru accesibilizarea sitului MCTI (stânga aspectul original, dreapta versiune accesibilă)

Componenta client a aplicației transmite componentei server textul conținut de paragraful curent; aceasta convertește textul (conform regulilor de citire în limba română) în fișier audio pe care apoi îl transmite înapoi componentei client. În sfârșit, fișierul audio este redat pe calculatorul utilizatorului.

APLICAȚII DESTINATE DEFICIENȚILOR DE VEDERE

Așa după cum s-a arătat, pentru a putea lucra la calculator, deficienții de vedere au nevoie de un cititor de ecran sau de un magnificator de ecran. Totuși există aplicații care pot fi utilizate de deficienții de vedere chiar și în absența unei asemenea tehnologii asistive. Desigur, acest lucru are sens atunci când informația oferită de aplicație nu poate fi redată în mod util de către cititorul sau magnificatorul de ecran.

Program educațional pentru învățarea alfabetului Braille

Un exemplu tipic de aplicație cu interfață vocală destinată deficienților de vedere este *Profu' de Braille*, care permite învățarea alfabetului Braille.

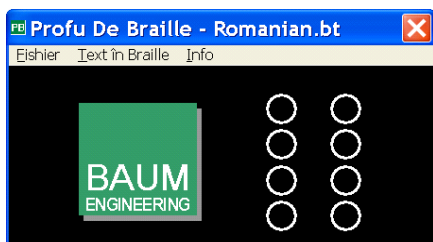


Figura 2. Interfața grafică pentru *Profu' de Braille* cu 6 sau cu 8 puncte

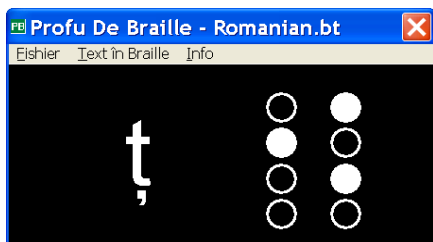


Figura 3. *Profu' de Braille* permite folosirea caracterelor specifice diverselor limbi

În modul implicit (Figura 2), *Profu' de Braille* permite folosirea unei tastaturi normale ca și tastatură Braille, iar

caracterul tastat în Braille este afișat pe ecran. În acest mod, programul anunță doar numele caracterului tastat.

În modul „Text în Braille” (Figura 3), pe lângă afișarea caracterului tastat, aplicația anunță prin voce sintetică atât caracterul cât și dispunerea punctelor în reprezentarea Braille.

În exemplul din Figura 4, programul anunță „a rond patru, șapte”, iar pentru exemplul din Figura 5 este anunțat „î din a majuscul unu, doi, șase, șapte”.



Figura 4. La tastarea oricărui caracter acesta este afișat împreună cu reprezentarea Braille



Figura 5. În modul „Text în Braille” aplicația afișează caracterul tastat și anunță dispunerea punctelor

Aplicații pentru citirea prin sinteză vocală a unui text

Un alt tip de aplicații, utilizate frecvent de nevăzători și deficienți de vedere sunt cunoscute sub numele generic de aplicații TTS (*Text To Speech*).

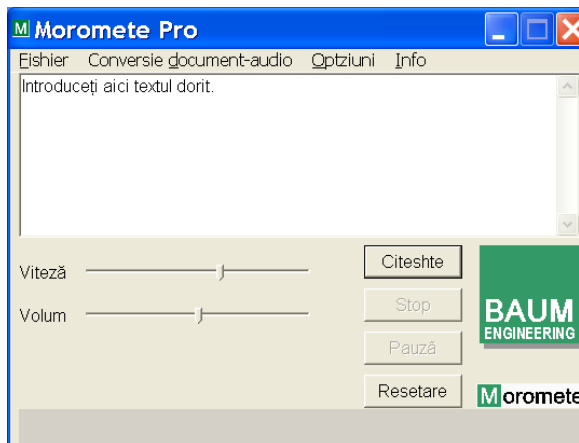


Figura 6. *Moromete Pro* - Exemplu de aplicație TTS

Aceste aplicații permit redarea vocală a unui text introdus în fereastra aplicației, sau a unui fișier document. Pe lângă redarea prin sinteză vocală a textului, aceste gen de aplicații permit de obicei și crearea unui fișier audio cu textul sintetizat.

Prin funcția de „Conversie document-audio”, aplicația *Moromete Pro* (Figura 6) permite realizarea unui document audio plecând de la un fișier document care

poate fi de tip pdf, doc, odt, rtf, txt, html sau chm. Fișierul audio obținut poate fi de tip mp3, wav, flac, spx sau aac. Utilizatorul are posibilitatea de a stabili calitatea fișierului audio, ceea ce influențează direct dimensiunea acestuia.

Prin funcția de conversie directă, Moromete Pro poate fi folosit pentru crearea rapidă de cărți audio (e-Book).

ECHIPAMENTE CU INTERFAȚĂ VOCALĂ PENTRU NEVĂZĂTORI

Nevăzătorii pot învăța să utilizeze diverse aparate, chiar dacă la proiectarea acestora nu s-a avut în vedere asigurarea accesului pentru persoane cu deficiențe de vedere. În general, nevăzătorii învață să distingă butoanele în funcție de poziția acestora și de forma lor.

Din păcate, tot mai des apar echipamente și aparate care nu pot fi folosite de către nevăzători, datorită faptului că se bazează pe ecrane tactile sau, așa cum se întâmplă în lifturi, au butoane a căror destinație nu poate fi intuită.

Pe de altă parte, există aparate special concepute pentru uzul deficienților de vedere care au panou de comandă tactil, uneori marcat cu caractere Braille și care pot avea și interfață vocală.

Imprimante Braille cu interfață vocală

Imprimantele Braille sunt concepute în general astfel încât să poată fi cu ușurință utilizate de către nevăzători sau de către persoane cu deficiențe de vedere. Acest lucru este asigurat în general printr-un panou de control cu butoane marcate Braille. Astfel este asigurat accesul inclusiv pentru nevăzătorii hipoacuzici.



Figura 7. Panou de control tactil pentru imprimanta Braille

În Figura 7 este prezentat un exemplu de panou de control tactil cu marcaj Braille pentru o imprimantă Braille. În zona centrală este vizibil difuzorul pentru interfața audio.

Fiecărei funcții îi este asociat un fișier audio, care este redat la apăsarea tastei sau combinației de taste corespunzătoare. De asemenea, imprimanta are o serie de mesaje înregistrate, care sunt redat când situația o cere. Acestea anunță terminarea imprimării, terminarea hârtiei, blocarea hârtiei în imprimantă etc.

Fișierele audio sunt disponibile pentru mai multe limbi, care la cerere pot fi instalate în memoria imprimantei.

Ca și în cazul altor dispozitive cu interfață vocală, nu există încă pe piață o imprimantă Braille care să preia comenzi de la utilizator prin recunoaștere vocală.

Agende electronice cu interfață vocală

O agendă electronică pentru nevăzători se bazează în general pe sinteza vocală pentru a permite accesul utilizatorului la funcțiile acesteia și la fișierele document.



Figura 8. Agendă electronică cu tastatură Braille, cu afișaj Braille și sinteză vocală

Figura 8 prezintă o agendă electronică pentru deficienți de vedere, care pe lângă sinteza vocală are și un afișaj Braille. Posibilitatea de a accesa informația prin două canale senzoriale (auditiv și tactil), are avantajul de a permite înțelegerea mai rapidă a informației și memorarea mai bună pe termen lung.

De asemenea, afișajul Braille permite utilizatorului un control mai bun asupra textului, deoarece caracterele care nu sunt citite ca atare (semnele de punctuație) sau cuvintele dificile sau străine, care nu sunt citite corect de către sinteza vocală pentru limba română pot fi imediat accesate, în paralel cu vocea sintetică.



Figura 9. Agendă electronică cu tastatură normală și sinteză vocală

În Figura 9 este prezentată o agendă electronică pentru deficienți de vedere care comunică cu utilizatorul exclusiv prin sinteză vocală. Utilizatorul accesează funcțiile și introduce datele cu ajutorul unei tastaturi normale.

În cazul cuvintelor dificile, sau a caracterelor speciale, la fel ca și în cazul lucrului cu un cititor de ecran fără afișaj Braille, utilizatorul poate citi caracter cu caracter.

În general, o agendă electronică pentru nevăzători are implementate o mulțime de funcții utile, care sunt disponibile în general pe orice agendă electronică (PDA). Dintre acestea, cele mai comune sunt agenda telefonică, editorul de texte, sincronizarea datelor cu un calculator personal, reportofon, calculator standard și calculator științific, navigator de internet, redare fișiere audio și cărți electronice și multe altele.

Dispozitive de citit prin sinteză vocală

Un dispozitiv de citire prin sinteză vocală încorporează într-un singur aparat un calculator și un scanner. Prin apăsarea unui singur buton, are loc scanarea documentului introdus în aparat, recunoașterea caracterelor, salvarea temporară a documentului și citirea prin voce sintetică a acestuia.



Figura 10. Dispozitiv de citire prin sinteză vocală

În Figura 10 este prezentat un aparat de citit prin sinteză vocală pentru deficienți de vedere.

Aparatul poate fi folosit exclusiv cu ajutorul unui singur buton, dar permite și accesarea unui meniu de funcții pentru utilizatorii avansați. Pe lângă redarea vocală, aparatul permite și conectarea unui afișaj Braille, ceea ce permite și accesul persoanelor hipoacuzice.

Orice text citit poate fi salvat în memoria aparatului în format document sau în format audio pentru a fi ulterior redat sau transferat în alt calculator sau dispozitiv de redare audio (de exemplu *MP3 player*); eventual fișierele pot fi salvate pe un disc optic sau alt mediu de stocare.

Pentru ca utilizatorul să poată regăsi ulterior fișierul, la salvare acesta este solicitat să pronunțe un nume de fișier care este înregistrat în format audio. Acest nume este asociat cu numele intern al fișierului salvat, iar la deschidere, utilizatorul poate asculta pe rând numele fișierelor salvate înainte de a se decide asupra fișierului care va fi deschis. Acest artificiu permite utilizarea aparatului fără a fi nevoie de tastatură.

Magnificatoare electronice cu sinteză vocală

Un magnificator electronic (numit și magnificator digital) permite mărirea electronică a imaginii unui document introdus sub camera aparatului. Mai mult decât atât, sunt posibile ajustări ale contrastului imaginii, alterarea moderată a culorilor pentru a compensa unele boli ale ochiului sau chiar afișarea imaginii în culori false. De exemplu, unii utilizatori preferă să citească alb pe negru, sau galben pe negru, dar și alte perechi de culori sunt disponibile, în funcție de preferințele utilizatorului și de afecțiunile de vedere ale acestuia.

Persoanele cu deficiențe de vedere nu au doar acuitate vizuală scăzută, dar pot avea (în funcție de afecțiune) alterată geometria câmpului vizual. Mai mult decât atât, citirea unui text, chiar și mărit, poate fi obositoare. Din acest motiv, unele magnificatoare permit descompunerea unui text în rânduri și derularea acestuia de la stânga la dreapta cu o viteză pe care utilizatorul o poate controla.

La fel ca și în cazul magnificatoarelor de ecran, care pentru a proteja restul de vedere al utilizatorului permite redarea unui text de pe ecran prin sinteză vocală, unele magnificatoare electronice permit redarea prin sinteză vocală a textului. Ca și în cazul dispozitivelor de citit, și în acest caz, are loc o recunoaștere optică a caracterelor și apoi redarea textului obținut prin sinteză vocală.

În Figura 11 este prezentat un magnificator digital care afișează dintr-o revistă un text mărit de aproximativ 25 de ori și în culori false, pentru o citire mai ușoară.



Figura 11. Magnificator electronic

Desigur, ca și în cazul dispozitivelor de citire, funcția de sinteză vocală este utilă în cazul textelor lungi, a căror citire poate fi obositoare. Spre deosebire de mașinile de citit, magnificatoarele cu sinteză vocală nu au deocamdată funcții avansate de stocare și conversie audio a textului citit. În curând vor fi disponibile magnificatoare care pe lângă sinteza vocală vor putea reda textul și cu ajutorul unui afișaj Braille.

ECHIPAMENTE DE UZ GENERAL CU INTERFAȚĂ VOCALĂ

Echiparea unor dispozitive de uz general cu o interfață vocală se poate dovedi deosebit de utilă atunci când utilizatorul este angajat în alte activități care nu îi permit să manipuleze eficient dispozitivul.

Conducerea unui vehicul este o activitate care necesită concentrarea atenției, iar utilizarea unor dispozitive, cum este telefonul mobil, este nu doar mai puțin eficientă, dar devine o activitate periculoasă, motiv pentru care în majoritatea țărilor folosirea telefonului fără un sistem *hands free* este interzisă de lege.

În cazul aparaturii de bord, în general, redarea vocală a unor informații sau chiar utilizarea comenzilor vocale, este de natură să crească nivelul de satisfacție a utilizatorului, datorită ergonomiei sporite. Desigur, aceste tehnologii sunt de natură să crească și siguranța în trafic.

Sisteme GPS

În funcție de modul în care un dispozitiv de poziționare globală (*GPS*) anunță mesajele necesare navigării, acestea pot fi de trei tipuri.

Cel mai simplu mod în care dispozitivul GPS anunță faptul că trebuie schimbată direcția de mers este prin tonuri. Acest sistem este potrivit pentru aparate simple, în general de uz pietonal, cu memorie redusă.

Cel mai comun sistem de interacțiune vocală a dispozitivelor de tip GPS este prin mesaje concatenate. Acest mod de interfață se bazează pe un număr redus de cuvinte preînregistrate, care în timpul funcționării sunt concatenate pentru a compune mesaje simple. Avantajul unei asemenea soluții constă în simplitate și în ușurința de funcționare; de asemenea, nu sunt necesare resurse hardware deosebite. Dezavantajul acestei soluții constă în faptul că numărul și complexitatea mesajelor sunt limitate.

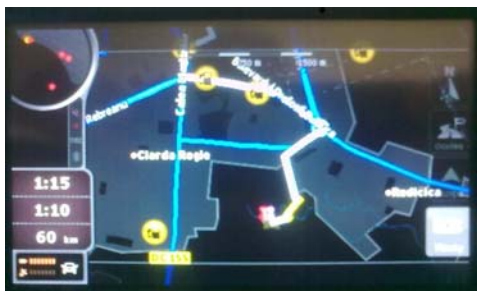


Figura 12. Reprezentarea traseului pe un sistem GPS

Figura 12 prezintă un sistem GPS clasic care afișează traseul care trebuie urmat, începând din poziția curentă. Pentru a accesa meniul dispozitivului, utilizatorul folosește ecranul tactil (Figura 13).

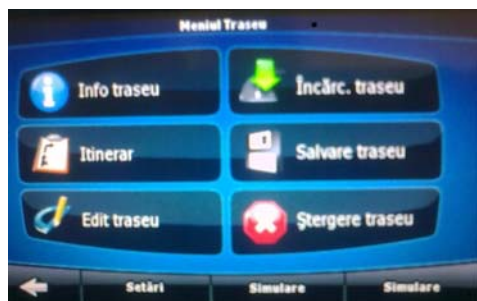


Figura 13. Meniul pentru trasee pentru un sistem GPS

Cel mai avansat sistem de interacțiune vocală constă în redarea mesajelor prin voce sintetică. O asemenea soluție permite redarea unor mesaje complexe, care pot include nume de străzi și localități. Desigur, așa ceva nu ar fi posibil prin mesaje preînregistrate. Pe lângă avantajul evident de a fi mult mai explicit în redarea indicațiilor de navigare, această soluție are dezavantajul unui echipament hardware mai pretențios, ceea ce ridică prețul total al dispozitivului

Deși există la ora actuală dispozitive GPS care oferă funcții de recunoaștere vocală, sistemul trebuie „antrenat” de către utilizator înainte de a putea fi folosit. Comenzile vocale sunt limitate și permit selectarea uneia dintre comenzile afișate de dispozitiv (Figura 14).

Desigur, acuratețea recunoașterii vocale este influențată de zgomotul de fond inerent traficului, de muzică și de diferența de timbru vocal al utilizatorului, care se poate datora unor afecțiuni temporare sau oboselii.



Figura 14. Interfața pentru comenzi vocale

Există două sisteme distincte de recunoaștere vocală: un sistem simplu, care permite selectarea dintr-un meniu cu un număr mic de opțiuni (ceea ce permite sistemului să fie destul de fiabil) și sistemele mult mai sofisticate de recunoaștere vocală, care permit (de exemplu) specificarea adresei de destinație sau căutarea unor puncte de interes de pe traseu (POI).

Aparatura de bord cu interfață vocală

Avantajele incontestabile ale interfeței vocale pentru creșterea securității traficului a făcut ca experiența acumulată la realizarea sistemelor GPS să fie extinsă pentru alte sisteme ale aparaturii de bord.



Figura 15. Un sistem de recunoaștere vocală capabil să distingă până la 300.000 de cuvinte în limba engleză

La fel ca și în cazul sistemelor GPS, există trei nivele distincte de interfață vocală.

Cel mai simplu sistem anunță prin mesaje preînregistrate diverse situații care solicită atenția utilizatorului. Acestea se pot referi la nivelul critic al carburantului, la temperatura exterioară sub o anumită valoare considerată critică, absența lichidului pentru spălarea parbrizului sau alte mesaje privind starea critică a unor sisteme.

În cazul unor sisteme mai complexe, pe lângă mesajele audio, utilizatorul poate controla prin comenzi vocale simple telefonul mobil sau sistemul audio, sistemul de navigație, precum și sistemul de climatizare.

În sfârșit, cele mai complexe sisteme de interfață vocală existente la ora actuală (Figura 15 prezintă un asemenea exemplu) permit controlul complet al sistemului de navigare, inclusiv introducerea adresei de destinație (așa cum s-a arătat mai sus), dictarea unui număr de telefon, sau specificarea numelui din agenda telefonului; de asemenea sistemele complexe permit dictarea de mesaje

SMS și trimiterea acestora numai prin comenzi vocale. În sfârșit, sistemele avansate permit controlul vocal al instalației de climatizare și controlul sistemului audio.

Tehnologia de control prin recunoaștere vocală a aparatelor de bord este aplicată și la elicopterele Gazelle ale armatei Marii Britanii. [3]. Sistemul nu depinde de utilizator, nu trebuie antrenat și are ca scop reducerea timpului în care piloții trebuie să consulte aparatura de bord. Pe lângă controlul anumitor sisteme, pilotul poate cere și informații, care îi sunt furnizate prin sinteză vocală.

Desigur, sistemul de control vocal este rezervat doar sistemelor necritice și în nici un caz nu implică armamentul și nici sistemele de siguranță.

Bancomate cu interfață vocală

Bancomatele cu interfață vocală oferă instrucțiuni audio, care permit unui utilizator cu deficiențe de citire să poată totuși utiliza aceste aparate, fără a fi asistat.

În general, bancomatele au o mufă pentru căști (Figura 16), ceea ce asigură nu numai accesibilitatea, ci și confidențialitatea utilizării acestora. Având în vedere aspectele legale din unele țări, care sunt foarte stricte cu privire la asigurarea accesibilității [6], precum și datorită faptului că tehnologia hardware și software necesară pentru redarea comenzilor vocale nu este scumpă, este de așteptat ca în câțiva ani să nu mai existe bancomate care să nu fie accesibile.



Figura 16. Bancomatele au în general o mufă standard pentru căști

Bancomatele actuale funcționează atât cu mesaje preînregistrate, cât și prin sinteză vocală.

CONCLUZII

Odată cu evoluția tehnologică apar tot mai multe aplicații informatice și echipamente cu interfață vocală. Rolul acestora nu este doar asigurarea accesibilității pentru utilizatorii cu deficiențe de citire (deficienți de vedere, dislexici, analfabeți etc.), ci și creșterea uzabilității și ergonomiei, mai ales în timpul conducerii unui vehicul, când atenția trebuie să fie concentrată în afara acestuia și mai puțin la instrumentele de bord.

La ora actuală problema recunoașterii vocale nu este complet rezolvată, deși s-a ajuns la rezultate destul de bune. În momentul în care recunoașterea vocală va depăși un anumit nivel, începând cu care sistemele vor putea fi utilizate cu încredere, răspândirea acestei tehnologii va fi aproape universală.

REFERINȚE

1. Lazar, J., Dudley-Sponaugle, A., Greenidge, K.-D. (2004) Improving web accessibility: a study of webmaster perceptions. *Computers in Human Behaviour* 20, 269-288.
2. Leuthold, S., Bargas-Avila, J., Opwis, K. (2008) Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users. *Int. J. Human-Computer Studies* 66, 257-270
3. McKeegan, N., Speech recognition technology allows voice control of aircraft systems, 2007. <http://www.gizmag.com/go/7484/>
4. Riga Ministerial Declaration (2006), Riga, Latvia. Available at: http://ec.europa.eu/information_society/events/ict_riga_a_2006/doc/declaration_riga.pdf
5. Takagi, H., Kawanaka, S., Kobayashi, M., Sato, D., Asakawa, C. (2009) Collaborative web accessibility improvement: challenges and possibilities. *Proceedings of ACM SIACCESS*, ACM, 195-202.
6. US Section 508 (1998). Available at: <http://www.section508.gov/>