

UNIVERSITATEA POLITEHNICA DIN BUCUREȘTI

Facultatea de Automatică și Calculatoare

Catedra de Calculatoare



REZUMAT TEZĂ DE DOCTORAT

**Interfețe Multimodale pentru Sisteme
Asistive pentru o Viață Activă**

CONDUCĂTOR ȘTIINȚIFIC:

Prof. Dr. Ing. Adina Magda FLOREA

AUTOR:

Imad Alex AWADA

București, 2019

Cuprins

Introducere.....	1
Motivația Alegerii Temei.....	1
Scopul Tezei	2
Structura Tezei	2
1. Definirea problemei	3
2. Stadiul Actual al Cercetării în Domeniu	4
3. CAMI – Un sistem pentru un Mod de Viață Activ și Asistat.....	8
4. Proiectarea Interfețelor Multimodale	8
5. O Interfață Multimodală Adaptivă pentru CAMI	9
6. Implementarea Interfeței	11
7. Experimente și evaluare	13
8. Concluzii	14
Bibliografie	17

Lista de Figuri

Figura 5.1. Arhitectura interfeței.	10
Figura 6.1. Pagina web principala a CAMI.....	12

Lista de Tabele

Tabelul 2.1. Comparație privind soluțiile legate de interacțiuni-voce	6
---	---

Lista de Abrevieri

2D	Bidimensional
3D	Tridimensional
AAL	Viață Activă și Asistată
AmI	Inteligența ambientală
API	Interfața de Programare a Aplicației
ASR	Recunoașterea Automată a Vocii
CAMI	Companion with Autonomously Mobile Interface
DM	Managementul Dialogului
GUI	Interfața Grafică cu Utilizatorul
HMI	Interacțiune Om-Mașină
ICT	Tehnologia Informației și Comunicării
JSON	Notăția Obiect JavaScript
NLP	Procesarea Limbajului Natural
NLU	Înțelegerea Limbajului Natural
SDK	Kitul de Dezvoltare Software
TTS	Text în Voce
UPB	Universitatea Politehnică din București

Introducere

Interfețele multimodale oferă multe modalități de interacțiune între utilizatori și sisteme, făcând interacțiunile mai naturale pentru utilizatori, așadar mai facile și mai atractive. Interfețele contribuie la eliminarea barierelor între tehnologie și utilizatorii fără cunoștințe tehnice sau cei cu nevoi speciale precum persoanele în vârstă.

În plus față de metodele tradiționale de interacțiune precum prin intermediul unei interfețe grafice utilizând un mouse și / sau o tastatură, interacțiunile multimodale permit utilizatorului să interacționeze cu sistemele prin cai naturale de interacțiune bazate pe vorbire, gesturi pe baza de atingere sau fără atingere, dar și prin alte modalități care sunt naturale pentru utilizatorii umani.

Motivația Alegerii Temei

Raportul ONU din 2017 privind îmbătrânirea populației arată că numărul de persoane cu vârste de peste 60 de ani a ajuns la 962 de milioane, adică mai mult decât dublul numărului din 1980. Mai mult, se estimează o dublare a numărului până în anul 2050 ajungându-se la 2,1 miliarde [1]. De asemenea, raportul arată că numărul de persoane în vârstă care trăiesc independent în casa lor este în creștere.

Numărul în creștere al oamenilor în vârstă este însoțit de o presiune mare asupra serviciilor sociale și sistemelor de îngrijire, care se confruntă cu probleme economice legate de costurile crescute ale serviciilor din domeniul sănătății și asistenței, dar și cu lipsa disponibilității personalului specializat. Pentru a face față acestor probleme, îngrijitorii, profesioniștii din domeniul sănătății, persoanele cu decizie politică dar și inginerii IT caută noi soluții care să permită persoanelor în vârstă să trăiască în casa lor cu un grad acceptabil de independență, asigurându-le calitatea vieții și siguranța. Una din soluțiile recomandate sunt sistemele de tip AAL (Viață Activă și Asistată).

Pentru ca interacțiunea cu sistemele AAL să fie accesibilă și mai ușoară pentru vârstnici și pentru a maximiza beneficiile acestora, orice sistem AAL trebuie să integreze o interfață multimodală. Aceasta are un rol esențial în determinarea acceptării sistemului de către utilizatori și, prin urmare, un rol esențial în succesul unui sistem AAL.

Fiecare utilizator al unui sistem AAL are o anumită experiență anterioară, un anumit buget precum și anumite cunoștințe și preferințe. De asemenea, fiecare utilizator are potențial o motivație diferită pentru folosirea sistemului pe care îl va utiliza în diferite condiții și medii. Este deci important ca interfața multimodală să integreze unele caracteristici adaptive (capabilități adaptive automate și caracteristici individualizabile).

Datorita necesitații mari pentru sistemele AAL și rolului crucial al interfețelor multimodale, inteligența ambientală (AmI) și interfețele multimodale sunt în prezent direcții de cercetare foarte active atât în mediul industrial cât și în cel academic.

Din aceste motive, actualul studiu prezintă problemele de interacțiune dintre vârstnici și tehnologie și propune o interfață multimodală care integrează diverse capacități adaptive pentru un sistem AAL care are ca utilizatori ținta persoanele vârstnice.

Scopul Tezei

Scopul prezentei cercetării este identificarea celor mai bune modalități de interacțiune a vârstnicilor cu calculatoarele și / sau dispozitivele mobile, a identifica practicile de proiectare adecvate pentru crearea interfețelor multimodale cu caracteristici adaptive și a dezvolta o interfață multimodală care poate fi integrată într-un sistem de viață asistat și activ. Interfața a fost creată, în mod specific, pentru CAMI, un sistem de asistare a utilizatorilor cu nevoi speciale în viața de zi cu zi, care include interacțiuni cu un robot umanoid și care poate fi ușor integrată în orice aplicație AmI.

Interfața multimodală propusă este compusă din 4 module: interfața grafică cu utilizatorul (GUI), modul de voce, modul de gesturi, modul de emoții. Modulele de voce și emoții cuprind câte 5, respectiv 2 componente. Modulul de gesturi are 2 submodule: gesturi pe baza de atingeri și gesturi fără atingeri. Fiecare submodule are 2 componente. În timp ce GUI și modulul de gesturi operează fără prezența unei conexiuni active la internet, modulele de voce și emoție au 2 moduri de operare: online (conexiune activă la internet) și offline (nu necesită o conexiune activă la internet).

Pentru a oferi utilizatorului flexibilitatea de a interacționa ușor cu sistemul de pe orice dispozitiv pe care utilizatorul îl deține sau dorește să folosească, interfața multimodală operează independent de orice dispozitiv sau sistem de operare.

Proiectarea interfeței multimodale prezintă recomandările privind proiectarea interfeței cu utilizatorul, dar și caracteristicile, respectiv comenzile care sunt constante pentru diverse dispozitive și module ale sistemului.

Interfața multimodală este multilingvistică și integrează diverse caracteristici particulare ale utilizatorului și capacități adaptive automate precum adaptarea la profilul utilizatorului (preferințele, problemele de sănătate etc.), activitatea utilizatorului, starea sa emoțională sau configurația sistemului și alți parametri.

Structura Tezei

Teza este structurată în 4 părți principale:

Prima parte, compusa din 2 capitole, prezintă aspectele teoretice din literatura de specialitate. Capitolul 1 prezintă aspectele generale ale domeniului inteligenței ambientale, ale sectorului vieții active și asistate, respectiv al interacțiunii multimodale. Capitolul 2 analizează studiile din domeniul interacțiunii om-mașină și din sectorul vieții active și asistate.

A doua parte a tezei prezintă introducerea în sistemul CAMI și prezintă o imagine de ansamblu privind proiectarea proceselor interfeței multimodale. Partea a doua cuprinde, de asemenea, 2 capitole. În capitolul 3 se face o scurtă prezentare a sistemului CAMI, subliniind facilitățile de interacțiune multimodală ale sistemului. Capitolul 4 oferă o imagine asupra proiectării interfețelor multimodale și ilustrează provocările care ar putea de luat în considerare în timpul proiectării interfeței multimodale pentru vârstnici, precum și recomandările din punct de vedere ergonomic, observate în timpul proiectării interfeței. Se prezintă un model al interacțiunilor multimodale.

A treia parte prezintă proiectarea și implementarea interfeței multimodale și a modulelor sale. și aceasta parte cuprinde 2 capitole. Capitolul 5 prezintă o imagine de ansamblu asupra proiectării interfeței multimodale CAMI. Sunt prezentate diferite funcționalități ale interfeței, precum și caracteristici ergonomice și adaptive ale acesteia. Acest capitol prezintă arhitectura interfeței și ilustrează elementele de proiectare ale fiecărui modul al acesteia. Capitolul 6 prezintă o imagine de ansamblu asupra implementării interfeței multimodale și ilustrează aspecte și caracteristici ale implementării diverselor module ale acesteia.

A patra parte, care cuprinde capitolele 7 și 8 prezintă diferite experimente și evaluarea interfeței, împreună cu rezultatele obținute, precum și concluziile, respectiv studiile viitoare planificate. Capitolul 7 prezintă experimentările efectuate și evaluarea rezultatelor obținute. Capitolul 8 prezintă concluziile generale ale tezei, precum și aspecte viitoare care privesc continuarea cercetărilor.

Capitolul 1

Definirea problemei

Capitolul prezintă aspectele generale din domeniul inteligenței ambientale, din sectorul vieții asistate și active, precum și din domeniul interacțiunii multimodale.

Inteligența ambientală se referă la o lume în care dispozitivele de calcul sunt integrate peste tot în mediu în mod transparent și formează o rețea de dispozitive inteligente ascunse care pot recunoaște și interpreta evenimentele umane, comunica între ele (și / sau cu utilizatorul) și se pot adapta mediul la nevoile utilizatorului, la obiceiurile sale, la gesturile și emoțiile acestuia. Un sistem AmI se prezintă următoarele particularități: este sensibil, oferă răspunsuri, este adaptiv,

transparent, ubicuu și inteligent. Aplicațiile AmI sunt diverse, de la senzori și comunicații fără fir la îngrijire personală și interacțiuni om-mașină noi. Unele domenii de aplicații ale AmI sunt domotică (automatizarea casnică), îngrijirea sănătății, munca, sport, turism, patrimoniu cultural, vehicule, transport, interacțiuni om-mașină. În domeniul îngrijirii sănătății, AmI oferă numeroase soluții care nu-i doar că îmbunătățesc starea de sănătate, ci și modifică modul în care se practică îngrijirea sănătății. Una din soluții se reprezintă prin sistemele pentru o viață activă și asistată.

Modul de viață activ și asistat este un concept în care soluțiile care includ una sau mai multe componente ICT (tehnologii de informații și comunicare) ajută vârstnicii să trăiască cu un anumit grad de independență în mediul dorit cât mai mult posibil [2]. Soluțiile au o valoare importantă pentru vârstnici, familiile și îngrijitorii lor pentru a le îmbunătăți sănătatea și starea de bine, dar și independența [3]. În plus, sistemele AAL au un rol important în reducerea numărului de solicitări privind sistemele de îngrijire. Conform asociației AAL, sectorul AAL acoperă 8 zone de servicii: sănătate și îngrijire, informații și comunicare, mod de viață și clădiri, siguranță și securitate, mobilitate și transport, vitalitate și abilități, timp liber și cultură, muncă și antrenare.

Întrucât vârstnicii sunt principalii utilizatori urmăriți ai sistemelor AAL, sistemul interacționează cu aceștia cât mai natural. Interfețele multimodale prezintă multiple modalități de interacțiune între utilizator și sistem precum vocea, gesturile, atingerile, în plus față de cele asigurate de interfețele tradiționale.

Capitolul 2

Stadiul Actual al Cercetării în Domeniu

Acest capitol analizează studiile din domeniul interacțiunii om-mașină (HMI) și sectorul AAL, la care se raportează cercetările efectuate în această teză.

Evoluția HMI nu a fost limitată nici pentru a îmbunătăți calitatea HMI, nici pentru proiectarea și implementarea interfețelor obișnuite. Există diverse concepte precum interfețele multimodale, interfețele adaptive și interfețele inteligente.

Există diverse soluții legate de vorbirea om-mașină. Unele sunt soluții comerciale și altele sunt de tip sursă deschisă. Unele soluții sunt gata de folosire, altele sunt o bază la construcția altora, în timp ce altele sunt proiecte de cercetare. Tabelul 2.1 compară diversele soluții existente, identificate.

Recunoașterea vocii integrează 3 modele care lucrează simultan: cel acustic, de pronunție și de limbă, în timp ce procesul de sinteză a vorbirii cuprinde 3 etape: etapa de preprocesare în care sistemul parcurge diversele cuvinte ale textului pentru a găsi cea mai adecvată modalitate de a citi respectivul text, urmata de etapa a doua, în care cuvintele sunt transformate în foneme, care

sunt componente ale sunetului fiecărui cuvânt vorbit. În a treia etapă fonemele sunt transformate în sunete.

Fata de soluțiile din tabelul 2.1, există multe studii privind aspectele interacțiunilor de tip voce între om și mașină și soluțiile problemelor identificate [4, 5, 6, 7] pentru recunoașterea vocii, [8, 9, 10] pentru sinteza dialogului, [11, 12, 13] pentru înțelegerea limbajului natural. În ceea ce privește managementul dialogului (DM) ale interacțiunilor de vorbire om-mașină, există 2 abordări: DM pe baza de cunoștințe precum sistemele propuse în [14, 15] și DM pe baza de principii precum studiile [16, 17, 18, 19, 20].

Există studii care analizează interacțiunile om-mașină prin gesturi bazate pe atingeri. Se identifică, pe de o parte, un singur gest pe baza de atingere în care utilizatorii folosesc un deget pentru a interacționa cu sistemul printr-un ecran tactil: pentru a izbi sau a muta [21, 22, 23], pentru a trage sau muta un obiect [24, 25], respectiv pentru a desena un șablon de gesturi [22, 26].

Pe de altă parte, există studii care prezintă gesturile multiple pe baza de atingere în care utilizatorii folosesc mai multe degete ale aceleiași mâini sau ale ambelor mâini pentru a interacționa cu sistemul printr-un ecran tactil pentru a roti [22, 24, 25, 27], pentru a redimensiona sau conduce. Unele aplicații care suportă gesturile cu mai multe atingeri și se adresează vârstnicilor, ca principalii utilizatori, sunt jocurile de antrenare cu mai multe atingeri din proiectul HERMES [28] și aplicația de e-mail proiectată de Hollinworth și Hwang [29].

Studii multiple analizează gesturile fără atingeri în care utilizatorii interacționează cu sistemul prin mișcarea corpului. Gesturile mâinii sunt cele mai analizate forme de gesturi fără atingeri, fiecare gest al mâinii fără atingeri poate fi împărțit în 5 faze: poziția de repaus, pregătirea, gestul de strângere, gestul de apucare/tinere și cel de retragere [30, 31]. Unele studii propun faze suplimentare ale gesturilor mâinii fără atingeri, cum ar fi faza de recul în [32]. O altă formă este gestul de mișcare a capului [33, 34]. Gesturile fără atingeri sunt folosite în timpul programelor de antrenare și reabilitare în care sistemul urmărește mișcările utilizatorilor și le reproduce pe ecran printr-un avatar în timp real precum jocurile "Voracy Fish" [35] și "Hammer & Planks" [36]. Gesturile fără mâini sunt folosite în activități de divertisment precum cele din jocurile Microsoft Xbox și Nintendo Wii.

Mai multe studii analizează interacțiunile om-mașină pe baza emoțiilor utilizatorului, recunoscute prin perceperea expresiilor faciale ale acestuia [37, 38, 39], limbajul trupului [40, 41], sau vocea utilizatorului [42].

Sistemele multiple care integrează o interfață multimodală există deja pe piață, printre exemple numărându-se: Memphis Intelligent Kiosk Initiative [43], MATCHKiosk Multimodal Interactive City Guide [44], sistemul Canesta [45], sistemul Touch'n'Speak [46], sistemul Put That There [47] și sistemul de direcționare Gaze pentru oamenii cu dizabilități [48], dar și asistenți personali care au apărut recent precum Siri, Cortana și Google Assistant.

Tabelul 2.1. Comparație privind soluțiile legate de interacțiuni-voce : *: perioada de testare gratuită și unele tranzacții gratuite lunare, ‡: unele caracteristici sunt gratuite, altele necesită plată, °: limbaj(e) predefinit(e), S-t-T: serviciu vorbire-text și T-t-S: serviciu text-vorbire.

Soluții	Serv.	Servicii		Limbaje								Observații
		Gratuit	Cu plată	EN	FR	RO	SW	DA	PL	IT	Adi.	
IBM Watson	S-t-T		✓	✓	✓						✓	Depinde de internet
	T-t-S									✓		
Dragon NaturallySpeaking	S-t-T		✓	✓	✓	✓	✓	✓	✓	✓	✓	Depinde de internet
	T-t-S											
Lexix	S-t-T		✓	✓								
	T-t-S											
Windows Desktop Speech Technology	S-t-T	✓		✓	✓						✓	Depinde de Microsoft Windows
	T-t-S					✓	✓	✓	✓	✓	✓	
Azure Speech Services	S-t-T		✓*	✓	✓		✓	✓	✓	✓	✓	Depinde de internet
	T-t-S					✓						
Google Cloud Speech Services	S-t-T		✓*	✓	✓		✓	✓	✓	✓	✓	Depinde de internet
	T-t-S											
Alexa Voice Service	All Services	✓		✓							✓	Dispozitiv dependent de internet și de compatibilitate
Julius	S-t-T	✓		✓°							✓°	Orice limbaj
Cmusphinx	S-t-T	✓		✓°	✓°						✓°	Orice limbaj
HTK	S-t-T	✓									✓	Orice limbaj
	T-t-S											
Kaldi	S-t-T			✓°							✓	Orice limbaj
Jasper	S-t-T			✓°							✓°	Depinde de placa Rpi Orice limbaj
Apple Speech Framework	S-t-T	✓		✓	✓		✓	✓		✓	✓	Depinde de iOS și internet
Ispeech	S-t-T		✓	✓	✓		✓	✓	✓	✓	✓	Depinde de internet
	T-t-S											
eSpeak	T-t-S	✓		✓°	✓°	✓°	✓°	✓°	✓°	✓°	✓°	
ResponsiveVoice.JS	T-t-S	✓‡	✓‡	✓	✓	✓	✓	✓	✓	✓	✓	Depinde de internet
LUIS	NLP		✓*	✓	✓					✓	✓	Depinde de internet
DialogFlow	NLP	✓‡	✓‡	✓	✓		✓	✓		✓	✓	Depinde de internet
RASA NLU	NLP	✓										Orice limbaj
Wit.ai	NLP	✓		✓	✓	✓	✓	✓	✓	✓	✓	Depinde de internet
Azure Bot Service	DM		✓*	N/A								Depinde de internet
Bot Builder SDK	DM	✓		N/A								Depinde de MS Windows

Ca urmare a eforturilor comunității AAL de a încuraja implementarea de sisteme care oferă la un loc mai multe soluții proiectate sau propuse în domeniul managementului sănătății, al stării de bine și al îmbătrânirii la domiciliu, exista diverse sisteme și proiecte AAL precum proiectul inCASA [49] platforma Persona [50], proiectul NITICS [51], platforma Sociable [52], proiectul PersonAAL [53], CareWell [54], proiectul healthy@work [55], proiectul Wellbeing [56] și proiectul EldersUP! [57].

În ceea ce privește soluțiile analizate legate de modulul de voce, majoritatea soluțiilor trebuie combinate cu alte soluții pentru a opera ca un modul de voce funcțional care poate interacționa cu utilizatorul fără a fi necesare alte soluție. Unele soluții sunt pe baza de cloud, iar altele pot fi implementate local. Majoritatea soluțiilor sunt în limba engleza, dar și în alte limbi. Performanța acestor soluții este foarte variată, iar fiecare performanță este diferită în funcție de limba pentru care este folosit. Performanța soluțiilor cloud este mai bună decât cea a soluțiilor implementate local. Totuși, aceasta depinde de disponibilitatea conexiunii la internet.

În ceea ce privește modulul de gesturi, diverse studii propun soluții pentru gesturile pe baza de atingeri, în timp ce alte studii propun soluții pentru gesturile fără atingeri.

În ceea ce privește modulul de emoții, cercetătorii folosesc diverse abordări de determinare a emoției utilizatorului precum determinarea emoției pe baza expresiei faciale, a limbajului corporal, sau a vocii utilizatorului.

În ceea ce privește sistemele multimodale, se menționează că modalități de intrare sistemele care integrează vorbirea cu alte modalități. Ele suportă o singură limbă și majoritatea sistemelor depind de dispozitive, nefiind compatibile pe toate platformele și nu oferă capacități adaptive.

În ceea ce privește sistemele AAL analizate, majoritatea integrează diverse servicii de asistare a vârstnicilor în viață de zi cu zi și să îi ajute să-și mențină un stil de viață sănătos. Majoritatea acestor sisteme permit interacțiunea cu utilizatorii prin interacțiuni multimodale. Cu excepția platformei Sociable, nici o soluție din setul de sisteme prezentate nu integrează un modul de exercițiu fizic care îi ajută pe vârstnici acasă, foarte important pentru a menține o viață sănătoasă.

Cu excepția proiectelor PersonAAL și EldersUP!, niciuna dintre soluții nu ia în considerare importanța integrării capacităților adaptive pentru interfețele utilizatorilor lor.

Toate soluțiile menționate (comerciale sau proiecte de cercetare) au funcționalități limitate și nu acoperă toate funcționalitățile urmărite de un sistem AAL. Din punct de vedere al arhitecturii sistemului, soluțiile sunt legate de o singură tehnologie dominantă care restricționează implementarea și necesită eforturi de a integra noile funcționalități și module.

Capitolul 3

CAMI – Un sistem pentru un Mod de Viață Activ și Asistat

O parte semnificativa a studiului prezentat în teza vizează proiectul european “CAMI: Artificial intelligent ecosystem for self-management and sustainable quality of life în AAL” în cadrul programului Uniunii Europene 2015-2018 “Viață activa și asistata” [58].

CAMI este un ecosistem de inteligența artificială pentru autoadministrare și o calitate sustenabilă a vieții în AAL. Consorțiul proiectului cuprinde 8 parteneri din 5 țări europene, coordonator fiind Universitatea Politehnică din București (UPB).

Acest capitol prezintă sistemele CAMI. CAMI cuprinde principalele funcționalități ale sistemelor AAL precum planificarea activității, detectarea eșecurilor, exerciții fizice supervizate, monitorizarea sănătății, automatizarea [59]. CAMI include o unitate robotică de teleprezență și integrează o interfață multimodală care permite interacțiunea ușoară a utilizatorului cu sistemul. Utilizatorii de bază ai sistemului sunt persoane în vârstă (55-75 de ani). Îngrijitorii și profesioniștii din sănătate pot accesa sistemul de la distanță, pot vizualiza datele privind starea de sănătate a persoanei și activitățile în timp real. De asemenea aceștia pot înregistra activitatea utilizatorului, precum și aspecte privind starea sănătății acestuia. Îngrijitorii și profesioniștii din sănătate, în funcție de poziție și de specializare, pot ajusta medicația, exercițiile zilnice și unele sarcini zilnice ale utilizatorului. Utilizatorul poate adăuga, șterge orice serviciu când dorește.

Capitolul 4

Proiectarea Interfețelor Multimodale

Acest capitol prezintă aspecte generale ale proiectării interfețelor multimodale. Înainte de a începe procesul de proiectare, este foarte importantă identificarea utilizatorilor țintă ai sistemului și unde vor folosi aceștia funcționalitățile sistemului. Doar după aceea, procesul de proiectare a interfeței multimodale poate începe.

De obicei, proiectarea unei interfețe a unui sistem complex este o sarcină foarte delicată, mai ales pentru un sistem de viață activă și asistată, în care interfața are un rol decisiv în acceptarea sistemului de către utilizatori, deci pentru asigurarea succesului întregului sistem. Provocările luate în considerare în timpul proiectării interfeței multimodale ale oricărui sistem se împart în 2 grupuri principale: provocările legate de factorii umani și cele legate de factorii tehnici. Provocările legate de factorii umani depind de utilizatorii țintă ai sistemului și se împart

în 3 categorii de nivele, în funcție de factorul cauza: probleme fizice, probleme cognitive, experiența cu calculatorul [60]. În afara provocărilor legate de factorii umani există și cele legate de factorii tehnici precum comenzile simultane contradictorii, fuziunea și fisiunea datelor, problemele privind siguranța și confidențialitate.

Pentru a crea o experiență facilă și plăcută a interacțiunii cu utilizatorii, proiectarea oricărei interfețe trebuie să urmeze câteva recomandări. În ceea ce privește conținutul interfeței care trebuie afișat pe ecran, unele aspecte trebuie să se comporte în mod constant pe diversele ecrane (pagini) ale interfeței. Terminologia interfeței și fontul folosit, dimensiunea sa, icon-urile și culorile trebuie să fie constante pe diversele ecrane. Interfața trebuie să fie simplă, icon-urile folosite să fie cunoscute de utilizatori, secvențele lungi și sarcinile complexe să fie fragmentate în pași separați și teme simple. Informația afișată trebuie organizată într-un număr limitat de alegeri. Ea trebuie să ofere indicii de explorare pentru că utilizatorul să știe pe ce ecran se află, trebuie să ofere răspunsuri informative despre ce se întâmplă sau despre ce s-a întâmplat și trebuie să genereze avertismente când este cazul. Cantitatea de informații trebuie minimizată, iar câmpurile cu datele de intrare trebuie să fie formate automat și validate pentru a ajuta utilizatorul, trebuie să fie evitate erorile simple, iar informațiile trebuie să fie afișate vizual, pe ecran o anumită perioadă. Interfața trebuie să aibă repere vizuale adecvate pentru că utilizatorul să exploreze interfața și să găsească mai ușor informațiile importante.

În ceea ce privește datele de intrare de tipul gesturilor (cu și fără atingeri) și sunetelor fonetice, comenzile trebuie să fie constante atât pe diverse dispozitive, cât și pentru serviciile și modulele sistemului. Datele de ieșire fonetice, mai ales în cazul răspunsului fonetic, trebuie să fie constante pe diversele ecrane ale interfeței. Dacă interfața are capacități adaptive diferite, fiecare din acestea trebuie introdusă în mod clar către utilizator la prima folosire. Utilizatorul trebuie să aibă o opțiune de a dezactiva fiecare caracteristică adaptivă, dar și întreaga capacitate adaptivă.

Capitolul 5

O Interfață Multimodală Adaptivă pentru CAMI

Capitolul prezintă aspecte generale privind proiectarea interfeței multimodale a sistemului CAMI. Întrucât sistemul urmărește vârstnicii ca principalii utilizatori, interfața este proiectată să îndeplinească necesitățile și cerințele acestora precum fonturile mari, butoanele mari, navigarea facilă, volumul sunetului, interacțiunea cât mai naturală cu sistemul. Interfața permite un acces facil la diversele caracteristici ale sistemului, întrucât utilizatorii sunt din diverse țări, au diverse istorii personale și culturi cu preferințe diferite, obiceiuri și cunoștințe, dar și probleme diferite. Interfața este multilinguală, suportând diverse limbi ale consorțiului CAMI (engleza, franceza, româna, suedeza, daneza, poloneza, italiana), cu excepția modulului

de voce (engleza, franceza, romana). Se pot adauga mai multe limbi la interfață, inclusiv în modulul de voce.

Interfața are funcționalități multiple. Permite utilizatorului să interacționeze ușor cu sistemul prin interacțiunile multimodale folosind GUI sau vocea, gesturile pe baza de atingeri și comenzile cu gesturi fără atingeri. Interfața permite sistemului să genereze rezultate vizuale și fonetice și să urmărească starea emoțională a utilizatorului, dar și evoluția sa în timpul diverselor sesiuni de exerciții fizice. Interfața permite utilizatorului accesul facil de pe orice dispozitiv pentru datele privind starea de sănătate (date în timp real), întâlnirile, memento, notificări inteligente, exerciții fizice și diverse caracteristici ale sistemului. Interfața permite utilizatorului să o poată individualiza (culori, ordinea afișării modulelor, forma răspunsului primit, limba interfeței etc.) și să pornească / oprească diverse module, capabilități adaptive sau module de intrare sau de ieșire. Interfața are o capacitate de adaptare automata. Interfața este proiectata ca sistemul să o adapteze automat, conform profilului și stării utilizatorului, dar și conform activității sale, configurației și altor parametri ai sistemului, precum condițiile de mediu. Interfața poate comunica preferințele utilizatorului aplicațiilor terțe. În plus, aceasta permite îngrijitorului să vadă, în timp real, de pe orice dispozitiv, datele și diversele rapoarte privind starea de sănătate a utilizatorului. Interfața permite îngrijitorului să particularizeze planul de medicație al utilizatorului pentru normalizarea valorilor variabilelor analizelor stării de sănătate. De asemenea, interfața permite îngrijitorului să ajusteze profilul stării de sănătate a utilizatorului și exercițiile care trebuie efectuate de către acesta, precum și alte opțiuni ca de exemplu înălțimea utilizatorului. În unele cazuri, foarte specifice, interfața permite îngrijitorului să adapteze orice dorește în sistemul utilizatorului.

Interfața operează pe mai multe platforme și dispozitive. Suporta interacțiuni multimodale și cuprinde 4 module: interfața grafica cu utilizatorul, voce, gesturi și emoții. Deși modulele operează împreună, fiecare este proiectat să funcționeze independent. Figura 5.1 prezinta arhitectura interfeței.

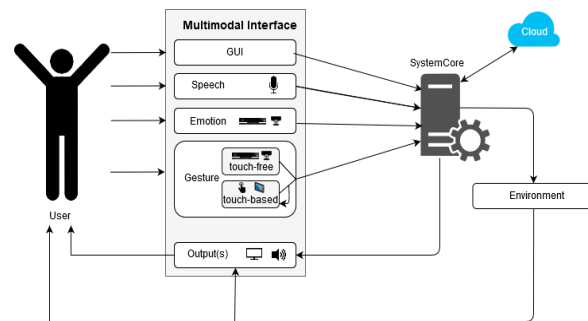


Figura 5.1. Arhitectura interfeței.

Interfața grafica cu utilizatorul ii oferă acestuia o alternativa clasica la alte module ale interfeței pentru a interacționa cu sistemul prin intermediul acestora.

Modulul de voce are 5 componente: preprocesarea audio, recunoașterea automata a vocii (ASR), înțelegerea limbajului natural (NLU), managementul dialogului (DM), sinteza text-voce (TTS). Se identifica doua moduri de operare: online și offline. Variabilele de intrare ale sistemului sunt: comenzile vocale (de vorbire) ale utilizatorului sau comenzile sintetizate de vorbire din sistem (inițializate de către sistem sau de către utilizator pentru alt modul de intrare).

Variabilele de ieșire sunt informațiile audio sintetizate în urma vorbirii sau informațiile vizuale. Modulul de gesturi are 2 submodule: gesturile pe baza de atingeri și gesturile fără atingeri. Primul submodul are 2 componente: achiziția de date și procesarea datelor; variabila de intrare este comanda de tip atingere data de utilizator, iar variabila de ieșire este variabila de ieșire a comenzii utilizatorului și / sau comanda trimisa la sistem. Al doilea submodul cuprinde 2 componente: achiziția și transformarea datelor. Variabila de intrare este comanda cu gesturi fără atingeri a utilizatorului și cea de ieșire este o comanda trimisa către sistem sau coordonatele 2D ale mișcării utilizatorului, reproduse pe ecranul sistemului în timpul sesiunilor de exerciții fizice.

Modulul de emoții are 2 componente: achiziția datelor și transformarea datelor. Are doua moduri de operare: online (direct) și offline (indirect). Modulul are ca variabila de intrare emoția utilizatorului și ca rezultat starea emoțională a utilizatorului, care se trimite către sistem.

Capitolul 6

Implementarea Interfeței

Capitolul prezinta aspecte generale privind implementarea interfeței multimodale.

Pentru a asigura o funcționare adecvata a interfeței pe diverse dispozitive și platforme, interfața a fost implementata cu limbajele HTML5, CSS3 și JavaScript. Totuși, unele module ale interfeței pot fi implementate folosind alte limbaje. Folosind Bootstrap, GUI se adaptează la dimensiunea ecranului dispozitivului. Sistemul de tip Grid al Bootstrap ajuta la alinierea elementelor diferite ale paginii pe baza rândurilor și coloanelor. Pagina este compusa din unul sau mai multe containere. Fiecare container cuprinde unul sau mai multe rânduri, un rând putând fi descompus în maxim la 12 coloane. Datorita variabilei booleene globale asociate fiecărui modul al sistemului, interfața verifica configurația sistemelor și daca un modul al sistemului nu este disponibil sau este inactiv, informațiile legate de modul vor fi ascunse în GUI și comenzile asociate modulului vor fi dezactivate.

Interfața extrage din fiecare profil al utilizatorului informația legata de preferințele GUI ale utilizatorului și adaptează GUI pentru a le satisface. Pentru a implementa aceasta caracteristica, fiecare grup de elemente are o culoare proprie asociata unei variabile. Valoarea variabilei se extrage din culoarea aleasa de utilizator pentru grupul memorat în profil. Fiecare poziție a modulului depinde de o variabila. Valoarea variabilei se extrage din ordinea modulelor alese de utilizator și memorate în profilul sau. Pagina web a GUI este prezentata în figura 6.1. Aceasta prezinta rezultatele ultimelor analize ale stării de sănătate, numărul zilnic de pași și durata somnului în acea zi. De asemenea, aceasta afișează și numărul de memento-uri, întâlnirile zilnice și unele notificări inteligente privind starea sănătății și activitatea. Toata comunicarea cu sistemul are loc pe baza JSON Web Tokens. Interfața extrage din fiecare profil al utilizatorului

informații despre înălțimea sa, gama acceptabila pentru fiecare valoare a indicilor sănătății, numărul de pași necesari zilnic și durata de somn necesara zilnica. Sistemul adaptează notificările conform informațiilor extrase pentru a se adapta la fiecare cerința privind sănătatea utilizatorului.

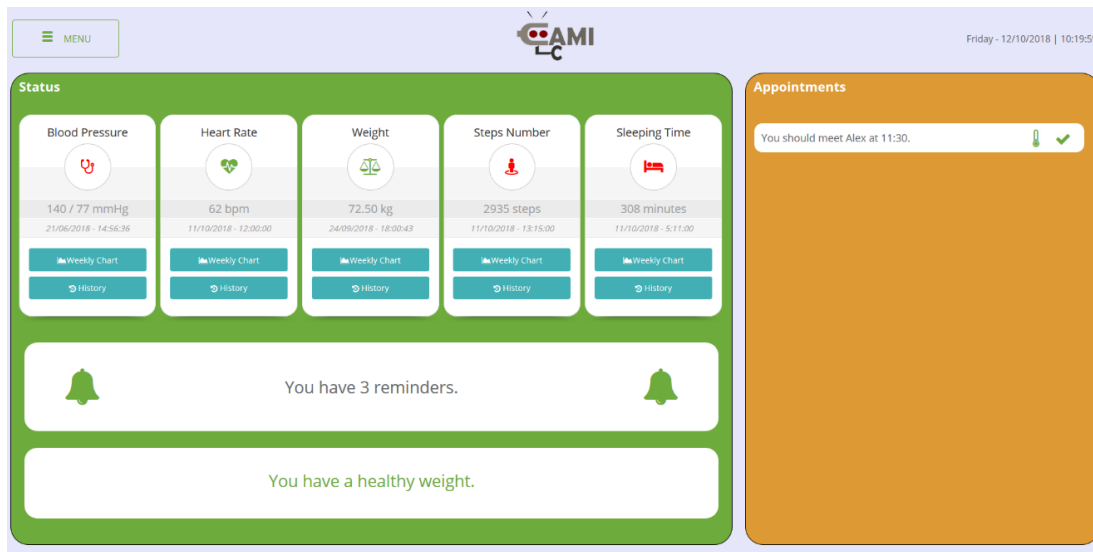


Figura 6.1. Pagina web principala a CAMI.

În ceea ce privește modulul de voce, componenta de preprocesare audio extrage datele utile modulului din profilul utilizatorului și din setările sistemului (precum limba, adâncimea în biți audio, frecvența audio adecvata, soluția TTS, soluția ASR etc.). Pentru modul online/direct de operare în cazul componentei de recunoaștere automată a vorbirii sunt folosite serviciile vorbire-text Microsoft sau Google pentru engleza și franceza (utilizatorul poate alege între ele). Pentru limba romana, se folosește serviciul Google vorbire-text. Sistemul captează comanda vocală a utilizatorului și trimite fișierul audio în cloud, la serviciul folosit, ambele servicii returnând traducerea textului în timp real. Pentru modul offline/indirect de operare, numărul de comenzi recunoscute este mai mic. Se folosește tehnologia Microsoft Windows Desktop Speech. Pentru modul direct/online de operare al componentei de înțelegere naturală a limbii, se folosește serviciul inteligent de înțelegere a limbii (LUIS) al Microsoft, pentru franceza și engleza, care permite sistemului să înțeleagă ce vor utilizatorii când se exprimă oral. Prin trimiterea textului rezultat de la componenta ASR la LUIS, sistemul primește informații detaliate și relevante despre cerințele utilizatorului într-un fișier JSON. Pentru limba romana, se folosește wit.ai API. Fluxul de operare și datele de ieșire sunt similare celor LUIS. Pentru modul indirect/offline de operare, se folosește RASA NLU, furnizat de Rasa Technologies GmbH. Pentru modul direct/online de operare al componentei text – vorbire, se folosesc serviciile Google și Microsoft de tip text-vorbire în engleza și franceza (utilizatorul le poate alterna). Pentru limba romana, se folosește ResponsiveVoice.JS API. Pentru modul indirect/offline de operare, se folosește tehnologia Microsoft Windows Desktop Speech.

În ceea ce privește submodul-ul de gesturi pe baza de atingeri, acesta suportă gesturi cu una sau mai multe atingeri. Acest submodul colectează datele introduse de utilizator prin ecranul tactil al dispozitivului.

În ceea ce privește submodul-ul de gesturi pe baza de atingeri, acesta este compatibil cu gesturile cu o singura atingere sau mai multe atingeri. Colectează datele de intrare ale utilizatorului prin ecranul tactil al dispozitivului de pe care se accesează interfață. Ecranul tactil este componenta de achiziție a datelor. Datele colectate sunt transmise la componenta de procesare a datelor, unde sunt procesate de către sistemul de evenimente al dispozitivului. Unele comenzi sunt executate direct de dispozitiv, altele fiind trimise la nucleul sistemului, unde sunt executate. Pentru a implementa acest modul, se folosește Hammer.js, o bibliotecă JavaScript, care permite crearea de gesturi pe baza de atingeri (doar pentru cele cu atingeri multiple).

În ceea ce privește submodul-ul cu gesturi fără atingeri, componenta de achiziție a datelor urmărește mișcările corpului utilizatorului prin senzorul Kinect V2 și primește de la acesta coordonatele mișcărilor corpului (coordonatele 3D), trimise la componenta de transformare a datelor în care sunt convertite în coordonate 2D folosind transformări afine și de vizualizare clasice. Coordonatele 2D sunt trimise în nucleul sistemului.

În ceea ce privește modul indirect de operare al modulului de emoții, componenta de achiziție a datelor urmărește chipul utilizatorului prin senzorul Kinect V2 de la care primește coordonatele 3D ale chipului. Coordonatele sunt trimise la componenta de transformare a datelor, unde sunt convertite în coordonate 2D folosind transformările afine și de vizualizare clasice. Coordonatele 2D sunt trimise la sistem.

În cazul operării directe/online a modulului de emoții, componenta de achiziție a datelor urmărește chipul utilizatorului cu o camera de la care primește un clip video, trimis apoi la componenta de transformare a datelor în care sunt extrase cadrele video, care sunt la rândul lor trimise la Microsoft Azure Face API unde sunt analizate, iar cloud trimite înapoi un fișier JSON cu rezultatul analizei.

Capitolul 6 prezintă detalii despre implementarea altor caracteristici și aspecte precum imaginea de profil, portabilitatea interfeței, securitatea și confidențialitatea, comunicarea cu aplicațiile terțe, fuziunea și fisiunea datelor, cazurile comenzilor simultane contradictorii.

Capitolul 7

Experimente și evaluare

Capitolul prezintă o prezentare generală a implementării interfeței multimodale.

Cele 2 capitole anterioare au prezentat proiectarea și implementarea interfeței multimodale. S-au testat diversele module ale interfeței în diferitele etape pentru a îmbunătăți soluția finală. Rezultatele testelor din etapele anterioare ale implementării sunt prezentate în

studiile [61, 62, 63, 64]. Modulele au fost testate în laborator, dar și la domiciliile utilizatorilor. La sfârșit, întreaga interfață a fost testată ca o soluție unitară.

Procesul de evaluare a determinat asigurarea unor funcționalități adecvate pentru diverse caracteristici și module ale interfeței integrate într-o singură soluție, dar și că o soluție autonomă pentru fiecare modul și caracteristica. Procesul de evaluare ia în considerare principalii utilizatori ai interfeței, adică persoanele vârstnice.

Rezultatele testelor fiecărui modul și caracteristici făcute în laborator și la domiciliul utilizatorilor sunt satisfăcătoare. Rezultatele testului care iau în considerare întreaga interfață ca o soluție unică sunt, de asemenea, satisfăcătoare. Detaliile testelor sunt prezentate în capitolul 7.

Capitolul 8

Concluzii

Capitolul 8 prezintă concluziile studiului din teza.

Studiul din teza a condus la crearea unei interfețe multimodale care poate fi integrată, în principal, în sistemele AAL, dar și pentru alte sisteme. Interfața multimodală integrează o interfață grafică cu utilizatorul, un modul de voce, unul de emoții și unul de gesturi cu sau fără atingeri. Interfața multimodală este în mai multe limbi, operează independent pe orice dispozitiv sau sistem de operare și integrează caracteristici individualizabile și capabilități adaptive automate.

Mai mult, teza prezintă o privire de ansamblu asupra domeniului inteligenței ambientale, asupra domeniului AAL precum și asupra proiectului CAMI. Se prezintă aspectele de ultimă oră privind interacțiunea om-mașină, respectiv privind domeniul AAL.

În cele ce urmează, sunt prezentate contribuțiile științifice la domeniul menționat mai sus:

- S-a creat o interfață multimodală pentru un sistem tip AAL care integrează 4 module și operează pe diverse platforme și dispozitive. Sistemele AAL integrează o interfață multimodală care, la rândul ei, integrează două module sau o interfață tradițională.
- S-a creat posibilitatea ca utilizatorul să poată modifica modul de operare pentru modulele de vorbire și emoție între direct/online și indirect/offline. Aceste module au un singur mod de operare în diverse sisteme.
- S-a creat posibilitatea ca utilizatorul să poată alege ce soluții de recunoaștere a vocii și sinteza a vocii dorește să folosească în modul de operare online/direct pentru modulul de vorbire. Pe alte sisteme este disponibilă doar o alegere implicită.

- S-au permis interacțiuni de vorbire cu modul de tip AAL în limba română. Nu exista alt astfel de sistem care să poată avea interacțiunile de vorbire în limba română.
- S-a creat o gamă largă de caracteristici adaptive pentru interfață (capabilități automate adaptive și caracteristici individualizabile). De obicei, sistemele integrează capabilități adaptive limitate.
- S-a creat un modul de recunoaștere a emoțiilor care face recomandări privind exercițiile fizice pe care un utilizator ar trebui să le practice bazându-se pe starea emoțională a utilizatorului în timpul fiecărei practici anterioare.
- S-a creat o interfață portabilă integrată atât pentru sistemele AAL, cât și pentru alte tipuri de sisteme. Interfețele multimodale de obicei nu sunt portabile.
- S-a creat posibilitatea ca interfața să poată comunica automat preferințele utilizatorului către orice aplicație terță care poate fi individualizată conform preferințelor utilizatorului.
- S-a testat extensiv diferitele module și caracteristici ale interfeței ca soluții autonome, dar și integrate împreună într-o singură soluție cu un număr important de utilizatori.
- S-a creat un set de reguli privind comenzile contradictorii primite simultan. S-a avut în vedere că setul de reguli să asigure neexecutarea vre-unei acțiuni ca urmarea a unor comenzi contradictorii cu orice parametru reglat de îngrijitor, în pofida gamei largi de capabilități adaptive automate și de caracteristici individualizabile.

Se vor proiecta și alte teme în viitor :

- La început, implementarea curentă a modului indirect/offline de operare a modului de emoții și a submodulului de gesturi fără atingeri depinde de prezenta senzorului Microsoft Kinect V2, astfel că prima temă este modificarea implementării modului și submodulului anterior astfel încât să poată fi eliminată dependența de Microsoft Kinect V2.
- O altă temă planificată care urmărește îmbunătățirea implementării modului de emoții este de a adăuga capacitatea de a recunoaște emoția din vocea utilizatorului. Implementarea extrage în afara de caracteristicile expresiei faciale a utilizatorului și pe cele ale vocii acestuia pentru a îmbunătăți detectarea emoției, dar și bazarea doar pe vocea utilizatorului pentru a detecta emoția acestuia când sistemul nu poate identifica chipul utilizatorului.
- O altă temă recomandată este compusă din 3 etape și urmărește modul indirect/offline de operare a modului de vorbire: în prima etapă, studiul dorește creșterea performanței recunoașterii automate a vocii și componentele text-vorbire, în a doua etapă se dorește extinderea numărului de comenzi recunoscute (pentru a

adapta pe cele recunoscute în modul direct/online de lucru), iar în a treia etapa se adaugă suportul tehnic în limba romana (si poate și în alte limbi).

- O altă temă planificată este îmbunătățirea implementării motorului de fuziune prin luarea în considerare a informațiilor contextuale care pot fi extrase din mediul înconjurător, precum și a beneficia mai mult de modulul de emoție.
- O temă recomandată este identificarea automata a utilizatorilor prin vocile și / sau chipurile lor în cazul sistemelor de tip AAL sau a altor sisteme folosite de mai mulți utilizatori. Odată ce utilizatorul este identificat, interfața are acces la informația legată de acesta și se va adapta conform preferințelor utilizatorului identificat.

Studiile efectuate în teza au stat la baza publicării a 9 lucrări de cercetare, din care 7 în conferințe indexate WoS și una într-o conferință în curs de indexare WoS. O lucrare de cercetare a fost acceptată a fi publicată în următoarele luni ca un capitol dintr-o carte în seria “Recent Advances in Intelligent Assistive Technologies: Paradigms and Applications”. În plus, studiile care au stat la baza acestei teze au fost prezentate la diverse evenimente științifice. În timpul lucrării, am participat la diverse proiecte de cercetare.

Bibliografie

1. United Nations World Population Ageing 2017 Report, New York: Department of Economic and Social Affairs, Population Division, 2017, available on the link: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf
2. J. Bohn, V. Coroamă, M. Langheinrich, F. Mattern and M. Rohs, “Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing”, in: Ambient Intelligence, Springer, Berlin, Germany, pp. 5-29, DOI: 10.1007/3-540-27139-2_2, 2005.
3. A.S. Crandall and D.J. Cook, “Current State of the Art of Smart Environments and Labs from an Ambient Assisted Living Point of View”, in: F. Florez-Revuelta & A.A. Chaaoui (Eds.) Active and Assisted Living: Technologies and Applications, pp. 11-28, DOI: 10.1049/pbhe006e_ch2, August 2016.
4. M. Mohri, F. Pereira and M. Riley, “Speech Recognition with Weighted Finite-State Transducer”, in: Computer Speech and Language Journal, Vol. 16, Iss. 1, pp. 69-88, DOI: 10.1006/csla.2001.0184, 2002.
5. L. Deng, J. Li, J.T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong and A. Acero, “Recent Advances in Deep Learning for Speech Research at Microsoft”, in: Proceedings of the 38th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, pp. 8604-8608, DOI: 10.1109/ICASSP.2013.6639345, IEEE, 2013.
6. Li D. and X. Li, “Machine Learning Paradigms for Speech Recognition: An Overview”, in: Transactions on Audio, Speech and Language Processing Journal, Vol. 21, Iss. 5, pp. 1060-1089, DOI: 10.1109/TASL.2013.2244083, IEEE, 2013.
7. S. Toshniwal, T.N. Sainath, R.J. Weiss, B. Li, P. Moreno, E. Weinstein and K. Rao, “Multilingual Speech Recognition with a Single End-to-End Model”, in: Proceedings of the 43rd International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, pp. 4904 - 4908, DOI: 10.1109/ICASSP.2018.8461972, IEEE, 2018.
8. D.W. Griffin and J.S. Lim, “Signal Estimation from Modified Short-Time Fourier Transform”, in: Proceedings of the 8th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Boston, United States of America, pp. 804+807, DOI: 10.1109/ICASSP.1983.1172092, 1983.
9. A.W. Black, “Multilingual Speech Synthesis” in: T. Schultz & K. Kirchhoff (Eds.) Multilingual Speech Processing, pp. 207-231, DOI: 10.1016/B978-012088501-5/50010-X, ISBN: 978-0-12-088501-5, 2006.

10. H. Zen, K. Tokuda and A.W. Black, "Statistical Parametric Speech Synthesis", in: *Speech Communication Journal*, Vol. 51, Iss. 11, pp. 1039-1064, DOI: 10.1016/j.specom.2009.04.004, 2009.
11. P.F. Brown, P.V. DeSouza, R.L. Mercer, V.J. Della Pietra and J.C. Lai "Class-Based n-Gram Models of Natural Language", in: *Computational Linguistics Journal*, Vol. 18, Iss. 4, pp. 467-479, Cambridge, United States of America, 1992.
12. S.F. Chen and J. Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling", in: *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, Santa Cruz, United States of America, pp. 310-318, DOI: 10.3115/981863.981904, 1996.
13. B. Roark, M. Saraclar and M. Collins, "Discriminative n-Gram Language Modeling", in: *Computer Speech & Language Journal*, Vol. 21, Iss. 2, pp. 373-392, DOI: 10.1016/j.csl.2006.06.006, 2007.
14. J. Peckham, "A New Generation of Spoken Dialogue Systems: Results and Lessons from the SUNDIAL Project", in: *Proceedings of the 3rd European Conference on Speech, Communication and Technology (Eurospeech)*, Berlin, Germany, pp. 33-40, 1993.
15. L. Lamel, S. Rosset, J.L. Gauvain and S. Bennacef, "The LIMSI ARISE System for Train Travel Information", in: *Proceedings of the 24th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Washington, United States of America, pp. 501-504, IEEE, 1999.
16. J. Chu-Carroll, "Form-based Reasoning for Mixed-Initiative Dialogue Management in Information-Query Systems", in: *Proceedings of the 6th European Conference on Speech, Communication and Technology (Eurospeech)*, Budapest, Hungary, pp. 1519-1522, 1999.
17. S. Seneff and J. Polifroni, "Dialogue Management in the Mercury Flight Reservation System", in: *Proceedings of the 2000 ANLP/NAACL Workshop on Conversational Systems*, Stroudsburg, United States of America, pp. 11-16, 2000.
18. V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T.J. Hazen and L. Hetherington, "Jupiter: A Telephone-Based Conversational Interface for Weather Information", in: *Transactions on Speech and Audio Processing Journal*, Vol. 8, pp. 85-96, IEEE, 2000.
19. O. Lemon and X. Liu, "Dude: A Dialogue and Understanding Development Environment, Mapping Business Process Models to Information State Update Dialogue Systems", in: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, Stroudsburg, United States of America, pp. 99-102, 2006.

20. S. Vargas, G. Riccardi and S. Quarteroni, “Persistent Information State in a Data-Centric Architecture”, in: Proceedings of the 9th SIGDial Workshop on Discourse and Dialogue, Stroudsburg, United States of America, pp.68-71, 2008.
21. A. Mertens, N. Jochems, C.M. Schlick, D. Dünnebacke and J.H. Dornberg, “Design Pattern TRABING: Touchscreen-Based Input Technique for People Affected by Intention Tremor”, in: Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS), Berlin, Germany, pp. 267-272, DOI: 10.1145/1822018.1822060, ACM, 2010.
22. A.M. Piper, R. Campbell and J.D. Hollan, “Exploring the Accessibility and Appeal of Surface Computing for Older Adult Health Care Support”, in: Proceedings of the 28th Conference on Human Factors in Computing Systems (CHI), Atlanta, United States of America, pp. 907-916, DOI: 10.1145/1753326.1753461, ACM, 2010.
23. C. Wacharamanotham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick and J. Borchers, “Evaluating Swabbing: A Touchscreen Input Method for Elderly Users with Tremor”, in: Proceedings of the 29th Conference on Human Factors in Computing Systems (CHI), Vancouver, Canada, pp. 623-626, DOI: 10.1145/1978942.1979031, ACM, 2011.
24. M. Kobayashi, A. Hiyama, T. Miura, C. Asakawa, M. Hirose and T. Ifukube, “Elderly User Evaluation of Mobile Touchscreen Interactions”, in: P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque & M. Winckler (Eds.) Human-Computer Interaction - INTERACT 2011, Lecture Notes in Computer Science, Vol. 6946, Springer, Berlin, Heidelberg, pp. 83-99, DOI: 10.1007/978-3-642-23774-4_9, 2011.
25. L. Findlater, J.E. Froehlich, K. Fattal, J.O. Wobbrock and T. Dastyar, “Age-Related Differences in Performance with Touchscreens Compared to Traditional Mouse Input”, in: Proceedings of the 31st Conference on Human Factors in Computing Systems (CHI), Paris, France, pp. 343-346, DOI: 10.1145/2470654.2470703, ACM, 2013.
26. C. Stöbel, H. Wandke and L. Blessing, “Gestural Interfaces for Elderly Users: Help or Hindrance?”, in: S. Kopp & I. Wachsmuth (Eds.), Gesture in Embodied Communication and Human-Computer Interaction - GW 2009, Lecture Notes in Computer Science, Vol. 5934. Springer, Berlin, Heidelberg, pp. 269-280, DOI: 10.1007/978-3-642-12553-9_24, 2010.
27. T. Apted, J. Kay and A. Quigley, “Tabletop sharing of digital photographs for the elderly”, in: Proceedings of the 24th Conference on Human Factors in Computing Systems (CHI), Quebec, Canada, pp. 781-790, DOI: 10.1145/1124772.1124887, ACM, 2006.

28. D. Facal, C. Buiza, M.F. González, J. Soldatos, T. Petsatodis, F. Talantzis, E. Urdaneta, V. Martínez and J.J. Yanguas, “Cognitive Games for Healthy Elderly People in a Multitouch Screen”, Proceedings of the 2009 International Congress on Digital Homes, Robotics and Telecare for All (IDRT4ALL), Barcelona, Spain, pp. 91-97, 2009.
29. N. Hollinworth and F. Hwang, “Investigating Familiar Interactions to Help Older Adults Learn Computer Applications More Easily”, in: Proceedings of the 25th British Computer Society Conference on Human Computer Interaction (BCS-HCI), Swinton, United Kingdom, pp. 473-478, 2011.
30. A. Kendin, “Some Relationships Between Body Motion and Speech” in: A. Seigman & B. Pope (Eds.) Studies in Dyadic Communication, pp. 177-216, 1972.
31. A. Kendin, “Gesture and Speech: Two Aspects of the Process of Utterance”, In: M.R. Key (Eds.) Nonverbal Communication and Language, The Hague: Mouton, pp. 207-227, 1980.
32. M. Kipp, “Gesture Generation by Imitation: From Human Behavior to Computer Character Animation”, Dissertation.com, Boca Raton, United States of America, ISBN: 1-58112-255-1, 2004.
33. M.E. Sargin, Y. Yemez, E. Erzin and A.M. Tekalp, “Analysis of Head Gesture and Prosody Patterns for Prosody-Driven Head-Gesture Animation”, in: Transactions on Pattern Analysis and Machine Intelligence Journal, Vol. 30, Iss. 8, pp. 1330-1345, DOI: 10.1109/TPAMI.2007.70797, IEEE, 2008.
34. G. Galanakis, P. Katsifarakis, X. Zabulis and I. Adami, “Recognition of Simple Head Gestures Based on Head Pose Estimation Analysis”, in: Proceedings of the 4th International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT), Rome, Italy, pp. 88-96, ISBN: 978-1-61208-356-8, 2014.
35. C. Lelardeux, D. Panzoli, J. Alvarez and P. Lagarrigue, “Serious Game, Simulateur, Serious Play: État de l'Art pour la Formation en Santé”, in: Proceedings of the 1st SeGaMed Conference, Nice, France, pp. 27-38, 2012.
36. I. Di Loreto, B. Lange, A. Seilles, S. Andary and W. Dyce, “Game Design for All: The Example of Hammer and Planks”, in: Proceedings of the 4th International Conference on Serious Games Development and Applications (SGDA), Trondheim, Norway, pp. 70-75, DOI: doi.org/10.1007/978-3-642-40790-1_7, 2013.
37. G.R. Vineetha, C. Sreeji and J. Lentin, “Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network”, in: Online Proceedings on Trends in Innovative Computing 2012 - Intelligent Systems Design, pp. 176-180, ISSN: 2150-7996, 2012.

38. A.E. Youssef, S.F. Aly, A.S. Ibrahim and A.L. Abbott, "Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid", in: *International Journal of Modeling and Optimization*, Vol. 3, Iss. 2, pp. 112-115, 2013.
39. P. Lemaire, L. Chen, M. Ardabilian and M. Daoudi, "Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients", in: *Proceedings of the 10th International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, China, pp. 1-7, DOI: 10.1109/FG.2013.6553821, IEEE, 2013.
40. F. Loi, J.G. Vaidya and S. Paradiso, "Recognition of Emotion from Body Language Among Patients with Unipolar Depression", in: *Psychiatry Research Journal*, Vol. 209, Iss. 1, pp. 40-49, DOI: 10.1016/j.psychres.2013.03.001, 2013.
41. L. Kiforenko and D. Kraft, "Emotion Recognition through Body Language using RGB-D Sensor", in: *Proceedings of the 11th International Conference on Computer Vision Theory and Applications (VISAPP)*, Vienna, Austria, pp. 398-405, DOI: 10.5220/0005783403980405, 2016
42. S. Yacoub, S. Simske, X. Lin and J. Burns, "Recognition of emotions in interactive voice response systems", in: *Proceedings of the 8th European conference on speech communication and technology (EUROSPEECH)*, Geneva, Switzerland, pp. 729-732, September 2003.
43. L. McCauley and S. D'Mello, "MIKI: A Speech Enabled Intelligent Kiosk", in: J. Gratch, M. Young, R. Aylett, D. Ballin & P. Olivier (Eds.) *Intelligent Virtual Agents (IVA)*, *Lecture Notes in Computer Science*, Vol. 4133, Springer, pp. 132-144, 2006.
44. M. Johnston and S. Bangalore, "MATCHKiosk: A Multimodal Interactive City Guide", in: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL) Interactive Poster and Demonstration Session*, Barcelona, Spain, pp. 222-225, DOI: 10.3115/1219044.1219077, 2004.
45. K. Saroha, S. Sharma and G. Bhatia, "Human Computer Interaction: An intellectual approach", in: *International Journal of Computer Science and Management Studies (IJCSMS)*, Vol. 11, pp. 147-154, 2011.
46. R. Raisamo, "A Multimodal User Interface for Public Information Kiosks", in: *Proceedings of the 2nd Perceptual User Interfaces (PUI) Workshop*, San Francisco, United States of America, pp. 7-12, 1998.
47. R.A. Bolt, "Put-that-there: Voice and gesture at the graphics interface", in: *ACM SIGGRAPH Computer Graphics Journal*, Vol. 14, Iss. 3, pp. 262-270, ACM, DOI: 10.1145/965105.80750, 1980.

48. P. Biswas and P. Langdon, “A new input system for disabled users involving eye gaze tracker and scanning interface”, in: *Journal of Assistive Technologies*, Vol. 5, Iss. 2, pp. 58-66, DOI: 10.1108/17549451111149269, June 2011.
49. G. Lamprinakos, E. Kosmatos, D. Kaklamani and I. Venieris, “An Integrated Architecture for Remote Healthcare Monitoring”, in: *Proceedings of the 14th Panhellenic Conference on Informatics*, Tripoli, Greece, pp. 12-15, DOI: 10.1109/PCI.2010.20, September 2010.
50. M.R. Tazari, F. Furfari, J.P.L. Ramos and E. Ferro, “The persona service platform for AAL spaces”, in: *Handbook of Ambient Intelligence and Smart Environments*, Springer, pp. 1171–1199, DOI: 10.1007/978-0-387-93808-0_43, 2010.
51. Active and Assisted Living Program: NITICS, retrieved from the project page on the Official AAL Website | URL: <http://www.aal-europe.eu/projects/nitics/>, accessed on: 2018-08-21.
52. Retrieved from the SOCIABLE project page on the JoinUp platform | URL: <https://joinup.ec.europa.eu/document/sociable-project-sociable>, accessed on: 2018-08-21.
53. C. Azevedo, C. Chesta, J. Coelho, D. Dimola, C. Duarte, M. Manca, J. Nordvik, F. Paterno, A. Sanders and C. Santoro, “Towards a Platform for Persuading Older Adults to Adopt Healthy Behaviors”, in: R. Orji, M. Reisinger, M. Busch, A. Dijkstra, M. Kaptein & E. Mattheiss (Eds.): *Proceedings of the 2nd International Workshop on Personalization in Persuasive Technology*, pp. 50-56, Amsterdam, Netherlands, 2017.
54. CareWell Official Website, retrieved from the URL: www.carewell-project.eu, accessed on: 2018-08-21.
55. Active and Assisted Living Program: healthy@work, retrieved from the project page on the Official AAL Website | URL: <http://www.aal-europe.eu/projects/healthywork/>, accessed on: 2018-08-21.
56. R. Planinc, M. Hödlmoser and M. Kampel, “Enhancing the Wellbeing at the Workplace”, in: *Proceedings of the 7th International Conference on eHealth, Telemedicine, and Social Medicine (eTelemed)*, pp. 213-216, Lisbon, Portugal, February 2015.
57. V. Giannoglou, K. Smagas, E. Valari and E. Stylianidis, “Elders-Up! An Adaptive System for Enabling Knowledge Transfer from Senior Adults to Small Companies”, in: *Proceedings of the 22nd International Conference on Virtual System and Multimedia (VSMM)*, Kuala Lumpur, Malaysia, pp. 17-23, DOI: 10.1109/VSMM.2016.7863163, October 2016.

58. Active and Assisted Living Program: CAMI (AAL-2014-1-087) Official Website, retrieved from the URL: www.camiproject.eu, accessed on: 2018-08-22.
59. I.A. Awada, I. Mocanu, A.M. Florea and B. Cramariuc, “Multimodal Interface for Elderly People”, in: Proceedings of the 21st International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, pp. 536-541, DOI: 10.1109/CSCS.2017.82, IEEE, 2017.
60. C. Dood, R. Athauda and M.T.P. Adam, “Designing User Interfaces for the Elderly: A systematic Literature Review”, in: Proceedings of the 2017 Australasian Conference on Information Systems (ACIS), Hobart, Australia, December 2017.
61. I.A. Awada, O. Cramariuc, I. Mocanu, C. Seceleanu, A. Kunnappilly and A.M. Florea, “An End-User Perspective on the CAMI Ambient and Assisted Living Project”, in: Proceedings of the 12th Annual International Technology, Education and Development Conference (INTED), Valencia, Spain, pp. 6776-6785, DOI: 10.21125/INTED.2018.1596, ISSN: 2340-1079, 2018.
62. I.A. Awada, I. Mocanu and A.M. Florea, “Exploiting Multimodal Interfaces în eLearning Systems”, in: Proceedings of the 14th eLearning & Software for Education Conference (eLSE), Bucharest, Romania, Vol. 2, pp. 174-181, DOI: 10.12753/2066-026X-18-094, 2018.
63. I.A. Awada, I. Mocanu, S. Jecan, L. Rusu, A.M. Florea, O. Cramariuc and B. Cramariuc, “Mobile@Old - An Assistive Platform for Maintaining a Healthy Lifestyle for Elderly People”, in: Proceedings of the 6th International Conference on E-Health and Bioengineering (EHB), Sinaia, Romania, pp. 591-594, DOI: 10.1109/EHB.2017.7995493, ISSN: 2575-5137, IEEE, 2017.
64. I.A. Awada, I. Mocanu, L. Rusu, R. Arba, A.M. Florea and B. Cramariuc, “Enhancing the Physical Activity of Older Adults Based on User Profiles”, in: Proceedings of the 16th RoEduNet Conference: Networking în Education and Research (RoEduNet), Targu Mures, Romania, pp. 120-125, DOI: 10.1109/ROEDUNET.2017.8123749, ISSN: 2068-1038, IEEE, 2017.