

Pregled modalnosti upravljanja elektronskih naprav

Jure Žilavec, Jože Guna in Matevž Pogačnik
Laboratorij za telekomunikacije, Fakulteta za elektrotehniko,
Univerza v Ljubljani, Tržaška cesta 25, 1000 Ljubljana, Slovenija
E-pošta: jure.zilavec@lfe.org

Review of interaction techniques for electronic devices

This paper is concerned with a review of interaction techniques used in the field of electronic device usage. The overview includes standard interaction, voice and gesture control, sight control and mind control. Different interaction techniques and technologies described here can be used for human-computer interactions in everyday life, entertainment, business applications as well as for people with special needs and disabilities.

Keywords: Human Computer Interaction, Multimodal Interactions, Speech, Gesture Recognition, Eye tracking

1 Uvod

Živimo v času, ko so internet in drugi elektronski mediji vse bolj pomembni za komunikacijo tako med ljudmi, kakor tudi v okviru uporabe naprednih multimedijskih storitev. Za uporabo računalnika, ali katerekoli druge naprave (npr. televizija, različni terminali, bankomat,...) potrebujemo uporabniški vmesnik. Poleg same grafične predstavitve uporabniškega vmesnika na zaslonu naprave, potrebujemo tudi eno ali več vhodno-izhodnih naprav, ki omogočajo interakcijo. Ključna naloga vhodnih naprav je prenesti informacijo od uporabnika do ciljne terminalne opreme in aplikacije ki jo uporabljamo.

Kljub temu, da na trgu znova in znova pojavljajo nove naprave, nove spletne strani in aplikacije, kakor tudi novi uporabniški vmesniki, pa je vse premalo pozornosti posvečene načinu komunikacije med človekom in napravo. Le tega želimo izboljšati in ga še bolj prilagoditi človeku, ga narediti intuitivnega in transparentnega ter s tem doseči lažjo komunikacijo in predvsem boljšo uporabniško izkušnjo za vse uporabnike. To še posebej velja za starejše in ljudi s posebnimi potrebami, ki so vse bolj prikrajšani v pridobivanju pomembnih informacij. Področje, ki se ukvarja z raziskovanjem načinov komunikacije se imenuje HCI (ang. Human-Computer Interaction).

Ko oblikujemo nove vmesnike in načine interakcije moramo biti pozorni na to, da čimbolj izkoristimo dane možnosti in da poiščemo najhitrejšo in najbolj naravno pot od uporabnika do naprave, pri čemer seveda upoštevamo kakšna vhodna naprava je na voljo v ciljnim sistemu. Izbor primerne načina interakcije oz. vhodne naprave je odvisen od tega, v kakšnem okolju se

bo naprava uporabljala in njen namen (igranje iger, vodenje robota, dostop in upravljanje multimedijskih vsebin, itd.). Kako uspešni smo pri naši izbiri nam pove uporabniška izkušnja in učinkovitost uporabnikov. Učinkovitost uporabnikov določa Fittov zakon (ang. Fitts' Law) [1], ki predvidi čas potreben za izvršitev določene akcije (recimo klik ikone).

V nadaljevanju si bomo pogledali različne tipe modalnosti oz. načinov interakcije ter naprave, ki se pri tem uporabljajo. Več se bomo posvetili alternativnim načinom, kot so sledenje očem, prepoznavanje govora in kretenj. Omenili bomo tudi klasične načine oz. tipe kot so tipkovnica, miška, daljinski upravljalnik in druge. Na kocu bomo omenili še upravljanje z mislimi in si pogledali primer uspešne kombinacije različnih modalnosti.

2 Klasične naprave, tipi

Tipična naprava, preko katere komuniciramo z računalnikom (in tudi drugimi napravami) je tipkovnica, ki se že od samega začetka do danes ni prav veliko spremenila. Interpretacija pritiska na tipkovnici je ponavadi prepuščena programski opremi. Računalnik razloči vsako tipko posebej in zazna vsak pritisk na tipkovnici ter to poroča tudi programskemu paketu, ki je namenjen interpretaciji.

Še ena, vsem zelo dobro poznana naprava je miška, ki je skozi leta postala nepogrešljiva za delo z računalniškimi sistemi. Po načinu delovanja ločimo tri glavne skupine mišk - mehanske, optomehanske in optične. Mehanske in optomehanske miške imajo v svoji notranjosti gumijasto ali jekleno, z gumo prevlečeno kroglico. Ker se kroglice in valjčki znotraj take miške mažejo, so izdelali optične miške, ki nimajo nobenih premičnih delov.

Daljinski upravljalniki so naprave, ki jih ravno tako uporabljamo kot eno izmed vhodnih enot. Delujejo z uporabo oddajanja IR (infrardeče) svetlobe ali pa preko RF - radijskega signala [2]. Prevladujoča tehnologija na področju daljinskih upravljalnikov je IR. Daljinski upravljalnik oddaja pulze IR svetlobe, ki predstavljajo določene binarne kode in so direktno povezane z različnimi ukazi. IR sprejemnik, ki je vgrajen v sprejemno napravo, dekodira sprejeti signal v binarne podatke, ki jih mikroprocesor te naprave razume in izvede ustrezen ukaz.

Poleg že omenjenih IR daljinskih upravljalnikov, so dokaj razširjeni tudi RF (radio frekvenčni) upravljalniki. RF upravljalniki oddajajo radijski signal, ki je povezan z določeno binarno kodo nekega ukaza. Sprejemnik na

drugi napravi, ki signal sprejme, ga dekodira in izvede ukaz. Prednost upravljalnikov na osnovi RF signala je domet, saj delujejo tudi že na 35 metrih razdalje, poleg tega pa jih stene in podobne prepreke ne motijo. Obstajajo tudi kombinirani daljinski upravljalniki.

Ekрани občutljivi na dotik so ekрани, ki so sposobni zaznati dotik in locirati pozicijo dotika. V večini primerov govorimo o dotiku z roko oz. prstom ali posebnim, prirejenim pisalom. Ti ekрани nudijo direktno interakcijo z napravo oz. direktno interakcijo z vsebino, ki je prikazana na ekranu. Poznamo več glavnih vrst tehnologij zaznavanja dotika na ekranu, in sicer: **uporovno, SAW** (ang. Surface Acoustic Wave), **kondenzatorsko in infrardečo**.

Uporovni zasloni na dotik so sestavljeni iz dveh fleksibilnih plasti iz uporabnega materiala, ki sta ločeni z zračno plastjo, na dotik se plasti stakneta skupaj in tako zaznamo točno pozicijo dotika [3]. Ti lahko zaznajo dotike praktično vsakega predmeta, imajo visoko ločljivost, dokaj nizko ceno v primerjavi z ostalimi tehnologijami in možnost zaznavanja večih dotikov hkrati.

SAW tehnologija [3], uporablja ultrasonične valove, ki grede skozi ekran, ob dotiku se del teh valov oz. žarkov absorbira in iz te spremembe v valovih sistem določi pozicijo dotika na ekranu.

Kondenzatorska tehnologija [3] zaslonov na dotik uporablja senzor, ki je ponavadi narejen kar iz stekla oblečenega v transparentno prevodno plast (recimo indium tin oxide - ITO). Z dotikom spremenimo kapaciteto kondenzatorja in s tem ugotovimo lokacijo dotika. Tudi ta tehnologija omogoča večtočkovno zaznavanje.

Infrardeča (IR) tehnologija uporablja polje IR svetlobe [4]. Z dotikom zmotimo polje svetlobe kar zaznajo fotosenzorji in preko tega se določi lokacija dotika. Poleg omenjenih obstaja še nekaj drugih tehnologij zaznavanja dotika na ekranu, kot na primer akustično prepoznavanje pulzov, vendar pa niso zelo razširjene.

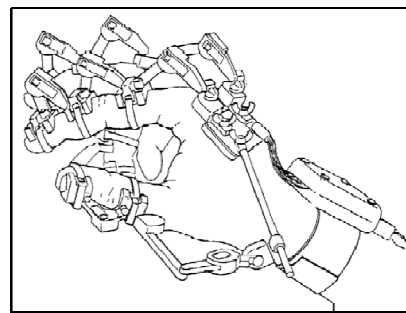
Poleg zgoraj naštetih klasičnih tipov modalnosti poznamo še druge, ki pa jih ne bomo obširneje opisovali. Naj za primer omenimo le igralne palice v računalništvu uporabljene predvsem za igranje iger, sledilne in igralne tablice, grafične tablice in še veliko drugih.

3 Kretnje

Upravljanje naprav s kretnjami je lahko uporabno v okolju, kjer je uporaba tradicionalnih vhodnih naprav (miške, tipkovnice) otežena (npr. zelo prašen prostor, upravljanje gibov robotskih rok pri delu z nevarnimi snovmi, itd.). Prednost uporabe kretenj za upravljanje računalnika se kaže tudi pri gluhih osebah, ki že uporabljajo govorico s kretnjami in lahko na ta način komunicirajo z računalnikom (pisanje). Prepoznavna kretenj se je pojavila tudi v virtualni resničnosti kjer uporabnik ob nošenju čelade za vizualizacijo virtualnega prostora, obleče še interaktivno rokavico, s

katero upravlja virtualni svet. Pokazalo se je, da trenutni sistemi še niso povsem zadovoljivi, saj lahko uporabnik mimogrede in nenamerno izvede določeno kretnjo, ki izvede neko neželeno nalogo, kar bistveno oteži interakcijo. Hkrati pa uporaba takega sistem ni popolnoma intuitivna, saj ljudje nismo vajeni virtualnega "letenja" skozi prostor.

Razlogi za uporabo prepoznavne kretenj [5] za interakcijo so torej: omogoča bolj naraven način komunikacije z računalnikom, dovoljuje da stopnjo natančnosti določi uporabnik, lahko jih uporabimo v okoljih kjer uporaba, recimo tipkovnice, ni mogoča, lahko se uporabi, če oseba ne more ali noče uporabiti klasičnega načina, deluje zelo intuitivno in zahteva le malo predhodnih izkušenj, lahko jih uporabimo tam kjer je tak način bolj primeren od klasičnega, omogoča lažjo komunikacijo z gluhih in omogoča razvoj novih aplikacij, ki prej niso bile možne.



Slika 1: Rokavica podjetja Exos Dexterous

Prepoznavo kretenj lahko izvajamo s pomočjo različnih naprav. Na trgu je prisotno že veliko številno podatkovnih rokavic, ki kretnje pretvarjajo v podatke, razumljive za naprave.

Poleg eksplicitnega zaznavanja je mogoče kretnje zaznavati tudi implicitno, z uporabo 3D kamer (TOF kamera, ang. Time Of Flight) [6]. Na tak način interakcijo približamo naravnemu gibanju človeka, s čimer tudi bistveno olajšamo uporabo in komunikacijo z različnimi napravami. Kamere za zajem 3D slike delujejo na podlagi merjenja časa odboja oddane IR svetlobe od objektov v prostoru, iz česar dobimo podatek o njihovi oddaljenosti od kamere. Na osnovi teh podatkov se nato izvede ujemanje dobljenih podatkov z znanimi podatki 3D modela ter z vnaprej določenim naborom kretenj.

Kot vhodno napravo za prepoznavo kretenj pa lahko uporabimo tudi stereovizijo [7], to je stereokamero oz. dve identični kameri, med katerima poznamo medsebojni odnos (razdaljo), ki nam pri obdelavi podobnih slik pomaga za določitev globine.

4 Prepoznavna govora

Avtomatska (računalniška) prepoznavna govora prevede govorno besedo v take vhodne podatke, ki jih razume računalnik. Največja težava oz. najslabši del v procesu prepoznavne govora je programska oprema. Lahko se

zgodí, da je za pravilno uporabo oz. prepoznavo potreben faktor človeške presoje, ki pa je trenutni avtomatski sistemi še ne premorejo.

Prepoznavo govora lahko delimo na dve kategoriji:

- sistem z majhnim besednim zakladom in velikim številom uporabnikov: taki sistemi so predvsem primerni za avtomatizirane sisteme, recimo za avtomatski telefonski odzivnik, ki naj razume različne naglase in narečja, upravljanje domačega multimedijskega sistema, itd.
- sistem z obsežnim slovarjem in omejenim številom uporabnikov: taki sistemi so najbolj primerni za uporabo v namenskih aplikacijah, ki jih uporablja le majhno število različnih uporabnikov. Tipičen primer je system za avtomatsko pisanje teksta preko diktiranja, itd.

Da iz govora dobimo tekst na ekranu ali pa ukaz, je potrebno, da analogno-digitalni pretvornik zvočne valove prevede v digitalne podatke, ki jih računalnik razume. Sistem zajete podatke filtrira, včasih tudi loči signal v različna frekvenčna področja. Zvok se še normalizira, včasih je potreba poravnava in ker ljudje ne govorimo vedno enako hitro je potrebno prilagoditi še hitrost govora. V naslednji fazi se signal razdeli na več manjših delov in vsak del posebej se primerja z znanimi fonemi v določenem jeziku. Zadnji korak je primerjava določenega fonema s fonemi v sistemski bazi in s fonemi v njegovi bližini. Preko statističnih modelov se zgradi kontekstualna zasnova fonemov, ki se nato primerja z znanimi besedami, frazami in stavki. Učinkovitost oz. uspešnost prepoznave govora se v glavnem meri s stopnjo točnosti prepoznave. Za to se uporablja mera WER (ang. Word Error Rate). [8]

Trenutno sta najbolj razširjena dva modela za prepoznavo govora, in sicer skriti markov model (Hidden Markov Model - HMM) [8] in nevronske mreže. Te metode uporabljajo kompleksne matematične funkcije, v principu pa delujejo tako, da analizirajo informacije oz. podatke, ki so sistemu znani ter iz tega ugotovijo informacije, ki so skrite. HMM je zelo popularen model tudi zato, ker ima sposobnost samoučenja, ker je relativno enostaven za uporabo ter ne zahteva veliko sistemskih resursov. Dostikrat se zgodi, da se za prepoznavo ne uporablja samo HMM ampak kombinacija tega in še katerega modela predvsem zato, da dobimo boljše rezultate.

Noben sistem za prepoznavo govora ne deluje popolnoma zanesljivo. Za slabšo prepoznavo je lahko krivo slabo razmerje signal-šum, prekrivanje govora, slaba procesorska moč računalnika ali pa homonimi (enakozvočnice). Z izboljševanjem algoritmov in naraščanjem procesorske moči naprav pa ima prepoznavo govora nedvomno lepo prihodnost v številnih aplikacijah.

5 Sledenje oči

Sledenje oči je tehnika [9], pri kateri skušamo ugotoviti točko gledanja v nekem trenutku in posledično sprožiti ustrezno akcijo, v primeru da je točka gledanja na interaktivnem elementu grafičnega vmesnika. Sledenje oči na področju HCI (ang. Human Computer

Interaction) nam poleg možnosti upravljanja naprav lahko zelo pripomore k izboljšanju razumevanja vizualne predstavitve različnih sistemov in uporabnosti grafičnih vmesnikov za uporabnike.

Sledenje očem izvajamo z uporabo kamere. Kamera je najprej usmerjena naravnost v oči, da dobimo močne odseve in s tem omogočimo lažje nadaljnje sledenje. Infrardeča kamera je uporabljena zato, da se izognemo premočni svetlobi usmerjeni v oči. Svetloba vstopi v oko kjer se je velik del odbije nazaj s čimer dobimo občutek svetlejšje zenice (poznano tudi kot efekt svetle zenice), ustvari pa se tudi odsev.



Slika 2: Odsev ob pogledu dol in desno od kamere

Ko programska oprema prepozna center zenice in lokacijo odseva, lahko izmeri vektor med tema dvema točkama. Z nadaljnjo trigonometrično obdelavo dobimo točko pogleda (ang. point-of-regard) očesa. Seveda je mogoče približno oceniti point-of-regard očesa že samo iz odseva. Pred uporabo sledenja očesa je potreba natančna kalibracija sledilca na oko določene osebe.

V primeru uporabe sledenja oči z namenom analize grafičnega vmesnika so najpomembnejši pojav, ki se meri fiksacije (ang. fixations) [9] in hitri premiki, ki se zgodijo med fiksacijami (ang. saccades). Opazujemo lahko tudi druge pojave, ki največkrat izhajajo iz strmenja (ang. gaze) in poti preleta (ang. scanpath).

Fiksacija je pojav, ko so oči relativno pri miru. V tem času se ponavadi dogaja absorbiranje informacije. Ti pojavi v povprečju trajajo 218 milisekund, z razponom med 66 in 416 milisekundami. [9]

Med hitrimi premiki med fiksacijami se ne dogaja nobeno sprejemanje informacij. Trajajo tipično neke med 20 in 35 milisekundami. Ti premiki so lahko premiki nazaj na prejšnjo točko, ti pa lahko služijo kot merilo za težavnost procesiranja med sprejemanjem informacij (fiksacijami).

Poti preleta (ang. scanpath) je opis poti očesa oz. celotne sekvence fiksacij in hitrih premikov med njimi. Optimalna pot je direktna, ravna pot do željenega objekta oz. tarče.

Frekvenco mežikanja in velikost zenic lahko uporabimo kot indikator spoznavnega procesa. V primeru, da je mežikanja malo, to kaže na več podatkov, ki jih je potrebno obdelati, hitrejše mežikanje lahko kaže na zasičenost s podatki.

Pri meritvah sledenja oči moramo biti pozorni na to, katere pojave želimo opazovati, saj je treba opremo prilagoditi in optimizirati za merjenje določenega pojava [9]. Seveda lahko opazujemo več pojavov hkrati, vendar pa podatki v tem primeru niso tako točni in zanesljivi, kot bi bili v primeru, da bi se koncentrirali samo na en pojav. Zelo pomemben je čas, ki ga določimo za minimalni čas fiksacije očesa.

Kot smo že omenili se meritve sledenja oči uporabljajo za upravljanje grafičnega vmesnika na ekranu računalnika. Uporabnik lahko postavi kurzor na določeno mesto na ekranu tako, da enostavno pogleda v tisto točko kjer želi imeti postavljen kurzor. Na določeno ikono lahko klikne s strmenjem v ikono ali pa z mežikanjem. S tem lahko tudi invalidom, ki ne morejo uporabljati rok, omogočimo uporabo računalnika.

6 Upravljanje z mislimi

Eden izmed novejših načinov komunikacije med človekom in napravo je upravljanje z mislimi. V resnici gre za direktno komunikacijo možgani-računalnik. Ta način komunikacije deluje tako, da bere nevronske spremembe v možganih (to so alfa, beta in gama valovi), EOG - elektro-okulogramske komponente (pozicijske spremembe med prednjim in zadnjim delom retine očesa, ki se spremeni glede na orientacijo očesa) in EMG - elektromiografijo (merjenje in spremljanje aktivacijskih signalov mišic) [10]. To so tri največje komponente, ki prispevajo k končnemu signalu, ki se opazuje. Signal se potem loči na individualne komponente s pomočjo matematičnih operacij, vključno s hitro fourierovo transformacijo.

Obstaja še drugi način zaznavanja sprememb v možganih, in sicer deluje preko branja EEG (elektroencefalografija) [11] majhnih nevronskega impulzov, ki se generirajo z določenimi aktivnostmi v možganih. Signal, ki ga naprava preko senzorja zazna, se pošlje najprej v obdelavo, kjer se le ta očisti nepotrebnega šuma, ojači in obdelava s pomočjo določenih algoritmov.

Pred uporabo je potrebno napravo še kalibrirati.

Tehnologija ni omejena le na uporabo z osebnimi računalniki, namenjena je tudi uporabi z igračami in podobnimi aplikacijami (nekatera podjetja, ki to tehnologijo uporabljajo so SegaToys, SquareEnix, Nokia, Mattel,...).

Trenutno sta na tem področju prisotna dva proizvajalca, OCZ s produktom NIA (ang. Neural Impulse Actuator) in podjetje NeuroSky s produktom MindSet, ki deluje z enim samim senzorjem.

Največja prednost uporabe takega vmesnika je, da se reakcijski časi uporabnika zmanjšajo za od (po oceni proizvajalcev) cca. 30 do 60 procentov v primerjavi z uporabo miške. Predvsem je namenjena uporabi pri igranju iger, predvsem tistih, kjer je pomemben reakcijski čas igralca (Unreal, Quake, Flight Simulator,...).

7 Primer kombinacije večih modalnosti

Lep primer uporabe kombinacije več načinov interakcije je projekt Natal [12], podjetja Microsoft, ki omogoča igranje in upravljanje brez uporabe fizičnih kontrolerjev. Škatlica senzorjev je velika približno 23 cm in jo namestimo pod ali nad televizor. Vgrajeno ima RGB kamero, senzor za zaznavanje oddaljenosti oz.

globine (deluje na osnovi 3D TOF kamre, ki so jo pri Microsoftu prevzeli od podjetja 3DV Systems) in mikrofona. Kombinacija tega omogoča prepoznavo govora, zaznavanje in sledenje človekovega telesa (ali delov telesa) v prostoru. Tako so dosegli, da naše telo postane velik upravljalnik oz. krmilna enota.

8 Zaključek

Poleg klasičnih načinov interakcije in vhodnih naprav smo spoznali kar nekaj alternativnih načinov komunikacije človek-naprava. Opazimo, da je vse več poudarka na teh načinih komunikacije, saj želimo, da je le ta čim bolj naravna in intuitivna za uporabnika, hkrati pa želimo uporabo različnih naprav (računalnika) omogočiti tako osebam s posebnimi potrebami kot tudi uporabo teh naprav v prostorih, kjer klasični načini (recimo tipkovnica in miška) niso primerni. Veliko inovativnih načinov se pojavlja predvsem za področje uporabe za igranje iger, kjer so udobnost in reakcijski časi še posebej pomembni.

Zahvala

Znanstveno-raziskovalno delo je bilo sofinancirano s strani Ministrstva za visoko šolstvo, znanost in tehnologijo Republike Slovenije.

Literatura

- [1] Fitts's UI Law Applied to the Web, <http://msdn.microsoft.com/en-us/library/ms993291.aspx>
- [2] RemoteControl, http://en.wikipedia.org/wiki/Remote_control
- [3] How do touch-screen monitors know where you're touching?, <http://computer.howstuffworks.com/question716.htm>
- [4] Build Your Own Multitouch Surface Computer, http://www.maximumpc.com/article/features/maximum_pc_builds_a_multitouch_surface_computer
- [5] Bill Buxton, Gesture Based Interaction, Haptic Input, December 14, 2008
- [6] Breuer P., Eckes C., Muller S.: Hand Gesture Recognition with a novel IR Time of Flight Range Camera - A Pilot Study, Universitat Koblenz-Landau, 2007
- [7] What is stereovision?, <http://www.vision3d.com/stereo.html>
- [8] Frederick Jelinek, Statistical Methods for Speech Recognition, The MIT Press, January 16, 1998
- [9] Alex Poole and Linden J. Ball, Eye Tracking in Human-Computer Interaction and Usability Research, 2004, Psychology Department, Lancaster University, UK
- [10] OCZ NIA, http://www.ocztechnology.com/products/ocz_peripherals/nia-neural_impulse_actuator
- [11] Neurosky MindSet, <http://www.neurosky.com/>
- [12] ProjectNatal, http://www.xbox.com/en-US/live/project_natal/