

Pregledni rad

Rukopis primljen 1. 10. 2019.

Prihvaćen za tisak 7. 12. 2020.

<https://doi.org/10.22210/govor.2020.37.02>

**Goran Paulin**

*gp@kreativni.hr*

Kreativni odjel d.o.o., Rijeka

Hrvatska

**Marina Ivašić-Kos, Ivo Ipšić**

*marinai@inf.uniri.hr, ivoi@inf.uniri.hr*

Sveučilište u Rijeci, Odjel za informatiku

Hrvatska

## **Mogućnost primjene govora u računalnim igrama temeljenim na lokaciji**

### **Sažetak**

Iako je govor u računalno sintetiziranom obliku postao dio računalnih igara već 1978. godine, njegova primjena, osobito u žanru računalnih igara temeljenih na lokaciji, slabo je istražena. U ovom radu predstavljen je pregled implementacije govora u računalnim igrama nastalim u razdoblju od 1978. do 2018. godine i dosadašnja iskustva njegove primjene. Fokus je stavljen na analizu mogućnosti korištenja govornih tehnologija u računalnim igrama temeljenim na lokaciji. Zaključak donosi odgovor na pitanje ima li smisla, s obzirom na specifičnosti žanra i aktualno stanje tehnologije, uvoditi govorne tehnologije u igre temeljene na lokaciji te koji su preduvjeti za to.

**Ključne riječi:** govorne tehnologije, računalne igre, igre temeljene na lokaciji, sinteza govora u igrama, automatsko raspoznavanje govora u igrama

## 1. UVOD

Računalne igre, zahvaljujući svojoj karakteristici generiranja i konzumiranja sadržaja u realnom vremenu, jedan su od najsofisticiranijih softverskih proizvoda današnjice (Cooper i Scacchi, 2015). Razvoj računalnih igara od njihovog nastanka, zbog potencijala komercijalne eksploatacije, usko prati razvoj tehnologije, a nerijetko ga i predvodi. Različiti aspekti percepcije pokušavaju se prenijeti u igre (od slike, preko zvuka do dodira), ali one tradicionalno ostaju fokusirane na vizualno – iz čega i proizlazi naziv "videoigra".

Povijesno, prisutna je tendencija prožimanja računalnih igara sa stvarnim svijetom – prvotno temeljenjem igara na temama iz stvarnog života, potom uranjanjem igrača u virtualnu stvarnost (engl. *virtual reality*) (Zyda, 2005), a u novije vrijeme kroz ideju prodiranja računalne igre u stvarni svijet (engl. *pervasive games*). U potonjem prednjače igre temeljene na lokaciji kao žanr koji je do sada najviše iskoračio u stvarni svijet, redefinirajući ga u okruženje u kojem se odvija igra (Benford, Magerkurth i Ljungstrand, 2005). Time grafika, do sada najznačajnija komponenta igara, pada u drugi plan, a primarnu ulogu preuzima fizička lokacija i sadržaj na njoj. Grafika, odnosno grafičko korisničko sučelje igračevog uređaja (npr. pametni telefon) ili sadržaji dopunjene stvarnosti (engl. *augmented reality*) promatrani kroz odgovarajući uređaj (npr. pametni telefon ili pametne naočale), u igrama temeljenim na lokaciji koriste se prvenstveno za pružanje informacija.

Govor je u računalne igre ušao davne 1978. godine, ali, unatoč brojnim pokušajima da se u njima i zadrži kao utjecajni čimbenik, uspio je opstati samo kao dekorativni element (van der Velde, 2018). Tek posljednjih godina, zahvaljujući neprilagođenosti dosadašnjih igračih kontrolera potrebama virtualne stvarnosti, prirodnosti korištenja govora kao sučelja u funkciji komunikacije čovjeka i stroja, ali i dobrim rezultatima koji se postižu na području govornih tehnologija kod sinteze i raspoznavanja govora, govor postaje ozbiljan konkurent da postane vodeći igrački kontroler sljedeće generacije. Da bi u tome uspio mora zadovoljiti dva osnovna predujeta: tehničku izvedivost (u skladu s potrebama igre i na razini koju korisnici očekuju) i ekonomsku isplativost (profitabilnost, kao opravdanost uvođenja u igru).

Tema govora u računalnim igrama slabo je istražena (Allison, Carter i Gibbs, 2017; Chatzidimitris, Gavalas i Michael, 2016). Iz tog razloga ovaj rad istražuje mogućnost primjene govornih tehnologija kao potencijalnog korisničkog sučelja u igrama – posebno igrama temeljenim na lokaciji i u govornim igrama. U radu su

---

predstavljene računalne igre i uloga zvuka, odnosno govora, u njima. Ukratko su prikazane govorne tehnologije značajne u kontekstu primjene u računalnim igrama i dan je pregled dosadašnjih istraživanja koja se odnose na analize igara, nastalih u razdoblju od 1978. do 2018. godine, koje koriste govor. Najveći dio rada posvećen je analizi mogućnosti korištenja govora u igrama temeljenim na lokaciji. Rad završava zaključkom u kojem se, s obzirom na dosadašnja iskustva primjene govora u računalnim igrama i trenutačnom stanju tehnologije, obrazlaže mogućnost uvođenja govornih tehnologija u igre temeljene na lokaciji. Uz to, rad pokušava odgovoriti i na pitanje može li se trend korištenja govora u računalnim igrama temeljenim na lokaciji promijeniti u korist govornih korisničkih sučelja te pod kojim uvjetima.

## 2. RAČUNALNE IGRE

Računalne igre su one koje se igraju koristeći računalo. Nerijetko je sinonim za računalnu igru "videoigra", čime se naglašava vizualna komponenta igre, odnosno način pružanja povratne informacije igraču putem grafičkog korisničkog sučelja. To je posljedica povijesnog razvoja računalnih igara čiji je nastanak 1950. godine (igra "Bertie the Brain") (Lourenço, Moura, Marques i Tavares, 2017) vezan uz novonastalu mogućnost korištenja ekranskog prikaza kao oblika interakcije između računala i korisnika – igrača.

Prve računalne igre crpile su ideje iz klasičnih igara (npr. križić-kružić, tenis), a autorima je bio cilj prenijeti ih u novi medij, računalo, testirajući pritom njegove mogućnosti, ali i ograničenja. Upravo kombinacija tih, primarno hardverskih, mogućnosti i ograničenja diktirala je njihovo oblikovanje (engl. *game design*), a s njim i razvoj žanrova. Iako je tehnička mogućnost korištenja zvuka u igrama nastala iste godine kada i prva računalna igra (McCann i Thorne, 2015), proteklo je trideset godina do prvog korištenja kontinuirane glazbe u nekoj igri i jednako toliko do korištenja prvog računalno sintetiziranoga govora (GamesRadar, 2010).

### 2.1. Klasifikacija i žanrovi

Ne postoji jedinstvena klasifikacija računalnih igara (Djaouti, Alvarez, Jessel, Methel i Molinier, 2007). Prema kriteriju omjera zabave i učenja u sadržaju mogu se podijeliti na zabavne (engl. *entertainment*) i ozbiljne (engl. *serious*). Zabavne igre, kao što im samo ime sugerira, namijenjene su primarno zabavi, dok ozbiljne igre najčešće simuliraju pojedine aspekte stvarnog svijeta kombinirajući u isto vrijeme podučavanje

---

igrača, trening i informiranje kroz komunikaciju, koristeći pritom mehaniku računalne igre (Zyda, 2005). I zabavne i ozbiljne igre karakterizira postojanje virtualnog svijeta u kojem vrijede određena pravila u okviru kojih igrač mora izvršavati zadane ciljeve. Žanrovska podjela omogućava kategorizaciju računalnih igara prema karakteristikama interakcije (Apperley, 2006) pri čemu, kombiniranjem elemenata pojedinih žanrova, dolazi do hibridizacije (Perry i DeMaria, 2009).

Razvoj tehnologije omogućio je međusobno prožimanje stvarnog i virtualnog svijeta na različite načine: od virtualne (engl. *virtual reality*), preko dopunjene (engl. *augmented reality*) do miješane stvarnosti (engl. *mixed reality*). Minijaturizacija uređaja temeljenih na računalnim komponentama, s naglaskom na pametne telefone (Löwgren, 2011), kao društveno prihvatljiv preduvjet njihovog korištenja u javnosti (Jegers, 2007), potaknula je, u okviru dopunjene i miješane stvarnosti, formiranje žanra "prodirućih igara" (engl. *pervasive games*) (Walz i Ballagas, 2007), a time i njihove najpopularnije izvedenice: igara temeljenih na lokaciji (Chatzidimitris i sur., 2016).

### 2.1.1. Igre temeljene na lokaciji

Tehnički preduvjet igara temeljenih na lokaciji je postojanje skupa IT usluga koje, pod zajedničkim nazivom "usluge temeljene na lokaciji" (engl. *location-based services*), pružaju informacije koje su kreirane, kompilirane, selektirane ili filtrirane uzimajući u obzir trenutačnu lokaciju korisnika, odnosno njihovih mobilnih uređaja (Küpper, 2005). Ove usluge, dostupne na pametnim telefonima i aplikacijama, omogućavaju pristup različitim informacijama vezanim uz lokaciju. Najčešće se koriste za navigaciju, praćenje kretanja i davanje preporuka o sadržajima u blizini korisnika.

Igre temeljene na lokaciji dizajnirane su kako bi se igrale u kretanju pri čemu se iskustvo igrača mijenja ovisno o lokaciji. Najčešće su višekorisničke i izvršavaju se na poslužitelju dok igračev uređaj (pametni telefon) služi kao klijent koji središnjoj aplikaciji prijavljuje svoju lokaciju u fizičkom prostoru i prenosi igraču zadatke koje je potrebno izvršiti. Razmjena informacija između igre i igrača odvija se koristeći grafičko sučelje pametnog telefona. Zvuku se ne pridaje posebna pozornost – ni od igrača, ni tijekom razvoja ovakvih igara (Chatzidimitris i sur., 2016).

### 2.1.2. Zvukovne igre

Na drugom kraju žanrovskog spektra razvile su se zvukovne igre (engl. *audio game*) – računalne igre koje umjesto slike koriste isključivo zvuk za prikazivanje stanja igre

---

---

(J. R. Parker i Heerema, 2008). Njihov razvoj započeo je početkom osamdesetih godina prošlog stoljeća u funkciji tehničke demonstracije 3D zvuka. Intenzivirao se na prijelazu stoljeća, nudeći mogućnost igranja na računalu slijepim i slabovidnim osobama, da bi do pojave digitalnih osobnih asistenata (npr. Alexa, Cortana, Siri) ugrađenih u različite uređaje razvojno stagnirao zbog nedovoljne uronjenosti igrača u igru. Digitalni osobni asistenti, koristeći metode umjetne inteligencije, unaprijedili su dotadašnju preferiranu metodu "pričanja priče" (engl. *storytelling*) u prirodan, konverzijski oblik komunikacije tijekom koje govorna igra govorom, zvukovima i glazbom igraču oslikava virtualni svijet, a igrač njome upravlja obraćajući joj se svojim glasom.

## 2.2. Uloga zvuka u igrama

Iako se računalne igre tradicionalno oslanjaju primarno na vizualno, čemu pridonosi i činjenica da čak 70 % – 80 % percepcije većina ljudi doživljava gledanjem, igre bez zvuka nikada nisu bile popularne. Povijesni izuzetak su jedino igre koje su, zbog tehničkih ograničenja, memorijske resurse morale prepustiti drugim komponentama igre (Weske, 2000). Danas je uobičajeno da računalne igre zvuk koriste kroz četiri različita aspekta: glazbu, zvučne efekte, govor i upravljanje (J. R. Parker i Heerema, 2008).

Računalne igre rano su obogaćene (pozadinskom) glazbom (Perry i DeMaria, 2009). Uloga glazbe bila je učiniti igrača manje usamljenim, povezati ga s virtualnim svijetom, uvesti kontinuitet i podsjetiti ga da se igra odvija čak i kada mu nešto drugo na trenutak odvuče pozornost. Uz to, glazba igri daje emociju, ali i uspostavlja tempo (J. R. Parker i Heerema, 2008).

Zvučni su efekti evolucijom računalnih igara postajali sve sofisticiraniji. Od prvotnog korištenja u ulozi auditivne povratne informacije, kojoj je cilj bio potvrditi izvršenost radnje koju su zatražili igrači (J. R. Parker i Heerema, 2008), razvili su se do stupnja u kojem realistično oslikavaju virtualni svijet (Perry i DeMaria, 2009). Često u igri, kao i u stvarnosti, prvo nešto čujemo i tako prepoznamo prije nego isto to ugledamo.

Govor u računalnim igrama omogućava prirodnu komunikaciju između čovjekolikih karaktera i igrača, odnosno njegovog avatara (J. R. Parker i Heerema, 2008). Zbog memorijskih ograničenja računala, sintetizirani govor bio je jedina opcija za implementaciju govora u ranim igrama. Njegova neprirodnost učinila ga je uspješnim primarno kada se koristio za davanje "robotškoga glasa" mehaničkim

---

karakterima u igri. S pojavom kompaktnog diska (engl. *CD-ROM*) omogućeno je korištenje prirodnijeg, digitaliziranoga govora čime se govor trajno uvriježio u računalnim igrama.

Upravljanje igrom pomoću zvuka gotovo uvijek podrazumijeva korištenje ljudskoga glasa. Možemo ga smatrati prirodnim sučeljem računalne igre ako tumačenje izrečenog u igri odgovara tumačenju u stvarnom svijetu (J. R. Parker i Heerema, 2008). Koristi se za navigaciju kroz izbornike i paralelno izvođenje više radnji – npr. igrač može mišem kontrolirati vlastitog avatara i pritom glasom delegirati zadatak računalno upravljanim suigračima (engl. *non-player characters*).

### 3. GOVORNE TEHNOLOGIJE

Govorne tehnologije obuhvaćaju tehnologije čija je svrha oponašanje ljudskoga govora i reagiranje na njega s ciljem interakcije čovjeka i stroja putem govorne komunikacije, odnosno govornog sučelja (Rabiner, 1994). Govorna sučelja, zahvaljujući pametnim telefonima i digitalnim asistentima, postaju važna komponenta modernih sustava komunikacije između čovjeka i stroja (Rabiner i Schafer, 2011) i sve raširenija u svakidašnjoj upotrebi (Porcheron, Fischer, Reeves i Sharples, 2018). Prikladna su i za korištenje u kretanju i vožnji jer ostavljaju korisniku slobodne ruke i oči, a omogućavaju i korištenje strojeva slijepim i slabovidnim te gluhim i nijemim osobama (Barlet i Spohn, 2012). U kontekstu primjene govora u računalnim igrama, značajne tehnologije su sinteza govora, automatsko raspoznavanje govora i prepoznavanje i potvrđivanje govornika.

#### 3.1. Sinteza govora

Sinteza govora je umjetna tvorba govora korištenjem računala. Kvalitetu sinteze određuju dvije karakteristike: razumljivost i dojam prirodnosti. Razumljivost se odnosi na stupanj razumijevanja generiranoga govora, a prirodnost ocjenjuje sličnost ljudskom govoru. Prozodija, odnosno parametri poput glasnoće, visine tona glasova, stanki u govoru, brzine izgovora i ritma, značajno utječu na obje karakteristike, a cilj sustava za sintezu govora je maksimizirati sličnost s prirodnim govorom.

Tradicionalno najzastupljenije metode sinteze su formantna sinteza i sinteza odabirom jedinica. Formantna sinteza koristi model izvora zvuka i filtra ugođenog na formantne frekvencije govornog trakta, a sinteza ulančavanjem jedinica oslanja se na bazu govornih jedinica (npr. fonema, difona, slogova, riječi ili rečenica) tvoreći

govorni signal spajanjem isječaka snimljenoga prirodnoga govora. Osnovni nedostatak formantne sinteze je neprirodnost govora (karakterizira se kao "robotski"), a prednosti su joj visoka razumljivost, manji računalni zahtjevi za memorijom i procesorskom snagom te mogućnost generiranja širokog raspona prozodije i intonacije, a s njima i oblikovanja različitih emocija u govoru. Prednost sinteze ulančavanjem jedinica je prirodnost govora, osobito u sustavima s uskom domenom primjene (npr. prognoza vremena ili vozni red), a nedostaci su joj veličina baze govornih jedinica i problem odabira odgovarajućih govornih jedinica u bazi, između većeg broja različito izgovorenih isječaka iste jedinice.

Statistička parametarska sinteza koristi velik korpus prirodnoga govora za kreiranje modela za tvorbu parametara izvora i filtera. Uglavnom se koriste Markovljevi modeli, a za razliku od sinteze odabirom jedinica, nakon učenja, baza uzoraka više se ne koristi. Sinteza na bazi dubokog učenja koristi duboke neuronske mreže, trenirane na velikom korpusu snimljenoga govora i zahtjeva značajne resurse procesorske snage. U stanju je proizvesti različite glasove, naglaske, realistične uzdahe i zvukove koje proizvode usnice tijekom govora, što doprinosi visokoj prirodosti govora (van den Oord i sur., 2016).

### 3.2. Automatsko raspoznavanje govora

Proces automatskog raspoznavanja govora u osnovi je inverzija problema sinteze govora, a cilj sustava za automatsko raspoznavanje govora je točno i učinkovito prevođenje govornog signala u tekst (Rabiner i Schafer, 2011). Tipičnu arhitekturu sustava za raspoznavanje govora čine obrada ulaznoga govornog signala koja izlučuje njegove značajke, akustični model koji ih ocjenjuje, jezični model koji formulira lingvistička pravila za testiranje hipoteze te samo testiranje hipoteze koje rezultira transkriptom raspoznatoga govora (Yu i Deng, 2014). Sustavi za raspoznavanje govora mogu zahtijevati učenje što ih čini ovisnima o konkretnom govorniku.

Tradicionalno dominantni sustavi za automatsko raspoznavanje govora oslanjali su se na Mel-Kepstralne koeficijente za izlučivanje značajki i skrivene Markovljeve modele za generiranje akustičkih modela (Yu i Deng, 2014). U novije vrijeme obje komponente zamjenjuju se obradom sirovoga zvučnog signala koristeći povratne neuronske mreže koje omogućavaju izgradnju modela s dugoročnim kontekstom (van den Oord i sur., 2016).

U prošlosti je temeljna komponenta sustava bio algoritam dinamičkog poravnanja vremena, koji mjeri sličnost između dvije sekvence govornog signala,

koje mogu varirati u vremenu i brzini, te omogućava nalaženje optimalne podudarnosti među njima. Ovaj algoritam, koji se oslanja na pretpostavljene statističke karakteristike značajki, koristi se i kod skrivenih Markovljevih modela. Skriveni Markovljevi modeli su učinkoviti za raspoznavanje fonema i pojedinih riječi, ali njihov osnovni nedostatak je nemogućnost modeliranja vremenskih ovisnosti. Problem vremenske ovisnosti uspješno rješavaju duboke unaprijedne neuronske mreže (engl. *deep feedforward*) i neuronske mreže s povratnim vezama (engl. *recurrent*), a njihov primarni nedostatak je brzina učenja. Točnost prevođenja govornog signala u tekst moguće je dodatno unaprijediti koristeći kombinaciju zvučno-vizualnih informacija i s time povećati robusnost sustava raspoznavanja govora (Huang i Kingsbury, 2013).

### 3.3. Prepoznavanje i potvrđivanje govornika

Za razliku od raspoznavanja govora, kod kojeg identificiramo što je izgovoreno, prepoznavanje govornika odnosi se na identifikaciju osobe prema karakteristikama njezina glasa. Pri prepoznavanju govornika određujemo tko govori, u procesu potvrđivanja govornika provjeravamo je li govornik onaj kojim se predstavlja (Rabiner i Schafer, 2011). U oba slučaja moguće je koristiti metode zavisne i nezavisne od konkretnog teksta (npr. lozinka). Iako prepoznavanje temeljeno na predefiniranom tekstu postiže bolje rezultate, preferira se metoda nezavisna od teksta jer govornici izbjegavaju izgovarati lozinke na glas u situaciji kada ih netko može čuti (Furui, 2008).

## 4. PREGLED RADOVA KOJI SU ANALIZIRALI KORIŠTENJE GOVORA U IGRAMA

U nastavku se donosi pregled radova koji su istraživali mogućnosti korištenja govora u računalnim igrama. Mogućnost korištenja govora u računalnim igrama temeljenim na lokaciji nije do sada bila ciljano istraživana. Štoviše, pokazalo se da je korištenje zvuka, općenito, u igrama temeljenim na lokaciji neistraženo (Chatzidimitris i sur., 2016), a isto vrijedi i za govor u igrama (Allison i sur., 2017).

Weske (2000) kao prekretnicu korištenja govora u igrama identificira Mattelovu igraću konzolu Intellivision koja je, za razliku od prethodnih igara, koje su govor tretirale prvenstveno kao zvučni dekor (Szczepaniak, 2014), omogućavala sintezu razumljivoga govora pomoću govornog modula (Intellivoice, 1982. god.). Igre rađene za Intellivoice govor su koristile kao ekstenziju korisničkog sučelja između igrača i igre



tako da se dio informacija važnih za donošenje odluka u igri davao govorom. To je omogućavalo igraču koncentriranje pogleda na akcijska događanja na ekranu, bez potrebe skretanja fokusa na različite statističke pokazatelje koje je potrebno pratiti tijekom igre. Uz to, govor je korišten za upozoravanje igrača na neposrednu opasnost u igri, što je omogućilo i promptnije reakcije igrača. Iako se ova novina potvrdila uspješnom jer je povećavala igračevu uronjenost u igru, komercijalni aspekt, odnosno trošak nabave dodatnog modula za igraču konzolu, pokazao se presudan u negativnom smislu – igrači nisu bili skloni tom trošku, a rezultat slabe prodaje Intellivoice modula su samo četiri Mattelove izdane igre koje se oslanjaju na govor.

Govor je potencijalno dominantna komponenta igara koje žanrovski pripadaju avanturama (Weske, 2000). Prije distribucije igara na kompaktnom disku, koji je omogućio pohranu velike količine snimljenoga govora, dijalozi u igri bili su predstavljeni isključivo tekstem na ekranu, što se u literaturi (J. R. Parker i Heerema, 2008) smatra "neprirodnim sučeljem".

Igarashi i Hughes (2001) se slažu da je raspoznavanje govora primjereno u mnogim slučajevima, poput konverzacije, ali ukazuju na činjenicu da je ono i ograničavajuće kada se koristi kao naredba koju je potrebno brzo izvršiti (npr. pucanje u igri) ili kontinuirana uputa (npr. pomicanje mape). Kao rješenje predlažu komplementiranje raspoznavanja govora neverbalnim značajkama poput visine tona, glasnoće i kontinuiranosti, preferirajući efikasnost takvoga govornog sučelja nad njegovom prirodnošću.

Foley (2005) naglašava da su igrača palica (engl. *joystick*) i dugmad na njoj brži načini za pokretanje karaktera u računalnoj igri u odnosu na govorno upravljanje, ali identificira situacije u kojima govoru treba dati prednost, poput konverzacije s drugim karakterima u igri, zapovijedanja karakterima ili korištenja izbornika. Zaključuje da je sljedeći logičan korak za povećanje uronjenosti u igre upravo raspoznavanje govora, a kao najbolji primjer navodi se igra "Lifeline" (2003. god.) koja je, sa svojim naprednim konverzacijskim sustavom temeljenim na umjetnoj inteligenciji, ostvarila značajan komercijalni uspjeh i privukla pozornost igrače zajednice na igre u kojima se govor koristi za upravljanje. Foley (2005) kao najveću prepreku uvođenju govornih tehnologija u igre ne vidi samu tehnologiju već potražnju, smatrajući da i korisnici i recenzenti računalnih igara, čiji zahtjevi upravljaju tržištem, prednost daju kvaliteti grafike, zvuka i priče, dok korištenje govora tretiraju tek kao povremenu inovaciju.

Problem potražnje za korištenjem govora u igrama i banalnost korištenja zvuka u dizajnu igara do 2007. godine, J. R. Parker i Heerema (2008) objašnjavaju

---

činjenicom da se tehnologija korištenja zvuka u okruženju interaktivnih medija ne razvija jednakim intenzitetom poput grafike ili umjetne inteligencije, a ne pogoduje joj ni mali broj istraživanja na temu korištenja govornog upravljanja u računalnim aplikacijama, kao ni na temu korištenja neverbalnoga glasovnog upravljanja u igrama. Autori su identificirali sedam različitih načina korištenja zvuka od čega se samo jedan odnosi na govor i to za upravljanje igrom. Igre koje za upravljanje koriste govor ograničene su na prepoznavanje specifičnih predefiniranih fraza, bez mogućnosti vođenja konverzacije izvan tog okvira. Preduvjet raspoznavanja govora je da računalo kontinuirano sluša i interpretira govornu uputu pri čemu bi bilo praktično da prepoznaje govornika ili da reagira tek na uputu izgovorenu nakon ključne riječi.

U istraživanju (Wadley, Carter i Gibbs, 2015) masivnih mrežnih igara (engl. *massively multiplayer online game*) u razdoblju od 2004. do 2014., koje omogućavaju igračima međusobnu glasovnu komunikaciju, ističe se da korištenje govora tijekom igranja radikalno transformira iskustvo *online* igranja čineći virtualne prostore socijaliziranim, pri čemu dijelom isključuje mogućnost skrivanja identiteta igrača, višezadačnost tijekom igranja i paralelno igranje više igara, a uvodi šum u komunikaciju jer nije uvijek jasno tko kome što govori. Također, ponavlja se zaključak da postoji vrlo malo istraživanja o utjecaju govora na iskustvo korištenja virtualnih okruženja.

Carter, Allison, Downs i Gibbs (2015) istražuju korištenje govora za upravljanje u igrama. Analizirajući skroman uzorak igara zaključili su da je govorna interakcija u igrama izuzetno polarizirana: kada funkcionira, omogućava značajnu uronjenost u igru, ali kada ne funkcionira, što je češće, izaziva frustraciju igrača. Povod frustraciji su kombinacija netočnosti raspoznavanja govornog upravljanja igrom, neučinkovitost i "društvena neugodnost" koja se manifestira ako se oko igrača nalazi publika koja ne sudjeluje u igri, ali percipira igračevu konverzaciju s igrom.

Računalne igre temeljene na lokaciji minimalno su istražene i u kontekstu korištenja zvuka općenito. Jedan od rijetkih radova posvećenih zvuku u njima je rad Chatzidimitrisa i suradnika (2016) u kojem se analizira učinak uvođenja zvukovne komponente u igre temeljene na lokaciji s ciljem mjerenja razine pripravnosti igrača u odnosu na kretanje virtualnih neprijatelja u igri. Autori, gradeći prototip igre, pokušavaju izbjeći korištenje ekrana igračevog uređaja (pametni telefon) kako bi se igrač mogao bolje fokusirati na okruženje ili lokaciju na kojoj se fizički nalazi. Pritom zvuk zatečen na lokaciji dopunjuju sintetiziranim pozicijskim zvukom. Dopuna zvuka, umjesto potpune supstitucije, značajno smanjuje rizik za igrača dok se kreće u

---

---

fizičkom svijetu jer igrač i tijekom igre može pravovremeno reagirati na različite zvukove upozorenja. Testiranje prototipa pokazalo je povećanu razinu pripravnosti igrača pri korištenju 3D zvuka.

Allison (2017) ukazuje na nagli razvoj govornih korisničkih sučelja tijekom proteklih nekoliko godina u kojima su govorne tehnologije postale sastavni dio pametnih telefona, vozila, domova i igara, omogućivši nam da se oslonimo na njih. To je potaknulo trend u kojem su govorna sučelja personificirana u obliku virtualnih asistentica: Appleove Siri i Amazonove Alexe. Na vrhu ljestvice najkorištenijih aplikacija Amazonove Alexe su govorne igre (Martin, 2018) i broj njihovih korisnika udeseterostručio se u protekloj godini.

Allison i suradnici (2017) istražuju povijest korištenja govora u digitalnim igrama i razloge njihovog nikad ostvarenoga globalnog uspjeha. Prema autorima, u kontekstu dizajna igara, tri su paralelna događanja utjecala na korištenje govora: prvi je nagli razvoj temeljnih tehnologija, drugi su virtualni asistenti, kojima se u međuvremenu pridružila i Microsoftova Cortana, a treći je značajan porast broja ugrađenih mikrofona i sustava za procesiranje govora u uređajima namijenjenim igranju računalnih igara – od pametnih telefona, preko kućnih računala do dodatne opreme igračih konzola poput Microsoftovog Kinecta. Identificirali su i sedam faza u kojima se, kao odgovor na pojavu novih platformi i podupirućih tehnologija, ciklički izmjenjivao uspon i pad primjene govornih tehnologija u igrama. Prva faza pokriva razdoblje od 1973. godine do ranih osamdesetih i odnosi se isključivo na razvoj nekomercijalnih istraživačkih prototipa. Slijedi razdoblje pokušaja komercijalne primjene tehnologije, ali urušavanje tržišta računalnih igara 1982. godine pokrenulo je "rat niskih cijena" igraće opreme uslijed kojeg dodatni trošak posebnih uređaja za govor nije dobro prihvaćen među igračima koji su, osim zbog troškova, frustrirani i primitivnošću tehnologije. Posljedično, u trećoj fazi, centar industrije računalnih igara privremeno se seli u Japan. Mikrofoni postaju standardna oprema igračih konzola koja omogućava glasovni razgovor (engl. *voice chat*), ali automatsko raspoznavanje govora još nije dovoljno upotrebljivo. U četvrtoj fazi oporavlja se zapadno tržište igara, a autori u njoj uočavaju i diferencijaciju žanrova: dok zapad preferira igre u kojima je uloga igrača svedena na zapovijedanje podčinjenima, nerijetko u ratnom okruženju, japansko tržište preferira igre koje počivaju na miroljubivoj konverzaciji između karaktera. Upravo ova diferencijacija imat će značajan utjecaj na mogućnost primjene

---

govornih tehnologija: zapadni žanrovi omogućili su relativno jednostavnu implementaciju sustava koji raspoznaje predefinirane govorne zapovjedi, dok je pred japanskim igrama bio daleko složeniji izazov razumijevanja prirodnoga govora kako bi se mogla ostvariti smisljena konverzacija između igrača i karaktera u igri na tragu one koju spominje Foley (2005). Zbog tehnoloških ograničenja, umjesto na smislenu konverzaciju, japansko tržište fokusira se na karaoke i tako definira do sada komercijalno najuspješniji žanr računalnih igara koji počiva na primjeni govorne interakcije. Predstavljanjem Kinecta započinje šesta faza u kojoj se čini da će zapadni pristup imati više uspjeha. Populariziraju se igre koje sadrže glasovno zapovijedanje, ali ono nije učinjeno nužnim kako bi iste igre mogli koristiti i igrači koji nemaju Kinect – time je učinjen kompromis koji nije pogodovao uvođenju govornih tehnologija u igre, na što je također već upozorio Foley (2005). Ovo će pokušati promijeniti, u sedmoj fazi, nezavisni autori računalnih igara koji, neopterećeni komercijalnim uspjehom, odlučuju govornu interakciju iskoristiti kao okosnicu svojih igara. Time mikrofon, koji je u međuvremenu postao općeprihvaćen, postaje nužan za igranje, a igrače, zbog iskustva koje im nude, takve igre privlače prirodnošću svojeg korisničkog sučelja – što je komponenta na koju je ukazao Weske (2000) osvrćući se na rani uspjeh Intellivoice igara.

Povod osvrtu (Ip, 2017) na stanje korištenja govora u igrama je nadolazeća igra "Starship Commander", namijenjena igranju u virtualnoj stvarnosti i kao takva idealan je kandidat za govorno upravljanje. Za razliku od ranijih igara, "Starship Commander" oslanja se na Microsoftovu "Custom Speech Service" tehnologiju koja prosljeđuje igračev govor u "oblak" gdje se analizira, određuje namjera igrača i vraća odgovor za nekoliko milisekundi. Testiranja na igračima stavljenim u ulogu zapovjednika svemirskog broda pokazuju da oni u visokom postotku (99 %) reagiraju kontekstualno i spontano odlučuju komunicirati s igrom u domeni teme – što, posljedično, olakšava odrediti namjeru igrača.

Van der Velde (2018) navodi potencijalne razloge za implementaciju govora u igrama: od dostupnosti invalidnim osobama, preko dubljeg uranjanja u igru kroz dodatni sloj integracije, do ublažavanja krivulje učenja početnicima, koji umjesto privikavanja na kontrole u igri mogu "jednostavno početi govoriti". Osvrće se i na problem kvalitetne implementacije govornih tehnologija jer igrači teže isključivo razumijevanju prirodnoga govora, a kao jedno od rješenja navodi Amazonov "Alexa

---

---

Skills Kit" (ASK) koji to omogućava kroz integraciju igre s Alexom. No, unatoč tomu, autorica zaključuje da je korištenje govornih tehnologija u domeni računalnih igara još uvijek zanemarivo i da igrači za komunikaciju i upravljanje karakterima u igrama preferiraju klasične metode upravljanja: igrače palice ili kombinaciju tipkovnice i miša.

## **5. ANALIZA MOGUĆNOSTI KORIŠTENJA GOVORA U IGRAMA TEMELJENIM NA LOKACIJI**

Za utvrđivanje mogućnosti korištenja govora u računalnim igrama temeljenim na lokaciji potrebno je analizirati dosadašnju praksu korištenja govornih tehnologija u igrama, izdvojiti inovacije koje u ukupnoj ocjeni pojedine igre djeluju pozitivno ili negativno te prepoznati koje su od tih značajki primjenjive na razvoj igara temeljenih na lokaciji, poštujući njihove specifičnosti. U tu svrhu formiran je uzorak od 50 igara koje uključuju igre različitog žanra analizirane u radovima čiji je pregled dan u 4. poglavlju, dopunjen s igrama pronađenim pretraživanjem interneta prema kriteriju korištenja bilo kakvog oblika govora specificiranog u recenzijama i dokumentaciji pojedinih igara. Pritom su izuzeti nastavci i klonovi pojedinih igara koji koriste prethodno uvedenu govornu tehnologiju, ako ne uvide inovacije.

Primjena govornih tehnologija u igrama započinje 1978. u SAD-u kada Texas Instruments predstavlja ručnu igraću konzolu Speak & Spell s edukativnom igrom "Say It" oglašavajući je kao "pravopisno pomagalo" (Edwards, 2008). Konzola je namijenjena isključivo engleskom govornom području i omogućava (formantnu) sintezu govora koju igra, kao nužnu komponentu, koristi za glasovno zadavanje slova i riječi koje igrač mora pravopisno pravilno utipkati. Evaluacija konzole i njene primjene u edukacijske svrhe pokazala je značajno poboljšanje pravopisa kod korisnika, ali isključivo za riječi iz ograničenog vokabulara igre i isključivo u razdoblju korištenja konzole (Terrell i Linyard, 1982). Iako korištenje Speak & Spell konzole ni na koji način nije vezano za lokaciju, činjenica da je riječ o ručnoj (prijenosnoj) igraćoj konzoli, koju je moguće koristiti na bilo kojoj lokaciji, definira jedan od preduvjeta za korištenje govora u igrama temeljenim na lokaciji – mobilnost uređaja. Uz to, činjenica da je govor nužna, a ne dekorativna komponenta igara ove konzole, pokazat će se presudna za (komercijalni) uspjeh budućih igara koje koriste govorne tehnologije.

---

S obzirom na to da Speak & Spell nije ostvario svoj edukacijski potencijal, japanski Sun Electronics, u namjeri korištenja sintetiziranoga govora na arkadnim mašinama, okreće se arkadnom žanru i dvije godine poslije (1980.) izdaje pucačku igru pod nazivom "Speak & Rescue" (u Japanu), odnosno "Stratovox" (u SAD-u). U njoj se govor koristi isključivo dekorativno, za obraćanje ljudskih karaktera (u igri) igraču. Ovaj komercijalni eksperiment potvrdio je da sinteza govora u igrama može funkcionirati i za druge jezike (japanski), ali ujedno je identificirao i problem prirodnosti sintetiziranoga glasa – iako se govor, u marketinškom kontekstu, pokazao kao dobrodošla inovacija, "robotski" govor ljudskih karaktera nije ispunio očekivanja igrača. Iste godine, također u Japanu, nastaje igra "Manbiki Shoujo" (engl. "*Shoplifting Girl*"), za prvo računalo tvrtke Commodore – PET 2001. Time se sintetizirani govor seli na kućna računala. Na žalost, "Manbiki Shoujo" slijedi primjer svoga japanskog prethodnika i govor (isključivo japanski) također koristi samo dekorativno. Pritom definira novi žanr – tzv. potajne (engl. *stealth*) igre, u kojima igrač izbjegava protivnika. Ovaj žanr će 35 godina kasnije (2015.), koristeći upravo govor kao temelj igrivosti, postati predvodnik nove renesanse korištenja govora u igrama. Kako problem prirodnosti nije bio tehnički rješiv, američka igra "Berzerk", za čak četiri različite platforme i do dva igrača, riješila ga je tako što je sintetizirani govor koristila isključivo za obraćanje robota (u igri) igraču – time su igrači bili zadovoljni, a formantna sinteza konačno komercijalno upotrebljiva u računalnim igrama.

Trend igara koji koriste sintezu govora prekinula je kanadska igra "Bear Hunt" – prva zvukovna igra. Korištena za demonstraciju prostornog (3D) zvuka omogućila je oslikavanje virtualnog okruženja igre isključivo zvukom i time utrla put zvukovnim odnosno govornim igrama koje će se pojaviti tek početkom 2000-ih godina.

Godinu 1982. obilježio je američki Mattel Electronics sa svojom Intellivision konzolom i Intellivoice govornim modulom za nju. Mattel je uveo govor u žanr simulacija i unaprijedio njegovu prirodnost što mu je omogućilo da ponudi igračima više različitih sintetiziranih glasova koji su varirali od robotskih do ljudskih (Orth, 2020). No, njegov najznačajniji doprinos bio je inzistiranje na govoru kao nužnoj, a ne samo dekorativnoj komponenti igre te su sve njihove igre izdane za Intellivoice ispoštovale taj uvjet. Igra "Space Spartans" bila je prva igra u kojoj je govor korišten kao komplement grafičkom korisničkom sučelju – dio bitnih informacija o događanjima u igri prezentiran je glasom i omogućio je igraču da ih percipira ostajući

---

---

fokusiran na (akcijsku) radnju. Matellova motivacija za korištenje govora kao nužne komponente bila je ekonomska – govor u igrama trebao je potaknuti igrače na kupovinu Intellivoice modula, međutim naglo urušavanje tržišta računalnih igara i "rat niskih cijena" proizvođača igraće opreme, u konačnici je odbio kupce koji nisu bili skloni dodatnim troškovima, a to je rezultiralo povlačenjem govornog modula iz prodaje već sljedeće godine.

Sličnu sudbinu dijelio je i Philipsov The Voice govorni modul za Magnavox Odyssey 2 konzolu. Za razliku od Mattela koji je eksperimentirao s uvođenjem govora u nove žanrove igara i pritom inzistirao na nužnosti govora u njima, Philips je odlučio ponoviti rani uspjeh Speak & Spell koncepta te je razvio edukativnu igru "Type & Tell!" s opcijom naizmjeničnog igranja do dva igrača za koju je govorni modul The Voice bio nužan. To je nažalost bila i jedina Philipsova igra s obaveznim korištenjem govornog modula pa ga igrači nisu bili motivirani kupiti (Goodman, 1983). Razdoblje igara u kojima je korištena sinteza govora završava pokušajem njenog uvođenja u kućna računala, također kroz dodatne hardverske module, a jedan od značajnijih primjera je Currah modul proizveden za ZX Spectrum, Commodore 64 i druga 8-bitna računala (Anson, 1983). Poučeni Matellovim i Philipsovim iskustvom, proizvođači igara za ove platforme nisu se htjeli ograničiti na malobrojne korisnike govornog modula te su sintetizirani govor koristili isključivo dekorativno, a većina igrača nije ga nikad imala prigodu čuti.

Japanski proizvođači ponovo se uključuju u utrku uvođenja govornih tehnologija u igre 1987. godine kada Bandai izdaje "Karaoke Studio" i njime utemeljuje karaoke žanr – do danas najprofitabilniji žanr korištenja govornih tehnologija u igrama, iako se uopće ne oslanja na govor u smislu njegovog raspoznavanja već isključivo na neverbalne značajke poput visine i trajanja tona. Komercijalni uspjeh karaoke žanra nije ugrozila ni činjenica da je za Nintendo Family Computer trebalo dokupiti ekspanziju za detekciju glasa. To je ujedno pokazalo da problem s kojim su se susreli Mattel i Philips nije bio isključivo ekonomske prirode već da tržište reagira na tzv. "nesklad identiteta igrača" (Carter i sur., 2015) – situacije u kojima se u dizajnu igre forsira upravljanje govorom u neprirodnom kontekstu (primjerice, korištenje govora za promjenu oružja korištenog u igri) čime se sprečava igračevo uživanje u ulogu što uzrokuje negativno iskustvo i negativnu ocjenu igre.

---

Najraniji primjer govorne interakcije, odnosno korištenja govora za upravljanje, je američka igra "Command: Aces of the Deep" za Windows platformu. Stavljanjem igrača u ulogu zapovjednika podmornice započinje trend koji su prepoznali Allison i suradnici (2017): "zapadne" igre nastavit će se oslanjati na korištenje govora primarno u funkciji zapovijedanja u (često ratnim) igrama, koristeći isključivo engleski jezik, dok će igre koje dolaze iz Japana, među kojima "Hey You, Pikachu!" i "Seaman", biti miroljubive simulacije koje naglasak stavljaju na prirodnu konverzaciju i najčešće su izdane dvojezično (japanski, ali i engleski jezik).

U razdoblju do 2002. godine dolazi do prepoznavanja potencijala korištenja zvukovnih igara za potrebe slijepih i slabovidnih osoba – time je, nakon dvadesetogodišnjeg mirovanja, prostorni zvuk dobio konkretnu primjenu u igrama. Na žalost, stupanj raspoznavanja govora još uvijek ne omogućava prirodnu komunikaciju isključivo govorom te ove igre zvuk i govor koriste samo za oslikavanje virtualnog okruženja pri čemu je igrač prisiljen za upravljanje umjesto glasa koristiti neki od klasičnih igračih kontrolera.

Godine 2002., američka igra "SOCOM: U.S. Navy SEALs", za različite PlayStation konzole (a kasnije i pametne telefone), uvodi glasovno, ali samo izborno, zapovijedanje računalno upravljanim suigračima (engl. *non-player characters*), ograničeno malim brojem predefiniраниh fraza. Problem ograničenog vokabulara već sljedeće godine rješava Sony Computer Entertainment Inc izdavači za PlayStation 2 dvojezičnu (japanski i engleski) igru "Operator's Side" (JP) / "Lifeline" (SAD). U ovom horor preživljavanju koristi se napredni konverzacijski sustav koji, po prvi put u igrama, počiva na umjetnoj inteligenciji i omogućava konverzaciju temeljenu na oko 500 glasovnih naredbi, u svrhu isključivo glasovnog upravljanja karakterom u igri. Igra se, upravo zbog načina na koji koristi govor, pokazala izuzetno uspješnom i u komercijalnom smislu.

Američki Draconis Entertainment 2003. godine izdaje zvukovnu igru u kojoj istovremeno koristi zvukove kao povratnu informaciju korisničkog sučelja, ali i (digitalizirani) govor kao prateći komentar igre. Iste godine francuski Ubisoft izdaje pucačku (taktičku) igru "Tom Clancy's Rainbow Six 3" u kojoj se govor koristi za upravljanje računalno upravljanim suigračima, ali i za međusobnu komunikaciju među igračima. Juang i Rabiner (2005) ukazat će na problem "šuma" u takvom

---



---

pristupu kojeg je još uvijek moguće riješiti isključivo razdvajanjem komunikacijskih kanala, što umanjuje prirodnost konverzacije.

Nakon duго vremena, 2005. godine japanski SCEJ vraća se edukativnom žanru. Igrificirajući aplikaciju za prevođenje, "Talkman" nudi glasovnu aktivaciju i prevođenje s i na japanski, engleski, korejski i kineski. Iako u novom, hibridnom žanru, Japanci ostaju dosljedni miroljubivosti i poticanju konverzacije. Nintendo iste godine predstavlja "Nintendogs", simulaciju života ljubimaca (pasa) u kojima igrač stupa u verbalnu komunikaciju sa svojim virtualnim ljubimcima dresirajući ih govorom. Kako platforma podržava i dodir ekrana, Nintendo ga koristi kako bi omogućio igračima "pogladiti" psa te na taj način korištenjem obiju komponenata (govor i dodir ekrana) na "prirodan način" predstavlja odličan primjer prirodnog sučelja (J. R. Parker i Heerema, 2008).

Microsoftov "(Project) Milo" 2009. godine nastavlja istraživanje prirodnih sučelja u igrama i, kao tehnički demo Kinecta, kroz konverzaciju s virtualnim karakterom (Milo) ujedinjuje igračev govor, pokret i različite akcije upravljanja koje koriste metode umjetne inteligencije. Uspjeh postignut na području raspoznavanja govora potaknuo je američki Bethesda Softworks 2011. godine na uvođenje glasovnih komandi u akcijsku igru "Skyrim" koja, osim engleskog, prva raspoznaje i umjetni jezik: Draconic.

Tek se 2012. godine pojavljuje prva govorna igra, nezavisnog proizvođača (engl. *indie*) Six to Start, "Zombies, Run!", namijenjena iOS platformi, koja u žanru dopunjene stvarnosti (engl. *augmented reality*) koristi glas za navigaciju igrača u pokretu (trčanje) zadavanjem konkretnih zadataka s obzirom na (fizičku) okolinu u kojoj se igrač nalazi. Ovo je jedini identificirani primjer igre na lokaciji koja koristi govorne tehnologije.

Kanadski Ubisoft 2013. godine koristi govor igrača za privlačenje pozornosti neprijateljskih karaktera u potajnoj igri "Tom Clancy's Splinter Cell: Blacklist". Zbog izrazite dinamike nogometne simulacije, kanadska igra "FIFA 2014", iste godine uvodi korištenje glasovnih komandi kako bi se paraleliziralo upravljanje karakterima u igri. Na taj način igraču je omogućeno koristiti klasične igrače kontrolere za pokretanje igrača (virtualnih nogometaša) po terenu dok istovremeno može, glasom, mijenjati formacije momčadi ili zatražiti zamjenu igrača. Kanađani 2013. godine dominiraju u govornim inovacijama pa tako North Side Inc. izdaje "Bot Colony", avanturu u prvom licu, koja omogućava prirodnu konverzaciju koristeći njihovu

---

vlastitu tehnologiju za obradu prirodnog jezika. Igra traži obavezno kreiranje akustičnog profila igrača i treniranje pretvorbe govora u tekst, što, u pozitivnom smislu, omogućava prepoznavanje i potvrđivanje govornika (igrača), ali u praksi odbija igrače koji nisu spremni na taj korak već preferiraju ono što van der Velde (van der Velde, 2018) navodi kao želju da "jednostavno počnu govoriti".

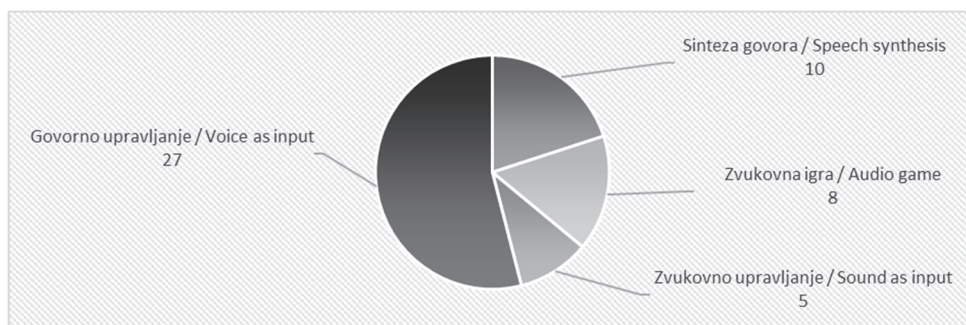
Igra "Ryse: Son of Rome" nema taj problem. U njoj je korištenje govora, odnosno glasovnih komandi, neobavezno, ali koncept igre, koja igrača stavlja u ulogu rimskog vojskovođe, osmišljen je tako da potiče igrače na spontano korištenje govora što joj je donijelo naklonost igrača. "Tomb Raider: Definite Edition", akcijska avantura u trećem licu, suprotnost je tome jer se prema ocjeni korisnika govor koristi izuzetno neprilagođeno. Glasovnim komandama moguće je, primjerice, obaviti promjenu oružja, što ne samo da je sporije nego putem tipkovnice, već je ocijenjeno i kao neprirodno (Carter i sur., 2015). Sličnu pogrešku, iste 2014. godine, napravila je i strateška igra "There Came an Echo" koja inzistira na isključivo glasovnom upravljanju jedinicama.

Godine 2015. objavljene su dvije potajne igre: američka "Mayday! Deep Space" i britanska "Plan Be" koje koriste govor za upravljanje karakterom u igri i svojevrzni su *homage* začetniku žanra (igra "Manbiki Shoujo") koji je koristio govor kao dekorativnu komponentu. Igra "Plan Be", odlično prihvaćena kod igrača, inzistira na govoru za izgradnju odnosa između igrača i karaktera u igri kojim upravlja (L. Parker, 2015). Zadnja kanadska govorna inovacija sadržana je u igri "Keep Talking and Nobody Explodes" iz 2015. godine. Iako se u govoru ne oslanja na računalo (ne koristi ni sintezu ni raspoznavanje govora), govor, odnosno dijalog između suigrača, sa svrhom razmjene informacija nedostupnih drugoj strani, nužan je element igre.

Pregled analiziranih igara završava s 2017. godinom. Tijekom nje pojavile su se dvije tematski srodne igre: objavljena "Star Trek Bridge Crew" i rani prototip igre "Starship Commander". Obje igre koriste napredne govorne tehnologije (Watson/IBM i Cortana/Microsoft) za zapovijedanje računalno upravljanim suigračima odnosno upravljanje igrom. Posebnost igre "Starship Commander" je u tome što pripada žanru virtualne stvarnosti, kojemu ne pristaju klasični igrači kontroleri već neposredna interakcija koja se, u svrhu komunikacije, najbolje ostvaruje upravo govorom. "Starship Commander", dijelom zahvaljujući i domeni same igre, može se pohvaliti razumijevanjem govora u rangu s čovjekovim (Roach, 2017): inicijalna testiranja pokazala su da je razina pogreške (engl. *word error rate*) u njoj ispod 6 % (Linn, 2016).

---

U nastavku su predstavljeni statistički podaci analiziranog uzorka igara.

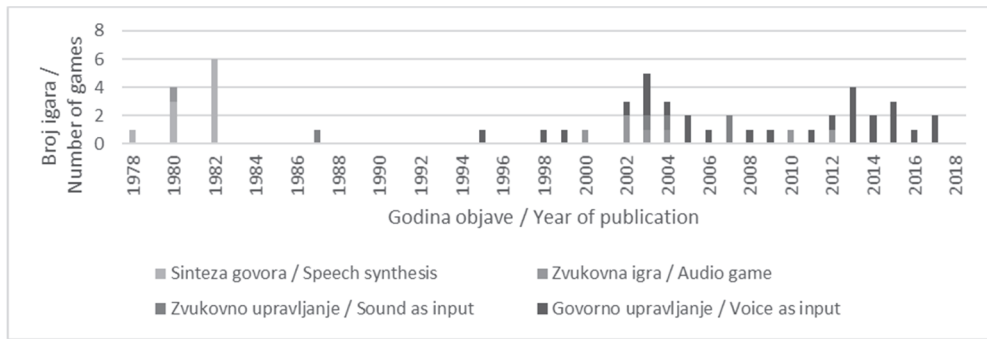


**Slika 1.** Distribucija broja objavljenih igara prema utvrđenim kategorijama primjene govornih tehnologija

**Figure 1.** Distribution of the number of published games according to the identified categories of speech technology application

Više od polovice (54 %) igara (Slika 1.) koristi govor za upravljanje igrom oslanjajući se na automatsko raspoznavanje govora. Samo 20 % koristi sintezu govora i to u najranijem razdoblju, do 1982. godine, što se može objasniti nedovoljnom prirodnošću sintetiziranoga govora, u kombinaciji s pojavom digitaliziranoga govora. Zvukovne igre u uzorku zastupljene su sa samo 16 % iako je njihov broj objektivno puno veći s obzirom na to da posljednjih godina raste trend proizvodnje zvukovnih igara za digitalne osobne asistente. No, unatoč rastućoj produkciji, ne uvode značajne inovacije pa su utoliko izostavljene iz analize u korist svojih prethodnika. Na začelju govornih tehnologija je korištenje zvuka za upravljanje igrom, s 10 % udjela, vezano isključivo za karaoke i druge glazbene žanrove. Izostalo je očekivano svrstavanje igara u više od jedne kategorije, pokazavši da se kategorije primjene govornih tehnologija koriste međusobno isključivo, npr. igre koje koriste zvukovno upravljanje ne spadaju u kategoriju zvukovnih igara, te ne koriste sintezu govora niti govorno upravljanje.

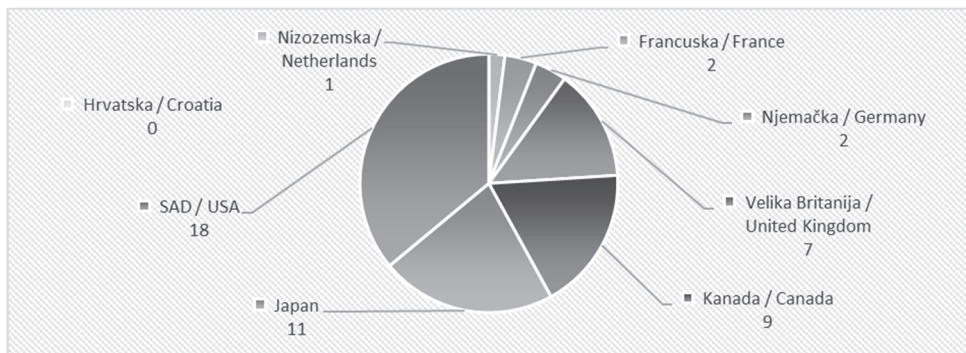
Prateći distribuciju broja objavljenih igara po godinama (Slika 2.) mogu se uočiti ciklusi na koje ukazuju Allison i suradnici (2017). Prvi vrhunac (1982. godine) obilježilo je korištenje sintetiziranoga govora, za drugi (oko 2003. godine) je zaslužno populariziranje glasovnog razgovora unutar masivnih mrežnih igara, a treći (oko 2013. godine) se odnosi na primjenu tehnologije za obradu prirodnog jezika.



**Slika 2.** Distribucija broja objavljenih igara po godinama i kategorijama primjene govornih tehnologija

**Figure 2.** Distribution of the number of published games by year and the identified categories of speech technology application

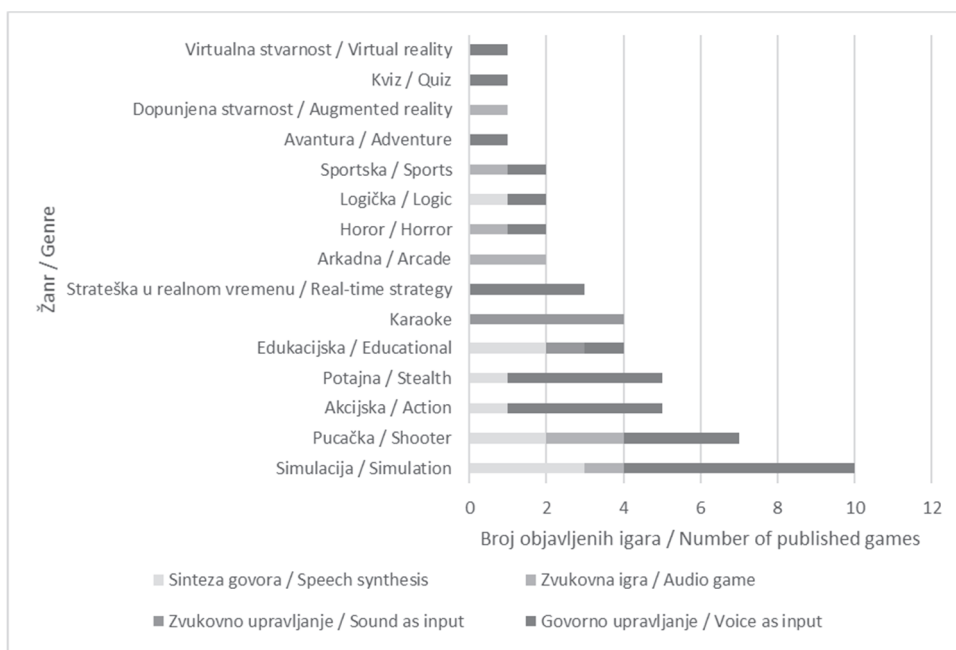
Kao proizvođači igara koje koriste govorne tehnologije, u analiziranom uzorku dominiraju Sjedinjene Američke Države s čak 36 % udjela (Slika 3.). Slijedi ih tradicionalni konkurent, Japan, s 22 %. Kanada (18 %) i Velika Britanija (14 %) također visoko kotiraju iako su im se pridružile tek početkom dvijetisućitih dok su Njemačka (4 %), Francuska (4 %) i Nizozemska (2 %) znatno manje zastupljene. Hrvatska (0 %) je dodana na popis kako bi se istaknulo da, prema podacima Klastera hrvatskih proizvođača računalnih igara (CGDA) iz 2018. godine, domaće tvrtke još ne koriste govorne tehnologije u igrama.



**Slika 3.** Distribucija broja objavljenih igara po državama proizvodnje

**Figure 3.** Distribution of the number of published games by production country

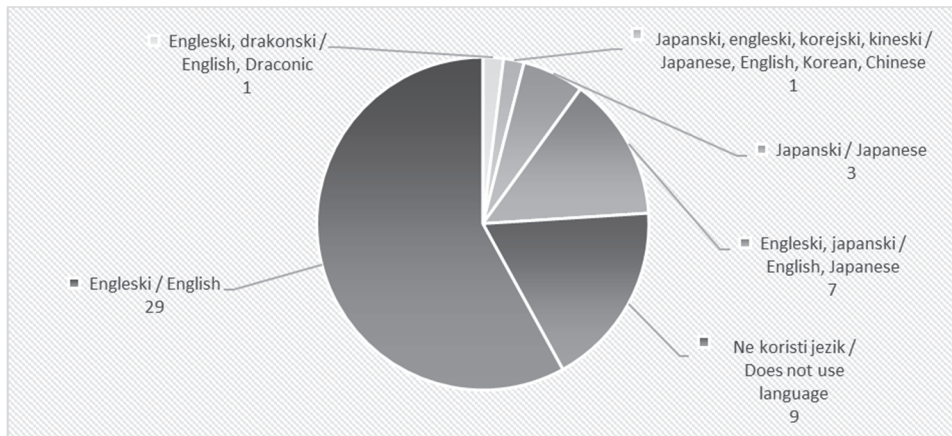
Distribucija po žanrovima je iznenađujuća (Slika 4.). Avantura, kao žanr koji tradicionalno počiva na konverzaciji i kao takav se čini idealan za primjenu govornih tehnologija, kao i kviz te noviji žanrovi dopunjene i virtualne stvarnosti, iako imaju izuzetni potencijal primjene govornih tehnologija, imaju jako malu zastupljenost govora (2 %). Korištenje govora najzastupljenije je kod simulacija (20 %) koju, zbog konverzacijskog odnosa prema igraču, karakterizira i visok udio (90 %) nužnosti govornog/zvukovnog upravljanja. Visoki postotak pucačkih (14 %) i akcijskih igara (10 %) odnosi se prvenstveno na dekorativno ili neobavezno korištenje govora u svrhu zapovijedanja, dok udio potajnih (10 %) igara raste u novije vrijeme zahvaljujući inzistiranju na nužnosti govornog/zvukovnog upravljanja.



**Slika 4.** Distribucija broja objavljenih igara po žanrovima i kategorijama primjene govornih tehnologija

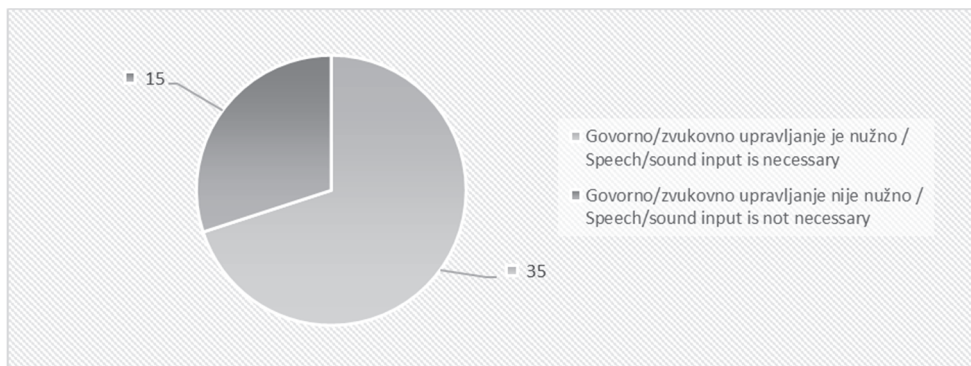
**Figure 4.** Distribution of the number of published games by genre and the identified categories of speech technology application

Među igrama koje koriste jezik (Slika 5.) dominantan je engleski s čak 92,68 %, što je očekivano s obzirom na to da je sinteza engleskog, kao jezika s većim broj govornika, tradicionalno puno bolja od sinteze drugih jezika.



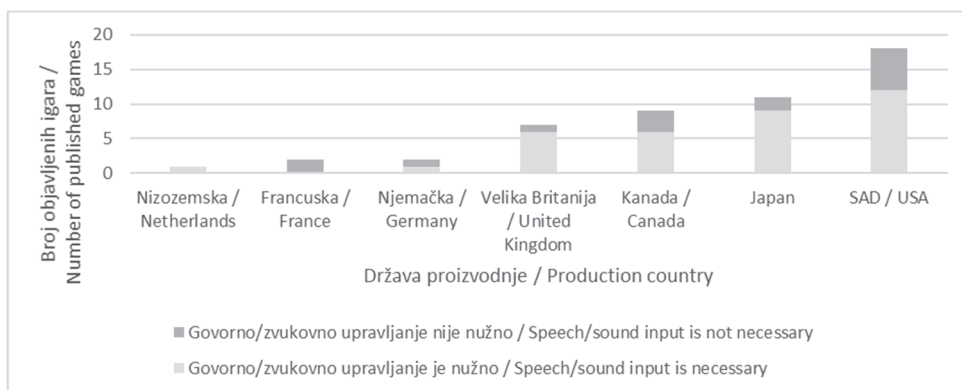
**Slika 5.** Distribucija broja objavljenih igara prema korištenim jezicima  
**Figure 5.** Distribution of the number of published games according to the languages used

Više od dvije trećine (70 %) igara u analiziranom uzorku govorno/zvukovno upravljanje tretira kao nužno (u smislu da ih je moguće igrati isključivo koristeći govorne tehnologije) (Slika 6.). Iako se to pokazalo komercijalno uspješnim receptom, ne ide u prilog osobama s govornim manama kojima nije ponuđena alternativa upravljanja igrom. Rješenje ovog problema mogla bi biti prilagodba sustava automatskog raspoznavanja govora ovom obliku hendikepa.



**Slika 6.** Distribucija broja objavljenih igara prema nužnosti govornog/zvukovnog upravljanja  
**Figure 6.** Distribution of the number of published games according to the necessity of voice/audio input

Kod primjene načela nužnosti govornog/zvukovnog upravljanja možemo pratiti i kako mu se proizvođači po državama prilagođavaju (Slika 7.): za Veliku Britaniju i Japan je karakteristično da nakon inicijalnog neinzistiranja na njemu, započinju dosljednu primjenu dok ga proizvođači iz SAD-a nakon početne gotovo isključive primjene, od 2011. godine kombiniraju kroz multimodalni pristup.



**Slika 7.** Distribucija nužnosti govornog/zvukovnog upravljanja po državama proizvodnje

**Figure 7.** Distribution of the necessity of voice/sound input by production country

## 6. ZAKLJUČAK

Specifičnost igara temeljenih na lokaciji je igranje u eksterijeru, u kretanju, koristeći društveno prihvatljive prijenosne uređaje za igru i potencijalnu konverzaciju s računalom putem pametnih telefona.

Na pametnim telefonima klasični igraći kontroleri zamijenjeni su dodirnom koji zahtijeva gledanje u ekran. Zahvaljujući standardno ugrađenim zvučnicima i mikrofonomima, izuzetak su modovi korištenja prilagođeni slijepim i slabovidnim osobama koji se oslanjaju na govorne i, u novije vrijeme, haptičke tehnologije. Pametni telefoni opremljeni su i kamerama koje mogu, kombiniranjem zvučno-vizualnih informacija, povećati robusnost sustava raspoznavanja govora. Uz to, snaga njihovog hardvera te mogućnost pristupa "oblaku" i integracije s digitalnim osobnim

asistentima, omogućavaju sve potrebne tehničke preduvjete za primjenu govornih tehnologija.

S obzirom na dosadašnja iskustva primjene govora u računalnim igrama, možemo zaključiti da ima smisla koristiti govorne tehnologije u igrama na lokaciji, ako je moguće poštovati sljedeće uvjete:

1. koncept igre, neovisno o žanru, potrebno je prilagoditi korištenju govornih tehnologija i učiniti ih nužnima za igru, pazeći pritom da se hendikepiranim korisnicima omoguće odgovarajuće alternative, ali u skladu s konceptom;
2. zbog sigurnosti igrača u kretanju i fokusa na događanja oko njega, korištenjem govora kao povratne informacije potrebno je pogledavanje na ekran svesti na minimum;
3. ako se koristi sinteza govora, mora rezultirati prirodnim govorom, a izuzetak može biti jedino govor robotskih karaktera;
4. ako je nužno koristiti prikaz podataka na ekranu, govorno upravljanje može se upotrijebiti za ubrzavanje tijeka igre paralelnim glasovnim komandama;
5. govorno upravljanje potrebno je koristiti kao prirodno sučelje, odnosno korištenje govora u igri mora odgovarati korištenju govora u stvarnom svijetu;
6. za uronjenost u igru potrebno je omogućiti prirodnu konverzaciju, pri čemu raspoznavanje govora mora biti u rangu s čovjekovim.

Uz navedeno, zbog komocije igrača, potrebno je izbjegavati forsiranje kreiranja akustičnog profila treniranjem. Ako je akustični profil potreban za prepoznavanje (zbog višeglasja na lokaciji) ili potvrđivanje igrača, preporuča se izgraditi ga pozadinski, tijekom korištenja igre, koristeći u međuvremenu druge metode identifikacije.

Istraživanje provedeno za potrebe ovog rada obuhvatilo je uzorak koji čini 50 reprezentativnih igara. Zbog nedostupnosti većeg dijela podataka o konkretnim korištenim igračim (engl. *game engine*) i govornim tehnologijama, te zbog potpunog izostanka podataka o troškovima implementacije govornih tehnologija i financijskog uspjeha pojedinih igara, izostala je analiza isplativosti uvođenja govornih tehnologija u igre koja bi svakako bila značajan doprinos zaključku. Ako navedeni podaci postanu dostupni, planira se ponoviti istraživanje, s većim uzorkom igara.

---



## REFERENCIJE

- Allison, F.** (2017). Voice interaction with virtual characters (Prošireni sažetak neobjavljene doktorske disertacije). *Proceedings of the 5th Annual CIS Doctoral Colloquium* (str. 34). Melbourne, Australija: The University of Melbourne.
- Allison, F., Carter, M. i Gibbs, M.** (2017). Word play: A history of voice interactions in digital games. *Games and Culture*, 15(2), 91–113. <https://doi.org/10.1177/1555412017746305>
- Anson, M.** (1983). *CURRAH  $\mu$ SPEECH programming manual*. SAD, Cleveland: Currah Computer Components Ltd.
- Apperley, T. H.** (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation and Gaming*, 37(1), 6–23. <https://doi.org/10.1177/1046878105282278>
- Barlet, M. C. i Spohn, S. D.** (2012). *Includification*. The AbleGamers Foundation. Dostupno na [https://accessible.games/wp-content/uploads/2018/11/AbleGamers\\_Includification.pdf](https://accessible.games/wp-content/uploads/2018/11/AbleGamers_Includification.pdf) [posljednji pristup 20. ožujka 2020.].
- Benford, S., Magerkurth, C. i Ljungstrand, P.** (2005). Bridging the physical and digital in pervasive gaming. *Communications of the ACM*, 48(3), 54–57. <https://doi.org/10.1145/1047671.1047704>
- Carter, M., Allison, F., Downs, J. i Gibbs, M.** (2015). Player identity dissonance and voice interaction in games. U A. Cox (ur.), *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play – CHI PLAY '15* (str. 265–269). New York, SAD: Association for Computing Machinery. <https://doi.org/10.1145/2793107.2793144>
- Chatzidimitris, T., Gavalas, D. i Michael, D.** (2016). SoundPacman: Audio augmented reality in location-based games. *18th Mediterranean Electrotechnical Conference (MELECON)* (str. 1–6). Lemesos, Cipar: IEEE. <https://doi.org/10.1109/MELCON.2016.7495414>
- Cooper, K. i Scacchi, W.** (2015). *Computer games and software engineering*. Boca Raton, SAD: Chapman & Hall/CRC.
- Djaouti, D., Alvarez, J., Jessel, J.-P., Methel, G. i Molinier, P.** (2007). Towards a classification of video games. U P. Olivier (ur.), *Proceedings of Artificial and Ambient Intelligence Conference (AISB'07)* (str. 1–6). Newcastle upon Tyne, UK: Culture Lab, Newcastle University.
-

- Edwards, B.** (16. prosinca 2008). VC&G interview: 30 years later, Richard Wiggins talks Speak & Spell development [Blog]. Dostupno na <http://www.vintagecomputing.com/index.php/archives/528>
- Foley, J.** (2005). The future of speech in games. *Game Informer*, 11, 73.
- Furui, S.** (2008). Speaker recognition. U *Scholarpedia*. Dostupno na [http://www.scholarpedia.org/article/Speaker\\_recognition](http://www.scholarpedia.org/article/Speaker_recognition) [posljednji pristup 21. listopada 2018.].
- GamesRadar.** (9. listopada 2010). *Gaming's most important evolutions*. Dostupno na <http://www.gamesradar.com/fl/gamings-most-important-evolutions/a-20101008102331322035/p-2>
- Goodman, D.** (1983). Video games update. *Video & Arcade Games*, 1(1), 32–44, 98–99.
- Huang, J. i Kingsbury, B.** (2013). Audio-visual deep learning for noise robust speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing* (str. 7596–7599). Vancouver, Kanada: IEEE. <https://doi.org/10.1109/ICASSP.2013.6639140>
- Igarashi, T. i Hughes, J. F.** (2001). Voice as sound: Using non-verbal voice input for interactive control. *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (str. 155–156). New York, SAD: Association for Computing Machinery. <https://doi.org/10.1145/502348.502372>
- Ip, C.** (23. lipnja 2017). *The next video-game controller is your voice*. Engadget. Dostupno na <https://www.engadget.com/2017/06/23/voice-based-gaming/>
- Jegers, K.** (2007). Pervasive game flow: Understanding player enjoyment in pervasive gaming. *Computers in Entertainment*, 5(1), 1–11. <https://doi.org/10.1145/1236224.1236238>
- Juang, B. H. i Rabiner, L. R.** (2005). Speech recognition, automatic: History (2. izdanje). U K. Brown (ur.), *Encyclopedia of language and linguistics* (str. 806–819). Amsterdam: Elsevier. Dostupno na <https://doi.org/10.1016/B0-08-044854-2/00906-8>
- Küpper, A.** (2005). Location-based services: Fundamentals and operation. Chichester, UK: Wiley.
- Linn, A.** (18. listopada 2016). Historic achievement: Microsoft researchers reach human parity in conversational speech recognition [Blog]. Dostupno na <https://blogs.microsoft.com/ai/2016/10/18/historic-achievement-microsoft-researchers-reach-human-parity-conversational-speech-recognition#sm.00000info02f8fihyc059kwdg4da>
-

- 
- Lourenço, T., Moura, R., Marques, J. i Tavares, P.** (2017). Outsider: A case study of the journey from design to meaning in videogames. *Proceedings of the 5th International Conference on Illustration and Animation CONFIA* (str. 523–530). Barcelos, Portugal: Instituto Politécnico do Cávado e do Ave.
- Löwgren, M.** (2011). *Immersion in location-based games* (Neobjavljena teza za magisterij). Skövde: University of Skövde.
- Martin, T.** (6. lipnja 2018). *25 fun Alexa games to play for your next dinner party*. CNET. Dostupno na <https://www.cnet.com/how-to/amazon-echo-fun-games-you-can-play/>
- McCann, D. i Thorne, P.** (2015). *The last of the first – CSIRAC: Australia's first computer*. Melbourne, Australija: The University of Melbourne.
- Orth, S. A.** (8. listopada 2020). *Mattel Electronics: Intellivoice games*. INTV Funhouse. Dostupno na <http://www.intvfunhouse.com/mattel/intellivoice.php>
- Parker, J. R. i Heerema, J.** (2008). Audio interaction in computer mediated games. *International Journal of Computer Games Technology*, 1–8. <https://doi.org/10.1155/2008/178923>
- Parker, L.** (13. siječnja 2015). *The voice-activated video game*. The Atlantic. Dostupno na <https://www.theatlantic.com/technology/archive/2015/01/the-voice-activated-video-game/384454/>
- Perry, D. i DeMaria, R.** (2009). *David Perry on game design: A brainstorming toolbox*. Boston: Charles River Media.
- Porcheron, M., Fischer, J. E., Reeves, S. i Sharples, S.** (2018). Voice interfaces in everyday life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (str. 1–12). New York, SAD: Association for Computing Machinery. <https://doi.org/10.1145/3173574.3174214>
- Rabiner, L. R.** (1994). Voice communication between humans and machines – An introduction. U D. B. Roe i J. G. Wilpon (ur.), *Voice communication between humans and machines* (str. 5–12). Washington, SAD: The National Academies Press. <https://doi.org/10.17226/2308>
- Rabiner, L. R. i Schafer, R. W.** (2011). *Theory and applications of digital speech processing* (1. izdanje). SAD, Upper Saddle River: Pearson.
- Roach, J.** (7. veljače 2017). Microsoft cognitive services push gains momentum [Blog]. Dostupno na <https://blogs.microsoft.com/ai/microsoft-cognitive-services-push-gains-momentum/>
-

- Szczepaniak, J.** (2014). The untold history of Japanese game developers. SMG Szczepaniak.
- Terrell, C. D. i Linyard, O.** (1982). Evaluation of electronic learning aids: Texas Instruments' "Speak Spell". *International Journal of Man-Machine Studies*, 17(1), 59–67. [https://doi.org/10.1016/S0020-7373\(82\)80008-4](https://doi.org/10.1016/S0020-7373(82)80008-4)
- van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... Kavukcuoglu, K.** (2016). *WaveNet: A generative model for raw audio*. arXiv. Dostupno na <https://arxiv.org/abs/1609.03499>
- van der Velde, N.** (23. svibnja 2018). Voice controlled games: The rise of speech technology in gaming [Blog]. Dostupno na <https://www.globalme.net/blog/voice-controlled-games>
- Wadley, G., Carter, M. i Gibbs, M.** (2015). Voice in virtual worlds: The design, use, and influence of voice chat in online play. *Human-Computer Interaction*, 30(3–4), 336–365. <https://doi.org/10.1080/07370024.2014.987346>
- Walz, S. P. i Ballagas, R.** (2007). Pervasive persuasive: A rhetorical design approach to a location-based spell-casting game for tourists. U A. Baba (ur.), *Situated Play: Proceedings of the 2007 Digital Games Research Association Conference*, vol. 4 (str. 489–497).
- Weske, J.** (2000). Digital sound and music in computer games. Dostupno na <http://3daudio.info/gamesound/> [posljednji pristup 10. kolovoza 2018.].
- Yu, D. i Deng, L.** (2014). *Automatic speech recognition: A deep learning approach*. London: Springer-Verlag. <https://doi.org/10.1007/978-1-4471-5779-3>
- Zyda, M.** (2005). From visual simulation to virtual reality to games. *Computer*, 38(9), 25–32. <https://doi.org/10.1109/MC.2005.297>
-

**Goran Paulin***gp@kreativni.hr*Creative Department Ltd, Rijeka  
Croatia**Marina Ivašić-Kos, Ivo Ipšić***marinai@inf.uniri.hr, ivoi@inf.uniri.hr*University of Rijeka, Department of Informatics  
Croatia

## Possibility of using speech in location-based computer games

### Summary

Although speech, in computer-synthesized form, became an integral part of computer games as early as 1978, its application, especially in the location-based computer game genre, has been barely explored.

In this paper, computer games and the role of sound and speech in them are discussed, focusing on location-based games and voice games. A brief introduction to speech technologies relevant in the context of applications in computer games is given, and current research relating to the analysis of computer games created in the period from 1978 to 2018, which uses speech, is presented.

To determine the possibility of using speech in location-based computer games, it is necessary to analyze the current practice of using speech technology in computer games in general and identify innovations that positively or negatively affect the overall rating of a particular game, recognizing which of these features apply to the development of location-based games, respecting their specifics. For this purpose, a sample of 50 games was formed that included games analyzed in previous research, supplemented with games found on the Internet by the criterion of using any form of speech specified in the reviews and documentation of individual games.

The conclusion answers the question of whether it makes sense, given the specifics of the genre and the current state of technology, to introduce voice technologies into location-based games and what are the prerequisites.

**Key words:** speech technology, computer games, location-based computer games, speech synthesis in computer games, automatic speech recognition in computer games

---