

## Katsaus

# Puheentunnistusteknologian nykytilanne ja mahdollisuudet elektronisissa peleissä

AKI HALONEN  
aki.halonen@utu.fi  
Turun yliopisto

SAMI HYRYNSALMI  
sami.hyrynsalmi@utu.fi  
Turun yliopisto

JOUNI SMED  
jouni.smed@utu.fi  
Turun yliopisto

### Tiivistelmä

Puheentunnistus on kypsynyt teknologisesti niin, että massamarkkinoille voidaan tuoda helppokäyttöisiä tuotteita. Peliteollisuus on jo aiemmin kunnostautunut ennakkoluuottomasti kokeilemaan uutta teknologiaa ja tuonut uusia innovaatioita markkinoille. Microsoftin Kinect-ohjelaitteen sisältämän valmiin puheentunnistusta tukevan ohjelmointirajapinnan myötä puheentunnistus saattaa olla seuraava merkittävä elektronisia pelejä uudistava ohjauselementti. Tässä katsauksessa tarkastellaan sekä puheteknologian kehittymistä, asiaa käsittelevää tieteellistä tutkimusta että vuosien saatossa julkaistuja puheohjausta hyödyntäviä kaupallisia pelejä. Lopuksi pohditaan lyhyesti tekniikan avoimia uusia mahdollisuuksia.

*Avainsanat: puheentunnistus, ohjaukseen liittyminen, käytettävyys*

### Speech Recognition in Video Games – Present and Future

#### Abstract

Speech recognition has reached the maturity needed for launching new, usable products for the mass market. In the past, the game industry has been open-minded in utilizing new technologies and brought new innovations to the market. A new auxiliary device in Microsoft's Kinect offers a usable application programming interface for speech recognition. Therefore, speech is a strong candidate for the next potential technological success in controlling interfaces. In this review, we study the development of speech technology, scientific research on the topic and commercially published games that utilize speech recognition. Finally, we briefly discuss new possibilities that the technology provides for the game industry.

*Key words: speech recognition, controlling interface, usability*

### Johdanto

Puheentunnistusta on teknologiana tutkittu ja kehitetty jo vuosikymmeniä ja sen teknisiä ongelmia voidaan pääosin pitää ratkaistuina (Juang & Rabiner 2005). Teknologinen kypsyys mahdollistaa käytettävien ja kaupallisesti elinvoimaisten ratkaisujen tuomisen massamarkkinoille. Uusien käyttöliittymäelementtien lisääminen olemassa oleviin elektronisiin peleihin ei ole kuitenkaan yksinkertaista, sillä niiden vaikutukset sekä käytettävyyteen että pelimekaniikkaan on tutkittava tarkoin. Uuden teknologian hyödyntämisen tarkoituksenmukainen tutkimus on yleisestikin tärkeää, sillä vuosikymmenten saatossa on nähty useita esimerkkejä epäonnistuneista teknologialanseerauksista. Esimerkiksi virtuaalitodellisuuskypärät ja -lasit

tuotiin 1990-luvulla markkinoille suurien mainosten saattamana, mutta laitteiden ensimmäiset sukupolvet katosivat nopeasti käyttökelttomina.

Nintendo Wii -konsolin liiketunnistusteknologian läpimurron myötä uudenlaiset, innovatiiviset ohjaustavat ovat saamassa jalansijaa video- ja tietokonepeleissä. Seuraava läpimurto ohjaustavoissa voi olla puheentunnistus. Liiketunnistusteknologia mahdollisti aivan uudenlaisten pelityyppien syntymisen; esimerkiksi erilaiset kuntoilu- ja liikuntapelit eivät olleet luontevia tai pelattavia perinteisillä käyttöliittymäelementeillä. Myös puheentunnistus avaa uudenlaisia mahdollisuuksia pelisuunnittelijoille. Yksinkertainen esimerkki tästä on karaokepeleissä sanojen tunnistaminen aiemmin käytetyn sävelkorkeuden lisäksi. Pidemmälle viedyissä sovelluksissa puheentunnistus voisi mahdollistaa suoran viestimisen pelihahmojen

kanssa.

Tässä katsauksessa esitellään puheentunnistusta videopeleissä sekä kaupallisten tuotteiden että tieteellisen tutkimuksen lähtökohdista. Aluksi esitellään lyhyesti puheentunnistustekniikan taustaa ja kehittymistä, minkä jälkeen tutustutaan puheentunnistusta käsittelevään akateemiseen tutkimukseen sekä käyttöliittymätutkimukseen että pelitutkimuksen näkökulmista. Esittelemme myös, millaisia puheen- tai äänen tunnistusta hyödyntäviä kaupallisia pelejä on julkaistu vuosikymmenten saatossa. Katsauksen lopuksi pohditaan lyhyesti, mitä uusi tekniikka voisi tarjota nykyisten kaltaisten pelien pelattavuudelle tai millaisia uusia pelityyppäjä se mahdollistaa.<sup>1</sup>

## Puheteknologian kehittyminen

**P**uhetta on pyritty tuottamaan koneellisesti ennen kuin puheentunnistus oli tekniikkana ajankohtainen. Jo 1700-luvulla tunnettiin äänenmuodostuksen periaatteet ja näitä tietoja hyödyntäen toteutettiin useampia puheenomaista ääntä tuottavia syntetisaattoreita. Aiemmin tehtyjä havaintoja puheen spektrin ominaisuuksista hyödynnettiin 1930-luvulla kehitetyssä manuaalisessa puhesyntetisaattori VODER:issa (Dudley, Riesz & Watkins 1939; Juang & Rabiner 2005). Nykyiset puheentunnistusjärjestelmät perustuvatkin pääosin juuri äänen spektrin mittaamiseen.

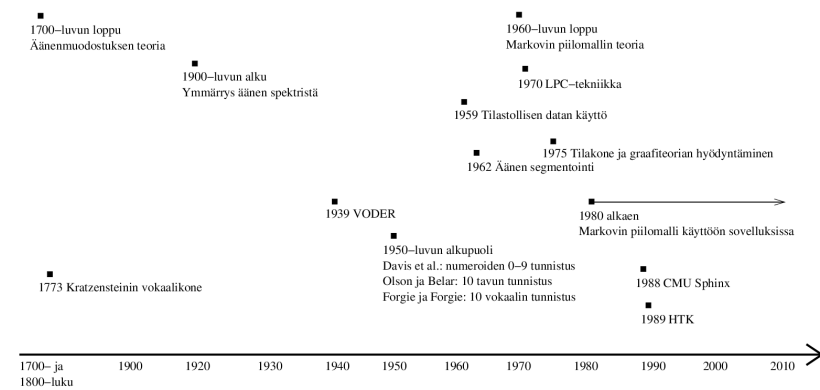
Ensimmäiset varsinaiset puheentunnistimet esiteltiin 1900-luvun puolivälissä. Davis, Biddulph ja Balashek (1952) kehittivät Bell-laboratoriossa järjestelmän, joka tunnistaa yksittäisen puhujan lausumat yksittäiset numerot nolasta yhdeksään. Olsonin ja Belarin (1956) järjestelmä tunnisti jo yksittäiseltä puhujalta kymmenen tavua. Samoihin aikoihin Forgie ja Forgie (1959) rakensivat puhujariippumattoman tunnistimen, joka kykeni erottamaan kymmenen vokaalia.

Nyky aikaisten tunnistusjärjestelmien kannalta Sakain ja Doshitan (1962) foneemitunnistin<sup>2</sup> toimi tärkeänä esiasteena äänen osiin jakamiselle ja näiden osien tunnistamiselle. Toinen merkittävä keksintö oli Fryn ja Denesin puheentunnistusjärjestelmään lisäämä tilastollinen tieto kielessä sallituista foneemijärjestyksistä. Tämä helpotti merkittävästi pidempien sanojen tunnistamista (Juang & Rabiner 2005).

Hahmontunnistusteknologian perusteet tulivat käyttöön puheentunnistuksessa 1970-luvulla. Äänneiden tunnistaminen puhesignaalista helpottui LPC-tekniikan (engl. Linear Predictive Coding) myötä. Tekniikalla puhenäytteen signaali muutetaan muotoon, jolla eri äänneiden tyypilliset piirteet on helpompi tunnistaa (Jurafsky & Martin 2000). Samoihin aikoihin esiteltiin myös ensimmäiset kaupalliset sovellukset. Alan kehittyminen sai aikaan sen, että Yhdysvaltojen puolustusministeriö ryhtyi rahoittamaan tutkimusta. Tuolloin saavutetuista tuloksista jäi elämään erityisesti tapa hyödyntää matemaattisen graafiteorian menetelmiä lauseiden tulkinnassa. Tässä lähestymistavassa pyrittiin mallintamaan kielioppia graafina, jonka pohjalta etsittiin lyhimpien polkujen menetelmällä lauseen todennäköisintä tulkintaa. (Juang & Rabiner 2005.)

Vaikka teoria Markovin piilomallin (engl. Hidden Markov Model) taustalla esiteltiin jo 1960-luvun lopulla, menetelmä saavutti suosiota vasta 1980-luvulla (Juang & Rabiner 2005). Nykyäänkin useat puheentunnistusjärjestelmät hyödyntävät juuri Markovin piilomallia, sillä 1980-luvulta alkaen painopiste on siirtynyt suoraviivaisesta hahmontunnistuksesta entistä enemmän tilastollisten mallien suuntaan.

Kuvassa 1 esitellään puheentunnistuksen merkittäviä virstanpylväitä ja sovelluksia aikajanalle sijoitettuna. Tekniikat ja teoriat on sijoitettu kuvan yläosaan, näiden sovellukset puolestaan lähemmäs alareunaa. Kuvasta voidaan havaita, että useimmat nykyäänkin sovellettavat merkittävät menetelmät kehitettiin hyvin lyhyen ajanjakson aikana 1960-luvun molemmin puolin.



Kuva 1.

Aikajanalle on sijoitettu puheentunnistuksen virstanpylväitä. Tekniikat ja teoriat on sijoitettu kuvan yläosaan ja sovellukset kuvan alaosaan.

Nykyaikainen puheentunnistus perustuu puhesignaalin, taajuustasossa ajan funktiona esitettävän analogisen signaalin, muuttamiseen diskreettiin muotoon. Tämä mahdollistaa sen käsittelemisen tietokoneella. Itse tunnistamista varten signaali pätkitään sopivan mittaisiksi palasiksi, eli *segmenteiksi*. Tämän jälkeen pyritään selvittämään, mitä foneemia tai foneemeja nämä segmentit vastaavat.

Segmenttien ja foneemien vastaavuutta selvitetään käyttäen spektrianalyysiä, jossa kustakin signaalisegmentistä aiemmin mainituilla käsittelymenetelmillä saatavaa spektriä verrataan tunnettujen foneemien tyypilliseen spektriin. Näin pyritään löytämään paras vastaavuus. Käytännössä siis signaalin energiaa mitataan eri taajuuksilla ja havaintoja verrataan tunnettuun näytekirjastoon. Koska puheen tapauksessa akustinen syöte ei kuitenkaan vastaa täydellisesti mallia, johon sitä verrataan, tarvitaan tunnistamisen avuksi tilastollisia algoritmeja, kuten aiemmin mainittua Markovin piilomallia.

## Puheentunnistusteknologian kehittyminen

Tiedeyhteisön parissa puheen- ja äänentunnistuksen yleistä toimivuutta peleissä on tutkittu jonkin verran, mutta tulokset teknologian hyödynnettävyydestä ovat olleet huomattavan vaihtelevia. Tutkimusten perusteella puheentunnistuksen hyödyt varsinkin toimintapelien ohjaamisessa ovat melko vaatimattomia, mikä johtuu lähinnä epätarkkuudesta ja vasteajasta. (Sporka, Kurniawan, Mahmud & Slavík 2006.) Toisaalta esimerkiksi Zhangin, Zhaon, Bain ja Hungin (2004) tutkimuksessa verkossa pelattavassa *Mahjong*-pelissä – jonka pyrkimyksenä oli saada toteutettua paremmin fyysistä peliä vastaava pelikokemus – puheentunnistus paransi pelijärjestelmän käytettävyyttä sekä *immersiota*, eli peliin eläytymisen tunnetta.

Puheentunnistuksen lisäksi Sporka ja kumppanit (2006) tutkivat mahdollisuutta käyttää pelin ohjaamiseen muita ääniä kuin puhetta. He testasivat esittelemäänsä menetelmää vertaamalla puhe- ja hyräilyohjausta *Tetrix*essä. Hyräily tuotti puheohjausta parempia tuloksia, mutta molemmat hävisivät selvästi tarkkuudessa perinteiselle näppäimistöohjaukselle. Samoin nykyisten hiiriohjaukseen perustuvien käyttöliittymien kontrollointi puheella on osoittautunut haastavaksi.

Puhtaasti pelejä käsittelevän tutkimuksen lisäksi on tutkittu myös erilai-

sia osoittimenkohdistusmenetelmiä puhe- ja ääniohjausta käyttäen: Harada, Landay, Malkin, Li & Bilmes (2006) saivat rohkaisevia tuloksia korvaamalla puhtaan puheohjauksen vokaaliohjauksella. Dain, Goldmanin, Searsin ja Lozierin (2004) tutkimuksessa perinteisen osoittimen korvaaminen vaihteittain tarkentuvalla ristikolla vaikutti tietyissä sovelluksissa toimivalta lähestymistavalta. Tässäkin tapauksessa puheohjauksen korvaaminen sanattomalla ääniohjauksella johti ohjauksen parantuneeseen nopeuteen ja tarkkuuteen (Chanjaradwichai, Punyabukkana & Suchato 2010).

Corradini, Hanneforth ja Bak (2007) sekä Corradini ja Samuelsson (2008) ovat esitelleet arkkitehtuuria *Pentomino*-pelin ohjaamiseen puheella. *Pentomino*-ongelmassa pyritään latomaan viiden neliön muodostamia kappaleita alustalle mahdollisimman tiiviisti. Neljän neliön kappaleilla pelattava *Tetris* on eräs *Pentominon* muunnos. Tuloksinaan he havaitsivat, että puhe on miellyttävä tapa ohjata peliä mutta aiheuttaa helposti ongelmia vasteajan ja verkkokommunikaation kanssa.

Hämäläinen, Höysniemi, Rouvi ja Turkki (2003) hyödynsivät äänentunnistusta lasten satupeli *KuKaKumma Muumaassa*, jossa pelaaja ohjaa hahmoa käyttäen koko kehoa ja ääntä – kamera tunnistaa liikkeen ja mikrofoni huutamisen. Teknologiaalintaa perusteltiin sillä, että fyysisten hallintalaitteiden puuttuminen auttaa lasta eläytymään peliin. Peliä on käytetty erityisesti lasten liikunnallisen kehityksen tutkimukseen.

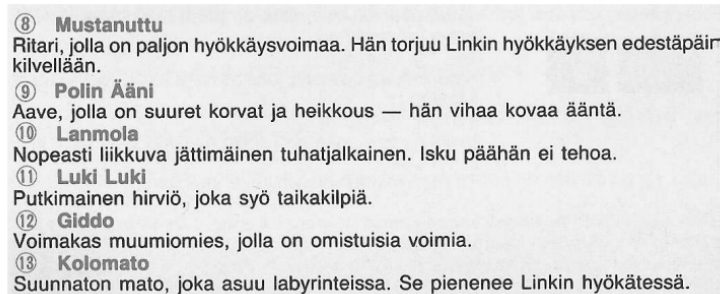
Yksi puheentunnistusteknologian haasteista on luonnollisten kielten laaja sanasto, joka vaikeuttaa puheen tulkitsemista. Rudra, Kavakli ja Bossomaier (2003), Rudra, Tien ja Kavakli (2005), sekä Rudra, Tien ja Bossomaier (2005) pyrkivät ratkaisemaan tämän ongelman kehittämällä tunnistimelle helpomman kielen. Lähtökohtana oli, että kieliopiltaan yksinkertaisempaa ja pienisanastoisempaa kieltä on yksinkertaisempi käsitellä. Rajoitettu sanasto ei ole käyttäjälle haasteellinen, kunhan järjestelmä toimii ja käyttäjä ymmärtää sen rajoitukset (Weinschenk & Barker 2000). Tällöin käyttäjän odotukset eivät nouse kohtuuttomiksi ja käyttökokemus on positiivinen.

Puheentunnistuksen vaikutuksia pelien käytettävyyteen on tutkittu vähän, vaikka puheteknologian yleistä käytettävyyssuunnittelua on ohjeistettu syvällisesti. Tosin Weinschenk & Barker (2000) huomauttavat puheohjattujen järjestelmien suunnittelua ohjeistavassa teoksessaan, että pelien tehtävä on haastaa käyttäjänsä

– toisin kuin hyötysovellutusten, joiden ainoa tarkoitus on auttaa käyttäjää toteuttamaan tehtävänsä. Tämän lisäksi käyttäjä pystyy vapaasti valitsemaan pelit, joita hän pelaa, mutta työpaikoilla käytettäviin sovelluksiin ei yleensä voi vaikuttaa. Tästä johtuen pelit vaativat erityistä huomiota käytettävyyden suunnittelussa. Halonen ja kumppanit (2012) ovat esitelleet alustavan käytettävyyssheuristiikan puheentunnistusta hyödyntävien pelien arvioimiseen.

## Puheen- ja äänentunnistus kaupallisissa peleissä

Puheen- ja äänentunnistusta on ennakkoluulottomasti kokeiltu tietokone- ja konsolipeleissä vuosien saatossa. Yksi ensimmäisistä äänentunnistusta hyödyntävistä peleistä olikin *The Legend of Zelda* alkuperäinen Japanissa julkaistu Nintendo Family Computer (*Famicom*) -versio (Nintendo EAD, 1986). Laitteen suorituskyky ei luonnollisesti riittänyt varsinaisen puheen käsittelyyn, minkä vuoksi peli tunnistikin vain mikrofoniin puhutun äänen voimakkuuden. Erityisiksi Famicom-pelit tekee se, että osa niistä julkaistiin läntisillä markkinoilla teknisesti käytännössä identtiselle Nintendo Entertainment Systemille (NES), jossa ei kuitenkaan ollut peliohjaimeen sisäänrakennettua mikrofonia. Tämä johti esimerkiksi monia amerikkalaisia ja eurooppalaisia pelaajia hämmäntäneeseen vihjeeseen *Legend of Zelda* ohjekirjassa (Kuva 2), jossa Polin Ääni -vihollisen heikkoudeksi mainitaan kovat äänet. Famicom-versiossa nämä viholliset saattoi kukistaa huutamalla peliohjaimeen mikrofoniin.



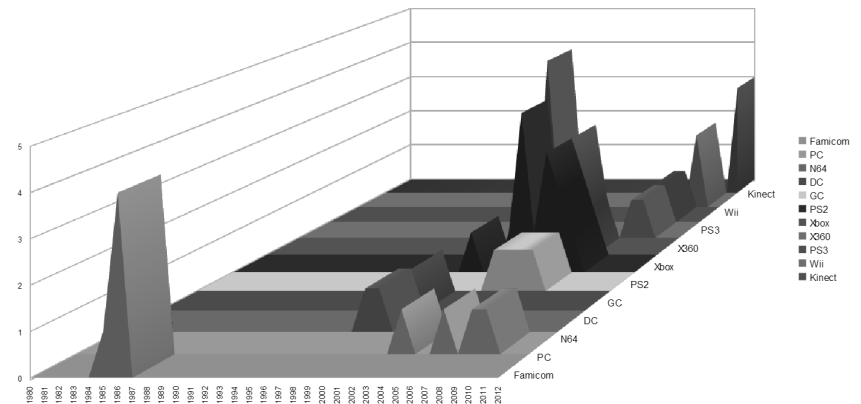
Kuva 2.

Polin Ääni -vihje hämmänsi 1980-luvulla pelaajia NES:n *Legend of Zelda* ohjevihkosessa. Teksti oli jäänyt japanista käännettyihin ohjeisiin, vaikka länsimaisissa konsoleissa ei ollut mikrofonia kuten japanilaisessa versiossa.

Puheentunnistuksen käyttö kaupallisissa tietokonepeleissä on jäänyt varsin vähäiseksi, mutta tästä huolimatta puheen- tai äänentunnistusta hyödyntäviä pelejä on julkaistu käytännössä kaikille 2000-luvun alun alustoille. Suosituimmissa ääniohjattavissa peleissä, kuten *Singstar*-sarjan laulupeleissä, tunnistaminen perustuu pääosin sävelkorkeuden tarkkailulle. Varsinaisesti puhetta tunnistavia pelejä on vain muutamia. Uudemmissa peleistä muutaman vuoden takaisessa *Tom Clancy's Endwarissa* (Ubisoft Shanghai 2009) on mahdollista korvata pitkätkin näppäinkomentosarjat äänikomennoilla. Yksinkertaisena esimerkkinä puheen käytöstä voidaan mainita myös Nintendo DS -käsikonsolille miljoonia myynyt koirankoulutuspelejä *Nintendogs* (Nintendo EAD 2005).

On olemassa myös joukko pelejä, kuten laulupelit, joita voi pelata ainoastaan äänen avulla (Hämäläinen, Mäki-Patola, Pulkki & Airas 2004). Kuten aiemmin mainitsimme, näistä kenties tunnetuin, SingStar-pelisarja (SCE London Studio 2004–), ei tosin tunnista pääosin niinkään sanoja vaan lähinnä äänenkorkeutta (Gamepro 2010). Myös pelkällä äänenkorkeudella voi ohjata peliä: näin toimii esimerkiksi vuoden 2006 Assembly-tapahtuman pelinkehityskilpailuun osallistunut *Racing Pitch* (Lauha 2006).

Kuvassa 3 on listattu julkaistuja puheen- tai äänentunnistusta hyödyntäviä pelejä sijoitettuna aikajanelle ja jaoteltuna eri pelialustojen mukaan. Taulukkoon 1 on poimittu muutamia esimerkkipelejä vuosien varrelta. Halonen (2012) on esittänyt täydellisemmän listauksen puheen- tai äänentunnistusta hyödyntävistä peleistä.



Kuva 3.

Puheen- tai äänentunnistusta hyödyntäviä pelejä sijoitettuna aikajanelle ja jaoteltuna eri alustojen mukaan

	Vuosi	Alusta	Julkaisija
Raid on Bungeling Bay	1985	Famicom	Hudson Soft
Atlantis no Nazo (lännessä Super Pitfall 2)	1986	Famicom	Sunsoft
Zelda no Densetsu (lännessä Legend of Zelda)	1986	Famicom	Nintendo
Hey You, Pikachu	1998	N64	Nintendo
Seaman	1999	DC, PS2	Sega
Densha de Go! 64	1999	N64	Taito
SOCOM	2002	PS2	Sony
Rogue Ops	2003	Xbox, PS2, GC	Kemco
Lifeline	2003	PS2	Konami
SWAT: Global Strike Team	2003	Xbox	Sierra
Rainbow Six 3	2003	Xbox	Ubisoft
Mario Party 6	2004	GC	Nintendo
Ghost Recon 2	2004	Xbox, PS2	Ubisoft
NASCAR o6 Total Team Control	2005	Xbox, PS2	EA
Mario Party DS	2007	DS	Nintendo
Seaman 2	2007	PS2	Sega
Tom Clancy's Endwar	2008	PS3, X360, PC	Ubisoft
Wheel of Fortune	2010	Wii	THQ

**Taulukko 1.**

Esimerkkejä vuosien varrella julkaistuista puheen- tai äänentunnistusta hyödyntävistä peleistä

On kehitetty myös apuohjelmia, jotka tulkitsevat puhekomennot pelin näppäinkomennoksi. Näin puhetta voidaan käyttää peleissä, jotka eivät itse tue sitä suoraan. Yhtä tällaista ohjelmaan, kaupallista VAC Systemiä<sup>3</sup>, arvioitiin *Pelit-*lehden testissä hieman hankalaksi käyttää mutta toimiessaan mielenkiintoiseksi lisäksi erityisesti simulaattoripeleissä (Ilomäki 2011).

Useimpien puheentunnistusta hyödyntävien pelien kaupallinen menestys on ollut melko vaatimatonta. Niiden haaste ovat olleet alustan rajoitukset, sillä usein ääniohjaus on vaatinut erillisen lisälaitteen hankkimisen. Vasta äskettäin julkaistu Microsoftin XBOX 360 Kinect -lisälaite on ensimmäisenä saavuttanut riittävän markkinaosuuden (Stevens 2011), että sille kannattaa toteuttaa kaupallisia pelejä. Lisäksi oheislaitte tarjoaa pelinkehittäjille valmiin puheentunnistus-rajapinnan, joka helpottaa uuden ohjaustavan integroimista jo olemassa oleviin

tuotekonsepteihin.

Vuoden 2011 E3-messuilla esiteltiin vihdoin puheentunnistusta hyödyn- täviä Kinect-pelejä, jotka ilmestyvät vuoden 2012 aikana. Näitä olivat esimerkiksi *Mass Effect 3* (BioWare/EA) ja *Tom Clancy's Ghost Recon Future Soldier* (Ubisoft).

Puheohjattavia pelejä on kuitenkin kaiken kaikkiaan kovin vähän. Lie- nee hedelmällistä pohtia tilanteeseen johtaneita syitä. Ennen 1990-luvun lopulla julkaistuja kuudennen sukupolven konsoleja laskentatehon puute saattoi aiheuttaa ongelmia tarkan puheentunnistuksen kanssa. Toisaalta uuden konse- pin yleistyminen vaatii onnistuneen ja menestyvän pioneerin, jotta useat kehittäjät uskaltavat lähteä tekemään ääniohjattavia pelejä. Liikeohjauksen puolella tällainen pioneeri oli selvästi Nintendo Wii.

Konsolipuolella syynä on varmasti myös se, että puheohjauksen vaatima mikrofoniohjain ei ole kuulunut käytännössä minkään laitteen vakiokokoonpa- noon. Näin potentiaalinen ostajakunta on ollut niin rajoitettu, että puheohjaukseen panostaminen ei ole ollut taloudellisesti järkevää. Poikkeus tähän ovat Famicom sekä tuoreemmat Nintendo DS -käsikonsolit, joissa on sisäänrakennettu mikro- foni. Tämän johdosta ääntä käytetäänkin ohjausmetodina selvästi useammin kuin muilla laitteilla.

Ääni- ja puheohjattavissa peleissä ei ole vielä nähty samanlaista tarjonnan räjähtämistä kuin liikeohjattavissa peleissä Wiin osoittauduttua valtavaksi myyn- timenestykseksi. Kinect saattaa toimia samankaltaisena pioneerina.

## Tulevaisuuden mahdollisuudet

Liiketunnistusohjauksen läpimurron myötä erilaiset liikuntapelit ja nopeatem- poiset fyysisiin suorituksiin perustuvat bilepelit ovat menestyneet, sillä tämän- kaltaisten pelien toteuttaminen perinteisellä peliohjaimilla ei ole ollut mielekästä. Toisaalta liiketunnistuksella ohjaaminen on yhdistetty osaksi perinteisempiä peli- tyyppisiä vaihtelevin tuloksin. Samoin on käsiteltävä puheentunnistusta: olemas- sa olevia ja vakiintuneita ohjausmenetelmiä ei kannata korvata puheohjauksella pelkästään siksi, että se on mahdollista, vaan on hedelmällisempää etsiä täysin uusia pelityyppejä ja pyrkiä toimimaan genrensä pioneerina.

Täysin uusi ohjaustapa mahdollistaa uudenlaiset pelikonseptit. Onneksi

peliteollisuus on historiansa aikana ottanut ennakkoluulottomasti käyttöön uusia teknologioita ja tekniikoita sekä tehnyt uusia avauksia. Pelaajille tämä antaa luvan odottaa tulevaisuudelta uusia pelityyppejä. Yksinkertaisimmillaan karaokepeleihin voidaan tulevaisuudessa liittää todellinen laulutunnistin, vierasta kieltä opiskeleville palautejärjestelmä voisi olla osa tietokonepeliä tai roolipeleihin voitaisiin lisätä mahdollisuus syventää immersiota kommunikoimalla suoraan pelihahmojen kanssa. Monimutkaisemmissa konsepteissa pelaaja pääsisi toimimaan esimerkiksi ralliauton kartanlukijana. Lisäksi teknologia voi avata uusia markkinasegmenttejä, joita peliteollisuus ei ole vielä saavuttanut, kuten näkö- tai toimintarajoitteisille tarkoitettut puhtaasti äänipelit, joissa sekä ohjaus että palaute toimivat äänellä (Rokosa 2011).

Puheentunnistusteknologian rajoitteena on sen vaatima suhteellisen häiriötön ja häiriintymätön peliympäristö. Perinteisillä ohjaimilla pelejä on voinut pelata ilman, että ne aiheuttavat ympäristöön minkäänlaista melua, mutta puhe- ja ääniohjauksen kanssa tämä ei ole mahdollista. Vastaavasti taustäännet saattavat johtaa virhetunnistuksiin, jotka haittaavat pelaamista.

Puheentunnistusteknologia on kuitenkin vain yksi osa kokonaisuutta, sillä olemassa olevien pelien puheentunnistus perustuu usein vain yksinkertaiseen avainsanojen havaitsemiseen. Luontevan kommunikaation saavuttamiseksi puheentunnistuksen tulisi olla monipuolisempaa kuin nykypeleissä. Tämä vaatii parempia luonnollisen kielen käsittelymenetelmiä (engl. Natural Language Processing, NLP), jolloin pelaajan puheen merkitys pystytään tulkitsemaan oikein, vaikka pelaaja ei käyttäisi juuri jotain tiettyä avainsanaa. Esimerkkejä vastaavasta teknologiasta ovat esimerkiksi Googlen hakukoneen käyttämän kaltaiset tekstiparserit tai Applen *Siri*-sovellus, jossa menetelmiä käytetään puheentunnistuksen yhteydessä. Tässä katsauksessa ei kuitenkaan syvennyttä NLP-menetelmiin, joiden kehittymistä ja soveltamista ovat käsitelleet laajasti Jurafsky & Martin (2000).

## Yhteenveto

Puheohjausta on tutkimuksissa pidetty miellyttävänä ja intuitiivisena ohjaustapana mutta haasteellisena tarkkuutta vaativissa nopeatempoisissa peleissä. Puheentunnistusteknologia kuitenkin mahdollistaa useamman asian hallitsemisen

samanaikaisesti, jolloin pelaaja voi toimia tehokkaammin kuin pelkien motoristen ohjainten avulla. Peleissä, joiden esittämissä tapahtumissa puhekommunikaatio on keskeisessä roolissa, järkevästi toteutetulla puheohjauksella on mahdollista parantaa pelin immersiota huomattavasti.

Kaupallisissa peleissä puhe- ja ääniohjausta on käytetty suhteellisen vähän, vaikka puheohjausta hyödyntäviä pelejä on kyllä julkaistu kaikille viime vuosikymmenen pelialustoille. Vaikuttavina tekijöinä ovat olleet epäilemättä rajallinen laskentateho ja erityisesti konsolipuolella puutteellinen laitteistotuki ja viime kädessä menestyvän pioneerin puuttuminen.

Puheentunnistus tarjoaa uusia mahdollisuuksia pelikonsepteille ja läpimurtoja ennen saavuttamattomissa olleissa kuluttajasegmenteissä. Uutena ohjausmenetelmänä puheentunnistus mahdollistaa myös uudenlaiset pelityypit, jotka eivät välttämättä suoraan kilpaile perinteisillä ohjaimilla pelattavien pelien kanssa vaan luovat oman tilansa markkinoille. Tekniset haasteet on suurimmalta osin ratkaistu. Seuraava siirto onkin pelinkehittäjiä.

## Viitteet

- 1 Katsaus perustuu suoraan Aki Halosen diplomityössä *Puheentunnistuksesta saatava lisäarvo tietokonepeleissä* tehtyyn tutkimukseen.
- 2 Foneemi on äänijärjestelmän pienin merkitystä erottelava yksikkö.
- 3 Voice activated commands, <<http://dwwac.com>>.

## Lähteet

### SANOMA- JA AIKAKAUSLEHDET

Stevens, Tim (2011). Microsoft sells 10 million Kinects, 10 million Kinect games. <http://www.engadget.com/2011/03/09/microsoft-sells-10-million-kinects-10-million-kinect-games/>, Luettu 27.3.2012.

Ilomäki, Antti (2011). VAC ON, VAC OFF. *Pelit*, 5, 55.

GamePro (2010). Interview with SingStar Producer Paulina Bozek. [http://www.gamepro.com.au/article/9528/interview\\_singstar\\_producer\\_paulina\\_bozek/](http://www.gamepro.com.au/article/9528/interview_singstar_producer_paulina_bozek/), Luettu 7.10.2010.



## PELIVIITTEET

Aleksei Pažitnov. *Tetris*. 1984.

Nintendo EAD. *Nintendogs*. Nintendo, 2005.

Perttu Hämäläinen, Johanna Höysniemi, Teppo Rouvi & Laura Turkki (2003). *KuKaKumma Muumaassa*. <http://www.cs.uta.fi/kukakumma/htmls/mitaih/frset.html>

Nintendo EAD (1986). *Zelda no Densetsu*. Nintendo.

Ubisoft Shanghai (2009). *Tom Clancy's Endwar*. Ubisoft. <http://endwargame.us.ubi.com>.

SCE London Studio (2004). *SingStar*. Sony Computer Entertainment. <http://www.singstargame.com>.

Jetro Lauha (2006). *Racing Pitch*, <http://secretexit.com/freeware/racingpitch>.

Bioware (2012). *Mass Effect 3*. Electronic Arts.

Ubisoft Paris ja Ubisoft Red Storm. *Tom Clancy's Ghost Recon: Future Soldier*. Ubisoft, (julkaistaan 2012)

## TUTKIMUSKIRJALLISUUS

Chanjaradwichai, Supadaech, Proadpran Punyabukkana & Atiwong Suchato (2010). Design and evaluation of a non-verbal voice-controlled cursor for point-and-click tasks. Teoksessa *Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology*. Singapore Therapeutic, Assistive & Rehabilitative Technologies Centre, 48:1–48:4.

Corradini, Andrea & Christer Samuelsson (2008). A generic spoken dialogue manager applied to an interactive 2D game. Teoksessa E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini & M. Weber (toim.): *Perception in Multimodal Dialogue Systems*. Heidelberg: Springer Berlin, 2–13.

Corradini, Andrea, Thomas Hanneforth & Adrian Bak (2007). A robust spoken language architecture to control a 2D game. Kirjassa *FLAIRS Conference'07*. 199–204.

Dai, Liwei, Rich Goldman, Andrew Sears & Jeremy Lozier (2004). Speech-based cursor control: a study of grid-based solutions. Teoksessa *Proceedings of the 6th International ACM SIGACCESS*

*Conference on Computers and Accessibility*. New York: ACM Press, 94–101.

Davis, K.H., R. Biddulph & S. Balashek (1952). Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, 24:6, 637–642.

Dudley, Homer, R.R. Riesz & S.S.A. Watkins (1939). A synthetic speaker. *Journal of the Franklin Institute*, 227:6, 739–764.

Forgie, James W. & Carma D. Forgie (1959). Results obtained from a vowel recognition computer program. *The Journal of the Acoustical Society of America*, 31:6, 844–844.

Halonen, Aki, Sami Hyrynsalmi, Kai K. Kimppa, Timo Knuutila, Jouni Smed & Harri Hakonen (2012). Towards Usability Heuristics for Games Utilizing Speech Recognition. Teoksessa M. Inaba, K. Hosoi, R. Thawonmas, A. Nakamura, M. Uemura (toim.): *4th Asian Conference on Simulation and AI in Computer Games & 4th Asian Simulation Technology Conference*. Eurosis-ETI, 65–69.

Halonen, Aki. (2012) *Puheentunnistuksesta saatava lisäarvo tietokonepeleissä*. Diplomityö. Turun yliopisto.

Harada, Susumu, James A. Landay, Jonathan Malkin, Xiao Li & Jeff A. Bilmes (2006). The vocal joystick: Evaluation of voice-based cursor control techniques. Teoksessa *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. New York: ACM Press, 197–204.

Hämäläinen, Perttu, Teemu Mäki-Patola, Ville Pulkki & Matti Airas (2004). Musical computer games played by singing. Teoksessa *Proceedings of the 7th International Conference on Digital Audio Effects*, Federico II University of Naples, Italia, 367–371.

Juang, B.H. & L.R. Rabiner (2005). Automatic speech recognition – a brief history of the technology. Teoksessa *Encyclopedia of Language and Linguistics*, toinen laitos, Elsevier.

Jurafsky, Daniel & James H. Martin (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Ensimmäinen laitos. Prentice Hall.

Olson, Harry F. & Herbert Belar (1956). Phonetic typewriter. *The Journal of the Acoustical Society of America*, 28:4, 767–767.

Roberts, Lawrence G (1967). Multiple computer networks and intercomputer communication. Teoksessa *Proceedings of the first ACM symposium on Operating System Principles*. New York: ACM, 3:1–3:6.

Rokosa, Ari (2011). *Äänipelit pelinkehittäjän näkökulmasta*. Pro gradu -tutkielma. Turun yliopisto.

Rudra, T., D. Tien & M. Kavakli (2005). Analysis of conversational game pidgin language. Kirjassa *TENCON 2005 2005 IEEE Region 10. IEEE*, 1–4.

Rudra, Tarashankar, David Tien & Terry Bossomaier (2005). Spoken communication with computer game characters. Teoksessa *Third International Conference on Information Technology and Applications*. IEEE, 494–497.

Rudra, Tarashankar, Manolya Kavakli & Terry Bossomaier (2003). A game pidgin language for speech recognition in computer games. Teoksessa G. Mitchell (toim.): *Proceedings of 3rd International Conference on Computational Semiotics for Games and New Media*. University of Tesside, UK: 90–98.

Sakai, T. & Shuji Doshita (1962). The phonetic typewriter. Teoksessa C.M. Popplewell (toim.) *Proceedings of IFIP Congress 1962*. IFIP, 445–450.

Sporcka, Adam J., Sri H. Kurniawan, Murni Mahmud & Pavel Slavík (2006). Non-speech input and speech recognition for real-time control of computer games. Teoksessa *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. New York: ACM Press, 213–220.

Weinschenk, Susan & Dean T. Barker (2000). *Designing effective speech interfaces*. New York: John Wiley & Sons, Inc.

Zhang, Jie, Ji Zhao, Shuanhu Bai & Zhiyong Huang (2004). Applying speech interface to mahjong game. Teoksessa Y-P.P. Chen (toim.) *10th International Multi-Media Modeling Conference*. IEEE Computer Society, 86–92.