

Tiina Männistö-Funk

Tiina Männistö-Funk, dosentti,  
ETH Zurich, Chair for Science  
Studies

# PUHETTA KONEISTA

## Ihmismäistä ääntä käyttävät teknologiat teoreettisessa tarkastelussa



*Koneellisesti tuotettu puhe on sekä teknisesti että sosiaalisesti erityinen teknologian ominaisuus. Puhuvia koneita olisi syytä tarkastella omana sosioteknisenä kategorianaan, koska ihmismäiseen puheeseen liittyy koneen ja käyttäjän suhteen näkökulmasta monia kysymyksiä. Nämä ovat toistaiseksi saaneet melko vähän tutkimuksellista huomiota osakseen. Artikkelissa määritellään puhuvien koneiden olemusta toimijuuden, toiseuden ja affektin käsitteiden avulla.*

Elämme tekniikan ympäröiminä, teknisesti muokatussa maailmassa ja ympäristössä. Käytämme erilaisia koneita ja laitteita lähes jatkuvasti, mutta yleensä emme kiinnitä itse laitteisiin erityisen paljon huomiota. Ne näyttävät välineinä, joita voimme hyödyntää ja joiden avulla voimme tehdä erilaisia asioita. Mutta mitä tapahtuu tälle välinsuhteelle, kun koneet puhkeavat puhumaan? Puhuvien koneiden pohtiminen johdattaa meidät miettimään tekniikkaan liittyviä suhteita ja odotuksia sekä keinotekoisien ihmisäänen herättämiä tunnereaktioita. Tässä artikkelissa määritellään puhuvien koneiden kategoria koneen ja käyttäjän suhteen näkökulmasta sekä pohditaan puhuvien koneiden herättämiä tunteita affektiivisena ilmiönä.

### Johdanto

Marraskuisena iltana vuonna 2019 ruumiiton, monotoninen ääni käskee minua ruokaostoksilla: "Laita tuote kassiin, ole hyvä." Olen Turun keskustassa sijaitsevan marketin automaattikassalla maksupuuhissa. Ympärilläni kaikuu kassojen puheen kakofonia, kun sekä omani että muita asiakkaita palvelevat kassat toivottavat tervetulleeksi, neuvovat ostosten käsittelyssä sekä maksamisessa ja hyvästelevät, kaikki samalla naisäänellä, samoin lausein. Vaihtelua yksitoikkoiseen puhemaisemaan tuo yksi kassoista, joka puhkeaa puhumaan käyttäjälleen englanniksi, erilaisella äänellä. Puhuvat kassat hallitsevat tilaa ja tilannetta konkreettisesti äänensä kautta. Niillä on myös kyky esimerkiksi paljastaa käyttäjänsä kielipreferenssit kaikille muille samassa tilassa oleville. Puheominaisuuden avulla automaattikassoille on pyritty antamaan inhimil-

lisen kassatyöntekijän piirteitä, mutta samalla niille on annettu tukku monimutkaisia toimijuuden mahdollisuuksia.

Tällä hetkellä meillä on arjessamme suhteellisen pieni määrä puheäänellä viestiviä koneita, vaikka puheominaisuus olisi teknisesti mahdollista liittää lähes mihin tahansa laitteeseen tai esineeseen. Verrattuna moneen muuhun koneelliseen ääneen koneiden puhe ei ole kovin hallitseva osa arkista ääniympäristöömme. Samalla se on kuitenkin tuttu osa äänimaisemaamme. Puhuvat kassat ja hissit tai juna-asemien ja muiden julkisten tilojen automaattikuulutukset eivät hätkähdytä meitä. Ne eivät saa meitä etsiskelemään ruumiittoman äänen lihallista lähdeä. Ymmärrämme, mitä ne ovat ja mikä niiden merkitys on. Termillä äänimaisema viitataan siihen, miten yksilö ja yhteisö ymmärtävät äänet osana arkista ympäristöään (Ampuja et al. 2005). Outi Ampujan (2007) mukaan teknologialla on ollut suuri rooli äänimaiseman murroksessa 1900-luvulta lähtien. Hänen mukaansa olemme erityisesti 1950-luvulta lähtien siirtyneet yhä voimakkaammin ”keinotekoiseen äänimaisemaan”, jota leimaavat teknologisesti tuotetut, jokapaikkaiset ja pitkäkestoiset äänet. Kaupungeissa tällaista ubiikkia koneääntä tuottavat tyypillisesti autot (Ampuja 2011, 11). Koneellisella puheella ei ole samanlaista akustisesti hallitsevaa roolia, mutta se on kuitenkin vakiintunut osaksi äänimaisemaamme tietyissä meille tutuissa yhteyksissä.

Koneäänet voidaan ymmärtää akustisena kommunikaationa, jossa Barry Truaxin (2001, 50) mukaan erityyppiset äänet ja hiljaisuudet muodostavat liukuvan jatkumon. Esimerkiksi ympäristön äänien, synteettisten äänien ja puheen välillä on kylläkin kommunikatiivisia eroja, mutta ei selkeästi määriteltäviä rajoja. Käyntiäänet ja muut koneiden toiminnan oheistuotteina syntyvät äänet osallistuvat ihmisen ja koneen suhteiden muokkaamiseen, saavat kulttuurisia merkityksiä ja voivat antaa tietoa esimerkiksi koneen toiminnasta, vaikkei se olisi ollut koneen suunnittelussa varsinaisesti tarkoituksena. Tämän lisäksi monissa koneissa on erilaisia merkkiääniä; piippauksia, hälytysääniä ja muita äänisignaaleja, jotka on erityisesti tarkoitettu kommunikaatioksi käyttäjän suuntaan. Kuulijan yksilöllisten valintojen lisäksi riippuu pitkälti kontekstista ja ympäristöstä, mille äänille annetaan huomiota, miten ne kuullaan ja miten niitä tulkitaan (Uimonen 2005, 67–75; Thompson 2002). Puhuva kone on kuitenkin tällä kuuntelun kommunikaation jatkumolla erityinen tapaus, sillä se puhuttelee aktiivisesti ihmistä, monissa tapauksissa myös kuuntelee ja kutsuu kommunikoimaan kanssaan. Tekniikkasuhteen käsittelyssä ihmisten ja koneiden välinen kommunikaatio on saanut osakseen verrattain vähän huomiota, vaikka se on jatkuva osa arkista elämäämme (Guzman 2017). Ihmismäisellä äänellä puhuvien koneiden tarkastelu auttaa meitä ymmärtämään paremmin paitsi koneellista puhetta kommunikaation muotona myös ihmisten ja koneiden suhdetta ja jaettua toimijuutta ylipäätään.

Koneellisen toimijuuden havainnointi kutsuu pohtimaan, mitä tapahtuu, kun koneet puhuvat ihmisille. Millaiseksi muodostuu ihmisen ja koneen suhde? Millainen kone on ihmismäisellä äänellä puhuva kone muihin koneisiin verrattuna? Mitkä ovat sen erityispiirteet? Kysymyksiä lähestytään tässä artikkelissa tarkastelemalla aluksi erilaisia puhetta tuottavien koneiden tyyppisiä, näiden tyyppien historiallisia kehityskulkuja sekä niiden ja käyttäjiensä suhdetta. Tämän jälkeen määritellään, millaiset puhetta tuottavat koneet voidaan luokitella puhuviksi koneiksi ja millä perusteilla puhuvat koneet muodostavat oman sosioteknisen kategoriansa. Termillä sosiotekninen tarkoitetaan sosiaalisten ja teknisten tekijöiden yhteisvaikutusta teknologioiden kehittämisessä ja käyttämisessä sekä näiden tekijöiden muokkautumista

vuorovaikutuksessa toistensa kanssa. Teknologian toimivuus on siis paitsi tekninen myös sosiaalinen kysymys, ja yhteiskunta ja teknologia ovat monimutkaisessa vuorovaikutussuhteessa toisiinsa. (Bijker 1997.)

Tarkastelen kirjallisuuden ja lehdistön tarjoamia esimerkkejä puhuvista koneista ja muotoilen näiden sekä tutkimuskirjallisuuden perusteella teoreettista pohjaa puhuvien koneiden sosiotekniselle tarkastelulle. Artikkelin ensimmäisessä osassa esittelen ihmismäistä ääntä hyödyntävien koneiden teknisen tyyppittelyn ja alalajit historiallisen tarkastelun perusteella, minkä jälkeen tarkastelen seuraavissa kahdessa osassa puhuvien koneiden kategorian rajoja, yhtäältä koneen toimijuuden ja käyttäjän sekä koneen suhteen kautta, toisaalta käyttäjän tunnereaktioiden kannalta. Tarkastelen ihmisten ja puhetta tuottavien koneiden suhdetta sosioteknisestä ja kulttuurisesta näkökulmasta esimerkiksi Bruno Latourin ja Karen Baradin teoreettista ajattelua hyödyntäen. Ehdotan, että puhuvien koneiden kategoriaa ei voi perustaa tietynlaisiin teknologisiin ratkaisuihin tai konemuotoihin, vaan tietynlaiseen koneen ja ihmisen suhteeseen, jolla on muovaavia vaikutuksia myös käyttäjäänsä. Tämän jälkeen käsittelen puhuvia koneita affektiivisuuden näkökulmasta. Affektit ovat ruumiiden rajat ylittäviä vaikutuksia ja tiloja, jotka saavat ilmiänsä tunnereaktioiden kautta. Affektiivinen vuorovaikutus ei rajaudu vain ihmisruumiisiin, vaan affektit voivat olla ihmisen ja jonkin ei-inhimillisen välille syntyvän välityssuhteen ilmentymä (Seyfert 2012). Väitän, että puhuvilla koneilla on omia erityispiirteitään sekä teknisen toimijuuden että affektiivisen välityssuhteen kannalta.

## Konepuheen historiaa ja tyypejä

Elokuussa 1846 *The Times* uutisoi itävaltalaisen Joseph Faberin kehittämästä ihmisenkaltaisesta, puhuvasta koneesta, Euphoniasta, jota lontoolaiset saattoivat katsella ja kuunnella Piccadillyn Egyptian Hall -näyttelytilassa. Lehden mukaan Faber oli onnistunut yli 25 vuoden kehitystyöllä ”luomaan instrumentin, joka puhuu selvästi, ymmärrettävästi ja tarkasti sanoja ja lauseita englanniksi, italiaksi, ranskaksi, saksaksi jne., laulaa kuningashymnin ‘God save the Queen’ sekä saksalaisen virren tai laulun ja nauraa iloisen hyväntuulisesti.” (*The Times* 12.8.1846.) Euphonia on yksi puhuvien koneiden historian merkkipaaluista ja ilmentää monia puhuviin koneisiin liittyneitä teknisiä ja kulttuurisia erityispiirteitä, joita käsittelen seuraavassa.

Äänen tuottamiseen tarkoitettuja koneita on erilaisia, musiikki-instrumenteista hälyttimiin ja signaalijärjestelmistä leluihin. Puhetta tuottavat koneet erottuvat muista ääntä tuottavista koneista historiallisesti omaksi kategoriakseen jo siinä mielessä, että niitä oli ennen 1800-luvun loppua hyvin vaikea valmistaa, kuten yllä kuvatun Euphonian poikkeuksellisuuskin osoittaa. Euphonia oli ihmisen kokoinen ja näköinen nukke, jonka kumisia huulia, kieltä ja suuonteloa ohjailtiin flyygeliä muistuttavalla koskettimistolla, samalla kun sen keuhkoihin pumpattiin ilmaa polkimilla. Sitä pidetään ensimmäisenä tunnettuna koneena, joka pystyi tuottamaan kaikki keskeiset ihmismäisen puheen äänneet ja yhdistämään ne ymmärrettäviksi lauseiksi (Riskin 2016). Ajatus puhuvista koneista ja halu kehittää niitä ovat kuitenkin hyvin vanhoja kulttuurisia trooppeja. Jo antiikin ajoilta ja keskiajalta tunnetaan sekä myyttejä puhuvista päistä että puhuvia patsaita, jotka saatiin äänteleämään tai puhumaan esimerkiksi johtamalla niihin ilmaa tai ihmisen puhetta erilaisia tekniikoita apuna käyttäen (Pettorino 2015, 32–38). Nämä puhuvat koneet

voidaan nähdä alalajina koneihmisille, joita on kuviteltu ja rakennettukin vuosisatojen ajan (ks. Parikka 2004, 18). Sekä tarinoiden että todellisuuden puhuviin päihin liitettiin maagisia merkityksiä (Truitt 2011). Arkeologit ovat havainneet äänitekniikoita käytetyn ylikuonnollisen vaikutelman aikaansaamiseen jo esihistoriallisina aikoina. Kalliomaalausten sijoittaminen paikkoihin, joissa oli kaikuisa akustiikka, mahdollisti vaikutelman puhuvista ja puheeseen vastaavista kuvista (Rainio et al. 2017).

1600-luvun puolivälissä merkittäväksi maailmanselitysmalliksi noussut mekanistinen luonnonfilosofia piti kaikkia eläviä olentoja pohjimmiltaan mekaanisina koneina, mutta älykkyyttä osoittavan puhekyvyn ajateltiin kuitenkin erottavan ihmiset eläimistä ja muista koneista. 1700-luvun suositut automaattit puolestaan jäljittelivät monimutkaisia inhimillisiä ja eläimellisiä toimintoja, mutta eivät pystyneet puhumaan. (Kang 2011.) Automaatteja rakentanut Wolfgang von Kempelen kehitti kuitenkin 1700-luvun loppupuoliskolla myös ihmisen äänielimistöä jäljitteleviä mekaanisia koneita, joilla saatiin aikaan ihmismäisen puheen ääniteitä. Kempelenin hyvin dokumentoitujen koneiden lisäksi onnistuneimpiin mekaanisen puhesynteesin ratkaisuihin luetaan Christian Gottlieb Kratzensteinin vuonna 1780 kehittämä urkukoneisto, jonka viidestä pillistä jokainen tuotti yhden ihmispuheen vokaaleista. Puhesynteesin historia 1700- ja 1800-luvuilla perustui urkupillien ja muiden soitinten sekä ihmisen äänielimistön jäljitelmien yhdistelmiin. Tätä perinnettä jatkoi myös Faberin Euphonia, joskin se oli rakenteeltaan huomattavasti monimutkaisempi kuin Kempelenin melko pienet ja yksinkertaiset laitteet. (Brackhane 2015, xiv, 26.)

1800-luvun viimeisistä vuosikymmenistä lähtien koneellisesta puheesta tuli huomattavasti yleisempää kuultavaa kuin aiemmin. Ensin keksittiin vahaliieriöille äänittävä ja niitä toistava fonografi, pian sen jälkeen prässättyjä äänilevyjä toistava gramofoni. Puhelimen ja radion sekä sähköisten kovaäänisten ja megafonien yleistyessä tuli teknisesti tuotetusta ihmisäänestä yhä tutumpi osa ääniympäristöä 1900-luvun alun vuosikymmeninä. Lisäksi 1920-luvun lopulla äänielokuvatekniikka mullisti jo vakiintuneen viihdemuodon laittamalla liikkuvat kuvat puhumaan. (Thompson 2002.) Teknistyvässä äänimaisemassa ihmispuhe oli yksi koneellisen äänen muodoista, vaikka esimerkiksi gramofonin keskeisimmäksi käyttötarkoitukseksi vakiintui musiikin, ei puheen toistaminen. Puhesynteesin kehitys laajensi koneellisen puheen käyttömahdollisuuksia entisestään.

Bell Telephone Laboratories -teollisuuslaboratorio esitteli 1930-luvun lopulla ensimmäisen sähköisen puhesyntetisaattorin, josta 1960-luvun alussa kehitettiin myös ensimmäinen tietokonepohjainen puhesynteesiratkaisu. Tietokonepohjaista puhesynteesiä käytettiin seuraavien vuosikymmenten aikana erityisesti kehitettäessä näkö- ja puhevammaisia auttavia teknisiä sovelluksia, mutta myös koneita ja tietokoneohjelmia, jotka avustivat kielen oppimisessa tai tekstien tuottamisessa. Lapsille suunnattu laite oli esimerkiksi Texas Instrumentsin vuonna 1978 markkinoille tuoma *Speak and Spell*, joka lausui siihen kirjoitettuja sanoja ja opetti näin oikeinkirjoitusta ja lukemista. (Rowe 2010, 11–14.) Puhesynteesillä luotua puhetta alettiin käyttää myös julkisten tilojen kuulutuksissa sekä joissain kuluttajille suunnatuissa laitteissa.

Ihmismäistä puhetta hyödyntävien koneiden rakentamisella on ollut koko niiden historian ajan erilaisia tarkoituksia ja motivaatioita. Taikurit ovat halunneet luoda vaikutelman puhuvista esineistä, kielen, fonetiikan ja fysiologian tutkijat ovat halunneet ymmärtää puheen synnyn mekanismeja, akustikot ovat pyrkineet tuottamaan ja analysoimaan ääniteitä, keksijät ovat



painineet äänen tallentamisen ja toiston kysymysten parissa ja muusikoita on kiinnostanut ihmisäänen värin ja sävyn toistaminen instrumenteilla. Samoilla tai eri koneilla on pyritty tutkimaan, tuottamaan, vahvistamaan ja toistamaan ihmisääntä eri yhteyksissä. (Hankins & Silverman 1995, 178–220.) Tekniset ratkaisut ovat osin jakautuneet eri tarkoituksia varten tehdyissä koneissa, mutta toisaalta myös kohdanneet ja yhdistyneet niissä.

Massimo Pettorino (2015, 31–32) on jakanut ihmismäisellä äänellä puhuvien koneiden historialliset ilmentymät kahteen päätyyppiin: äänensiirtoon perustuviin laitteisiin, joissa ihmispuhe on siirretty erilaisia tekniikoita käyttäen koneen puheeksi, sekä keinotekoisia puhetta tuottaviin koneisiin, siis erilaisiin puheesynteesin muotoihin. Jonathan Sterne (2003, 81, 94–96, 98–99) puolestaan on esittänyt, että puhuvien koneiden kehittyessä kiinnostus ja painopiste siirtyi inhimillisen puheentuotannon mekanismeista kuulemisen ja kuuntelun mekanismeihin. Tekniikan myötä syntyi hänen mukaansa uudenlainen keskittyvän teknisen kuuntelun tapa, jota jo stetoskooppi ja lennätin pohjustivat 1800-luvulla. Kuunneltavat tekniikat olivat silti usein myös puhuvia tekniikoita. Täydentäisin Pettorinon luokittelua kaksijakoisesta kolmijakoiseksi: puheesynteesin ja äänensiirron lisäksi monet puhuvat koneet ovat hyödyntäneet tallennettua puhetta, joko sellaisenaan tai synteettisesti muokattuna. Tämä puhuvien koneiden tekninen muoto on uusin, fonografeista alkavan laitelinjan mahdollistama.

Tämän hetkisessä koneellisessa puheessa kaikki kolme teknistä toteutusmuotoa elävät rinnan, joskus jopa yhdessä ja samassa laitteessa tai järjestelmässä. Vaikkapa junien ja juna-asemien kuulutusjärjestelmästä voidaan kuulla



Sokos Wiklundin Facebook-päivitys puhuvasta pikakassasta: "PIKAKASSA, eli itsepalvelukassa, avattu Wiklund Herkkuun, ensimmäinen Turussa! #SokosWiklund". Lähde: Sokos Wiklundin Facebook-päivitys, 29.5.2013. Saatavilla: <https://www.facebook.com/sokoswiklund/photos/a.200012316696417/583415765022735/?type=3&theater> (linkki tarkistettu 6.4.2020).

sekä nauhoitettua puhetta, sähköisesti välitettyä reaaliaikaista kuulutuspuhetta että puhesynteesillä aikaansaatuja tiedonantoja. Auton navigaattori ja digitaaliset assistentit hyödyntävät puhesynteesiä, mutta monissa laitteissa, vaikkapa hisseissä tai pienelektronikassa tietty toiminto aktivoi aina tietyn saman puhetallenteen, joka kertoo esimerkiksi ovien sulkeutumisesta tai virran kytketymsistä päälle ja pois. Ovipuhelimista tai muun muassa poliisiautoista ja tietynlaisista valvontakameroista voimme puolestaan kuulla välitettyä puhetta, jota toinen ihminen samanaikaisesti tuottaa. Kaikkiin kolmeen konepuheen tekniseen muotoon voi lisäominaisuutena liittyä vuorovaikutteisuus käyttäjän mahdollisuutena vastata koneen puheeseen puhumalla itse. Tämä voi perustua siihen, että käyttäjän puhe nauhoitetaan tai välitetään tai siihen, että kone sisältää puheentunnistustekniikkaa ja voi sen myötä reagoida käyttäjänsä sanomisiin.

Ehkä eniten käyttämässämme arkisessa laitteessa, puhelimessa, ovat käytössä kaikki konepuheen tekniikat ja kaikkien kohdalla myös vuorovaikutteisuus. Puhesynteesiä voi Applen Sirin kaltaisten assistenttien puheen lisäksi kuulla monissa puhelinpalveluissa ja nämä vastavuoroisesti myös hyödyntävät puheentunnistusta. Puhelinvastaajissa ja ääniviesteissä puhe on nauhoitettua ja toisen ihmisen kanssa puhelimessa puhuttaessa välitettyä. Voidaan kuitenkin kysyä, onko puhelinta jälkimmäisessä käytössä sosioteknisesti perusteltua lukea puhuvaksi koneeksi, kun kyseessä on teknisavusteinen kahden (tai useamman) ihmisen keskinäinen ja tasavertainen puhekommunikaatio. Esitän, että käyttäjäsuhteessa erottavaksi tekijäksi ei silti niinkään nouse koneellisen puheen teknisen toteutuksen tyyppi kuin se, millaista roolia puhe näyttää käyttäjän ja koneen suhteessa. Puhesynteesi konepuheen toteuttamistapana yhdistyy usein sellaisiin erityisiksi koettuihin koneisiin kuin tekoälyihin ja robotteihin, mutta myös nauhoitettu ja välitetty puhe voivat tietyissä teknisissä yhdistelmissä ja tilanteissa olla perustana sellaiselle sosiotekniselle ilmiölle, jota kutsun puhuvaksi koneeksi. Seuraavassa luvussa määrittelenkin puhuvien koneiden kategorian rajoja tarkastelemalla erilaisia teknologisia toimijuuksia ja suhteita.

### **Puhuva kone toimijuuksena, suhteena ja asetelmana**

Maaliskuussa 2018 kansainvälinen media julkaisi uutisia aavemaisesta naurusta, joka oli säikäyttänyt Amazonin älykaiuttimien käyttäjiä (Chokshi 2018; Wong 2018). Digitaalisen assistentin, Alexan, kerrottiin purskahtaneen nauruun äkillisesti ja ilman syytä, toisinaan jopa kesken täydellisen hiljaisuuden. Amazon kiiruhti selittämään naurun virhetoiminnoksi ja poistamaan Alexan nauruominaisuuden.

Puhe on monella tapaa erityinen toimijuuden piirre verrattuna muihin tapoihin, joilla elottomien esineiden ajatellaan toimivan. Alexan nauru on äärimmäinen esimerkki siitä, miten puhekyky ylittää yleensä koneelle mahdolliset toimijuuden rajat, toisinaan myös yllättävällä tai epämiellyttävällä tavalla. Puhuvia koneita on usein tarkoituksella käytetty kyseenalaistamaan elollisen ja elottoman rajoja sekä ällistyttämään yleisöjä (Smith 2008, 189). Kiehtova efekti tuntuu kerta toisensa jälkeen pohjautuvan puheeseen ihmillisyyden merkinä, jolla voidaan luoda illuusio ajattelukyvyistä. Jussi Parikan (2004, 115–116) mukaan ”ajattelevat koneet”, eli laskevat ja analysoivat koneet, merkitsivät 1800-luvulta lähtien muutosta koneiden ja ihmisten suhteessa: Koneiden ei ollut enää tarkoitus hämmästyttää matkimalla ihmistä,

vaan vastata rationaaliin kontrollin ja tehokkuuden tarpeisiin. Puheominaisuuksia tarkasteltaessa käy kuitenkin ilmi, että ihmismäisyyden matkiminen on jatkunut muiden pyrkimysten rinnalla kaiken aikaa. Usein koneista on nimenomaan puheen avulla annettu todellisuutta pystyvämpi kuva ja täten hämärretty käsitystä koneiden mahdollisuuksien rajoista.

Moniin 1900-luvun robotteihin liitettiin puheominaisuus, edellisessä luvussa esitellyn jaottelun mukaisesti joko äänensirrolla tai nauhoituksena. 1920-luvun brittiläinen Eric-robotti piti julkisia puheita siten, että siinä olevaan vastaanottimeen siirrettiin ihmispuhetta radiosignaalin avulla (Riskin 2016). 1930-luvulla Westinghouse Electric Corporationin Elektro-robotissa esittelijän äänikäskyjen tavumäärä aktivoi tiettyjen releiden avulla sopivan puheäänityksen ja näin päästi tapahtumissa esittämään keskustelija robotin kanssa. Yleisön kanssa tavalla tai toisella jutustelevat robotit muodostuivat 1900-luvun jälkipuoliskolla teknologiamessujen vakiovetonauloiksi. Ihmismäinen puhe oli spektaakkeli ja merkki teknologisesta edistyksellisyydestä. Myös ensimmäinen Macintosh-tietokone laitettiin puhumaan, kun Steve Jobs julkisti sen vuonna 1984.<sup>1</sup> Kone esitteli itsensä ja vitsaili yleisölle käyttäen tekstin puheeksi kääntävää puhesynteesiä. Käytännössä kesti kuitenkin vielä 20 vuotta, ennen kuin Mac-tietokoneissa oli käytettävissä laaja puhesynteesiominaisuus (Rowe 2010, 14).

Kathleen Richardsonin (2016) mukaan robottifiktio ja tosimaailman robottien välillä ei ole koskaan ollut selkeää rajaa. Fiktio on luonut odotuksia puhuvista, älykkäistä roboteista, ja näyttelyrobotit ovat eri vuosikymmenillä antaneet tälle fiktiolle enemmän tai vähemmän onnistuneen tosimaailman muodon. Näiden todellisten ja fiktiivisten robottien varaan on muodostunut ajatus eräänlaisesta koneellisesta henkilöydestä, jota Richardson kutsuu teknologiseksi animismiksi. Se saa meidät ajattelemaan, että koneet ovat, tai voisivat olla, elollisia persoonia. Viime vuosina keskustelua on käyty muun muassa Sophia-robotista, joka esiintyy puhuvana, ihmismäisenä tekoälynä ja on muun muassa saanut Saudi-Arabian kansalaisuuden. Todellisuudessa Sophia on kaukana ajattelevasta koneesta. Robotin valmistanut Hanson Robotics kommentoi *The Verge* -verkkolehden mukaan (Vincent 2017), että on kuitenkin vain hyvä, jos illuusio Sophian älykkyydestä menee täydestä, koska se saa yleisön uskomaan vahvan tekoälyn mahdollisuuteen.

Ihmisenveroisien tietokoneälyn nopean kehittämisen mahdollisuutta tai jopa todennäköisyyttä on pidetty esillä 1950-luvulta asti, vaikka kehitys on useita kertoja osoittautunut arveltua huomattavasti hitaammaksi ja hankalammaksi (Nivala 2019, 21). Tieteenfilosofi Karen Baradia (2003) mukaillen voidaan ajatella, että rajoitetulla sanavarastolla ja erilaisilla teknisillä trikeillä keskustelemaan laitettut robotit eivät ole vain kömpelö heijastus diskursiivisesta tekoälyn ideasta. Sen sijaan ne tekevät ideaa todeksi käytäntöinä, jotka ovat samanaikaisesti sekä materiaalisia että diskursiivisia. Barad on muun muassa Michel Foucault'n ja Judith Butlerin poststrukturalistisen ajattelun pohjalta kehittänyt todellisuuden luonteen selittämiseksi agentiaalisen realismin mallin,<sup>2</sup> jossa merkitystä luovia perusyksiköitä eivät ole sanat vaan materiaalis-diskursiiviset ilmiöt. Niissä materiaaliset olennot ja merkitykset muotoutuvat sekä ontologisesti että semanttisesti. Thomas Hankins ja Robert Silverman (1995) ovat tutkineet erilaisia 1600–1800-luvun tieteellisiä instrumentteja ja havaintoesineitä, joista huomiota herättäneimmät paljastuivat usein jonkinlaisiksi huijauksiksi. Tästä huolimatta ne saattoivat olla merkittäviä materialisoidessaan ja vahvistaessaan tieteellisiä tai teknologisia ideoita ja periaatteita. Messuja ja yleisötapahtumia viimeiset 50 vuotta kiertäneiden

1 Videotallenteen esittelytilaisuudesta voi katsoa osoitteessa <https://www.youtube.com/watch?v=2B-XwPjn9YY>.

2 Alkukielinen ilmaus on "agential realism", vaihtoehtoisena suomennoksena esiintyy joissain yhteyksissä "toiminnallinen realismi".

puhuvien robottien voidaan ajatella edustavan samantapaisia materiaalisia ideoita. Ne ovat pönkittäneet ja ruumiillistaneet ajatusta puhuvista, ajattelevista koneista mahdollisimman huomiohakuisella tavalla.

Puheominaisuus vaikuttaa siihen, miten koneista ajatellaan, mutta se vaikuttaa myös koneiden käytännön toimijuuteen. Tekniikan kulttuurisessa ja sosiaalisessa tutkimuksessa ajatus koneiden ja laitteiden omasta aktiivisuudesta on suhteellisen yleinen. Bruno Latour (1990) on esittänyt tekniikan olevan yhteiskunnan suhteiden konkreettinen ja näkyvä muoto, johon on latautunut toimijuutta. Tämä toimijuus voi olla erilaisten inhimillisten ja ei-inhimillisten toimijuuksien yhteensovittamista. Esimerkiksi turvavyön on toiminnallaan ratkaistava yhtälö, joka syntyy ihmisten huolimattomasta ajotoiminnasta ja autojen taipumuksesta murskata ajajansa onnettomuustilanteissa. Turvavyöllä on mekaanista toimijuutta ja myös sosiaalinen mandaatti toimintaansa. Latour (1992) on myös kertonut omasta autostaan, joka käytännössä pakottaa hänet kiinnittämään turvavyön toistamalla sietämätöntä merkkiääntä vyön ollessa auki. Näin auto toteuttaa yhteiskunnassa vallitsevaa turvavyöpakkoa, johon ihminen voi toiminnallaan mukautua tai jota vastaan hän voi kapinoida. Samalla sosiaalisella periaatteella toimii puhuva kassa, joka käskee asiakasta maksamaan ostoksensa ja kyselee, onko tämä varmasti skannannut kaikki tuotteet.

Teknologian käyttäjiä tarkastelevassa sosiologisessa tutkimuksessa nähdään tekniikkaan rakennetun valmiiksi tiettyjä käyttäjäoletuksia tai -konfiguraatioita, joiden mukaan käyttäjän oletetaan toimivan ja joilla on suoraa vaikutusta siihen, miten tekniikkaa voidaan käyttää ja miten käyttäjä voi sen kanssa toimia (ks. Woolgar 1990; Akrich 1995; Hyysalo 2010). Toimijaverkostoteoriassa korostuu kollektiivinen toiminta, johon osallistuvat niin ihmiset kuin esineetkin, mutta Latourin (2005, 80–81) mukaan materiaalisten ”täysin hiljaisten välikäsien” toimijuutta voi usein olla vaikea havaita. Samanlaista näkökantaa on nostettu esiin esimerkiksi, kun antropologisessa tutkimuksessa on koetettu antaa materiaaliselle maailmalle ja esineille suurempi painoarvo. Esineet eivät puhu puolestaan, minkä takia suhtautuminen niihin on väistämättä ihmislähtöistä (ks. Holbraad 2011). Mutta päteekö sama niihin välikäsiin, jotka ovat kaikkea muuta kuin hiljaisia? Niihin, jotka puhuvat?

Puhuvissa koneissa erityistä on, ettei niiden aktiivisuutta yleensä ole vaikea havaita. Ne tekevät konetoimijuuden korvin kuultavaksi. Toisaalta ne voivat myös kätkeä osia konemaisesta toimijuudestaan ihmismäisen toimijuuden verhon taakse. Lisäksi ne tuovat kuuluviin olettamansa käyttäjän, siis puheen kuulijan ja siihen reagoivan tai vastaavan käyttäjän. Voidaan jopa ajatella, että nauravan Alexan tapauksessa nimenomaan käyttäjäoletus on koettu kauhistuttavana: koneen nauru tekee sen kuulijasta naurun kohteen. Muun muassa tämän kammottavuuden potentiaalini laimentamiseksi kuluttajille suunnatut puhuvat koneet on huomattavan usein laitettu puhumaan naisten äänillä tai niistä on tehty muilla keinoilla mahdollisimman harmittomia ja ensisijaiseksi oletettua (mies)käyttäjää sosiaalisesti alemmassa asemassa olevia (Männistö-Funk & Sihvonen 2018). Puhuville koneille tyypillistä on toimijuussuhteen neuvottelu, jossa koneellisesti tuotettu, toistettu tai välitetty puheääni horjuttaa ihmisen ja koneen toiminnallista jakoa ja vaikuttaa myös ihmiseen toimijana määritellään tätä kuuntelijaksi tai puhekumppaniksi, rakentaessaan puheen ja kuuntelun kommunikaatiota ihmisen ja koneen välille.

Toimijuuteen perustuvan määrittelyn kautta voidaan hahmottaa puhuvien koneiden kategorian sisältöä ja rajoja. Konetoimijuuden määrittelystä tarkasteluna esimerkiksi puhelin ei ole puhuva kone silloin, kun sitä käytetään



tavalliseen puhelinkeskusteluun kahden ihmisen välillä, mutta voi saada omaa puhuvaa toimijuutta, kun siitä kuuluu puhelinvastaajan nauhoitettu viesti tai puhesynteesiä hyödyntävä puhelinpalvelun ohje, tai kun puhelimen kaiuttimesta kuuluva puhe ottaa yksityistä tai julkista tilaa haltuunsa. Nauhoitteiden kuunteluun käytetyt soittimet tai lähetyksen seuraamiseen käytetty radio, televisio tai tietokone vaikuttavat aktiivisilta, mutta niiden ääntely on milloin tahansa käyttäjän pysäytettävissä tai käynnistettävissä. Niistä kuuluva puhe ei myöskään liity itse koneen käyttämiseen tai kutsu käyttäjää tietyntilaiseen fyysiseen toimintaan muissa kuin poikkeustilanteissa, kuten lähetyshäiriöissä tai vaaratiedotuksissa, jolloin radiosta tai televisiosta tulee ikään kuin lähetyjärjestelmän tai viranomaisten puolesta puhuva kone.

Mielenkiintoisen rajatapauksen muodostavat mainokset ja muut tiedotukset, jotka katkaisevat valitun lähetyksen. Esimerkiksi Spotifyn ilmaisversion voidaan ajatella muuttuvan puhuvaksi koneeksi siinä vaiheessa, kun mainospuhe katkaisee kuunneltavaksi valitun musiikin. Myös laivan hytissä kuulutuksella herättävä radio tai televisio ja kaupan äänentoistojärjestelmä tai koulun keskusradio, tietyissä tapauksissa myös julkisten tilojen televisiot ja muut puhuvat monitorit, määrittävät toimijuudeltaan puhuviksi koneiksi. Ne toimivat tyypillisesti suuremman teknisen kokonaisuuden osina ja äänitorvina, jolloin esimerkiksi rakennus tai kulkuväline kokonaisuudessaan voidaan hahmottaa puhuvana koneena. Puhuvien koneiden kategoriaa voidaan piirtää siis esiin toimijuussuhteen perusteella, mutta kategorian rajoja ei voi kiinnittää tiettyihin teknologioihin ja koneisiin, sillä olennaista on tilanteen eri osapuolien, koneiden, ihmisten ja muiden osallisten vuorovaikutus.

Teknologianfilosofi Don Ihde on määrittellyt kolme erilaista ihmisen ja teknologian suhdetta: välityssuhteet, toiseuden suhteet ja taustasuhteet.<sup>3</sup> Välityssuhteessa teknologia toimii väylänä maailmaan eli välittäjänä maailman ja ihmisen välillä. Taustasuhteessa teknologia muovaa suhdetta maailmaan lähes huomaamattomasti ja itsekseen. Toiseuden suhteessa keskiöön nousee ihmisen ja koneen suhde, toisin kuin kahdessa muussa. (Verbeek 2001.) Esitän, että puhuvissa koneissa on aina jollain tapaa kyse tällaisesta toiseuden suhteesta. Joissakin koneissa puheominaisuus aiheuttaa lähes aina toiseuden suhteen käyttäjään, eli kone on puhuva kone, siis puhuva toinen, kaiken aikaa, jopa silloin kun se vaikenee tai kuuntelee. Sen sijaan joillain koneilla ja järjestelmillä, sellaisilla kuin yllä käsitellyt radio, televisio sekä suoratoistosovellukset, suhde käyttäjään vaihtelee Ihden kolmen kategorian välillä, ja kone määrittyy puhuvaksi koneeksi vain toiseuden suhteen kytkeytyessä päälle. Edellisessä alaluvussa esitellyistä konepuheen tuottamisen teknologioista puhesynteesi esiintyy tyypillisimmin puhuviksi koneiksi luokiteltavissa soioteknisissä tapauksissa, mutta sopivissa olosuhteissa ja yhdistelmissä myös välitettyä tai nauhoitettua puhetta hyödyntävä kone voi olla määriteltävissä puhuvaksi koneeksi.

Uusmaterialistinen teoriasuuntaus on viime vuosikymmeninä nostanut esiin kaiken toiminnan materiaalisuutta. Sen näkökulma alleviivaa, ettei jonkin olennon tai esineen toimijuus riipu ihmisenkaltaisuuudesta, eikä inhimillisen ja ei-inhimillisen toimijuuden välille voi vetää selvää rajaa. Toimijuuksia on erilaisia, ja missä tahansa tilanteessa on aina tavalla tai toisella kysymys erilaisten toimijuuksien parven yhteistoiminnasta (Bennett 2010, 31, 34), jonka tapaiseen myös Latour on viitannut kirjoittaessaan toimijuudesta jaettuna (Latour 1993, 138). Tältäkin näkökannalta puhuminen on erityinen toimijuuden muoto, jonka roolia eri tilanteissa on aiheellista tarkastella. Uusmaterialistiksi luokiteltu Barad (2007, 136–141) menee vielä Latouria pidemmälle: Toimijat

3 Alkukielellä "relation of mediation", "alterity relation" ja "background relation".

ja toimijuus eivät pelkästään muokkaudu toimintatilanteissa ja vuorovaikutuksissa. Toimijat eivät edellä toimintaa, vaan saavat olemuksen vasta siinä, suhteessa toisiinsa, tämän suhteen itseensä sisällyttäen. Parvimaisen toimijuuden näkökulmasta voidaankin ajatella, että puhuvat koneet muokkaavat käyttäjänsä toimijuutta tai koneen ja käyttäjän toimijuuden tasapainoa perustavammalla tavalla kuin vain sisältämällä tiettyjä ohjaavia oletuksia. Ihmisestä tulee koneen toiminnan kohde, kun hän kuulee sen puhetta, ja kone voi olla jaetun toiminnan aktiivisempi osapuoli, joka tietyissä tapauksissa myös kuuntelee ihmisen puhetta. Postfenomenologisessa tekniikan filosofiassa ollaan kiinnostuneita nimenomaan ihmisen ja tekniikan suhteesta, sekä siitä, miten tekniikka muokkaa ihmisen suhdetta maailmaan (Verbeek 2016). Esimerkiksi Sirin kaltaisten konetoimintojen käyttäminen muokkaa ihmisen suhdetta maailmaan hyvin monimutkaisilla tavoilla. Maailmassa olemisesta ja arjen toiminnasta tulee puhuvan koneen kanssa kommunikoimisen välittämää ja kanavoimaa (Guzman 2017).

Jos otetaan todesta Baradin kuvaama toimijoiden hahmottuminen toiminnan hetkissä, voidaan pohtia myös sitä, millainen on konetta kuunteleva tai koneen kanssa puhuva ihminen. Parikan (2004, 115–116) ajatus ihmismäisistä konemaisemmiksi muuttuvista koneista voi johdattaa ajattelemaan peilikuvana myös koneiden käyttäjän vastaavaa muutosta. Puhuvien koneiden kuunteleminen, ja ehkä varsinkin niille puhuminen vie kommunikaatiota kohti konemaista logiikkaa, jossa käyttäjän pitää osata tulkita konepuhetta ja pyrkiä itse puhumaan mahdollisimman selkeästi ja tarkasti, siis ”konemaisesti”. Vain muuntumalla konemaisemmaksi ihminen saa kommunikaation esimerkiksi Sirin, navigaattorin tai automaattisen puhelinpalvelun kanssa onnistumaan. Tällainen keskinäistä muovautumista sisältävä suhde vaikuttaa toimijuuden lisäksi monella tapaa myös tunteisiin, joita tarkastelen seuraavaksi.

### **Puhuvat koneet menevät tunteisiin**

*Scientific American* -lehti (25.7.1896) nosti vuoden 1896 keksintönumerossaan fonografin yhdeksi suurista keksinnöistä ja kertoi tarinan siitä, miten nuori Edison oli tullut esittelemään laitetta ennen sen patentointia lehden toimitukseen loppuvuodesta 1877. Hän oli nostanut koneen pöydälle ja vääntänyt sen kahvasta. Lehden toimittajien ällistykseksi kone oli puhunut: ”Hyvää huomenta. Mitä kuuluu? Mitä pidätte fonografista?” Kuvaus on malliesimerkki siitä, mitä Tom Gunning (2003, 46–47) kutsuu outouden tunteeksi (”sense of uncanny”). Hänen mukaansa erityisesti uudet teknologiat ovat omiaan aiheuttamaan outouden tunteen, mutta tietyt teknologiat ovat omiaan tuottamaan myös toistuvan tunteen siitä, että niiden toiminta on jollain tapaa ihmeellistä, salaperäistä ja maagista.

Monilla on kokemusta hätkähdyttävästä tunteesta, jonka itsekseen puhumaan puhkeava lelu ja vaikkapa navigaattori voi aiheuttaa. Esitänkin, että puheominaisuus antaa erityisen hyvät edellytykset outouden tunteen syntymiselle. Robotteihin ja tekoälyyn liittyvät epämiellyttävät vierauden ja outouden tunteet kiinnostavat esimerkiksi tekoälyn, robotiikan ja vaikkapa chatbottien tutkijoita ja kehittäjiä, mutta koneellisen puheen rooli näiden tuntemusten synnyssä on jäänyt yllättävän vähälle huomiolle (ks. esim. Skjuve et al. 2019).

Robotiikan ja tekoälyn noustua yhä vahvemmin teknisen ja taloudellisen kiinnostuksen kohteiksi 2000-luvulla monet tutkijat ja tekniikan kehittäjät ovat pohtineet japanilaisen robotitutkijan Masahiro Morin alkuaan jo vuonna

1970 määrittelemää outolaakson ("uncanny valley") ongelmaa. Morin (2012) mukaan elottomat esineet miellyttävät meitä sitä enemmän, mitä ihmis-  
mäisempiä ne ovat. Kuitenkin, kun jokin esine tai asia on lähes täydellisen  
ihmismäinen, tämän miellyttävyyden tunteen sijaan koetaan äkillistä  
outouden tunnetta ja vieroksunta. Esimerkiksi hyvin ihmismäinen proteesi  
tai robotti onkin miellyttävän sijasta kammottava. Morin mukaan liike lisää  
jäljitelmien kammottavuutta, mutta hän ei ota kantaa puheen vaikutukseen.  
Voidaan kuitenkin arvella, että puhe erityisen ihmismäisenä ominaisuutena  
voisi lisätä kammottavuutta vielä liikettäkin enemmän.

Kuten Varsovan yliopiston ja MIT:n tutkijat toteavat tuoreessa tutkimuk-  
sessaan (Ciechanowski et al. 2019), humanoidien olioiden aiheuttamista re-  
aktioista tiedetään yhä varsin vähän, vaikka Morin kuvailema outolaakso on  
suunnannut paljon huomiota outouden tuntemuksiin. Omassa tutkimukses-  
saan he havaitsivat, että testihenkilöt reagoivat huomattavasti negatiivisemmin  
chatbottiin, joka puhui ihmisäänellä ja jolle oli animoitu mahdollisimman  
realistinen ihmisavatar, kuin samaan chatbottiin silloin, kun se kommunikoi  
pelkästään kirjoitetun tekstin välityksellä ja ilman avataria. Puhuva avatar  
aiheutti testihenkilöissä huomattavaa psykofyysistä kiihtymystä ja epämiellyt-  
täviä tunteita, joita pelkällä tekstillä viestinyt chatbot ei juurikaan aiheuttanut.  
Tutkijat arvelevat tämän voineen johtua käytetyn avatarin ja puheäänen epä-  
täydellisyydestä, tai siitä, että ihmismäisempi chatbot lisäsi testihenkilöiden  
odotuksia muttei vastaavasti täyttänyt niitä.

Minsoo Kang (2011) on esittänyt koneiden aiheuttavan outouden tunteita,  
jos ne imitoivat ihmistä hyvin mutta eivät täydellisesti. Sen sijaan kauhun  
tunne syntyy, jos imitointi on niin täydellistä, että vastapuoli ei pysty var-  
muudella luokittelemaan olentoja ihmisiksi tai koneiksi. Bottitutkijoiden  
mukaan chatboteissa ei kannattaisikaan tavoitella ihmismäisyyttä, ainakaan  
jos uskottavaa ihmismäisyyttä ei voida saavuttaa (Ciechanowski et al. 2019).  
Myös Mori (2012, 100) suosittelee, ettei esimerkiksi proteeseissa ylipäätään  
pyrittäisi mahdollisimman suureen ihmismäisyyteen, jossa riskinä ovat kam-  
motuksen tunteet, vaan suosittaisiin tunteiden mukavuusalueelle mahtuvaa  
lievää ihmisenkaltaisuutta.

Keinotekoisien puheen voisi olettaa olevan potentiaalisesti vielä vahvempaa  
outouden tunnetta tai kammotusta herättävä piirre kuin keinotekoisien raajan.  
Toisaalta taas koneellinen puhe on arjessa yleisempi ja tutumpi ilmiö kuin  
koneellinen raaja. Konepuhe on osa ääniympäristöämme ja osaamme odottaa,  
tunnistaa ja tulkita sitä ainakin tietyissä fyysisissä tiloissa ja tilanteissa. Kone-  
puheen vaikutuksista on kuitenkin vaikea löytää tutkimusta. Esimerkiksi yllä  
mainitussa chatbot-tutkimuksessa ei tarkasteltu pelkän puheäänen vaikutusta  
tai mietitty, millainen rooli sillä itsessään oli tunnereaktioiden aiheuttajana,  
vaikka tutkijat totesivatkin puhuvan avatarin aiheuttavan selvän negatiivisen  
reaktion testihenkilöissä. Kammotuksen ja outouden lisäksi kyse voi olla ar-  
kisesta ärsytyksestä ja turhautumisesta. Science fictionin ja tiedeutopioiden  
älykkään koneen lupausta materialisoivat arjen puhuvat koneet osoittautuvat  
kerta toisensa jälkeen turhauttavan älyttömiksi, ja puhe vivahteikkaan kom-  
munikaation muotona voi pahimmillaan alleviivata koneen älyttömyyttä ja  
tilanteen vaivaannuttavuutta.

Kulttuuritutkimuksessa affektin käsitettä on käytetty erityisesti viime  
vuosituhannen vaihteesta lähtien sellaisten, usein ruumiillisten, kokemusten  
tarkasteluun, joita kielellisten representaatioiden paradigma ei tavoita. Affekti  
käsitetään tällöin voimaksi, joka ylittää ruumiilliset rajat ja liittää yhteen eri-  
laisia fyysisen maailman yksiköitä, inhimillisiä ja ei-inhimillisiä. Se on eri asia

kuin tunne, mutta voi manifestoitua tuntemuksina tai tunteina (Blackman & Venn 2010). Affektin käsitteellä on juurensa sekä filosofisessa että psykologisessa perinteessä ja sen määritelmät vaihtelevat. Tässä artikkelissa nojaan erityisesti Gilles Deleuzen kirjoituksista kumpuaviin ajatuksiin affektista ruumiillisena tilasta toiseen siirtymisenä, joka lävistää ja sekoittaa sosiaalisen tulkinnan tason (ks. Hemmings 2005, 552; Deleuze 1997). Tästä näkökulmasta korostuu affektin luonne eräänlaisena energiana tai virtauksena, joka syntyy erilaisten ruumiiden ja fyysisten objektien asetelmissa (Newell 2018, 2).

Affektien tutkimuksessa asetelmat, eri toimijoiden, fyysisten entiteettien ja tilojen muodostelmat ovatkin tärkeä tarkastelun kohde (Slaby 2018). Tällaisista muodostelmista, asetelmista tai kietoutumista teoreettisesti ja metodologisesti kiinnostunutta ajattelua edustaa keskeisenä hahmona aikaisemmin mainittu Karen Barad. Yhteismuotoutuva toimijuus ja virtaavat affektiiviset voimat liittyvät siis yhteen ja tarjoavat hedelmällisen näkökulman puhuviin koneisiin. Affektien virtaamisen ajatus on saanut tutkijat kysymään, voisiko virtauksen voimakkuutta eri tilanteissa mitata tai arvioida (Blackman & Venn 2010, 17). Koneiden puhetta voisi mahdollisesti lähestyä tällaisena ihmisiä ja koneita sisältävissä asetelmissa virtaavana intensiteettinä, joka yllä mainitun chatbot-esimerkin mukaisesti voisi olla myös mitattavissa ja arvioitavissa.

Sosiologi Amparo Lasén (2010) on tarkastellut matkapuhelinta affektiivisena teknologiana. Matkapuhelin on osallisena tunteisiin, jotka syntyvät ihmisten ja muiden entiteettien jaetussa toimijuudessa. Puhuvien koneiden kohdalla saatetaan tarkoituksella rakentaa vaikutelmaa myös koneen omista mahdollisista tunteista. Itkevä tai jokelteleva vauvanukke on ilmeinen esimerkki tällaisesta äännelevästä tunnekoneesta, mutta julkisuudessa on maalailtu myös ajatuksia todellisista tuntevista koneista. *Tekniikan Maailma* -lehdessä (Tekniikan Maailma 21.2.2018, 101) kerrottiin helmikuussa 2018 laveasti kullattateknologian CES-messuilla Las Vegasissa esitellyistä roboteista. Jutussa kuvaillaan puhuvia ja liikkuvia viihderobotteja seuraavasti: ”Palvelualttiisiin pikkuapulaisiin on helppo kiintyä, etenkin kun niiden tekoälykapasiteetti kasvaa kaiken aikaa. Emoshape on kehittänyt tunne-mikropiirin, joka auttaa koneita ymmärtämään puhutun äänen vivahteita ja todellista merkitystä. Jos (tai kun) roboteista tulee tuntevia, kiintymyssuhteen rajat koneen ja elollisen välillä alkavat hämärtyä, etenkin kun robotit ovat omaksuneet sukupuolen.”

Tuntuu paksulta väittää, että koneet olisivat tulossa tunteviksi, kun niiden todellinen älykkyykskään ei ole näköpiirissä ja kun tunteita voidaan pitää ajattelun yhtenä muotona (vrt. Minsky 2006). Koneiden mahdollisista tunteista voidaan siis kirjoittaa lähinnä spekulatiivisesti. Sen sijaan koneiden kyky simuloida tunteita ja reagoida ihmisen tunnetiloihin on ollut käytännöllisen tutkimuksellisen kiinnostuksen kohteena viimeisten parin vuosikymmenen ajan. Tunteet ovat tulleet yhä vahvemmin teknisen kiinnostuksen kohteiksi, koska niiden on tunnistettu vaikuttavan teknologian käytettävyyteen ja hallittavuuteen (Lasén 2010).

*Affective computing* etsii tapoja, joilla tietokoneet voivat tunnistaa käyttäjänsä tunteita ja reagoida niihin. Se ei nimestään huolimatta pohjaa affektin käsitteen filosofisiin ja kulttuurintutkimuksellisiin perinteisiin vaan sijoittuu insinööritieteisiin ja yhdistää psykologiaa ja tietotekniikan tutkimukseen. Tutkimussuuntauksen suosio 2000-luvulla osoittaa, että tunteet ovat ainakin joiltain osin tulleet osaksi teknologiatutkimusta. Tutkimussuuntauksen pioneeri Rosalind Picard (2004) on käyttänyt Microsoftin Office-pakettiin vuonna 1996 ilmestynyttä paperiliittimen muotoista animoitua avustajaa Clippyä esimerkkinä koneen epäonnistuneesta tunneviestinnästä. Clippy ei voinut tunnistaa



tietokoneen käyttäjän tunteita, eikä reagoida niihin. Sen sijaan se käyttäytyi itse tilanteessa kuin tilanteessa huomiota herättävän veikeästi, esimerkiksi tanssahteli ja iski silmää, mikä oli omiaan lisäämään sen käyttäjissä aiheuttamaa ärsytystä. Yhtenä ratkaisuna koneiden ja käyttäjien välisiin tunneongelmiin on tutkittu ja kehitetty koneiden kykyä hankkia tunneinformaatiota käyttäjiensä puheesta (Picard 2004). Kuunteluominaisuus, joko puheominaisuuteen yhdistettynä tai siitä erillään, vaikuttaa jälleen omalta osaltaan huomattavasti koneen ja käyttäjän suhteeseen sekä käyttäjän tunnereaktioihin.

Microsoftin klemmariavustaja ei äännellyt muuten kuin koputtamalla tai päästelemällä esimerkiksi suhahduksia ja pingahduksia kaahatessaan paikalle tai muuttaessaan muotoaan. Puhuvien ja interaktiivisuuteen pyrkivien koneiden ärsytyspotentiaalin voi olettaa olevan vielä korkeampi kuin kesken kirjoitustöitä paikalle syöksyvän mykän klemmarin. Toiminnallisesti ja tunteellisesti puhuvat koneet erottuvat omaksi kategoriakseen juuri siinä, miten ne vaikuttavat ihmisiin. Ihmismäinen puhe ei ole koneissa vain ominaisuus muiden joukossa, vaan hyvin erityinen ominaisuus koneen toimijuuden ja käyttäjän tunnereaktioiden kannalta. Kognitiivisessa tarkastelussa on havaittu luonnon äänien mahdollistavan keskittymisen ajatuksiin sekä toimintaan ja edistävän lepoa, koneiden ja ihmisten tuottamien äänten puolestaan hajottavan huomiokykyä ja aiheuttavan varuillaan oloa, mikä voi olla rasittavaa ja haitallista terveydelle (Andringa & Lanser 2013). Kulttuurintutkimuksellinen näkökulma voisi esimerkiksi äänentutkimuksen ja affektin käsitteen kautta kuitenkin vielä lisätä ymmärrystä siitä, miten millaisiinkin ääniin reagoidaan erilaisissa tilanteissa ja ympäristöissä, miten niitä tulkitaan ja millaisiin rooleihin konepuhe asettuu ääniympäristössä ja ihmisen sekä koneen suhteissa.

On hyödyllistä tarkastella puhuvia koneita omana erityisenä sosioteknisenä kategorianaan, jotta saisimme paremmin selkoa koneellisen puheen vaikutuksista. Näiden vaikutusten tarkastelun kautta voimme ymmärtää myös laajemmin ihmisten ja koneiden toimijuuden neuvottelua ja keskinäistä muokkautumista, affektien virtaamista ja aivan arkipäiväistä kanssakäymistä erilaisissa materiaalisissa ja äänellisissä suhteissa ja asetelmissa. Voimme myös terävämmin kiinnittää huomiota siihen, millaisia teknologisia kehityssuuntia tehdään todeksi laittamalla koneet puhumaan, millaisia oletuksia konepuhe tekee ihmisistä ja koneista tai haluammeko koneiden ylipäätään puhuvan meille.

## Lopuksi

Olen tässä artikkelissa hahmotellut puhuvien koneiden sosioteknisen kategorian sisältöä ja rajoja eri näkökulmista. Esitän, että historiallisen tarkastelun perusteella voidaan erottaa kolme ihmismäistä puhetta hyödyntävien koneiden teknistä tyyppiä: puhesynteesiin perustuvat koneet, ääntä välittävät koneet ja tallennettuun puheeseen perustuvat koneet. Näihin kolmeen tyyppiin lukeutuvat myös kaikki nykyiset puhuvat koneet, joko eri tyyppien yhdistelminä tai yhteen tiettyyn tyyppiin kuuluvina. Tekoälyä soveltavat puhuvat koneet hyödyntävät erityisesti puhesynteesiä, mutta erilaiset konepuheen ratkaisut voivat luoda ajattelun tai tunteiden illuusiota myös sellaisten koneiden kohtaamisessa, joihin ei sisälly tekoälyä.

Konepuhe asettaa arjen käytäntöjen osana neuvottelunalaiseksi ihmisen ja koneen toiminnallisen jaottelun. Kaikki ihmismäistä puhetta hyödyntävät koneet eivät kuitenkaan ole sosioteknisesti määriteltävissä puhuviksi koneiksi.

Tekniikan sosiaalisen tutkimuksen näkökulmasta olen tässä artikkelissa esittänyt, että puhuville koneille tunnusomaista on toimijuussuhteen neuvottelu ja sellainen suhde käyttäjänsä, jossa kone ei pelkästään välitä tai muokkaa todellisuutta vaan nostaa puheominaisuudella toiminnan keskiöön ihmisen suhteen koneeseen tai teknologiseen järjestelmään. Tunteiden ja affektiivisuuden näkökulma tuo vielä oman lisänsä puhuvien koneiden ja ihmisten suhteisiin. Esitän, että puheominaisuus luo erityisen otolliset olosuhteet outouden mutta myös ärsytyksen tunteiden syntymiselle ihmisen ja koneen affektiivisissa kohtaamisissa. Konepuheen affektiivisuus on kuitenkin jäänyt yllättävän vähälle huomiolle, vaikka tietokoneiden, tekoälyjen ja erilaisten robottien vaikutuksia käyttäjiin on tutkittu viime vuosina innokkaasti.

Puhuvat koneet eivät pelkästään välitä puhetta, vaan niiden puhe muokkaa käyttäjän ja koneen suhdetta tai ohjailee käyttäjää. Uusmaterialistisen teorian valossa voidaan ajatella, että käyttäjä saa koneita ja muita olioita sisältävissä asetelmissa hahmonsä, jota puhekommunikaatio koneen kanssa omalta osaltaan leimaa. Koneen puhe voi olla yksisuuntaista, kuten monissa puhuvissa leluissa, pienelektronikassa ja kuulutuksissa, tai vuorovaikutteista, kuten erilaisissa ääniohjauksella toimivissa laitteissa, roboteissa ja sovelluksissa. Äänentoistojärjestelmät voivat tehdä kokonaisista rakennuksista tai liikennevälineistä puhuvia koneita, joiden sisä- tai ulkopuolella oleskelevista ihmisistä tulee koneellisesti vahvistetun tai tuotetun puheen kohteita ja tietynlaisen puhekommunikaation osallisia. Äärimmäisiä esimerkkejä tällaisista arkkitehtonisista puhekoneista ovat tieteiselokuvien ja -kirjallisuuden kuvittelemat puhuvat ja ajattelevat avaruusasemat tai -alukset.

Mediakulttuurin rooli puhuviin koneisiin liittyvien odotusten luomisessa vaikuttaa merkittävältä, vaikka myös tätä puolta olisi tarpeen tutkia lisää. Mediakuvaston ruokkimat suuret ja runsaat odotukset puhuvista ja ajattelevista koneista elävät rinnan todellisuuden melko niukkojen ja usein turhauttavien tai mitäänsanomattomien konepuhekohtaamisten kanssa. Jotta niitä voitaisiin kunnolla vertailla, tarvitaan kuitenkin lisää kulttuurista ja historiallista tietoa sekä konepuheen roolista käyttökoneikassa että sen herättämistä reaktioista ja synnyttämistä tilanteista. Fiktio ja mediadiskurssien puhuvat koneet eivät elä erillään käyttöön kehitellyistä teknisistä puhe-sovelluksista, vaan määrittelevät kuviteltavissa olevan rajoja ja sitä, mikä on toivottavaa. Ne myös tapaavat nostaa toistuvasti esiin tiettyjä teemoja, kuten robotteihin kohdistuvia pelkoja tai visioita huippuälykkäistä, inhimillisistä koneista, kun taas arjen konekohtaamisissa esiin nousevat tunteet ja käytöt saattavat jäädä vähemmälle huomiolle.

Ääniympäristön ja affektiivisuuden käsitteet avaavat mahdollisuuksia sen pohtimiseen, miten puhuvat koneet asettuvat osaksi sosioteknisesti muokatun arjen monimutkaista kudosta. Kulttuurisesti opimme tulkitsemaan ja kuuntelemaan hyvin monenlaisia ääniä, jos ne ovat meille merkityksellisiä, mutta puheäänellä saattaa olla myös erityisiä affektiivisia ominaisuuksia suhteiden, toimijuuden ja vuorovaikutuksen luojana sekä toimijuuden jakautumisessa. Puheella on ääniympäristössä usein erilainen rooli kuin esimerkiksi musiikilla tai merkkiäänillä. Puhe on vahvaa toimijuutta ja vallan väline, joten puhuvien koneiden kohdalla esiin on nostettava myös kysymys siitä, kenen äänillä, millaisista positioista ja kenelle ne puhuvat.

*Kirjoittaja haluaa esittää lämpimät kiitokset artikkelin molemmille anonyymeille vertais-arvioijille, joiden kommentit ja ehdotukset auttoivat merkittäväällä tavalla kehittämään artikkelin argumentaatiota.*

## Lähteet

- Ampuja, Outi; Järviluoma, Helmi; Kilpiö, Kaarina & Uimonen, Heikki (2005) Muuttuva äänimaisema – johdatus tutkimukseen. Teoksessa Outi Ampuja & Kaarina Kilpiö (toim.) *Kuultava menneisyys: Suomalaista äänimaiseman historiaa*. Turku: Turun historiallinen yhdistys, 9–20.
- Ampuja, Outi (2007) *Melun sieto kaupunkielämän välttämättömyytenä: Melu ympäristöongelmana ja sen synnyttämien reaktioiden kulttuurinen käsittely Helsingissä*. Helsinki: SKS.
- Ampuja, Outi (2011) Meluongelman muotoutuminen ja urbaanin äänimaiseman arviointikriteerit Helsingissä 1950-luvulta nykyaikaan. *Yhdyskuntasuunnittelu* 49(1), 8–23.
- Andringa, Tjeerd C. & Lanser, J. Jolie L. (2013) How pleasant sounds promote and annoying sounds impede health: A cognitive approach. *International Journal of Environmental Research and Public Health*, 10(4): 1439–61.
- Barad, Karen (2003) Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs: Journal of Women in Culture and Society* 28(3), 801–831.
- Barad, Karen (2007) *Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.
- Bennett, Jane (2010) *Vibrant Matter: A Political Ecology of Things*. Durham: Duke University Press.
- Bijker, Wiebe E. (1997) *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge, MA: The MIT Press.
- Blackman, Lisa & Venn, Couze (2010) Affect. *Body & Society* 16:1, 7–28.
- Brackhane, Fabian (2015) "Kann was natürlicher, als Vox humana, klingen?" Ein Beitrag zur Geschichte der mechanischen Sprachsynthese. *Phonus*, No. 8. Saarbrücken: Institut für Phonetik, Universität des Saarlandes.
- Chokshi, Niraj (2018) Amazon Knows Why Alexa Was Laughing at Its Customers. *The New York Times* 8.3.2018. Saatavilla: <<https://www.nytimes.com/2018/03/08/business/alexlaugh-echo.html>>. (Linkki tarkistettu 6.4.2020.)
- Ciechanowski, Leon; Przegalinska, Alexandra; Magnuski, Mikolaj & Gloor, Peter (2019) In the Shades of the Uncanny Valley: An Experimental Study of Human–Chatbot Interaction. *Future Generation Computer Systems* 92, 539–548.
- Deleuze, Gilles (1997) *Essays Critical and Clinical*. Minneapolis: University of Minnesota Press.
- Gunning, Tom (2003) Renewing Old Technologies: Astonishment, Second Nature, and the Uncanny in Technology from the Previous Turn-of-the-Century. Teoksessa David Thorburn & Henry Jenkins (toim.) *Rethinking Media Change: The Aesthetics of Transition*. Cambridge: The MIT Press, 39–60.
- Guzman, Andrea (2017) Making AI safe for humans: A conversation with Siri. Teoksessa Robert W. Gehl & Maria Bakardjeva (toim.), *Socialbots and their friends: Digital media and the automation of sociality*. New York, NY: Routledge, 69–85.
- Hankins, Thomas L. & Silverman, Robert J. (1995) *Instruments and the Imagination*. New Jersey: Princeton University Press.
- Hemmings, Claire (2005) Invoking affect: cultural theory and the ontological turn. *Cultural Studies* 19(5), 548–567.
- Holbraad, Martin 2011. Can the Thing Speak? *Open Anthropology Cooperative Press*. Working Papers Series #7. Saatavilla: <<http://openanthcoop.net/press/http://openanthcoop.net/press/wp-content/uploads/2011/01/Holbraad-Can-the-Thing-Speak2.pdf>>. (Linkki tarkistettu 21.3.2020.)
- Kang, Minsoo (2011) *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Cambridge: Harvard University Press.
- Lasén, Amparo (2010) Mobile Media and Affectivity: Some Thoughts about the notion of Affective Bandwidth. Teoksessa Joachim Höflich, Georg F. Kircher, Christine Linke & Isabel Schlotte (toim.) *Mobile Media and the Change of Everyday Life*. Frankfurt am Main: Peter Lang, 131–154.
- Latour, Bruno (1990) Technology is Society Made Durable. *The Sociological Review* 38(1\_suppl), 103–131.
- Latour, Bruno (1992) Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. Teoksessa Wiebe E. Bijker & John Law (toim.) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, Mass.: MIT Press, 225–258.
- Latour, Bruno (1993) *We have never been modern*. Cambridge, Mass.: Harvard University Press.

- Latour, Bruno (2005) *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
- Minsky, Marvin (2006) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster.
- Mori, Masahiro (2012) The Uncanny Valley. Kääntäneet Karl F. MacDorman & Norri Kageki. *IEEE Robotics & Automation Magazine* 19(2), 98–100.
- Männistö-Funk, Tiina & Sihvonen, Tanja (2018) Voices from the uncanny valley: How robots and artificial intelligences talk back to us. *Digital Culture and Society* 4(1), 45–64. Saatavilla: <<http://digicults.org/files/2019/11/dcs-2018-0105.pdf>>. (Linkki tarkistettu 21.3.2020.)
- Newell, Sasha (2018) The Affectiveness of Symbols: Materiality, Magicity, and the Limits of the Antisemiotic. *Current Anthropology* 59(1), 1–22.
- Nivala, Asko (2019) Onko tekoälyä olemassa? *Niin & näin* 3/2019, 19–25.
- Parikka, Jussi (2004) *Koneoppi: Ihmisen, teknologian ja median kytkennät*. Turku: Turun yliopisto.
- Pettorino, Massimo (2015) The History of Talking Heads: The Trick and the Research. *First International Workshop on the History of Speech Communication Research*, Dresden, Germany, September 4–5, 2015. ISCA Archive. Saatavilla: <<https://pdfs.semanticscholar.org/ce62/d996d50d23d4138f6d99410105d18612ac71.pdf>>. (Linkki tarkistettu 21.3.2020.)
- Picard, Rosalind W. (2004) Towards Machines with Emotional Intelligence. *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics*, Setúbal, Portugal, August 25–28, 2004.
- Rainio, Riitta; Äikäs, Tiina; Lahelma, Antti & Lassfolk, Kai (2017) Nauravat kalliot: Pohjois-Suomen pyhien paikkojen kaikututkimus. Teoksessa Heikki Uimonen, Meri Kytö & Kaisa Ruohonen (toim.) *Muuttuvat suomalaiset äänimaisemat*. Tampere: Tampereen yliopistopaino, 167–199.
- Richardson, Kathleen (2016) Technological Animism: The Uncanny Personhood of Humanoid Machines. *Social Analysis* 60(1), 110–128.
- Riskin, Jessica (2016) *The Restless Clock: A History of the Centuries-long Argument over what Makes Living Things Tick*. Chicago: University of Chicago Press.
- Rowe, Debbie A. (2010) From Wood to Bits to Silicon Chips: A History of Computer Synthesized Speech. Teoksessa John Mullenix & Steven Stern (toim.) *Computer Synthesized Speech Technologies: Tools for Aiding Impairment*. Hershey: Medical Information Science Reference, 9–27.
- Scientific American* (1896) The Phonograph. 25.7.1896, Vol. 75, No. 4, 65–66.
- Seyfert, Robert (2012) Beyond Personal Feelings and Collective Emotions: Toward a Theory of Social Affect. *Theory, Culture & Society*, 29(6), 27–46.
- Skjuve, Marita; Haugstveit, Ida Maria; Følstad, Asbjørn & Brandtzaeg, Petter Bae (2019) Help! Is my Chatbot Falling into the Uncanny Valley?: An Empirical Study of User Experience in Human-Chatbot Interaction. *Human Technology*, 15(1), 30–54.
- Slaby, Jan (2018) Affective Arrangement. Teoksessa Jan Slaby & Christian von Scheve (toim.) *Affective Societies: Key Concepts*. New York: Routledge, 109–118.
- Smith, Jacob (2008) Tearing Speech to Pieces: Voice technologies of the 1940s. *Music, Sound, and the Moving Image* 2, 183–206.
- Sterne, Jonathan (2003) *The Audible Past. Cultural Origins of Sound Reproduction*. Durham: Duke University Press.
- Taylor, Jessica (2009) Speaking Shadows: A History of the Voice in the Transition from Silent to Sound Film in the United States. *Journal of Linguistic Anthropology*, 19(1), 1–20. Saatavilla: <<https://doi.org/10.1111/j.1548-1395.2009.01016.x>>. (Linkki tarkistettu 21.3.2020.)
- Tekniikan Maaailma (2018) Koneiden äly lähentyy ihmismieltä. *Tekniikan Maaailma* 4/2018, 21.2.2018, 100–103.
- Thompson, Emily (2002) *The Soundscape of Modernity. Architectural Acoustics and the Culture of Listening in America, 1900–1933*. Cambridge: The MIT Press.
- The Times (1846) The Speaking Automaton. *The Times*, 12.8.1846, 3.
- Truax, Barry (2001) *Acoustic Communication*. Lontoo: Ablex Publishing.
- Truitt, E. R. (2015) *Medieval Robots: Mechanism, Magic, Nature, and Art*. Philadelphia: University of Pennsylvania Press.
- Uimonen, Heikki (2005) *Ääntä kohti. Ääniympäristön kuuntelu, muutos ja merkitys*. Acta Universitatis Tamperensis 1110. Tampere: Tampere University Press.



Verbeek, Peter-Paul (2001) Don Ihde: The Technological Lifeworld. Teoksessa Hans Achterhuis (toim.) *American Philosophy of Technology: The Empirical Turn*. Bloomington: Indiana University Press, 119–146.

Verbeek, Peter-Paul (2016) Toward a Theory of Technological Mediation: A Program for Post-phenomenological Research. Teoksessa Jan Kyrre Berg Olsen Friis & Robert P. Crease (toim.) *Technoscience and Postphenomenology: The Manhattan Papers*. Lontoo: Lexington Books, 189–204.

Vincent, James (2017) Sophia the robot's co-creator says the bot may not be true AI, but it is a work of art. *The Verge* 10.11.2017. Saatavilla: <<https://www.theverge.com/2017/11/10/16617092/sophia-the-robot-citizen-ai-hanson-robotics-ben-goertzel>>. (Linkki tarkistettu 21.3.2020.)

Wong, Julia Carrie (2018) Amazon working to fix Alexa after users report random burst of 'creepy' laughter. *The Guardian* 7.3.2018. Saatavilla: <<https://www.theguardian.com/technology/2018/mar/07/amazon-alexa-random-creepy-laughter-company-fixing>>. (Linkki tarkistettu 21.3.2020.)