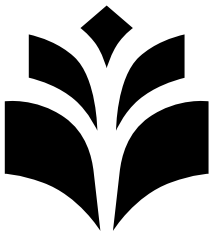
Puheteknologiat masennuksen diagnosoinnissa

Rene Ruotsalainen

Kandidaatintutkielma



UNIVERSITY OF
EASTERN FINLAND

Tietojenkäsittelytieteen laitos
Tietojenkäsittelytiede
Helmikuu 2022

ITÄ-SUOMEN YLIOPISTO, Luonnontieteiden ja metsätieteiden tiedekunta, Joensuu
Tietojenkäsittelytieteen laitos
Tietojenkäsittelytiede

Ruotsalainen, Rene: Puheteknologiat masennuksen diagnosoinnissa
Kandidaatintutkielma, 18 s.
Ohjaaja: Tomi Kinnunen
Helmikuu 2022

Tiivistelmä: Tässä tutkielmassa muodostetaan kirjallisuuskatsaus jonka tarkoitus on selvittää millaisia puheteknologisia keinoja on hyödynnetty masennuksen varhaisessa diagnosoinnissa. Kirjallisuuskatsauksessa kuvataan perustavanlaatuinen määritelmä masennuksesta mielialahäiriönä, sekä kuinka kyseisen sairauden diagnosointi voisi hyötyä puheteknologisista menetelmistä. Tätä seuraa perinteisten diagnosointikeinojen kuvaus. Luvussa 2 määritellään ne puheen ominaisuudet joita analysoimalla masennuksen havaitseminen on mahdollista. Tätä seuraa katsaus useista erilaisista puheteknologisista lähestymistavoista sekä millaisia tuloksia niiden sovelluksista on saatu. Tutkielman tuloksena voisi todeta että puheteknologiset menetelmät voisivat toimia jo olemassa olevien menetelmien rinnalla, keinona aikaistaa masennuksen diagnoosia. On kuitenkin epätodennäköistä että ne täysin korvaisivat perinteiset menetelmät.

Avainsanat: puheteknologia; puheentunnistus; masennus

Sisällys

1. Johdanto

1.1 Sairauden kuvaus

Maailman terveysjärjestön mukaan masennus eli *depressio* on yleinen mielialahäiriö, josta kärsii maailmanlaajuisesti arviolta noin 264 miljoonaa ihmistä (*Mental Disorders*, 2017). Se on yleisempää naisilla kuin miehillä. Masentuneelle henkilölle tyypillistä on yleinen mielenkiinnon tai hyvänolon tunteen puute, matala itsetunto, ruokahaluttomuus, väsyneisyys ja huono keskittymiskyky. Masennus on herkästi uusiutuva tila ja noin puolella ensimmäisestä masennusjaksostaan parantuneella sairausjakso toistuu (Huttunen, 2018).

Masennuksen vakavuus voi vaihdella ajoittain. Lievät masennustilat voivat olla vain pieni haitta työkyvykkyydelle tai sosiaaliselle kanssakäymiselle, mutta pitkittynyt vakava masennus voi tehdä ihmisestä jo täysin työkyvyttömän. Masennus on myös merkittävä tekijä kohonneelle itsemurhariskille. Jonkin asteisesta masennuksesta kärsivien joukossa itsemurhariski on arviolta noin 5 %, pitkään vakavasta masennuksesta kärsineiden kesken jopa 15-20 %. (Huttunen, 2018)

Koska masennus on hyvin yleisesti esiintyvä mielialahäiriö ja se aiheuttaa yksilön normaalia toimintaa lamauttavia oireita, sen hoito on tärkeää pelkästään jo yleisen kansanterveyden kannalta. Työkyvyttömyys ja sosiaalisesta kanssakäymisestä vetäytyminen tarkoittaa ettei yksilö pysty toimimaan rakentavana osana yhteiskuntaa, joka aiheuttaa vuorostaan taloudellista taakkaa. Ongelma ei kuitenkaan ole helppo ratkaista. Koska masennuksesta kärsivillä ihmisillä on taipumus vetäytyä sosiaalisista kontakteista, se voi täysin lähipiirin huomaamatta syventyä vaikeaksikin masennukseksi.

1.2 Perinteiset tavat diagnosoida masennus

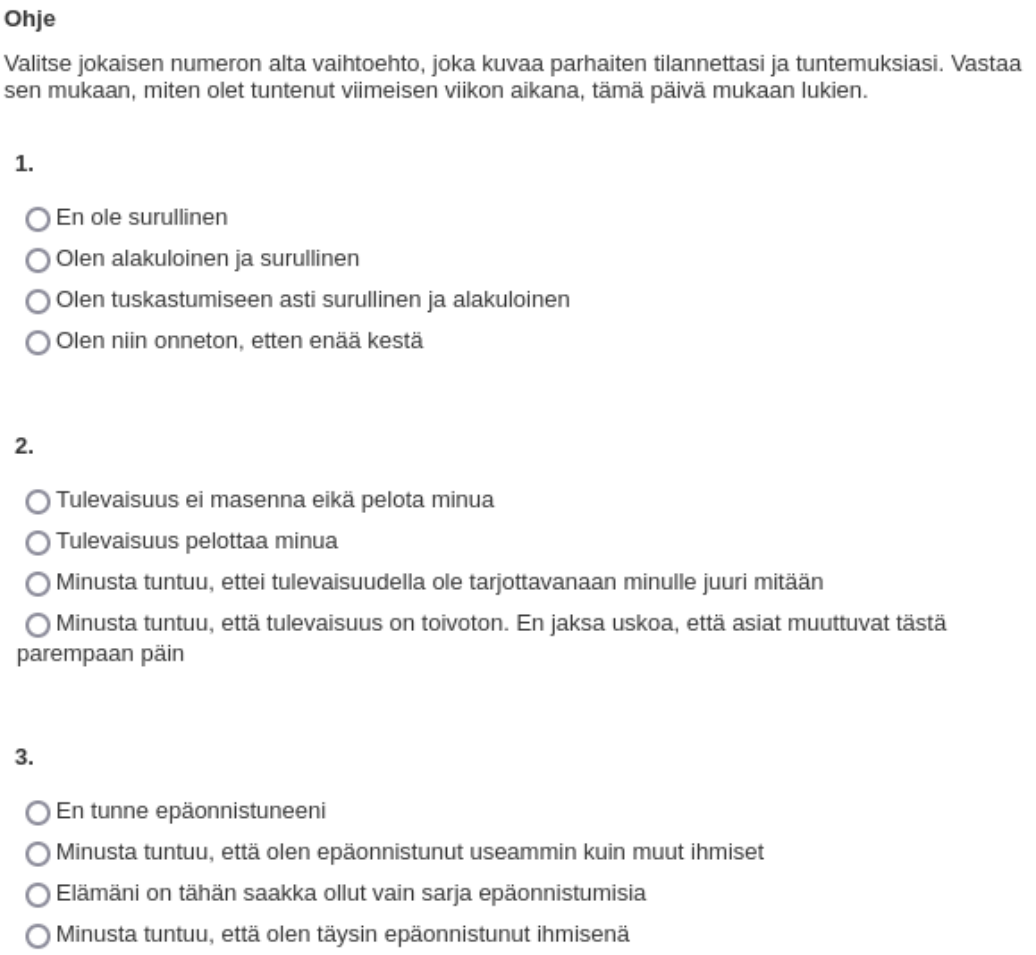
Masennuksen diagnoosi alkaa tavallisesti seulontaprosessista. Seulontaan kuuluvat erilaiset kyselyt jotka arvioivat yksilön mielentilaa sekä pyrkivät havaitsemaan viitteitä masennuksesta (*Depressio: Käypä hoito -suositus*, 2020). Kyselyt voivat olla esimerkiksi täysin julkisia verkkokyselyitä tai terveydenhuollon ammattilaisen laatimia kyselyitä joilla pyritään kartoittamaan jonkin tietyn väestönosan hyvinvointia.

Perinteisellä kyselyihin perustuvalla seulonnalla on kuitenkin omat heikkoutensa. Koska seulonta ei ole yksilöityä, ja sen perusteella saadut tulokset perustuvat vapaaehtoisesti luovutettuihin tekstipohjaisiin vastauksiin, on hyvinkin todennäköistä että moni kyselyihin vastanneista ei saa tarvitsemaansa apua. Lisäksi on hyvin mahdollista että osa vastanneista johdatetaan hoidon pariin, vaikka he eivät tarvitsisi sitä. Yleinen masennuksen seulontaan käytetty työkalu on *Patient Health Questionnaire-9* (PHQ-9). PHQ on diagnostinen työkalu henkilön terveydentilan arviointiin (Kroenke, Spitzer & Williams, 2001). PHQ-8 ja PHQ-9 ovat sen osia, joissa kysymykset on laadittu masennuksen arviointia varten. Nimen perässä oleva numero viittaa kyselyn kysymyksien lukumäärään. Wittkampfin ja muiden (2007) tekemässä tutkimuksessa PHQ-9 -kyselyn todettiin olevan hyvä työkalu perusterveydenhuollossa, jos sitä voidaan käyttää kohderyhmälle jossa masennuksen tiedetään olevan keskimääräistä yleisempää. Manean ja muiden (2015) mukaan PHQ-9 ei ole riittävän herkkä erottamaan lievästi masentuneita henkilöitä, joka voi johtaa vääriin negatiivisiin tuloksiin. Tämä tarkoittaa että PHQ-9 saattaa olla hyödyllinen työkalu masennuksen vakavuuden luokittelussa mutta ei sen tunnistamisessa.

Seulonnan tulosten perusteella yksilöitä voidaan suositella tarkempaan tutkimukseen. Viimekädessä masennuksen diagnoosi tehdään kuitenkin aina lääkärin haastattelun pohjalta, eikä tätä tulisikaan täysin korvata millään automatisoidulla järjestelmällä.

1.3 Puheteknologian tuomat mahdollisuudet

Masennukseen sairastuu elämänsä varrella 10-15 % ihmisistä (Huttunen, 2018). Sairaus onkin hyvin yleinen, ja monioireisen luonteensa takia varsin hankalasti diagnosoitavissa ilman että oireista kärsivä tunnistaa ne itse ja hakee omatoimisesti apua. Tästä johtuen uskon että puheteknologisia keinoja käyttävät menetelmät voisivat madaltaa kynnystä hakea ammattiapua. Lisäksi käyttäessä puheteknologisia menetelmiä masennuksen havaitsemiseen, pyritään tutkimuksesta poistamaan tietynlainen inhimillinen elementti, joka voisi vääristää seulonnalla suoritettua tutkimusta.



Ohje

Valitse jokaisen numeron alta vaihtoehto, joka kuvaa parhaiten tilannettasi ja tuntemuksiasi. Vastaa sen mukaan, miten olet tuntenut viimeisen viikon aikana, tämä päivä mukaan lukien.

1.

- En ole surullinen
- Olen alakuloinen ja surullinen
- Olen tuskastumiseen asti surullinen ja alakuloinen
- Olen niin onneton, etten enää kestä

2.

- Tulevaisuus ei masenna eikä pelota minua
- Tulevaisuus pelottaa minua
- Minusta tuntuu, ettei tulevaisuudella ole tarjottavanaan minulle juuri mitään
- Minusta tuntuu, että tulevaisuus on toivoton. En jaksa uskoa, että asiat muuttuvat tästä parempaan päin

3.

- En tunne epäonnistuneeni
- Minusta tuntuu, että olen epäonnistunut useammin kuin muut ihmiset
- Elämäni on tähän saakka ollut vain sarja epäonnistumisia
- Minusta tuntuu, että olen täysin epäonnistunut ihmisenä

Kuva 1.1: Verkossa täytettävä seulontakysely (Terveyden ja hyvinvoinnin laitos, 2021)

Oireista kärsivää osaa väestöstä voi olla hankala tavoittaa, ja laaja tavoittaminen voi olla myös taloudellisesti rasittavaa. Hämäläisen ja muiden (2008) tutkimuksen mukaan vain noin 50 % erilaisista masennuksen oireistä kärsivistä ihmisistä hakee aktiivisesti apua terveydenhuollosta. Vaihtoehtoisia keinoja masennuksen oireiden havaitsemiseen tulisi kokeilla mahdollisuuksien mukaan. Tässä puheteknologiset keinot voisivat tarjota lievitystä yhteiskunnalliseen rasitukseen. Quatierin ja Malyskan (2012) suorittamassa tutkimuksessa masennuksen todettiin vaikuttavan ihmisen tuottamaan puheeseen, ja pystyttiin havaitsemaan erilaisia biomarkkereita joiden avulla masentuneet henkilöt voitaisiin tunnistaa. Puheteknologisten metodien paikka masennuksen diagnosoinnissa olisi kuitenkin sen ensioireiden havaitsemisessa.

1.4 Tutkielman sisältö ja tavoitteet

Seuraavassa tutkielmassa tuodaan esille erilaisia näkökulmia ja menetelmiä masennuksen aikaiseen diagnosointiin. Tarkastelussa on nimenomaan puheteknologisia keinoja hyödyntävät lähestymistavat. Esiteltävistä menetelmistä osa on täysin uusia ja osa pyrkii parantamaan aiemman menetelmän tarkkuutta. Menetelmät käydään perustavan laatuisesti läpi muodostaen lukijalle selkeän mielikuvan siitä, miten tutkimukset on toteutettu, ja mitä tuloksia niissä on saavutettu.

Tutkielman luvussa 1.2 kuvataan perinteinen menetelmä masennuksen diagnoosin saavuttamiseen. Tämä on lähtötilanne, jonka rinnalle pyritään löytämään mahdollisia vaihtoehtoja. Luvussa 2 annetaan pohjustava käsitys siitä kuinka puhetta analysoidaan, ja mitä puheen osia voidaan hyödyntää masennuksen diagnosoinnissa. Alaluvuissa siirrytään tarkastelemaan eri toteutustavoilla suoritettuja tutkimuksia. Tutkimukset on eriltelty laadun ja menetelmien perusteella omiksi alaluvuikseen.

Tavoitteena on muodostaa tilannekuva puheteknologian tämänhetkisistä sovelluksista sekä niiden mahdollisuuksista masennuksen diagnosoimisessa. Koska puheen analysointi puheteknisesti masennuksen havaitsemisessa on varsin nuori tutkimuskohde, on mielestäni tärkeää myös pitää yhtenä tavoitteena koota yhteen ja tuoda esille erilaisia vaihtoehtoisia tapoja tehostaa masennuksen varhaista diagnosointia.

2. Puheteknologiset lähestymistavat diagnosointiin

Ennen eri lähestymistapojen tarkempaa tarkastelua, on hyvä ymmärtää yksinkertaisimmillaan kuinka puhetta analysoimalla voidaan havaita masennukseen viittaavia tekijöitä. Tunnetun datan analysoinnin avulla on voitu havaita tiettyjä yhtymäkohtia masentuneiden henkilöiden puheessa. *Shimmer* eli puheen amplitudin epävakaus on yksi ominaisuus jonka on havaittu kasvavan masennuksen vaikeutuessa (Quatieri & Malyska, 2012). Stasak, Epps ja Goecke (2017) huomasivat vuorostaan tutkimuksessaan että käyttäessään analysoinnissa lauseita jotka sisältävät enemmän vaivaa vaativia puheen ominaisuuksia, luokittelu tuottaa parempia tuloksia verrattuna kaikkien lauseiden käyttöön. Tästä voi päätellä että masennus vaikeuttaa hankalaa artikulaatioita tai monimutkaisempaa kieltä vaativien sanojen tuottamiseen.

Syväoppimisen menetelmät ovat tavallisia puheentunnistuksen sovelluksissa. Tässä katselmassa esiteltävissä tutkimuksissa syväoppimista käytetään pääasiassa datan luokitteluun, joka on yksi merkittävistä syväoppimisen menetelmien käyttökohteista.

2.1 Neuroverkko

Keinotekoinen neuroverkko (artifical neural network) tai lyhyemmin vain *neuroverkko* (neural network) on yksi koneoppimisen malleista. Neuroverkoista puhuessa saatetaan käyttää myös termiä *syväoppiminen* (deep learning). Niitä hyödynnetään mm. tiedon luokittelussa ja ryhmittelyssä. Neuroverkot muodostavat keskeisen osan monista tässäkin tekstissä esitellyistä tutkimuksista.

Yksinkertainen neuroverkon malli muodostuu joukoista neuroneita jotka muodostavat kerroksia. Näiden kerroksien neuronit ovat kytketty jokaiseen seuraavan kerrokseen neuroniin. Ensimmäinen kerros muodostuu syötteestä ja viimeinen kerros sisältää tulosteen. Välissä on vähintään yksi piiloitettu kerros. Neuronit ovat kytkettynä toisiinsa

painotetuilla kaarilla jotka muokkaavat syötettä. Neuronit aktivoivat seuraavia neuroneita näiden laskutoimitusten perusteella. Funktioon lisätään tavallisesti jokin raja-arvo jonka sen täytyy ylittää aktivoidakseen seuraavan neuronin. Esimerkki kuvastaa yleistä monikerrosperseptronia (Multilayer Perceptron, MLP).

Konvoluutioneuroverkko (convolutional neural network, CNN) on MLP joka sisältää konvoluutiokerroksia. Ne ovat erityisen tehokkaita analysoimaan kuvia ja muita signaaleita. Tehokkuus perustuu reunojen ja muotojen havaitsemiseen eri suodattimien avulla mahdollistaen tiedon tiivistämisen menettämättä tärkeitä ominaisuuksia. Suodattimien koot ja niiden sisältämät arvot ovat vapaasti määriteltävissä. Konvoluutiokerroksen tuloste on sen syötteen sekä sen suodattimien konvoluutio, jossa suodatin liukuu syötteen yli ottaen pistetulon suodattimen arvojen sekä suodattimen alle jäävien syötteen arvojen välillä. Tämä arvo asetetaan tulosteeseen, sama toistuu kunnes kaikki syötteen arvot ovat käyty läpi.

Pistetulon kaava on

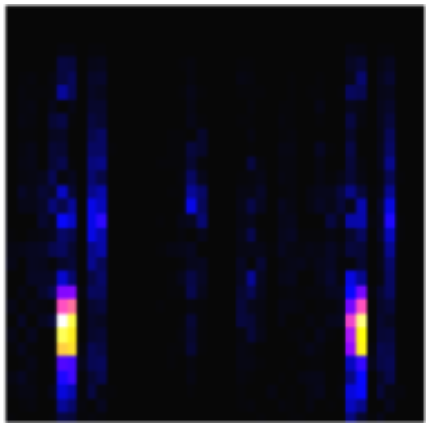
$$a \cdot b = (a_1 \cdot b_1) + ... + (a_n \cdot b_n),$$

missä a ja b ovat vektoreita. a sisältää suodattimen arvot ja b sisältää ne syötteen arvot jotka jäävät suodattimen alle.

2.2 Neuroverkkojen hyödyntäminen

Chlasta, Wołk ja Krejtz (2019) toteuttivat tutkimuksessaan uuden sovelluksen masennuksen oireiden havaitsemiseen, joka hyödyntää konvoluutioneuroverkkoa. Tutkimus pohjautuu vuoden 2017 AVEC "Real-life Depression and Affect Recognition-työpajan sekä 2018 Interspeech -konferenssin julkaisemiin vakuuttaviin tuloksiin automaattisesta masennuksen havaitsemisesta (ADD, Automatic Depression Detection). Tulokset antavat osviittaa ADD -metodien tehokkuudesta ja tarkkuudesta, varsinkin multimodaalisissa lähestymistavoissa.

Tutkimuksessa käytetyt tietojoukot muodostettiin DAIC-WOZ -tietokannan (*DAIC-WOZ Database*, 2016) sisältämistä tallenteista. Jokainen tallenne sisältää äänitteen tapaamisesta, transkription äänitteestä sekä yhdellä arvolla esitetyn masennusarvon PHQ-8 -asteikon mukaisesti. Asteikon mukaan 10 tai suurempi arvo viittaa masentuneeseen koehenkilöön ja sitä pienempi arvo viittaa ei-masentuneeseen koehenkilöön. Tietokannan tallenteista valittiin samat 107 tallennetta joita käytettiin aiemmin mainitun 2017 AVEC -työpajan julkaisuissa, joka helpottaa vertailua tulosten välillä. Äänitteiden



Kuva 2.1: Esimerkki CNN:lle syötetystä spektrogrammista. (Chlasta et al., 2019)

esiprosessoinnissa niistä muodostettiin kaksi tietojoukkoa. Tutkimuksessa käytettiin 15 sekuntin mittaisia näytteitä jotka leikattiin käyttämällä Python ja Sox -ohjelmia.

Tietojoukko A sisälsi 107 puhenäyttettä, jotka leikattiin 60 sekunnin kohdalta jokaisesta äänitteestä. Näytteitä ei otettu äänitteiden alusta jotta niiden laatu olisi mahdollisimman korkea. Toisinsanoen näytteiksi haluttiin valita kohta joka sisältäisi mahdollisimman paljon puhetta ja mahdollisimman vähän taustamelua. Toinen tietojoukko B luotiin jotta menetelmää voidaan testata huomattavasti suuremmalla määrällä näytteitä. Se täytettiin näytteillä samoista tallenteista, jotka leikattiin 15 sekuntin mittaisiksi näytteiksi 60 sekuntin kohdalta aina seitsemänteen minuuttiin asti. Tämä muodosti 2568 näytettä joista 720 oli masentuneilta koehenkilöiltä.

WAV-muodossa olevat tallenteet prosessoitiin niin että niitten näytteenottotaajuus pieneni. Prosessoiduista tiedostoista tallennettiin 224x224 pikseliä kooltaan olevat spektrogrammit. Näitä muodostettuja kuvia käytettin konvoluutionneuroverkon koulutuksessa sekä ennustuksessa. Myös suuremman resoluution kuvia testattiin, mutta niiden ei huomattu vaikuttavan merkittävästi tuloksiin.

Tutkimuksessa käytettiin useampaa, jo valmiiksi koulutettua jäännösneuroverkkoa (residual neural network, ResNet), joita muokattiin sopimaan tutkimuksessa käytettyyn menetelmään. Tutkimuksessa käytettyjen mallien nimessä oleva luku kertoo alkuperäisten kerrosten lukumäärän. Mallien viimeinen kerros muokattiin sisältämään kaksi ulostuloa, yksi kumpaakin ennustusluokkaa kohden (masentuntut/ei-masentunut). Keskeinen idea oli kouluttaa valmista mallia, mutta myös viimeistä kouluttamatonta kerrosta erillään. Tämä tehtiin tarkkailemalla häviöfunktiota ja keskeyttäen koulutuksen kun oppimisnopeus on korkeimmillaan, häviöfunktion arvon kuitenkin vielä laskiessa. Kaikki paitsi viimeinen konvoluutiokerros jäädytettiin ennen kuin koulutusta jatkettiin. Tutkimuksen tulosten mukaan testit joissa käytettiin ResNet-34 ja ResNet-50 -malleja

tuottivat parhaimmat tulokset. ResNet-34 -järjestelmää testattiin tietojoukolla A ja ResNet-50 -järjestelmää tietojoukolla B.

Myös toinen tutkimusryhmä hyödynti samaa DAIC-WOZ -tietokantaa tutkiessaan konvoluutioneuroverkkojen käyttömahdollisuuksia masennuksen tunnistamisessa. Srimadhur ja Lalitha (2020) käyttivät tutkimuksessaan kolmea erilaista mallia, joista kaksi pyrki muodostamaan binääriluokituksen syötetyistä näytteistä, kun taas kolmas pyrki tarkentamaan luokittelua jakamalla näytteet neljään eri luokkaan masennuksen vakavuuden perusteella. Siinä tietokannan tallenteista muodostettiin 2240 seitsemän sekuntin pituista näytettä joista 1107 oli ei-masentuneilta ja 1133 oli masentuineilta koehenkilöiltä.

Ensimmäinen malli käytti syötteenä näytteistä muodostettuja spektrogrammeja. Malli sisältää kolme konvoluutiokerrosta ja käyttää Max-pooling -kerroksia ominaisuuksien terävöittämiseen. Mallin yleistämiseksi jokaisen pooling-kerroksen jälkeen siihen lisättiin drop out -kerros todennäköisyydellä 0.25. Tämän todennäköisyyden mukaan drop out -kerros hylkää osan neuroverkon neuroneista. Srimadhurin ja Lalithan (2020) mukaan spektrogrammeista voi havaita että ei-masentuneessa tilassa puheen intensiteetti kohdistuu matalemmille taajuuksille, kun taas masentuneessa tilassa myös korkeamman taajuuden komponentit ovat intensiivisiä. Lisäksi intensiivisyys esiintyy lyhyemmissä ajanjaksoissa. Näiden ominaisuuksien perusteella konvoluutioneuroverkko lajitteli näytteet masentuneiksi ja ei-masentuneiksi.

Toinen malli luokittelee näytteet niiden raa'an aaltomuodon perusteella spektrogrammin sijaan. Suodattimien koot muutettiin konvoluutiokerroksittain kokoihin 32, 18 ja 12 jotta ne soveltuisi paremmin aaltomuotoisen syötteen ominaisuuksien tunnistamiseen. Useista tutkimuksessa testatuissa binääriluokittelevista end-to-end -malleista parhaiten suoriutui malli 6.

Kolmas malli laajensi toista mallia luokittelemalla näytteet kahden luokan sijaan neljään eri luokkaan PHQ-8 -luokittelun mukaan siten, että PHQ-8 -tulokset 10-13, 14-16 ja 17-20 muodostivat kolme masentunutta luokkaa. Ei-masentuneen luokan muodostivat näytteet joiden PHQ-8 -tulos oli alle 10. Tutkimuksessa todettiin että datan määrä ei ole riittävä kouluttamaan mallia joka pystyy luokittelemaan näytteet edes kohtalaisin tuloksin.

Taulukossa 2.1 on listattu Chlastan ja muiden (2019) sekä Srimadhurin ja Lalithan (2020) tutkimuksissa parhaiten suoriutuneiden neuroverkkomallien tulokset. Nopeasti tarkasteltuna ResNet 34 -malli vaikuttaisi parhaalta esitellyista tuloksista, mutta täytyy muistaa että sitä testattiin tutkimuksessa mainitulla tietojoukolla A, joka oli suhteellisen suppea kooltaan. Kyseinen malli luokitteli 27 näytteestä 26 ei-masentuneeksi,

Taulukko 2.1: Parhaiten suoriutuneiden mallien tulokset

Malli	F-tulos	Sisäinen tarkkuus	Saanti	Ulkoinen tarkkuus	Tietojoukko (Julkaisu)
Jäännösneuroverkko 34	0.62	0.57	0.67	81%	Tietojoukko A (Chlasta et al., 2019)
Jäännösneuroverkko 50	0.31	0.33	0.29	67%	Tietojoukko B (Chlasta et al., 2019)
Spektrogrammimalli 1	0.66	0.56	0.80	59%	Koko DAIC-WOZ -tietokanta (Srimadhur & Lalitha, 2020)
Spektrogrammimalli 2	0.66	0.58	0.77	61%	
End-to-end -malli 6	0.77	0.79	0.74	75%	Koko DAIC-WOZ -tietokanta (Srimadhur & Lalitha, 2020)

ja kun tietojoukko oli niin epätasapainossa että 20 näytettä oli ei-masentuneilta, vääristää tämä mielestäni tuloksia aika merkittävästi. ResNet-50 -malli, jota testattiin suuremmalla tietojoukolla vaikuttaa paremmalta, mutta sekin luokittelee näytteet hyvin herkästi ei-masentuneiksi. 469 ei-masentuneesta näytteestä se luokitteli 384 onnistuneesti, mutta toisaalta 183 masentuneesta näytteestä myös 114 luokiteltiin väärin ei-masentuneeksi. Srimadhurin ja Lalithan (2020) esittämät spektrogrammimallit saavuttivat keskivertotuloksia, mutta tutkimuksen mukaan niitä haittasivat näytteiden laatuun liittyvät epätasaisuudet. End-to-end -koulutetun mallin oli tarkoitus paikata tätä puutteellisuutta, ja se pystyikin luokittelemaan näytteet suuremmalla tarkkuudella. Tämän mahdollisti suuri määrä näytteitä tilanteessa missä näytteet jaettiin vain kahteen luokkaan. Tarkempaan luokitteluun pyrkivän mallin end-to-end -kouluttamiseen ei kuitenkaan tietojoukon laajuus riittänyt.

2.3 Sukupuoliriippuvainen vokaalien analysointi

Vlasenkon ja muiden (2017) julkaisemassa tutkimuksessa sen sijaan tarkasteltiin kvalitatiivisesti kuinka vokaaliäänteiden analysoiminen voisi auttaa masennuksen oireiden tunnistamisessa. Vokaali on äänne jossa lausuttaessa ääniväylä on avoin eivätkä kieli, hampaat tai huulet rajoita ilman virtausta äännettä muodostaessa. Vokaaliäänteiden tunnistaminen pohjautuu ensimmäisen ja toisen formantin taajuuksien analysointiin. Formantit ovat puhesignaalin taajuuksia joissa äänenpaine on selvästi suurempi muuhun signaaliin verrattuna (*What are formants?*, 2005). Tutkimuksessä käytetyssä DAIC-WOZ -tietokannassa tallenteet on luokiteltu masentuneeksi tai ei-masentuneeksi PHQ-8 -pisteytyksen perusteella. Tietokannan tallenteet ovat englanniksi, joten tutkimus hyödyntää englannin kielen fonetiikkaa.

Tietokannan tallenteista muodostettiin 32249 näytettä. Näytteiden vokaaliäänteistä arvioitiin niiden foneettiset rajat käyttämällä *pakotettua linjausta* (forced alignment),

minkä lisäksi niille muodostettiin foneettiset käännökset. Koska tietokanta ei sisältänyt käännöksiä, ne otettiin CMU Pronouncing Dictionary -sanakirjasta. Hyödyntämällä PRAAT -ohjelmistoa (Boersna & Weenink, 2022) formanteille muodostettiin ylä- ja alarajat, jonka jälkeen jokaisesta vokaaliosuudesta otettiin formanttien keskiarvot. Tämä tehtiin jokaiselle 15 painottamattomalle ARPAbet-joukon foneemille sukupuolikohtaisesti. Lopputuloksena saatiin sukupuoliriippuvaiset keskimääräiset formanttien taajuudet, sekä niiden keskihajonta. Tutkimukseen valittiin 2x10 äännettä jotka poikkesivat eniten todennäköisyysjakaumasta. Äänteitä valittaessa Vlasenko, Sagha, Cummins ja Schuller (2017) huomasivat selkeän sukupuolten välisen eron poikkeavissa foneemeissa, mikä tukee sukupuoliriippuvaista analyysiä.

Tutkimus näyttää että vokaalien ensimmäisen ja toisen formantin keskimääräiset arvot eroavat toisistaan masentuneilla ja ei-masentuneilla puhujilla. Myös sukupuolten välinen ero on huomattava. Miehillä on havaittavissa selkeä siirtymä alaspäin ensimmäisessä formantissa, kun taas naisilla siirtymä on ylöspäin. Näiden tutkimuksessa tehtyjen havaintojen perustella Vlasenko ja muut (2017) halusivat verrata uuden menetelmänsä suoriutumiskykyä jo laajemmin tunnistettuun eGeMAPS-menetelmään. *Extended Geneva Minimalistic Acoustic Parameter Set* (eGeMAPS) (Eyben ym. 2016) on akustisten ominaisuuksien joukko jota on käytetty laajalti eri paralingvistisissa tehtävissä, mukaan lukien masennuksen tunnistamisessa (Vlasenko ym. 2017).

Taulukko 2.2: Eri menetelmien F_1-tulos masentuneille (ei-masentuneille) luokille.

	eGeMaps		VL-Formants	VL-Formants & eGeMaps
	Sukupuoliriippumaton	Sukupuoliriippuvainen		
Mies	0.14(0.73)	0.48(0.62)	0.53(0.71)	0.52(0.70)
Nainen	0.78(0.86)	0.83(0.90)	1.00(1.00)	1.00(1.00)
Molemmat	0.55(0.79)	0.65(0.77)	0.75(0.87)	0.74(0.86)

Suoriutumiskykyjen vertailu tehtiin lineaarisella luokittelulla käyttämällä *Liblinear*-kirjastoa. Testi suoritettiin kahdelle eri joukolle näytteitä, joista laajempi ja paremmin suoriutunut on esitelty taulukossa 2.2. Näytteitä karsittiin niin että jäljelle jäi vain lausahdukset jotka sisälsivät aiemmin valittuja todennäköisesti poikkeavia äänteitä. Tutkimuksessa tehty eGeMaps-luokittelu tehtiin yleisesti kummallekkin sukupuolelle, sekä ottamalla huomioon erot sukupuolien välisissä ominaisuuksissa.

Merkittävin parannus huomattiin miesten tuloksissa kun siirryttiin sukupuoliriippuvaiseen luokitteluun. Vlasenko ja muut (2017) kuitenkin huomauttaa että sukupuoliriippumattoman

luokittelun F_1-tulos on varsin huono miesten tapauksessa. Tutkimuksessa esitellyt vokaalitason formantit (VL-formants) johdonmukaisesti suoriutuvat paremmin kuin eGeMaps, korostaen niiden kykyä havaita masennuksen aiheuttamia muutoksia puheesta. Menetelmä saavutti täydellisen F_1-tuloksen naisten masentuneelle ja ei-masentuneelle luokille. Vlasenkon ja muiden (2017) mukaan tämä tulos antaa vahvaa näyttöä siitä että olennaiset masennuksen vaikutukset esiintyvät naispuhujien formanteissa. Tutkimuksessa myös todetaan että sukupuolten eroavan suorituskyvyn suurus oli hieman yllättävä, mutta kuitenkin odotettavissa ottaen huomioon erot masennuksen vaikutuksissa miesten ja naisten välillä. Menetelmien yhdistäminen ei aiheuttanut huomattavaa eroa tuloksissa.

Tutkimuksessa mainittiin kuinka *VL-Formants* päihitti useat muut puhetekniset lähestymistavat jotka käyttivät saman tietokannan näytteitä. Lupaavat tulokset laajalti erilaisiin ratkaisuihin verrattuna kertovat siitä että puheen vokaaliformanttien muutokset ovat hyvä lähestymistapa masennuksen havaitsemiseen. Ymmärrettävää kuitenkin on että kun aihetta ei ole tutkittu paljoa, täytyy tässä tutkimuksessa saadut tulokset pystyä varmentamaan saavuttamalla samankaltaisia tuloksia myös muiden tietokantojen tallenteita hyödynnettäessä. Jos tämä onnistuu, voisi vokaalien formanttien tarkastelua ajatella potentiaalisena vaihtoehtona muodostamaan ratkaisu joka pystyy saavuttamaan hyviä tuloksia myös tuntemattomalle datalle.

3. Käyttöönoton mahdollisuudet

Kasvava määrä tutkimuksia luo paljon uusia mahdollisuuksia hyödyntää puheteknologiaa terveydenhuollon työkaluna. Jotta puheteknologisia menetelmiä voisi ottaa käyttöön terveydenhuollossa, täytyy niiden tarjota jotain uuta masennuksen tunnistamisprosessiin. Tavoitettavuus on yksi elementti jossa puhetekniset ratkaisut voisivat tarjota suuren edun perinteisiin kyselyihin verrattuna. Puhenäytteiden keräämisen voisi olettaa olevan helpompaa kuin saada kukin täyttämään monisivuinen kyselylomake. Pitkien tallenteiden nauhoittaminen voi kuitenkin olla haastavaa, siksi olisikin tärkeää että käytettävä menetelmä pystyisi luokittelemaan puhujan lyhyen näytteen perusteella.

Aloshban, Esposito ja Vinciarelli (2020) tarkastelivat yhden multimodaalisen sovelluksen tarkkuutta masennuksen tunnistamissa alle kymmenen sekunnin näytteestä. Menetelmä sisälsi sekä audio- että tekstipohjaisen luokittelun joita testattiin myös yksitellen. Tutkimuksessa luokittelun tarkkuus raportoitiin lausekekohtaisesti. Tulokset viitaisivat siihen että muutamien lausekkeiden perusteella tehdyn luokittelun tarkkuus on verrattavissa, tai jopa parempi, kuin koko näytteen luokittelun tarkkuus. Multimodaalinen lähestymistapa suoriutui unimodaalisia paremmin kun koko puhekorpus oli käytössä. Lauseketasolla audiopohjaisen ratkaisun tarkkuus oli kuitenkin verrattavissa multimodaaliseen ratkaisuun. Ainakin tämän tutkimuksen valossa jopa tämänhetkisillä lähestymistavoilla olisi potentiaalia toimia osana terveydenhuollon menetelmiä.

Tavoitettavuuteen liittyen on tärkeää kartoittaa mahdollisia menetelmiä äänitteiden keräämiseen joiden perusteella puhujat voitaisiin luokitella. Dineley ja muut (2021) tutkivat kuinka älypuhelimia voisi käyttää etämenetelmänä puhenäytteiden keräämisessä. Lisäksi he tilastoivat sosiodemograafista tietoa tutkimukseen kutsutuista henkilöistä ymmärtääkseen kuinka yksilöiden status vaikuttaa heidän halukkuuteensa luovuttaa puhenäytteitä ja vastata siihen liittyviin kysymyksiin.

Tutkimukseen kutsuttiin 384 masennuksesta kärsivää henkilöä, joista 54 % vastasi kutsuun. Tarkastelluista tekijöistä vain iällä oli merkittävä vaikutus kyselyyn osallistumisen todennäköisyydessä. Osallistujien tuli tallentaa puhettaan kahden viikon välein täyttäen

samassa yhteydessä *Inventory of Depressive Symptomatology – Self-Reported (IDS-SR)* -masennuskyselyn. Nämä suoritettiin vartavasten kehitetyllä puhelinsovelluksella. Äänitteitä tuli tehdä kaksi, joista toinen oli käsikirjoitettu teksti ja toisessa käyttäjä sai puhua vapaasti. Tutkimukseen osallistuneista 51 % suoriutui käsikirjoitetun tekstin tehtävästä ja 44 % vapaan puheen tehtävästä. Osallistujat raportoivat tuntevansa olonsa huomattavasti mukavammaksi käsikirjoitettua puhetta nauhoittaessa.

Kun näytteitä kerätään henkilöiden omaehtoisella äänittämisellä, käsikirjoitetun tekstin lukeminen voi tuottaa enemmän vastauksia koska luovan puheen tuottaminen voi olla tavallista hankalempaa masennuksesta kärsivälle yksilölle. Tätä tukee Dineleyn ja muiden (2021) havainto, että masennuksen vakavuus vaikutti merkittävästi osallistujan mukavuuteen äänitteitä tehdessä. Tämä kuvaa erinomaisesti masennuksen diagnosoinnin hankaluutta: ne, ketkä tarvitsevat eniten apua kärsivät kognitiivisista vaikeuksista jotka heijastelevat kyselyiden ja äänitteiden laatimiseen.

Kiss, Sztahó ja Tulics (2021) esittelivät oman esimerkkinsä valmiista sovelluksesta joka voisi toimia erilaisten puheeseen vaikuttavien sairauksien diagnosoinnin apuna. Sovellus analysoi käyttäjän antaman puhenäytteen tekstistä *"The North Wind and and the Sun"* ja pyrkii löytämään siitä viitteitä masennuksesta, käheydestä sekä Parkinsonin taudista. Sovellus luokittelee puhenäytteen yhteen neljästä luokasta; masentunut, Parkinsonin tauti, terve sekä käheä ääni. Lisäksi se kertoo todennäköisyyden millä puhenäyte kuuluu kuhunkin luokkaan. Tutkimuksen mukaan sovelluksen käyttämä luokittelumalli luokitteli oikein 73 % masentuneista näytteistä ja 93 % terveistä näytteistä. Kaikkien luokkien tarkkuus kokonaisuutena oli 81.1 %.

Sovelluksen tarkoitus ei ollut kuitenkaan vain luokitella henkilöitä, vaan myös arvioida mahdollisesti esiintyvien terveysongelmien vakavuusasteet. Masennuksen vakavuuden luokitteluun käytettiin *Beck Depression Inventory II* -asteikkoa, puheen käheyden luokitteluun *RBH*-asteikkoa ja Parkinsonin taudin luokitteluun *Hoehn & Yahr* -asteikkoa. Puhenäytteen prosessoinin jälkeen tulokset on esitelty havainnollistavina kuvioina käyttöliittymässä. Tämän kyseisen sovelluskokeilun mallit on muodostettu unkarin kielen puheen pohjalta, mutta Kiss et al. (2021) mukaan samat mallit ovat todettu toimivaksi myös englanninkielisille puhenäytteille.

Puheentunnistus on isossa roolissa osana nykyteknologiaa, ja tutkimuksia lukiessa pystyi huomaamaan että tietyt tutkijat ovat selvästikin todenneet että puheteknologioilla on tulevaisuudessa paikka myös masennuksen oireiden tunnistamisessa. Vaikuttaisi siltä, ettei kysymys ole enää pystyykö puheen analysointia hyödyntämään tällä osa-alueella, vaan *milloin* ja *kuinka* sitä pystyisi hyödyntämään. Tämä näkyy myös tutkimussuuntien

kehityksessä kohti käytännön ratkaisuja ja niihin liittymiä ongelmia jotka ovat toistaiseksi olleet puheteknologian käyttöönoton tiellä.

4. Yhteenveto

Tämän katsauksen perusteella voisi todeta että tämän tutkimuskohteen suosio on ollut viime vuosina nousussa. Lukuisat tutkimusryhmät ovat esittäneet omia ratkaisujaan, joka ei ole yllättävää ottaen huomioon masennuksen merkittävät yhteiskunnalliset vaikutukset. Tämänhetkinen tutkimustyö vaikuttaisi olevan hyvin hajanaista ja selkeästi pyrkii tutkimaan laajalti eri osa-alueita löytääkseen tehokkaimman tavan lähestyä ongelmaa.

Iso osa tutkimuksista, kuten muutama tässäkin katsauksessa esitellyistä tutkimuksista, kohdistivat tutkimuksensa tarkasti suunniteltujen neuroverkkomallien pohjalle. Niissä keskityttiin puheen audiosignaalin kokonaisvaltaiseen analysointiin. Ainakin Chlastan ja muiden (2019) sekä Srimadhurin ja Lalithan (2020) tekemissä tutkimuksissa tulokset olivat vaihtelevia. Vähemmän tutkittu osa-alue oli tiettyjen ominaisuuksien eristäminen audiosignaalista. Vlasenko ja muut (2017) tarkastelivat muutoksia vokaaliäänteissä, ja saavuttivatkin suhteellisen hyviä tuloksia. Tämän lisäksi myös Kiss ja muut (2021) hyödyntivät sovelluksessaan vokaalien analysointia, saavuttaen samaan tapaan suhteellisen korkean tarkkuuden näytteiden luokittelussa. Tämän työn puitteissa tehdyn tarkastelun perusteella vaikuttaisi siltä että puheen tiettyjen osien tarkastelu tuottaisia tarkempia tuloksia verrattuna koko signaalin analysoimiseen esimerkiksi konvoluutioneuroverkkoa hyödyntämällä. Mitään menetelmää ei voi kuitenkaan vielä yksittäisten tutkimusten perusteella linjata muita menetelmiä paremmaksi tai huonommaksi.

Tutkimustyö puheteknologioiden hyödyntämisessä masennuksen havaitsemisessa on kuitenkin tällä hetkellä vielä kehittyvässä vaiheessa. Osassa tutkimuksista pohditaan vielä teoreettisia kysymyksiä kun taas toiset tutkimusryhmät ratkaisevat jo käytännönläheisempiä kysymyksiä. Näiden tuloksien valossa jonkin asteinen käyttöönotto voisi olla jo mahdollista. Ei ole kovin vaikea ajatella että Kissin ja muiden (2021) esittelemän sovelluksen tapainen työkalu olisi käytössä esimerkiksi osana peruskoulun terveystarkastusta. Se on ei-tungetteleva ja maltillisia resursseja vaativa ratkaisu joka voisi tarjota arvokasta dataa puheteknologisten menetelmien tehokkuudesta. Täytyy toki ymmärtää että

muutokset terveydenhuollon menetelmissä tapahtuvat hitaasti, ja ilman tilastoihin perustuvaa näyttöä niiden käyttöönotto voi olla epätodennäköistä.

Kuten oletettu, tämä katsaus ei muuttanut sitä käsitystä että mahdolliset puheteknologiset menetelmät kuuluvat kuitenkin vain masennuksen varhaisten merkkien havaitsemiseen, sekä laajempimittaiseen väestön kartoitukseen. Lisääntynyt tarkkuus ja laajempi ulottuvuus vain helpottavat eniten terveydenhuollon palveuilta tarvitsevan väestönosan löytämistä. Vastuu yksilöllisessä arvioinnissa ja diagnoosin antamisessa on kuitenkin koulutetuilla terveydenhuollon ammattilaisilla.

Tässä tutkimuskatsauksessa esille tuodut tutkimukset saavuttivat vaihtelevia tuloksia, eivätkä antaneet suoraa ratkaisua ongelmaan. Ne kuitenkin esittivät mielenkiintoisia ja tieteellistä tutkimustyötä edistäviä menetelmiä. Heikkojen tulosten parantamiseksi sekä jo lupaavien tulosten todentamiseksi, tutkimustyötä on jatkettava. On kuitenkin mahdollista että puheteknologiset menetelmät ovat yksi tulevaisuuden tekijöistä masennuksen yhteiskunnallisten haittavaikutusten lieventämisessä.

Viitteet

Aloshban, N., Esposito, A. & Vinciarelli, A. (2020). Detecting depression in less than 10 seconds: Impact of speaking time on depression detection sensitivity. Teoksessa *Proceedings of the 2020 international conference on multimodal interaction* (s. 79–87). New York, NY, USA: Association for Computing Machinery. Lainattu saatavilla https://doi.org/10.1145/3382507.3418875

Boersna, P. & Weenink, D. (2022). *Praat: doing phonetics by computer [computer program].* http://www.praat.org/. (Luettu: 2022-02-17)

Chlasta, K., Wołk, K. & Krejtz, I. (2019). Automated speech-based screening of depression using deep convolutional neural networks. *Procedia Computer Science*, *164*, 618 - 628. Lainattu saatavilla http://www.sciencedirect.com/science/article/pii/S1877050919322756 (CENTERIS 2019 - International Conference on ENTERprise Information Systems / ProjMAN 2019 - International Conference on Project MANagement / HCist 2019 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2019) doi: https://doi.org/10.1016/j.procs.2019.12.228

Daic-woz database. (2016). https://dcapswoz.ict.usc.edu/. (Luettu: 2022-02-18)

Depressio: Käypä hoito -suositus. (2020). (Suomalaisen Lääkäriseuran Duodecimin ja Suomen Psykiatriyhdistys ry:n asettama työryhmä. Helsinki: Suomalainen Lääkäriseura Duodecim, 2020 (viitattu 30.11.2020). Saatavilla internetissä: www.kaypahoito.fi)

Huttunen, M. (2018). *Lääkärikirja duodecim.* https://www.terveysportti.fi/terveyskirjasto/tk.koti?p_teos=dlk. (Luettu: 2020-11-28)

Kiss, G., Sztahó, D. & Tulics, M. G. (2021). Application for Detecting Depression, Parkinson's Disease and Dysphonic Speech. Teoksessa *Proc. interspeech 2021* (s. 956–957).

Kroenke, K., Spitzer, R. L. & Williams, J. B. (2001). The phq-9: validity of a brief

depression severity measure. *PubMed*, *16*(9), 606-13. doi: https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Mental disorders. (2017). https://www.who.int/news-room/fact-sheets/detail/mental-disorders. (Luettu: 2020-11-27)

Quatieri, T. F. & Malyska, N. (2012). Vocal-source biomarkers for depression: a link to psychomotor activity. Teoksessa *Proc. interspeech 2012* (s. 1059–1062). Lainattu saatavilla https://www.isca-speech.org/archive/interspeech_2012/i12_1059.html

Srimadhur, N. & Lalitha, S. (2020). An end-to-end model for detection and assessment of depression levels using speech. *Procedia Computer Science*, *171*, 12-21. Lainattu saatavilla https://www.sciencedirect.com/science/article/pii/S1877050920309662 (Third International Conference on Computing and Network Communications (CoCoNet'19)) doi: 10.1016/j.procs.2020.04.003

Stasak, B., Epps, J. & Goecke, R. (2017). Elicitation design for acoustic depression classification: An investigation of articulation effort, linguistic complexity, and word affect. Teoksessa *Proc. interspeech 2017* (s. 834–838). Lainattu saatavilla http://dx.doi.org/10.21437/Interspeech.2017-1223 doi: 10.21437/Interspeech.2017-1223

Terveyden ja hyvinvoinnin laitos. (2021). *Masennuskysely.* https://www.mielenterveystalo.fi/aikuiset/itsearviointi/Pages/BDI.aspx. (Luettu: 2022-01-25)

Vlasenko, B., Sagha, H., Cummins, N. & Schuller, B. (2017). Implementing gender-dependent vowel-level analysis for boosting speech-based depression recognition. Teoksessa *Proc. interspeech 2017* (s. 3266–3270). Lainattu saatavilla http://dx.doi.org/10.21437/Interspeech.2017-887 doi: 10.21437/Interspeech.2017-887

What are formants? (2005). https://person2.sol.lu.se/SidneyWood/praate/whatform.html. (Luettu: 2022-02-19)