

Hädänalaisen puheen prosodia

Foneettisia tutkimuksia suomalaisista hätäpuheluista

LAURI TAVI

Väitöksenalkajaisitelmä Itä-Suomen yliopistossa 26. kesäkuuta 2020

Artikkelimuotoisessa väitöskirjassani olen tutkinut niitä akustis-foneettisia mittareita, joita voisi hyödyntää hädän- eli stressinalaisen soittajan tunnistamisessa hätäpuheluissa. Väitöstutkimustani voidaan pitää eksploratiivisena, sillä foneettisia tutkimuksia emotionaalisen stressin tunnistamisesta ei ole aiemmin toteutettu suomalaisista hätäpuheluista. Kuitenkin emotionaalinen puhe yleisesti on kiinnostanut eri alojen puheen tutkijoita niin ihmistieteen kuin teknologian aloilla jo vuosikymmenten ajan. Syy kiinnostukseen löytyy tunnetilojen ilmaisemisen suuresta merkityksestä ihmisten välisessä luonnollisessa vuorovaikutuksessa. Tunteiden tunnistamisella puheesta on kuitenkin myös merkittäviä sovellusmahdollisuuksia sekä viranomais- että kaupallisissa palveluissa. Aidon hädänalaisuuden tunnistaminen hätäpuheluissa onkin esimerkki tunteiden tunnistamisen (elin)tärkeystä yhteiskunnallisessa viestinnässä, sillä sen avulla viranomaisten rajallisia resursseja voidaan kohdentaa mahdollisimman tehokkaasti. Vaikka hätäpuheluissa soittajan hädänalaisuus pyritään aina selvittämään sanallisesti, väitöstutkimuksessani osoitan, että hädänalaisuus on myös mahdollista tunnistaa soittajan puheen prosodian avulla.

Puheen prosodia

Fonetiikka tutkii puheäänteiden lisäksi puheen prosodisia piirteitä. Prosodia käsittää puheen korkeuden, voimakkuuden, rytmin ja nopeuden vaihtelut. Akustis-foneettisesti tällaiset puheen piirteet mitataan perustaajuuden (f_0), äänienergiatasojen sekä kielellisten yksiköiden kestojen vaihteluina. Esimerkkinä niin sanotusta puutteellisesta prosodiasta voidaan mainita 1900-luvun puhesyntetisaattorit, joista kuuluisin lienee fyysikko

Stephen Hawkingin käyttämä syntetisaattori. Tällainen synteettinen puhe kuulosti monotoniselta ja robottimaiselta, sillä vielä viime vuosisadalla prosodisen vaihtelun mallintaminen oli hyvin puutteellista.

Yksi prosodian tutkimuksen ydinajatuksista on, että puheen prosodia informoi kuulijaa sen sanallista sisältöä enemmän. Puheen prosodisella vaihtelulla on vaikutus puheen sisältöön, mutta sillä voidaan ilmaista muun muassa puhujan persoonallisuutta ja tunnetiloja. Olennaista on, että nämä ilmaisut voivat tapahtua myös puhujan tahdosta riippumatta. Näin ollen prosodian tutkimuksella voidaan paljastaa puheäänestä puhujakohtaista tietoa, jota ovat esimerkiksi puhujan tunnetila, ikä, kansalaisuus tai jopa päihtymys.

Forensinen fonetiikka

Forensinen fonetiikka on puheen tutkimuksen ala, jossa foneettisia menetelmiä hyödynnetään puhetta sisältävissä rikos- tai muissa vastaavissa aineistoissa. Forensisen fonetiikan perusajatus on, että puhe mielletään biometriseksi eli ihmisen fysiologiaan perustuvaksi tunnisteksi. Kyseessä on vahvasti poikkitieteellinen ala, sillä siinä hyödynnetään muun muassa kielitiedettä, akustiikkaa, tilastotiedettä, koneoppimista, matematiikkaa, lääketiedettä ja psykologiaa. Nykyään tilastotiede ja koneoppiminen ovat nousseet tärkeiksi erityisesti puheanalyysien luotettavuuden arvioinnissa sekä automatisoinnissa. Psykologia tulee taas kyseeseen, kun arvioidaan nauhoitteessa esiintyvän puheen emotionaalista sisältöä (Hollien 2012).

Emotionaalinen puhe tarkoittaa esimerkiksi vihaista, rentoutunutta tai stressaantunutta puhetta. Erityisesti negatiivisia emootioita ilmentävää puhetta esiintyykin usein forensisissa puhemateriaaleissa, joihin kuuluvat myös hätäpuhelut. Puhujan tunnetilan tunnistaminen tällaisista materiaaleista voi tuoda lisävaloa puhujan motiiveihin tai esitetyn asiointilan todenmukaisuuteen. Merkittävää on myös, että emotionaalinen puhe heikentää puhujantunnistusjärjestelmien ja muiden puheteknologisten sovellusten tarkkuutta (Hansen & Patil 2007).

Väitöstutkimukseni käsittelee stressin foneettista tunnistamista hätäpuheluista, sillä stressin tai vastaavan tunteen tunnistaminen voi tuottaa tietoa ilmoitetun tilanteen vakavuudesta ja hyödyttää äänen akustiseen analyysiin perustuvia psykologisia arviointoja. Stressin foneettinen tunnistaminen perustuu siihen, että stressi aiheuttaa muun muassa lihasjäykkyyttä ja hengityksen nopeutumista, joilla on vaikutus puheen tuottoon (esim. Murray, Baber & South 1996). Stressi(ääne)n tarkka määrittely on kuitenkin osoittautunut haastavaksi, sillä stressin tai muidenkaan tunteiden kategorioille ei ole olemassa tarkkaa rajausta. Aikaisemmissa tutkimuksissa stressiksi onkin määritelty kognitiivinen eli laboratoriossa tuotettu, simuloitu eli näytelty tai emotionaalinen eli aidoissa tilanteissa yllättäen ilmenevä stressi. Tässä väitöskirjassa stressi on määritelty emotionaaliseksi stressiksi, joka esiintyy aidoissa tilanteissa ja sen on aiheuttanut suora henkeen tai terveyteen kohdistuva uhka.

Tämän väitöstutkimuksen aineistona olleet noin 300 hätäpuhelutallennetta kytkevät väitöskirjani forensiseen fonetiikkaan. Hätäpuhelutallenteet oli vastaanotettu Kuo-

pion hätäkeskuksessa vuonna 2016, mutta soittoja vastaanotettiin myös muista kaupungeista. Soittajista vain noin 10–20 % oli hädänalaisia soittajia, sillä suurin osa puheluista oli muun muassa viranomaisten välisiä soittoja sekä niin sanottuja taskupuheluita. Koska hätäpuhelutallenteet ovat varsin arkaluotoista aineistoa, niitä on saatavilla tutkimuskäyttöön varsin rajallisesti. Hätäkeskuslaitos myönsi tutkimusluvan kyseiseen aineistoon tarkoilla ehdoilla vuonna 2016.

Keskeisimmät tutkimustulokset

Artikkelimuotoinen väitöskirjani käsittää johdannon lisäksi neljä kansainvälistä vertaisarvioitua tutkimusta, joista kolme on julkaistu ja neljäs on julkaistavaksi hyväksytty käsikirjoitus. Tutkimuksista kaksi ensimmäistä (Tavi 2017, Tavi 2019) käsitelivät stressiäänän akustis-foneettisia korrelaatioita naispuhujilla. Tulokset osoittivat, että tietyt akustis-foneettiset piirteet, kuten korkea f_0 ja energian painottuminen korkeille taajuuksille, assosioituvat stressaantuneen puhujan ääneen. Vaikka tulokset tukivat vahvasti aiempia tutkimuksia, niissä ilmeni myös uusia havaintoja. Vokaalien akustisissa analyyseissä paljastui, että esimerkiksi *i*-vokaalin tuotto on siirtynyt stressaantuneilla puhujilla keskemälle suuta. Lisäksi kolmen perinteisen koneoppimisalgoritmin (l. päätöspuun, diskriminanttianalyysin ja logistisen regression) testaus osoitti, että stressiäänän ja neutraaliäänän luokittelemisessa akustis-foneettinen analyysi kannattaa keskittää tiettyyn vokaaliin sekalaisen vokaaliryhmän sijasta, jotta luokittelutarkkuus parantuu.

Kolmannessa tutkimuksessa (Tavi, Alumäe & Werner 2019) kehitettiin automaattinen narinaäänänlaadun tunnistin hätäpuheluille. Tunnistin perustui syväoppimismenetelmiin, jotka ovat koneoppimisen eräs kehittynyt alalaji. Narinatunnistimen kehittämisen motivaationa oli se, että narinan on huomattu assosioituvan tunnetiloihin, joista olennaisin tämän väitöstutkimuksen kannalta on stressin vastakohta eli rentoutuneisuus (Ishi, Sakakibara, Ishiguro & Hagita 2008). Toinen merkittävä syy narinan tunnistamiselle on sen negatiivinen vaikutus akustis-foneettiseen analyysiin: f_0 :n mittauksiin voi aiheutua virheitä, sillä narinaäänän äänihuulijaksot ovat usein epä-säännöllisiä. Kun narinaa tunnistettiin hätäpuheluista, tutkimuksessa kehitetty tunnistin osoittautui paremmaksi kuin viime vuosikymmenellä kehitetty narinatunnistin, jota on käytetty laajasti aiemmissa tutkimuksissa. Lisäksi mitatut narinamäärät paljastivat, että hätäpuheluissa narinan määrä on selvästi pienempi kuin Aalto-yliopiston keskustelukorpuksessa, johon ei liity stressiä.

Neljäs tutkimus (Tavi & Werner 2020) toteutettiin tapaustutkimuksena, jossa analysoitiin kahden itsemurha-aikeistaan ilmoittavan miessoittajan prosodista vaihtelua hätäpuhelun aikana. Tutkimuksessa tarkasteltiin erityisesti soittajan ja hätäkeskustyöntekijän välistä vuorovaikutusta akustis-foneettisilla mittauksilla. Mittaukset paljastivat, että hätäkeskustyöntekijöiden verbaalisella avuntarjouksella on vaikutusta niihin soittajien prosodisiin piirteisiin, jotka on aiemmissa tutkimuksissa liitetty itsetuhoisuuteen tai depressioniin (ks. Cummins, Scherer, Krajewski, Schnieder, Epps & Quatieri 2015). Lisäksi tutkimuksessa kehitettiin uusi prosodisen prominenssin mittari, SPI (sylla-

bic prosodic index), jonka toimivuus varmistettiin suurella ranskalaisella korpuksella, C-PROM:lla. SPI:tä voi käyttää tavun painokkuuden mittaamiseen missä tahansa puheaineistossa, joka sisältää tavusegmentoinnit. Tutkimuksessa soittajien SPI:n arvo suureni avuntarjoamisen jälkeen, mikä osoittaa itsetuhoisten soittajien puheen painokkuuden kasvamisesta avunsaannin jälkeen.

Yhteenveto

Emootioiden lisäksi niin hätäpuheluissa kuin muissakin aineistoissa puheen prosodiaan vaikuttavat useat tekijät, kuten vuorovaikutus ja puhujien monenlaiset fysiologiset tilat. Tästä syystä emotionaalisen puheen analyysissä olennaista on aineiston mahdollisimman tarkka määrittely. Myös yksittäisten tutkimustulosten yleistämisen erilaisiin aineistoihin tulee olla varovaista. Väitöstutkimukseni on kuitenkin osoittanut, että hädänalainen puheääni on mahdollista tunnistaa akustis-foneettisesti myös huonolaatuisista hätäpuheluista ja että prosodialla (erityisesti f₀:lla) on siinä keskeinen rooli. Väitöskirjassani olen myös halunnut tuoda esille, että foneettinen ja tietojenkäsittelytieteellinen näkökulma tarjoavat yhdessä tehokkaan tavan emotionaalisen puheen analysointiin.

Lähteet

- CUMMINS, NICHOLAS – SCHERER, STEFAN – KRAJEWSKI, JAREK – SCHNIEDER, SEBASTIAN – EPPS, JULIEN – QUATIERI, THOMAS F. 2015: A review of depression and suicide risk assessment using speech analysis. – *Speech Communication* 71 s. 10–49. <https://doi.org/10.1016/j.specom.2015.03.004>.
- HANSEN, JOHN H. – PATIL, SANJAY 2007: Speech under stress. Analysis, modeling and recognition. – Müller, Christian (toim.), *Speaker classification I* s. 108–137. Berlin: Springer. https://doi.org/10.1007/978-3-540-74200-5_6.
- HOLLIN, HARRY 2012: About forensic phonetics. – *Linguistica* 51 (1) s. 27–53. <https://doi.org/10.4312/linguistica.52.1.27-53>.
- MURRAY, IAIN R. – BABER, CHRIS – SOUTH, ALLAN 1996: Towards a definition and working model of stress and its effects on speech. – *Speech Communication* 20 s. 3–12. [https://doi.org/10.1016/S0167-6393\(96\)00040-4](https://doi.org/10.1016/S0167-6393(96)00040-4).
- ISHI, CARLOS TOSHINORI – SAKAKIBARA, KEN-ICHI – ISHIGURO, HIROSHI – HAGITA, NORIHIRO 2008: A method for automatic detection of vocal fry. – *IEEE Transactions on Audio, Speech, and Language Processing* 16 (1) s. 47–56. <https://doi.org/10.1109/TASL.2007.910791>.
- TAVI, LAURI 2017: Acoustic correlates of female speech under stress based on /i/-vowel measurements. – *International Journal of Speech, Language, and the Law* 24 s. 227–241. <https://doi.org/10.1558/ijsl.32506>.
- TAVI, LAURI 2019: Classifying females' stressed and neutral voices using acoustic-phonetic analysis of vowels: an exploratory investigation with emergency calls. – *International Journal of Speech Technology* 22 s. 511–520. <https://doi.org/10.1007/s10772-018-09574-6>.

- TAVI, LAURI – ALUMÄE, TANEL – WERNER, STEFAN 2019: Recognition of creaky voice from emergency calls. – Proceedings *INTERSPEECH 2019*, Graz, Austria, s. 1990–1994. <https://doi.org/10.21437/Interspeech.2019-1253>.
- TAVI, LAURI – WERNER, STEFAN 2020: A phonetic case study on prosodic variability in suicidal emergency calls. – *International Journal of Speech, Language, and the Law* 27 s. 59–74. <https://doi.org/10.1558/ijssl.39667>.

Lauri Tavi: *Prosodic cues of speech under stress: Phonetic exploration of Finnish emergency calls*. Dissertations in Education, Humanities, and Theology; 154. Joensuu: Itä-Suomen yliopisto, 2020. Väitöskirja on luettavissa osoitteessa <http://urn.fi/URN:ISBN:978-952-61-3403-1>.

Kirjoittajan yhteystiedot:
etunimi.sukunimi@uef.fi