Automaattinen puheentunnistus ja puheen fysikaaliset perusteet

Kandidaatintutkielma, 28.8.2023

Tekijä:

Nicolas Richterich

Ohjaaja:

Tuomas Grahn

JYVÄSKYLÄN YLIOPISTO
FYSIIKAN LAITOS

Tiivistelmä

Richterich, Nicolas
Automaattinen puheentunnistus ja puheen fysikaaliset perusteet
Kandidaatintutkielma
Fysiikan laitos, Jyväskylän yliopisto, 2023, 85 sivua

Tässä tutkielmassa käsitellään puheen tuottamisen fysiikkaa ja automaattista puheentunnistamista. Tutkielma jakautuu neljään osaan, joista ensimmäisessä tutustutaan puheen fyysisiin perusteisiin Lagrangelaisen mekaniikan ja termodynamiikan keinoin sekä käydään läpi pitkittäisten aaltojen seuraksena ilmeneviä akustisia ominaisuuksia fluideissa. Toisessa luvussa esitellään analogisen aaltosignaalin digitoinnin periaatteet ja tutustutaan eri muunnoksiin signaalinkäsittelyn työkaluina, jonka jälkeen syvennytään signaalin lyhytaikakäsittelyyn ja sen sovelluksiin puheenkäsittelyssä. Kolmannessa luvussa puheentuotantoa käsitellään foneettisesti lähde-suodin-mallin avulla sekä käydään läpi yksittäisten foneemien laskennallisia tunnistamismetodeja. Tutkielman neljännessä luvussa perehdytään ensin neuroverkkojen ja koneoppimisen perusteisiin, jonka jälkeen käydään läpi niille perustuvia nykyaikaisessa puheentunnistuksessa käytetyttyjä kielimalleja ja akustisia malleja. Lopuksi esitellään suomenkielinen puheentunnistumalli.

Tutkielman tavoite on tutustuttaa lukija puheen aaltomekaaniseen luonteeseen sekä luoda yleiskatsaus puheentuotannon ja -tunnistamisen mekanismeihin.

Avainsanat: Puheteknologia, puheentunnistus, akustiikka, signaalinkäsittely, fonetiikka, puheakustiikka

Abstract

Richterich, Nicolas
Automatic Speech Processing and the Physical Basis of Speech
Bachelor's thesis
Department of Physics, University of Jyväskylä, 2023, 85 pages.

This thesis is concerned with the physics of speech production and the mechanisms of automatic speech recognition. The thesis is split into four parts, of which the first investigates the physical basis of speech through Lagrangian mechanics and thermodynamics and the acoustic phenomena caused by longitudinal waves in fluids. The second chapter presents the principles of analog signal discretization and examines different transformation methods as tools for signal processing, after which short-term signal processing and its applications in speech processing are delved into. In the third chapter speech production is explained with the source-filter-model of phonation and different ways of phoneme recognition are discussed. Neural networks and machine learning are presented in the fourth chapter, after which language models and acoustical models based on them are investigated. Finally, a Finnish language speech recognition model is presented.

The goal of the thesis is to familiarize the reader with the wave mechanical nature of speech and to give an overview in the mechanisms of speech production and recognition.

Keywords: Speech processing, speech recognition, acoustics, signal processing, phonetics, speech acoustics

Sisällys

1 Johdanto

Puhetta on esitetty eri kirjoitusjärjestelmillä noin 5500 vuoden ajan mesopotamialaisesta nuolipääkirjoituksesta lähtien ja länsimaisen kielentutkimuksen juuret ulottuvatkin noin 2000 vuoden taa välimerelliseen antiikkiin. Kreikkalaiset ja roomalaiset grammaatikot sekä reetorit sivusivat äänteiden fonetiikkaa tutkimuksissaan ja lukuisat koulukunnat pytagoralaisista stooalaisiin tutkivat erikokoisten äänilähteiden suhteita. Arkkitehti Vitruvius tunnisti äänen paineen palloaaltomaiseksi vaihteluksi ilmassa [1] ja antiikin arkkitehtuuriset teokset temppeleistä teattereihin osoittavatkin edelleen erinomaista ymmärrystä ääniaaltojen luonteesta ja niiden etenemisestä. Keskiajalla arkkitehtuurista akustiikkaa kehitettiin kirkoissa ja hoveissa perustuen lähinnä kirkollisiin ohjenuoriin pyhän musiikin ja palvomisen rajoissa, kunnes renessanssin aikana musiikillinen ja arkkitehtuurinen kokeilu alkoi vapautua Etelä- ja Länsi-Euroopassa. Tieteellisen vallankumouksen aikana monet antiikinaikaiset näkemykset äänesta osoitettiin matemaattisesti oikeiksi Mersennen (1636) [2] toimesta ja Newton esitti äänennopeudelle matemaattisen ilmaisun väliaineen ominaisuuksien perusteella (1687) [3]. Aaltoteoriat saivat matemaattisen pohjan Eulerin, Lagrangen ja d'Alembertin töistä 1700-luvulla ja nykyaikainen akustinen teoria kehitettiin 1800-luvulla Helmholtzin, lordi Rayleighin ja monien muiden voimin. Samaan aikaan fonetiikalle muodostui tieteenä teollista kysyntää äänentallennuslaitteiden ja puhelimen keksimisen myötä ja fonetiikantutkimus hyödynsikin monia teollisen vallankumouksen hyödykkeitä yhdistämällä uuden akustisen teorian kehittyvään lääketieteeseen. Telekommunikaatiot loivat tarpeen puheen matemaattiselle tulkinnalle jota Bellin laboratorioiden, Shannonin, Fantin ja muiden tutkimus on kehittänyt nykyiselle tasolleen 1900-luvulla. [4]

Keinotekoisten neuroverkkojen kehitys alkoi 1950-luvulla yrityksenä mallintaa ihmisen hermojärjestelmää matemaattisesti. Tietokoneiden ja koneoppimisen kehittyessä oltiin 2000-luvulle saapuessa pystytty hyödyntämään neuroverkkoja monissa tehtävissä, kuten kuvaluokittelussa ja sanaluokittelussa. Nykyaikaiset puheentunnistusmallit ovat neuroverkkopohjaisia järjestelmiä, jotka koostuvat kuvaluokitteluun perustuvaan akustiseen malliin ja sanaluokitteluun perustuvaan kielimalliin. Nykyai-

kaiset akustiset mallit ja kielimallit pohjautuvat huomiomekanismiin (2016) [5], joka tarkastelee foneemeja ja sanoja niiden koko kontekstissa rajallisen analyysin sijaan.

Tutkielmassa tutustutaan ääniaaltoihin ensin yksiulotteisesti Lagrangen mekaniikan keinoin ja sitten kolmiulotteisesti pallovärähtelijän ja eulerilaisen virtausmekaniikan avulla. Fysikaalisen osuuden jälkeen käydään läpi ääniaaltojen digitalisointi ja muuntaminen spektrogrammeiksi, joista pystytään tunnistamaan puheominaisuuksia. Puheominaisuuksien syntymekanismeihin ja puheominaisuuksien tunnistamismetodeihin tutustutaan foneettisesti lähde-suodin-mallin avulla, jonka jälkeen siirrytään käsittelemään neuroverkkoja, koneoppimista ja niiden avulla tuotettuja kielimalleja ja akustisia malleja.

2 Akustiikka

Tässä luvussa johdetaan puheen akustiset perusteet energiasta lähtien. Alaluvussa 2.1 johdetaan yksiulotteiselle pitkittäiselle värähtelylle d'Alembertin aaltoyhtälö mukaillen lähteitä [6] ja [7], jonka jälkeen alaluvussa 2.2 käydään läpi muutamia etenevien aaltojen ominaisuuksia lähteen [6] mukaisesti. Alaluvun 2.3 pääasiallisena lähteenä toimii [8] ja siinä johdetaan energialähtöisesti kaasun ominaisuuksia väliaineena. Alaluvussa 2.4 käsitellään kolmiulotteisen kaasun dynamiikkaa johtamalla virtausmekaaninen liikemääräyhtälö lähdettä [9] mukaillen ja pallomainen aaltofunktio lähteiden [7][10] mukaisesti. Viimeisessä alaluvussa 2.5 käydään läpi ääniaaltojen akustisia ominaisuuksia pääasiassa lähteiden [6], [9], [11] ja [12] avulla.

2.1 Ääniaallot

Värähtely on pitkittäisesti tai poikittaisesti tapahtuvaa liikettä väliaineessa ajan funktiona. Värähtely syntyy herätteestä, vahvistuu resonanssista ja vaimenee häviön johdosta. Energian säilymislain mukaisesti värähtelyä ei häviä tai synny tyhjästä, vaan värähtelyliikkeen energia saapuu ja poistuu jonakin toisena energian muotona. Värähtelyliikettä voidaan mallintaa liike- ja potentiaalienergian vuorotteluna. Lagrangen formalismissa mekaniikkaa käsitellään energiakeskeisesti siten, että käsitelty järjestelmä minimoi vaikutuksen [6]

$$S = \int \mathcal{L} \mathrm{d}t \tag{1}$$

eli se noudattaa pienimmän vaikutuksen periaatetta. Lagrangen yhtälön $\mathcal{L}$ termit ovat liike-energia T ja potentiaalienergia U ja yhtälö on muotoa

$$\mathcal{L} = T - U. \tag{2}$$

Vaikutus S minimoituu kun $\mathcal{L}$ toteuttaa Euler-Lagrangen yhtälön

$$\frac{\partial \mathcal{L}}{\partial q_j} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}_j} = 0, \tag{3}$$

jossa alaindeksi $j = 1,2,3,...,n$ on järjestelmän vapausasteiden määrä. Vapausasteita on $n = s - m$ jossa s kertoo kuinka monta yhtälöä järjestelmää kuvaa ja m on järjestelmän rajoitteiden määrä. Toisen asteen yhtälön tapauksessa tarvitaan $2s$ alkuehtoa kuvaamaan järjestelmää singulaarisesti. Euler-Lagrangen yhtälöstä voidaan johtaa järjestelmän liikeyhtälöt.

Väliaineita joissa aaltoliikettä tapahtuu voidaan kuvata yksinkertaistetusti jousimassa-järjestelminä, jotka koostuvat ideaalisista massoista m joiden keskinäistä vuorovaikutusta mallinnetaan jousivoiman κ omaavina jousina. Oletetaan värähtelijäjärjestelmän olevan konservatiivinen ja sen omaavan yleiset koordinaatit q_j ja yleiset nopeudet $\dot{q}_j$, jotka kuvaavat järjestelmän tilaa jokaisella ajan hetkellä t. Liitteessä A esitellään yleisistä nopeuksista riippuva yhtälö liike-energialle ja johdetaan yleisistä koordinaateista riippuva yhtälö potentiaalienergialle. Energiat voidaan sijoittaa Lagrangen yhtälöön (2)

$$\mathcal{L} = \frac{1}{2}\sum_{j,k}^{n} m_{jk}\dot{q}_j\dot{q}_k - \frac{1}{2}\sum_{j,k}^{n} C_{jk}q_jq_k, \tag{4}$$

jolloin Euler-Lagrangen yhtälö (3) saa muodon

$$\frac{\partial U}{\partial q_i} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i} = 0,$$

jossa

$$\begin{cases} \frac{\partial U}{\partial q_j} = \sum_{j,k}^{n} C_{jk}q_j \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{j,k}^{n} m_{jk}\dot{q}_j = \sum_{j,k}^{n} m_{jk}\ddot{q}_j \end{cases}.$$

eli

$$\sum_{j}^{n}(C_{jk}q_j + m_{jk}\ddot{q}_j) = 0. \tag{5}$$

On nähtävissä että kukin summa-alkio j koostuu massa- ja jousiosasta, jolloin voimme todeta potentiaalienergian vakion C vastaavan jotakin muutosta vastustavaa tekijää κ. Kun kulmataajuus on [6]

$$\omega_0^2 \equiv 2\pi f_0 \equiv \frac{\kappa}{m} \tag{6}$$

saamme kullekin summa-alkiolle harmonisen värähtelijän yhtälön

$$\sum_{j}^{n}(\ddot{q}_j + \omega_0^2 q_j) = 0. \tag{7}$$

Tämän toisen asteen differentiaaliyhtälön ratkaisu on tunnetusti [6] Eulerin identiteetti joka voidaan esittää siniaaltomuodossa

$$q_j(t) = A_j e^{i(\omega_0 t - \delta)} = A_j(\cos(\omega_0 t - \delta) + i\sin(\omega_0 t - \delta)). \tag{8}$$

Sijoittamalla ratkaisu aiempaan summa-alkioiden yhtälöön saadaan

$$\sum_j^n (\kappa_{jk} - m_{jk}\omega_0^2) A_j e^{i(\omega_0 t - \delta)} = 0, \tag{9}$$

josta voidaan ottaa yhteinen sinimuotoinen tekijä pois

$$\sum_j^n (\kappa_{jk} - m_{jk}\omega_0^2) A_j = 0. \tag{10}$$

Nyt olemme saaneet liikeyhtälöt muotoon jossa ne ovat n-suuruinen joukko lineaarisia homogeenisia algebrallisia yhtälöitä, jotka kertoimien A_j tulee toteuttaa. Jotta yhtälölle on olemassa epätriviaali ratkaisu, tulee sulkujen sisällä olevan termin determinantin hävitä

$$\det(\kappa_{jk} - m_{jk}\omega_0^2) = 0. \tag{11}$$

Kyseessä on siis ominaisarvo-ongelma, jonka ratkaisuna olevat ominaisarvot ovat järjestelmän ominaistaajuuksien neliöt ω_0^2. Ominaiskulmataajuudet vastaavat järjestelmän resonanssikulmataajuuksia, jolloin identiteetin (6) avulla järjestelmän resonanssitaajuus on

$$f_0 = \frac{v}{\lambda} = \frac{1}{2\pi}\sqrt{\frac{\kappa_{jk}}{m_{jk}}} \tag{12}$$

jolla on periodi

$$T_0 = 2\pi\sqrt{\frac{m_{jk}}{\kappa_{jk}}}. \tag{13}$$

Tällöin kun harmonista värähtelijää poikkeutetaan tasapainopisteestään ja järjestelmän annetaan värähdellä vapaasti, värähtely asettuu omalle ominaistaajuudelleen joka määräytyy jousivakion ja massan suhteesta. Yhtälöstä saatavat ominaisvektorit, niiden ortonormaalius osoitetaan liitteessä B.

Pitkittäistä värähtelyä varten tarkastelkaamme värähtelyjärjestelmää yksiulotteisena hilana jossa on n-määrä massoja m_j jotka sijaitsevat yleisissä koordinaateissa q_j joiden etäisyys tasapainopisteistä q_{j0} määräytyy häiriön $\psi_j(q_j,t)$ mukaan. Kullakin jousella on jousivakio κ_j ja pituus $l_j(q_j,t)$, jolloin jonkin massan m_j pieni siirtymä

aiheuttaa muutoksen $\Delta l_j = \psi_j - \psi_{j-1}$ pituudessa. Kun yksikkökoordinaatti on Δq_j, voidaan todeta massatiheyden olevan $M_j = m_j/\Delta q_j$, yksikkökoordinaatin jousivakion olevan $K_j = \kappa_j \Delta q_j^2$ ja pituuden muutostiheyden olevan häiriön paikkaderivaatta $\Delta l_j/\Delta q_j = \partial\psi_j/\partial q_j = \psi'$. Määritelkäämme myös häiriön aikaderivaatta $\dot{\psi} = \partial\psi/\partial t$. Värähtelijäjärjestelmän Lagrangen yhtälö voitaisiin siis kirjoittaa

$$\mathcal{L}(\psi_j, \frac{\partial\psi_j}{\partial t}) = \frac{1}{2}\sum_j^n m_j\dot{\psi}_j^2 - \frac{1}{2}\sum_j^n \kappa_j \Delta l_j^2, \tag{14}$$

mutta koska nyt käsittelemme jatkuvana approksimoitavaa hilaa ja tekijää jolla on kaksi itsenäistä muuttujaa, jaamme Lagrangen yhtälön yksikkökoordinaatilla ja saamme Lagrangen tiheyden [7]

$$\frac{\mathcal{L}}{\Delta q_j} = \mathcal{L}\Big(\psi_j, \frac{\partial\psi_j}{\partial q_j}, \frac{\partial\psi_j}{\partial t}\Big) = \frac{1}{2}\sum_j^n M_j\dot{\psi}_j^2 - \frac{1}{2}\sum_j^n K_j(\psi_j')^2 \tag{15}$$

jolloin Euler-Lagrangen yhtälössä Lagrangen tiheyttä tulee derivoida sekä ajan että paikan suhteen

$$\frac{\partial\mathcal{L}}{\partial\psi_j} - \frac{\partial}{\partial\psi_j}\frac{\partial\mathcal{L}}{\partial\psi_j'} - \frac{\partial}{\partial t}\frac{\partial\mathcal{L}}{\partial\dot{\psi}_j} = 0. \tag{16}$$

Voimme kirjoittaa tämän myös muodossa

$$M_j\ddot{\psi}_j - K_j\psi_j'' = 0 \tag{17}$$

josta saamme häiriön joukkonopeuden v_j käyttämällä identiteettiä [7]

$$\frac{K_j}{M_j} \equiv \frac{\omega_j^2}{k_j^2} = v_j^2. \tag{18}$$

Nyt yhtälö voidaan kirjoittaa yksiulotteisena d'Alembertin aaltoyhtälönä häiriölle ψ_j

$$\frac{\partial^2\psi_j}{\partial q_j^2} = \frac{1}{v_j^2}\frac{\partial^2\psi_j}{\partial t^2}. \tag{19}$$

2.2 Ääniaaltojen eteneminen

Yksiulotteisen aaltoyhtälön ratkaisut ovat muotoa

$$\psi = Ae^{-ik(x\pm vt)} = Ae^{i(\omega t\pm kx)}, \tag{20}$$

joiden ylemmät merkit kuvaavat vasemmalle etenevää aaltoliikettä ja alemmat merkit oikealle etenevää aaltoliikettä.

Jos asetamme kaksi erisuuntaisesti etenevää aaltofunktiota superpositioon

$$\psi = \psi_1 + \psi_2 = Ae^{-ik(x+vt)} + Ae^{-ik(x-vt)}$$

$$= Ae^{-ikx}(e^{i\omega t} + e^{-i\omega t}) = 2Ae^{-ikx}\cos(\omega t) \tag{21}$$

niin superpositioaaltofunktion reaaliosa on

$$\psi = \mathrm{Re}(2Ae^{-ikx}\cos(\omega t)) = 2A\cos(kx)\cos(\omega t). \tag{22}$$

Huomataan että kahden vastakkaisesti etenevän aaltofunktion superpositiona syntynyt aaltofunktio on staattinen ja kosinitermien johdosta sillä on solmukohtia joissa aaltofunktiot kumoavat toisensa. Ilmiön mukaisia paikallaan värähteleviä aaltoja kutsutaan seisoviksi aalloiksi.

Esittämällä oikealle etenevän aaltofunktion kulma-aaltonumeron kompleksisena $k = \alpha - i\beta$, aaltofunktio saa muodon

$$\psi = Ae^{i(\omega t-\alpha x)-\beta x} = Ae^{-\beta x}e^{i(\omega t-\alpha x)} \tag{23}$$

jossa $e^{-\beta x}$ on etäisyyden mukana aaltoliikettä vaimentava tekijä.

Jos puolestaan kulmataajuus merkitään kompleksiseksi $\omega = \alpha + i\beta$ ja kulma-aaltonumeron reaaliseksi saadaan aaltofunktion muotoon

$$\psi = Ae^{i(\alpha t+i\beta t-kx)} = Ae^{-\beta t}e^{i(\alpha t-kx)} \tag{24}$$

jossa $e^{-\beta t}$ on ajan mukana aaltoliikettä vaimentava tekijä. Tällöin siis kompleksinen kulma-aaltonumero on olennainen eteneviä aaltoja käsitellessä ja kompleksinen kulmataajuus seisovia aaltoja käsitellessä. [6]

Aaltofunktion eksponentin argumenttia $\phi = \omega t \pm kx$ kutsutaan funktion vaiheeksi.

Liikkumalla etenevän aaltoliikkeen mukana siten että aaltomuoto näyttää pysähtyneeltä, etenemme aallon vaihenopeudella v_ϕ. Vaihenopeus on vakio kun vaihe ei muutu

$$\mathrm{d}\phi = 0 \tag{25}$$

jolloin vaiheen muutosten tulee olla yhtäsuuria

$$\omega \mathrm{d}t = k\mathrm{d}x \tag{26}$$

jolloin vaihenopeus on

$$v_\phi = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\omega}{k}. \tag{27}$$

Jos asetamme kaksi samaan suuntaan etenevää mutta hieman eroavilla kulmataajuuksilla ja kulma-aaltonumeroilla varustettua aaltofunktiota superpositioon saamme yhtälön

$$\psi = \psi_1 + \psi_2 = Ae^{i(\omega t - kx)} + Ae^{i((\omega+\Delta\omega)t-(k+\Delta k)x)}$$
$$= A\Big(e^{i(\omega+\Delta\omega/2)t}e^{-i(k+\Delta k/2)x}\Big)\Big(e^{i((\Delta\omega t-\Delta kx)/2)}e^{-i((\Delta\omega t-\Delta kx/2))}\Big) \tag{28}$$

jonka reaaliosa on

$$\psi = 2A\cos\Big(\frac{\Delta\omega t - \Delta kx}{2}\Big)\cos\Big(\Big(\omega + \frac{\Delta\omega}{2}\Big)t - \Big(k + \frac{\Delta k}{2}\Big)x\Big). \tag{29}$$

Tämä aaltofunktio kuvaa huojuvaa aaltoliikettä jossa jälkimmäinen kosinitermi on kantaja-aalto ja sitä edeltävä kosinitermi yhdessä amplituditermin kanssa on modulointiaalto. Kantaja-aalto määrittää superpositioaallon vaiheen ja aaltomuodon, kun taas modulointiaalto määrittää superpositioaallon huojunnan taajuuden ja amplitudin. Korkeataajuista kantaja-aaltoa voidaan siis amplitudimoduloida (AM) tai taajuusmoduloida (FM) matalataajuisella modulointiaallolla. Modulointiaallon etenemisnopeutta kutsutaan joukkonopeudeksi

$$v = \frac{\Delta\omega}{\Delta t} \tag{30}$$

joka on epädispersiivisissä aineissa yhtäsuuri kuin vaihenopeus eli $v = v_\phi$. [6]

Aaltofunktioiden superpositioperiaate voidaan yleistää n-määrälle aaltofunktioita superpositioaaltofunktiona

$$\psi(x,t) = \sum_{r=1}^{n} Ae^{i(\omega_r t - k_r x)} \tag{31}$$

joka voidaan kirjoittaa integraalina kun n lähestyy ääretöntä

$$\psi(x,t) = \int_{-\infty}^{\infty} A(k)e^{i(\omega t - kx)}, \tag{32}$$

jossa $A(k)$ on eri taajuuksien amplitudien jakauma eli spektraalijakauma. Kun spektraalijakaumalla on merkittäviä arvoja vain tietyn kulma-aaltoluvun k_0 ympäristössä, voidaan aaltofunktio kirjoittaa aaltopakettina

$$\psi(x,t) = \int_{k_0-\Delta k}^{k_0+\Delta k} A(k)e^{i(\omega t - kx)}, \tag{33}$$

joka sisältää aaltofunktioita vain pieneltä taajuuskaistalta. Katkeamattomalla etenevällä aallolla ei voi kantaa informaatiota pisteestä toiseen, mutta moduloimalla siitä aaltopaketteja voidaan aallosta muodostaa signaali joka etenee joukkonopeudella. [6]

Tietyn aaltopakettijonon todennäköisyys signaalissa on

$$p = \left(\frac{1}{x}\right)^{I} = x^{-I}, \tag{34}$$

jossa x on aaltopaketin mahdollisten tilojen määrä ja I on aaltopakettien lukumäärä eli signaalin informaatiosisältö. Informaatiosisältö voidaan siis kirjoittaa

$$I = -\log_x(p) \tag{35}$$

ja informaatiosisällön odotusarvo on signaalin entropia [13]

$$\langle S \rangle = \sum_i p_i I = -\sum_i p_i \log_x(p_i). \tag{36}$$

2.3 Ilman termodynaamisia ominaisuuksia

Puheessa ääniaaltojen pääasiallinen väliaine on ilma, joka on noin 78% typpeä N_2 ja 21% happea O_2 [14]. Jos tarkastelemme ilmaa diatomisena ideaalikaasuna suljetussa laatikossa jossa on N-määrä energian E omaavia molekyylejä tilavuudessa V, voimme käsitellä sitä statistisen mekaniikan keinoin kanonisena joukkona. Energian tapoja jakautua molekyylien kesken kuvaa mikrotilojen lukumäärä $\Omega(E,V,N)$, joka maksimoituu tasapainotilassa. Mikrotiloilla voidaan liitteen C mukaisesti johtaa kylmyystekijäksi kutsuttu suure [15]

$$\beta = \frac{\partial \ln \Omega}{\partial E}. \tag{37}$$

Kylmyystekijä kertoo siis tilojen jakautuvan logaritmisesti energiatasoille. Järjestelmän tiloihin liittyy myös järjestelmän entropia S, joka kuvaa tietyn tilan odotusarvoa. Koska entropian muutos energian suhteen määrittää järjestelmän lämpötilan [8]

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T}, \tag{38}$$

voidaan todeta samoista muuttujista riippuvan kylmyystekijän liittyvän myös lämpötilaan. Kylmyystekijän termodynaaminen tulkinta voidaankin esittää käänteisen lämpötilan T sekä kaasuvakion R ja Avogadron vakion N_A tai Boltzmannin vakion k_B avulla [8]

$$\beta = \frac{\partial \ln \Omega}{\partial E} \equiv \frac{N_A}{RT} = \frac{1}{k_B T}. \tag{39}$$

Logaritmin derivoinnin avulla saamme mikrotilojen lukumäärälle esityksen

$$\frac{\partial \ln \Omega}{\partial E} = \frac{\partial \Omega}{\partial E}\frac{1}{\Omega} = \beta \implies \int_0^\Omega \frac{1}{\Omega}\partial\Omega = \beta \int_0^E \partial E \iff \ln \Omega = \beta E \implies \Omega = e^{\beta E}, \tag{40}$$

jolloin voimme todeta tietyn mikrotilan todennäköisyysjakauman olevan eksponentiaalinen

$$p = \frac{1}{\Omega} = e^{-\beta E} = \left(\frac{1}{e}\right)^{\beta E}, \tag{41}$$

ja kylmyystekijällä skaalatun energian olevan järjestelmän informaatiosisältö

$$\beta E = \frac{E}{k_B T} = \frac{E}{k_B}\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{S}{k_B} = -\ln(p) \tag{42}$$

jonka odotusarvo on yksikötön informaatioentropia

$$\left\langle \frac{S}{k_B} \right\rangle = -\sum_i p_i \ln(p_i) \tag{43}$$

joka saa termodynaamisen tulkinnan Gibbsin entropiana [8]

$$\langle S \rangle = -k_B \sum_i p_i \ln(p_i). \tag{44}$$

Termodynamiikan toisen pääsäännön mukaan järjestelmä saavuttaa maksimientropian tasapainotilassa, eli saamme järjestelmän tasapainotilan kun löydämme entropian maksimoivan todennäköisyysjakauman p. Liitteessä D on johdettu entropian maksimoivaksi jakaumaksi Boltzmannin jakauma

$$p_i = \frac{1}{\sum_i e^{-\beta E_i}} e^{\beta E_i} = \frac{1}{Z} e^{\beta E_i}, \tag{45}$$

jossa Z on jakauman partitiofunktio. Boltzmannin jakauman avulla voimme määrittää järjestelmän Helmholtzin vapaan energian, joka kertoo energiamäärän jolla järjestelmä voi isotermisesti tehdä työtä [8]

$$F = U + TS = \sum_i p_i E_i - T k_B \sum_i p_i \ln(p_i) = -\frac{1}{\beta} \ln(Z). \tag{46}$$

Helmholtzin vapaan energian muutos tilavuuden suhteen määrittää työtä tekevän suureen eli järjestelmän paineen

$$P = \left(\frac{\partial F}{\partial V} \right)_{T,N} \tag{47}$$

jonka avulla saamme yleisen puristuskertoimen [8]

$$K = -V \frac{\partial P}{\partial V}. \tag{48}$$

Rajoitteiden avulla saamme myös järjestelmän isotermisen kokoonpuristuvuuden

$$\kappa_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T \tag{49}$$

ja isentrooppisen kokoonpuristuvuuden

$$\kappa_S = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_S. \tag{50}$$

Kokoonpuristuvuuksien suhde määrittää järjestelmän adiabaattisen vakion

$$\gamma = \frac{\kappa_T}{\kappa_S} \tag{51}$$

jonka avulla voidaan määrittää järjestelmän isentrooppinen puristuskerroin

$$K_S = \gamma P. \tag{52}$$

2.4 Virtaukset ja aallot kolmiulotteisessa avaruudessa

Kun käsiteltävä järjestelmä on suuri (hiukkasmäärä $N >> V$), on mielekkäämpää seurata yksittäisten hiukkasten liikeen sijaan virtauselementtejä $u(\mathbf{r},t)$, $v(\mathbf{r},t)$ ja $w(\mathbf{r},t)$ jotka kuvaavat hiukkasten keskiarvoisia nopeuksia infinitesimaalisissa tilavuuksissa. Järjestelmän nopeuskenttävektori on tällöin

$$\mathbf{V} = u\hat{\imath} + v\hat{\jmath} + w\hat{k} \tag{53}$$

ja kokonaiskiihtyvyys

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V}\cdot\nabla)\mathbf{V} = \frac{\partial \mathbf{V}}{\partial t} + \left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}\right)\mathbf{V}, \tag{54}$$

jossa ensimmäinen termi on paikallinen kiihtyvyys ja toinen termi on konvektiivinen kiihtyvyys. [9]

Kertomalla nopeuskenttävektoria $\mathbf{V}$ pinta-alan A normaalilla $\mathbf{n}$ ja integroimalla pinnan S yli saadaan pinnan läpi virtaavan aineen tilavuusvuo

$$Q = \int_S (\mathbf{V}\cdot\mathbf{n})\mathrm{d}A, \tag{55}$$

jota kertomalla aineen massatiheydellä saadaan pinnan läpi virtaava massavuo

$$\dot{m} = \int_S \rho(\mathbf{V}\cdot\mathbf{n})\mathrm{d}A. \tag{56}$$

Voimme kirjoittaa tilavuuden läpi virtaavan aineen minkä vain ominaisuuden B määrän kontrollitilavuudessa CV

$$B_{\mathrm{CV}} = \int_{\mathrm{CV}} \left(\frac{\mathrm{d}B}{\mathrm{d}m}\right) \mathrm{d}m = \int_{\mathrm{CV}} \left(\frac{\mathrm{d}B}{\mathrm{d}m}\right) \rho \ \mathrm{d}\mathbb{V}, \tag{57}$$

jossa $\mathrm{d}\mathbb{V} = (V \cdot \mathbf{n})\ \mathrm{d}A\ \mathrm{d}t$. Suureen B muutos ajassa on koko järjestelmälle Reynoldsin jatkuvuuslauseeksi kutsuttu yhtälö [9]

$$\frac{\mathrm{d}}{\mathrm{d}t} B_{\mathrm{syst}} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\mathrm{CV}} \left(\frac{\mathrm{d}B}{\mathrm{d}m}\right) \rho \mathrm{d}\mathbb{V} \right) + \int_{\mathrm{CS}} \left(\frac{\mathrm{d}B}{\mathrm{d}m}\right) \rho (V \cdot \mathbf{n}) \mathrm{d}A \tag{58}$$

jossa ensimmäinen termi kuvaa kontrollitilavuuden CV sisäistä muutosta ja toinen termi kontrollitilavuudesta kontrollipinnan CS kautta sisään ja ulos virtaavaa vuota.

Liitteessä E on johdettu Reynoldsin jatkuvuuslauseen avulla virtaavan aineen infinitesimaalisen elementin differentiaalinen liikemääräyhtälö

$$\rho \mathbf{g} - \nabla P + \nabla \cdot \vec{\tau}_{ij} = \rho \frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}, \tag{59}$$

jossa ensimmäinen termi vastaa elementin kokema painovoimaa, toinen painetta, kolmas elementin viskositeettia ja yhtälön oikea puoli kertoo elementin tiheyden kerrottuna elementin kokonaiskiihtyvyydellä eli elementin kokeman kokonaisvoiman. [9]

Alaluvussa 2.1 johdettu d'Alembertin aaltoyhtälö voidaan yleistää kolmiulotteiseen avaruuteen muodossa [10]

$$\nabla^2 \psi(\mathbf{r}) = \frac{1}{\mathbf{v}^2} \frac{\partial^2 \psi_j}{\partial t^2} \tag{60}$$

jossa

$$\mathbf{r} = (x, y, z) \quad \text{ja} \quad \nabla = \frac{\partial}{\partial x}\hat{\imath} + \frac{\partial}{\partial y}\hat{\jmath} + \frac{\partial}{\partial z}\hat{k}, \tag{61}$$

jonka ratkaisut ovat muotoa

$$\psi(\mathbf{r}) = A e^{i(\omega t \pm \mathbf{kr})} \tag{62}$$

jossa

$$\mathbf{k}^2 = k_x^2 + k_y^2 + k_z^2. \tag{63}$$

Herätteen ollessa pistemäinen sen tuottamat aallot ovat pallomaisia, jolloin on

hyödyllistä käyttää pallokoordinaatteja aaltoyhtälölle

$$\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} = \frac{1}{\mathbf{v}^2}\frac{\partial^2\psi_j}{\partial t^2}, \quad (64)$$

jolloin aaltoyhtälö voidaan separoida muuttujittain

$$\psi(r,\theta,\phi,t) = R(r)\cdot\Theta(\theta)\cdot\Phi(\phi)\cdot T(t). \quad (65)$$

Liitteessä F on johdettu pallovärähtelijän aaltofunktio

$$\psi(r,\theta,\phi,t) = R(r)\cdot\Theta(\theta)\cdot\Phi(\phi)\cdot T(t) = (ABC)j_\ell(kr)P_\ell^m(\cos(\theta))\cos(m\phi)\cos(\omega t)$$

$$= Dj_\ell(kr)Y_\ell^m(\theta,\phi)\cos(\omega t). \quad (66)$$

Pallovärähtelijän aiheuttama häiriö leviää väliaineessa aaltofunktion [7]

$$\psi(r,\theta,\phi,t) = E(j_\ell(kr) + iy_\ell(kr))Y_\ell^m(\theta,\phi)e^{-i\omega t} \quad (67)$$

kuvaamalla tavalla, jossa y_ℓ on Besselin pallofunktio muotoa

$$y_\ell(x) = \sqrt{\frac{\pi}{2x}}Y_{n+\frac{1}{2}}(x) = (-1)^{n+1}\sqrt{\frac{\pi}{2x}}J_{-n-\frac{1}{2}}(x). \quad (68)$$

2.5 Ääniaaltojen akustisia ominaisuuksia

Jos alaluvussa 2.3 kuvailtuun järjestelmään kohdistetaan hetkellinen heräte, väliaine kokee puristuksen ja paineeseen syntyy muutos. Tämä tasapainotilasta poikkeava paine on äänipaine

$$p = P - p_0, \quad (69)$$

jossa P on järjestelmän kokonaispaine ja p_0 on paine tasapainotilassa. Edetessään väliaineessa äänipaine muodostaa alaluvussa 2.1 käsittelemäämme häiriötä vastaavan ääniaallon. Olettaen ääniaallon etenevän nopeammin kuin lämmönsiirron, voimme todeta ääniaallon etenemisen olevan adiabaattinen prosessi ja äänennopeus voidaan ilmoittaa paineen muutoksen suhteena väliaineen tiheyden $\rho = m/V$ isentrooppiseen muutokseen tai ainekohtaisen isentrooppisen puristuskertoimen (52) suhteena tiheyden muutokseen [8]

$$v^2 = \left(\frac{\partial P}{\partial\rho}\right)_S = \frac{\partial K_S}{\partial\rho}. \quad (70)$$

Putkessa ääniaallot etenevät tasoaaltoina ja paineaaltoa kuljettavat hiukkaset värähtelevät edestakaisin yksiulotteisesti tasapainopisteensä ympärillä. Tällöin hiukkassiirtymää kuvaa aaltofunktion fyysinen reaaliosa

$$y = \mathrm{Re}(y_0 e^{i(\omega t \pm kx)}) = y_0 \cos(\omega t - kx), \tag{71}$$

jonka aikaderivaatta on akustinen hiukkasnopeus, joka kuvaa ääniaallon johdosta liikkuvien hiukkasten nopeutta

$$u = \frac{\mathrm{d}y}{\mathrm{d}t} = -y_0 \omega \sin(\omega t - kx). \tag{72}$$

ja joka kerrottuna aaltoliikkeen poikkipinta-alalla A on akustinen hiukkasvuo [12]

$$U = Au = -Ay_0 \omega \sin(\omega t - kx). \tag{73}$$

Yhtälöistä (48) ja (69) näemme että

$$p = P - p_0 = dP = -K\frac{\mathrm{d}V}{V} = -K\frac{A\ \mathrm{d}y}{A\ \mathrm{d}x} = -K\frac{\mathrm{d}y}{\mathrm{d}x} = -Ky_0 k \sin(\omega t - kx), \tag{74}$$

jolloin hiukkasnopeus on siis hiukkassiirtymän aikaderivaatta ja paine on hiukkassiirtymän paikkaderivaatta. Huomataan myös että hiukkassiirtymäaalto y ja paineaalto p omaavat vaihe-eron $\pi/2$ eli hiukkassirtymä saavuttaa ääriarvonsa paineaallon ollessa nolla ja paineaalto saavuttaa ääriarvonsa hiukkasten ollessa tasapainotilassa.

Äänenvoimakkuus eli ääniaallon intensiteetti muodostuu äänipaineen ja akustisen hiukkasnopeuden tulosta

$$I = pu = Kk\omega y_0^2 \sin^2(\omega t - kx), \tag{75}$$

ja intensiteettitasoa mitataan

$$L_I = 10 \log_{10}\left(\frac{I}{I_0}\right) \tag{76}$$

jossa referenssi-intensiteetti on usein $I_0 = 1\ \mathrm{pW/m^2}$ [12]. Äänenvoimakkuutta mitataan kuitenkin äänipainetasolla

$$L_p = 20 \log_{10}\left(\frac{p}{p_0}\right), \tag{77}$$

jossa referenssiäänipaineeksi valitaan usein ihmiskuulon alaraja $p_0 = 20\ \mu$Pa [12].

Alaluvussa 2.2 nähtiin aaltoliikkeen vaimenevan sen edetessä avaruudessa ja ajassa. Ääniaaltoa väliaineessa vaimentavaa tekijää kutsutaan akustiseksi impedanssiksi ja se riippuu väliaineen ominaisuuksista. Levitessään paineaalto vaimenee $p \propto 1/r$, kun taas intensiteetti vaimenee $I \propto 1/r^2$. Ääniaallonnopeuden ja tiheyden avulla voidaan määrittää väliaineen akustinen ominaisimpedanssi [11]

$$z = \rho v, \tag{78}$$

kun taas paineen ja hiukkasvuon avulla saamme väliaineen yleisen akustisen impedanssin

$$Z = \frac{p}{U} = \frac{p}{Au} = \frac{z}{A}. \tag{79}$$

Kun ääniaalto ψ_A saapuu kahden eri impedanssin omaavan alueen rajapinnalle, se jakautuu kahtia rajapinnalta takaisinheijastuvaan aaltoon ψ_B ja rajapinnan läpäisevään aaltoon ψ_C, jolloin impedanssin Z_1 alueella on siis aaltofunktio $\psi_1 = \psi_A + \psi_B$ ja impedanssin Z_2 alueella aaltofunktio $\psi_2 = \psi_C$. Koska paineaallon ja hiukkasvuon on oltava jatkuvia rajapinnalla $x = 0$ saamme rajaehdot

$$\psi_1(0) = \psi_2(0) \qquad \text{ja} \qquad \frac{\partial \psi_1(0)}{\partial x} = \frac{\partial \psi_2(0)}{\partial x}. \tag{80}$$

Näistä saamme yhtälöparin

$$\begin{cases} p_A + p_B = p_C \\ U_A - U_B = U_C \iff p_A/Z_1 - p_B/Z_1 = p_C/Z_2 \end{cases} \tag{81}$$

yhtälön (79) avulla. Yhtälöparista voidaan ratkaista heijastus- ja läpäisykertoimet

$$R = \frac{p_B}{p_A} = \frac{Z_2 - Z_1}{Z_1 + Z_2} \tag{82}$$

ja

$$T = \frac{p_C}{p_A} = \frac{2Z_2}{Z_1 + Z_2} \tag{83}$$

sekä absorptio

$$A = 1 - R^2. \tag{84}$$

Impedanssierojen ollessa suuria eli impedanssien ollessa epäsopivia keskenään, heijas-

tusilmiö on voimakas ja tuottaa alaluvussa 2.2 kuvailtuja seisovia aaltoja. Jos ääniaaltosignaali on lyhyt suhteessa sen kulkemaan matkaan, heijastusilmiö tunnetaan kaikuna. [6] [12]

Kun ääniaalto kulkee sylinterimäisessä putkessa jonka pää on avonainen, putken sisällä ja ulkopuolella olevalla väliaineella on yhtälön (79) mukaisesti impedanssiero ja aalto heijastuu osittain suuaukolta. Kun putken pituus on L ja putken molemmat päät ovat avonaisia, putkeen syntyy seisovia aaltoja taajuuksilla [16]

$$f_n = \frac{v}{\lambda_n} = n\frac{v}{2\pi L}. \tag{85}$$

Jos putkessa on kaventumia tai laajenemia eli sen läpileikkauspinta-ala vaihtelee, tulee virtausnopeuden muuttua massan säilymislain mukaisesti [9]

$$\dot{m}_1 = A_1\rho\mathbf{V}_1 = A_2\rho\mathbf{V}_2 = \dot{m}_2, \tag{86}$$

eli pinta-alan kasvaessa virtausnopeus pienenee ja pinta-alan pienentyessä virtausnopeus kasvaa. Kun väliaineen hiukkaset virtaavat hitaasti ($u < 0{,}3$ Ma), voidaan kokoonpuristuvaa väliainetta approksimoida kokoonpuristumattomana ja todeta virtauksen noudattavan liitteessä G johdettua Bernoullin yhtälöä

$$P + \frac{1}{2}\rho|\mathbf{V}|^2 + \rho g z = C, \tag{87}$$

jossa ensimmäinen termi on painetermi, toinen termi on kineettinen energia, kolmas termi on potentiaalienergia ja C on vakio. Massan säilymisen (86) ja Bernoullin lain (87) avulla näemme, että nopeuden kasvaessa putken läpileikkauspinta-alan pienentymisen johdosta väliaineen paine pienenee ja virtausnopeuden laskiessa pinta-alan kasvamisen johdosta väliaineen paine kasvaa.

Tarkastellessa kuutiomaista huonetta kolmiulotteisena laatikkona jonka rajoilla on suuri impedanssiero, voidaan huomata yhdenkin jatkuvan ääniherätteen heijastuksien täyttävän laatikon ja muodostavan huoneeseen interferenssikuvion. Kolmiulotteinen aaltofunktio (62) voidaan esittää karteesisessa koordinaatistossa muodossa

$$\psi = A\cos\left(\frac{n_x\pi x}{L_x}\right)\cos\left(\frac{n_y\pi y}{L_y}\right)\cos\left(\frac{n_z\pi z}{L_z}\right)e^{i\omega t}, \tag{88}$$

jolloin huoneeseen syntyy seisovia aaltoja taajuuksilla [12]

$$\mathbf{f}_n = \frac{v}{2\pi}\mathbf{k} = \frac{v}{2\pi}\sqrt{\left(\frac{n_x\pi}{L_x}\right)^2 + \left(\frac{n_y\pi}{L_y}\right)^2 + \left(\frac{n_z\pi}{L_z}\right)^2}. \tag{89}$$

Jos huoneen äänikenttää tarkastellaan k-avaruudessa, voimme todeta yhtälön (63) olevan pallon yhtälö jossa kukin k-avaruuden hilapiste edustaa resonanssitaajuutta. Saadaksemme resonanssitaajuuksien tiheyden huoneessa, tulee tutkia vain positiivisia aaltolukuja jolloin aaltolukupallon tilavuudeksi saadaan $(4/3\pi\mathbf{k}^3)/8 = 1/6\pi\mathbf{k}^3$. Kun jokaisen hilapisteen etäisyys naapuripisteeseen on π/L_i, voidaan k-tilavuus per hilapiste kirjoittaa muodossa $\pi^3/L_xL_yL_z = \pi^3/V$ ja saamme resonanssitaajuus- eli mooditiheyden pallolle säteen $\mathbf{k}$ sisällä

$$\mathbf{N} = \frac{1/6\pi\mathbf{k}^3}{\pi^3/V} = \frac{4\pi}{3}V\left(\frac{\mathbf{f}}{v}\right)^3 \tag{90}$$

jolloin mooditiheys per taajuus $\mathbf{f}$ on

$$\frac{\mathrm{d}\mathbf{N}}{\mathrm{d}\mathbf{f}} = 4\pi V\frac{\mathbf{f}^2}{v^3}. \tag{91}$$

Seisovat aallot muodostavat huoneeseen liitteessä H esitellyn äänienergiatiheyden E_0 joka vaimenee ajassa. Äänienergiatiheystaso voidaan kirjoittaa

$$L_E = 10\log_{10}\left(\frac{\langle E\rangle}{E_0}\right) = 4{,}34\ln\left(\frac{\langle E\rangle}{E_0}\right) \tag{92}$$

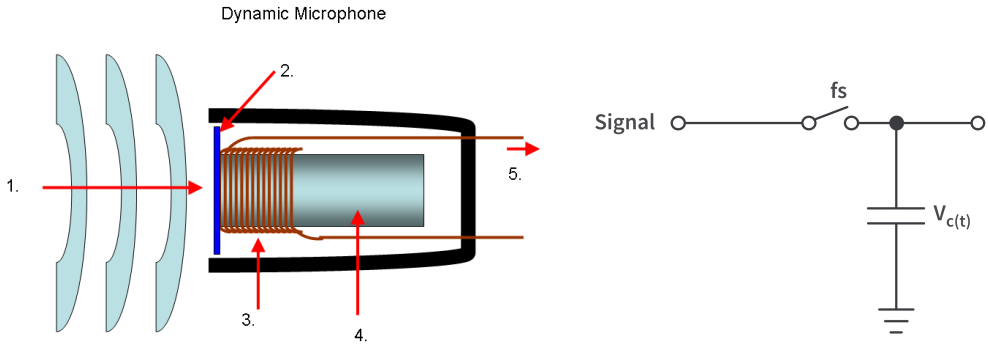
ja kaiun vaimenemista huoneessa voidaan mitata äänienergiatiheystason muutoksella ajassa

$$\dot{L}_E = 4{,}34\frac{\langle\dot{E}\rangle}{\langle E\rangle}. \tag{93}$$

3 Signaalinkäsittely

Luvussa esitellään digitaalisen signaalinkäsittelyn perusteita ja käydään läpi puheen matka analogisesta signaalista spektriksi. Digitaalinen signaalinkäsittely tutkii erilaisia diskreeteistä näytepisteistä saatuja datajoukkoja, joita voidaan mallintaa aaltomuotoisina funktioina. Alaluku 3.1 perustuu lähteeseen [17] ja kertoo signaalin digitoinnista ja käsittelystä eri muunnoksin, kun taas alaluku 3.2 on lähteen [18] mukainen kuvaus signaalin ikkunoinnista ja spektriominaisuuksista.

3.1 Signaalin diskretisointi



(a) Dynaamisen mikrofonin rakenne. Kuva: [19]

(b) Yksinkertaisen A/D-muuntimen piirikaavio. Kuva: [20]

Kuvio 1. Kuvan (a) dynaaminen mikrofoni koostuu joustavasta kalvosta (2.), käämistä (3.) ja kestomagneetista (4.). Mikrofoni muuntaa ääniaaltojen (1.) signaalin sähköiseksi signaaliksi (5.), joka voidaan tuoda kuvan (b) piiriin. A/D-muuntimessa sähkösignaalia näytteistetään avaamalla ja sulkemalla kytkintä taajuudella f_s ja kytkimen liikkeiden välissä kondensaattori V pitää jännitteen tasaisena.

Ilmassa etenevä paineaalto voidaan muuntaa sähköiseksi signaaliksi mikrofonilla, jonka toiminta perustuu magneetin ympäröivään käämiin kuvion 1 mukaisesti. Ilman hiukkasten värähtely työntää käämiä edestakaisin magneetin ympärillä, jolloin magneettikenttä indusoi käämiin ääniaaltoa vastaavan sähkövirran. Analogisen sähkösignaalin diskretisointi tapahtuu ottamalla signaalista n-kappaletta näytteitä

näytteenottotaajuudella f_s. Näytteistäminen tapahtuu A/D-muuntimella (Analog-to-Digital), jonka näytteenotto- ja pitopiiri yksinkertaisimmillaan koostuu kuviossa 1 esitetystä digitaalikellon kontrolloimasta kytkimestä sekä kondensaattorista. Kytkin sulkeutuu näytteistämisen aikana ja avautuu pitoa varten, kun taas kondensaattori varautuu portin ollessa kiinni ja pitää jännitteen vakiona portin ollessa auki. Tällöin piiri muuntaa analogisen signaalin portaistetuksi diskreetiksi signaaliksi. [17]

Jos periodisesta funktiosta otetaan liian harvoin näytteitä, korkeataajuuksiset aallot voivat vaikuttaa matalataajuisilta aalloilta ja niiden etenemissuunta voi vääristyä. Ilmiötä kutsutaan aliasioitumiseksi tai laskostumiseksi ja siltä voidaan välttyä kasvattamalla näytteenottotaajuutta korkeammaksi. Voimme havaita diskretisoidun sekvenssin muodostaman aaltomuodon katoavan eli sekvenssin jokaisen alkion saavan saman arvon, kun näytteenottotaajuus on kaksinkertaisesti näytteistetyn signaalin korkeimman taajuuskomponentin taajuus $f_s = 2f_{\max}$. Tämä on siis näytteenottotaajuuden alaraja ja pienin aliasoitumisen estävä näytteenottotaajuuden määräytyy Nyquist-Shannonin näytteenottolauseesta [21]

$$f_s > 2f_{\max}. \tag{94}$$

Aikadiskreetin signaalin muuttumista voidaan tutkia saapuvan ja lähtevän signaalin ominaisuuksien avulla sen kulkiessa eri järjestelmien läpi esittämällä tuleva signaali sekvenssinä $x[n]$ ja lähtevä signaali sekvenssinä $y[n]$. Jos signaalien relaatio on lineaarinen ja aikainvariantti (Linear time invariant, LTI) eli järjestelmä muuntaa signaaleja lineaarisesti eikä muuta signaalin vaihetta, niin järjestelmää voidaan kuvata impulssivasteella $h[n]$ eli

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = x[n] * h[n], \tag{95}$$

jossa $*$ on konvoluutio-operaatio ja $x[k]$ on saapuvan signaalin

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \tag{96}$$

painosekvenssin k:es arvo. [17]

Jotta signaalia voidaan tarkastella aika-alueen sijaan taajuusalueessa eli ajan sijaan taajuuden funktiona, tulee signaalille suorittaa z-muunnos jota merkitään

operaattorilla $\mathcal{Z}\{\cdot\}$. Muunnos ottaa diskreetin sekvenssin $x[n]$ ja palauttaa jatkuvan funktion $X(z)$

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n]z^{-n}, \tag{97}$$

jossa $z \in \mathbb{C}$. Taajuusalueella voimme nyt esittää systeemiin saapuvaa ja siitä lähtevää signaalia jatkuvina funktioina $X(z)$ sekä $Y(z)$ ja lineaarista aikainvarianttia järjestelmää siirtofunktiona $H(z)$

$$Y(z) = H(z)X(z). \tag{98}$$

Toisin kuin aika-alueen impulssivaste, taajuusalueen siirtofunktio on analyyttinen ja helposti ratkaistava. Siirtofunktio ja impulssivaste linkittyvät relaatiolla

$$H(z) = \sum_{n=N_1}^{N_2} h[n]z^{-n}. \tag{99}$$

Kun muuttuja on polaarimuotoa $z = re^{i\omega}$ ja pituus $r = 1$, z-muunnos pelkistyy aikadiskreetiksi Fourier-muunnokseksi (Discrete-time Fourier transformation, DTFT), jota merkitään operaattorilla $\mathcal{F}_{DT}\{\cdot\}$

$$X(e^{i\omega}) = \mathcal{F}_{DT}\{x[n]\} = \sum_{k=-\infty}^{\infty} x[n]e^{-i\omega n}. \tag{100}$$

Kvantittamalla taajuuden $\omega = 2\pi k$ jossa $k = 0,1,...,N-1$ saamme aikaan diskreetin Fourier-muunnoksen (Discrete Fourier transform, DFT) jota merkitään operaattorilla $\mathcal{F}_D\{\cdot\}$

$$X[k] = \mathcal{F}_D\{x[n]\} = \sum_{n=0}^{N-1} x[n]e^{-i(2\pi kn/N)}, \tag{101}$$

joka palauttaa diskreetin sekvenssin. [17]

3.2 Signaalin lyhytaikakäsittely

Jos käsiteltävä systeemi on lineaarinen mutta aikavariantti, se ei omaa impulssivastetta $h[n]$. Edellisessä alaluvussa kuvattuja työkaluja voidaan kuitenkin käyttää ajallisesti lyhyisiin osiin signaalia, jotka ovat aikainvariantteja. Ikkunasekvenssi $w[m]$ on lyhyellä intervallilla määritelty funktio, joka leikkaa signaalisekvenssistä $x[n]$

lyhyen ikkunan kohdassa k [18]

$$x[k,m] = w[k - Lm/2]x[k]. \tag{102}$$

Peräkkäisillä ikkunoilla $x[k,m-1]$ ja $x[k,m]$ on päällekkäisyyttä alueella $k \in [Lm/2,\, L(m+1)/2[$. Kun ikkunat ynnätään yhteen, saadaan

$$\begin{aligned} x[k,m-1] + x[k,m] &= w[k - L(m-1)/2]\; x[k] + w[k - Lm/2]\; x[k] \\ &= (w[k - L(m-1)/2] + w[k - Lm/2])\; x[k] \end{aligned} \tag{103}$$

jolloin signaalisekvenssin $x[k,m-1] + x[k,m] = x[k]$ rekonstruktio on täydellinen vain ja vain jos

$$w[k + L/2] + w[k] = 1, \qquad k \in [0,\, L/2[. \tag{104}$$

Eräitä usein käytettyjä ikkunointifunktioita ovat kolmiofunktio [18]

$$w[k] = 1 - \left| \frac{k - \frac{L}{2}}{\frac{L}{2}} \right|, \quad 0 \le k \le L \tag{105}$$

ja Hannin ikkunafunktio

$$w[k] = \frac{1}{2}\left(1 - \sin\left(\frac{2(k+0.5)\pi}{L}\right)\right) = \sin^2\left(\frac{\pi(k+0.5)}{L}\right), \quad 0 \le k \le L. \tag{106}$$

Signaali-ikkunoita käsiteltäessä on kätevää käyttää aikadiskreetin Fourier-muunnoksen impulssivasteen sijaan lyhytaikaista Fourier muunnosta (short-time Fourier transform, STFT), jota merkitään operaattorilla $\mathcal{F}_{ST}\{\cdot\}$

$$X(m, e^{i\omega}) = \mathcal{F}_{ST}\{x[m]\} = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-i\omega n} \tag{107}$$

ja joka on diskreetissä muodossaan $\mathcal{F}_{DST}\{\cdot\}$ [18]

$$X[m,k] = \mathcal{F}_{DST}\{x[m]\} = \sum_{m=0}^{L-1} x[m]w[n-m]e^{-i(2\pi k/n)m}. \tag{108}$$

Kun signaalin kullekin ikkunalle on suoritettu Fourier-muunnos, voidaan signaali jälleenrakentaa taajuus-aika-spektrogrammina asettamalla muunnetut ikkunat yhteen peräkkäin. Koska kuuloaisti on logaritminen, on puhesignaalin spektrogrammista

mielekästä ottaa logaritmi puheominaisuuksia tarkastellessa jolloin spektrogrammin taajuuksille asettuvasta tehosuureesta tulee desibeleinä mitattu äänekkyys. Jos nyt palautamme signaalin aika-alueelle käänteisellä Fourier-muunnoksella, saamme signaalin tehokepstrin (Power cepstrum) [18]

$$\text{Tehokepstri} = |\ \mathcal{F}^{-1}\{\ \log(\ |\ \mathcal{F}\{\ x[k]\ \}\ |\ )\ \}\ |^2 \tag{109}$$

jonka suureina ovat aikaa kuvaava kvefrenssi (Quefrency) ja tehoa kuvaava absoluuttinen äänekkyys. Nyt signaalia on käsitelty ottamaan huomioon äänekkyydeltään puheominaisuuksille tärkeitä ominaisuuksia.

Koska ihmiskuulo kokee myös taajuuden logaritmisena, puheenkäsittelyä varten on empiirisesti määritelty Mel-taajuusskaala

$$m = 2595\ \log_{10}\left(1 + \frac{f}{700}\right), \tag{110}$$

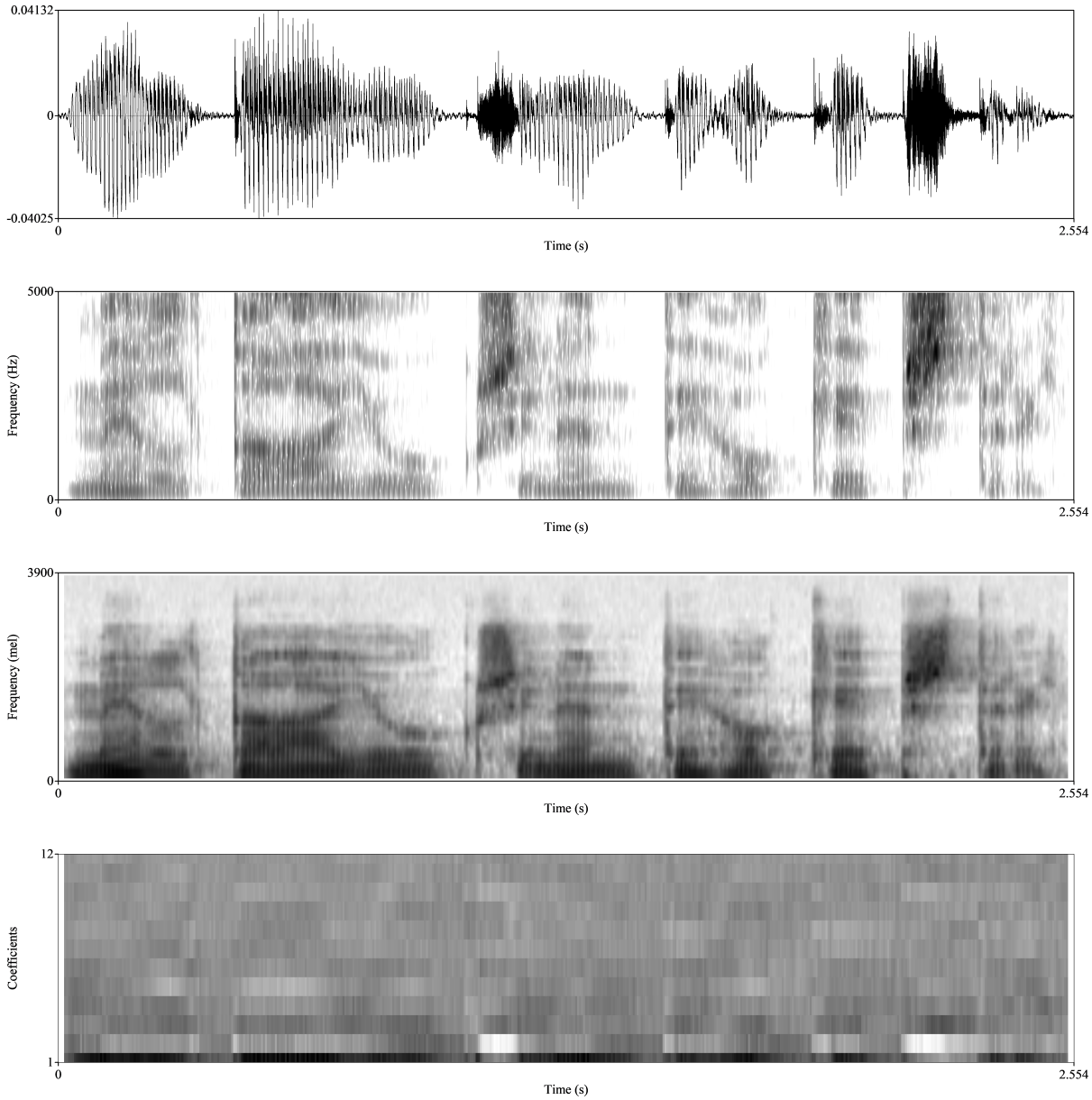
jossa m on Mel-taajuus ja f on Hz-taajuus [18]. Käyttäen Mel-taajuuksin eteneviä kolmiofunktioita painokertoimina

$$w_{k,m} = \begin{cases} \frac{m - f_{k-1}}{f_k - f_{k-1}} & \text{kun} \quad f_{k-1} < m \leq f_k, \\ \frac{f_{k+1} - m}{f_{k+1} - f_k} & \text{kun} \quad f_k < m \leq f_{k+1}, \\ 0 & \text{muuten} \end{cases} \tag{111}$$

voimme kertoa signaalia $x[n]$ kohdassa k saaden puheelle ominaisille taajuuksille painotetun signaalin

$$u[k] = \sum_{m=f[k-1]+1}^{f[k+1]-1} w[k,m]|x[m]|^2 \tag{112}$$

josta voidaan rakentaa Mel-taajuuksinen spektrogrammi. Sijoittamalla Mel-painotetun signaalin tehokepstriin saamme Mel-taajuuksisen tehokepstrin (Mel-frequency cepstral, MFC), josta voidaan tunnistaa Fourier-muunnoksen reaalisina kertoimina Mel-taajuuksisia kepstrikertoimia (Mel-frequency cepstral coefficients, MFCC), jotka kertovat akustisen energian sijoittumisesta eri taajuuksille. MFC-kertoimet ovat puheominaisuuksia, joiden perusteella puhe-esityksestä voidaan tunnistaa puheen eri äänteitä. Äänisignaalin signaalinkäsittelyä eri vaiheissa esitellään kuviossa 2.

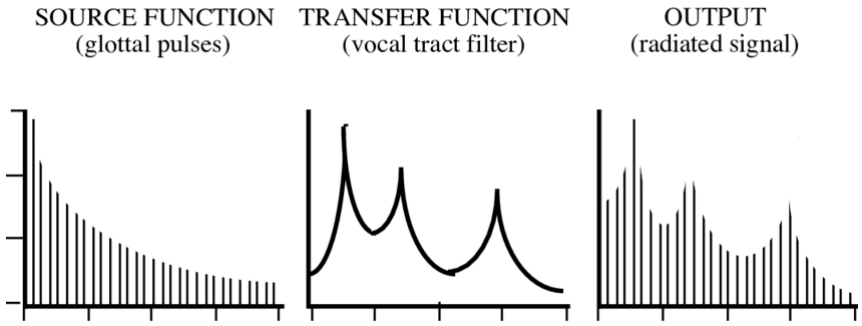


Kuvio 2. Neljä eri esitystä äänisignaalista. Ylimpänä äänitaso-aika-kuvaaja, seuraavaksi taajuus-aika-spektrogrammi hertseinä, toiseksi alimpana taajuus-aika-spektrogrammi meleinä ja alimpana Mel-taajuuksiset tehokepstrikertoimet (MFCC). Spektrogrammit muodostettiin käyttämällä 5 ms Hann-ikkunoita ja MFC-kerroinvektorit rajattiin 12-ulotteisiksi jotta ne erottuisivat, mutta käytännössä MFC-kertoimia käytetään lähinnä 32- tai 39-ulotteisina.
Kuvat tehtiin Praat-ohjelmalla. Kuva: Nicolas Richterich.

4 Fonetiikka

Luvussa käsitellään puheentuotantoa foneettisesta näkökulmasta. Alaluku 4.1 esittelee lähde-suodin-mallin fonaatioteorian ja artikulatorisen järjestelmän mukaillen lähteitä [22], [23] ja [24]. Alaluvun 4.2 aluksi tarkastellaan IPA-taulukkoa, jonka jälkeen käydään useiden lähteiden mukaisesti läpi foneemien laskennallisia tunnistamistapoja.

4.1 Foneemien tuottaminen



Kuvio 3. Lähde-suodin-malli. Äänihuulet muodostavat vasemmalla kuvatun lähdefunktion ja ääntöväylä suotimena muodostaa keskellä kuvatun siirtofunktion. Näiden lineaarikombinaatio on oikealla esitetty ääniaaltoina leviävä signaali, jonka huiput ovat signaalin resonanssitaajuuksia eli formantteja. Kuva: [25]

Ihmisen äänentuotantojärjestelmä koostuu keuhkoista, äänihuulista ja ääntöväylästä joiden sisällä virtaavaa ilmaa järjestelmässä voidaan alaluvun 2.5 mukaisesti kohdella kokoonpuristumattomana väliaineena. Lähde-suodin-malli (Source-filter model) jakaa fonaation kuvion 3 mukaisesti äänihuulien muodostamaan signaalilähteeseen ja ääntöväylän muodostamaan signaalisuotimeen [22]. Puheentuotantoon tarvittava ilmavirta voidaan synnyttää yksin kielellä eli velaarisesti tai äänihuulilla eli glottaalisesti, mutta valtaosa puheeseen tarvittavasta ilmavirrasta muodostetaan pulmonisesti keuhkojen tilavuutta muuttamalla. Ilmavirran voi synnyttää vetämällä ilmaa sisään eli ingressiivisesti tai puhaltamalla ilmaa ulos eli egressiivisesti, kuten yleisesti puheessa tehdään. [24]

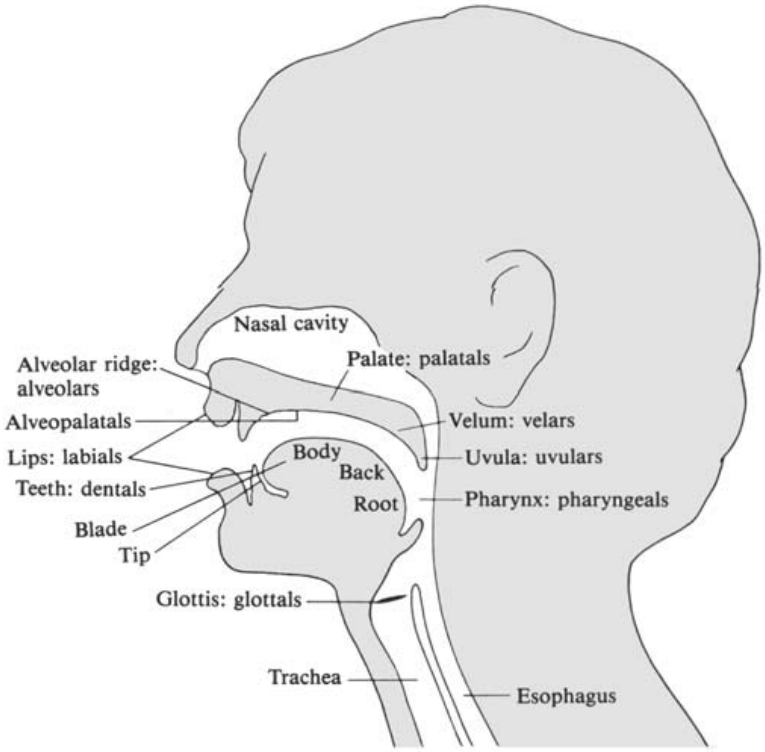
Lähdesignaali muodostuu ilmavirtauksen kulkiessa äänihuulien muodostaman ääniraon läpi. Hengittäessä äänihuulet ovat sivussa eivätkä ne vaikuta kurkun kautta kulkevaan ilmavirtaan, mutta ääntäessä kurkunpään lihakset painavat äänihuulia yhteen. Puristaessa äänihuulia yhteen äänirako pienenee ja egressiivisessä prosessissa äänihuulten alapuolelle muodostuu ylipaine, joka työntää äänihuulet auki. Ilman virtaus henkitorvessa vastaa alaluvussa 2.5 käsiteltyä virtausta vaihtelevan läpileikkauspinta-alan omaavassa putkessa, jolloin ääniraon kohdalla syntyy Bernoullin lain mukaisesti korkean nopeuden omaava ilmavirta. Nopean ilmavirran muodostama alipaine vetää äänihuulet jälleen yhteen, jolloin ilmavirta katkeaa tai supistuu jonka jälkeen prosessi toistuu [23]. Näin ääntöväylään muodostuu paineaalloista koostuva lähdesignaali, joka koostuu alimmasta resonanssitaajuudesta f_0 ja sen yläsävelsarjasta $\sum_n f_{n+1}$. Alin resonanssitaajuus eli ensimmäinen formantti f_0 määrää ihmisen äänenkorkeuden ja vaihtelee miehillä välillä $100 - 500$ Hz ja naisilla välillä $130 - 800$ Hz äänihuulten pituuden mukaan [26]. Äänihuulia voidaan mallintaa yksinkertaisesti kahtena vastakkain värähtelevänä jousi-massa-järjestelmänä, jolloin järjestelmä noudattaa Hooken lakia ja äänihuulten värähtelyn nopeutta kuvaa huulten jännitteen $T = \kappa(L - L_0)$ ja lineaarisen tiheyden $\mu = m/L$ suhde [23]

$$v^2 = \frac{T}{\mu} = \frac{TL}{m}. \tag{113}$$

Tällöin äänihuulten tuottamat resonanssitaajuudet ääntöväylässä ovat molemmista päistä auki olevan putken resonanssitaajuuksien (85) mukaan ja taajuudet voidaan esittää Mersennen lailla

$$f_n = \frac{v}{\lambda_n} = n\frac{1}{2\pi L}\sqrt{\frac{TL}{m}} = n\frac{1}{2\pi}\sqrt{\frac{TL}{mL^2}} = n\frac{1}{2\pi}\sqrt{\frac{T}{mL}} = n\frac{1}{2\pi}\sqrt{\frac{\kappa}{m}\left(1 - \frac{L_0}{L}\right)}. \tag{114}$$

Lähdesignaali saapuu ääntöväylään, jossa artikulaatio tapahtuu. Artikulaatiossa ilmavirran aerodynaaminen energia muuttuu akustiseksi energiaksi. Ääntöväylää voidaan mallintaa molemmista päistä avoimena putkena, jonka poikkipinta-ala vaihtelee. Äänihuulien tuottamien paineaaltojen nähdään muodostavan ääntöväylään seisovia aaltoja alaluvun 2.2 mukaisesti. Ääntöväylän supistumat aiheuttavat jälleen Bernoullin lain mukaisesti nopeita ilmavirtoja ja painevaihteluita ääntöväylässä, jotka moduloivat lähdesignaalia ääntöväylän muodon mukaisesti luoden suotimen siirtofunktioon kuviossa 3 nähdyt formanttipiikit.

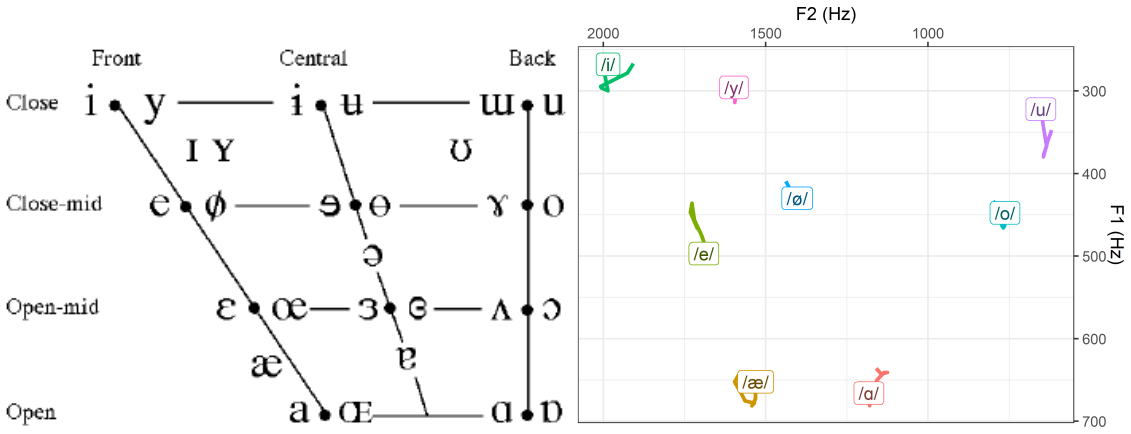


Kuvio 4. Ääntöpaikat ääntöväylässä. Kuva: [27]

Artikulatoriset elimet jakautuvat aktiivisiin elimiin, kuten huuliin ja kieleen, sekä passiivisiin elimiin, kuten hampaisiin ja kitalakeen. Elimet ohjaavat ja estävät ilmavirran kulkua, jolloin kullekin foneemille muodostuu oma ääntöpaikka eli foneemin laatuun eniten vaikuttava sijainti ääntöväylässä sekä ääntötapa eli artikulatoristen elinten asento ja tapa koskettaa toisiaan ääntöväylässä. Kuviossa 4 esitellään ääntöpaikat ja niiden mukaan nimetyt foneemikategoriat. Kuviosta voidaan myös yhdistää kategorioihin liittyvät artikulatoriset elimet eli huulet (labiaalit), hampaat (dentaalit), kitalaki (palataalit ja velaarit), kitakieleke (uvulaarit), nielu (faryngeaalit), äänihuulet (glottaalit) ja kielen eri osat. [24]

4.2 Foneemien tunnistaminen

Ihmiskielten yleisimmät foneemit esitetään usein koottuna kansainväliseen foneemiaakkostoon (International phonetic alphabet, IPA), joista kuviossa 5 esitellään yleisimmät vokaalit ja kuviossa 6 pulmoniset konsonantit. Taulukossa foneemit ovat luokiteltu niiden artikulatoristen ominaisuuksien avulla, joiden perusteella voidaan tutkia niiden yhteyksiä signaalin puheominaisuuksien kanssa. Foneemit voidaan jakaa



(a) IPA-vokaalikartta. Kuva: [28]

(b) Suomen kielen vokaalit. Kuva: Nicolas Richterich

Kuvio 5. Kuvassa (a) on esiteltynä IPA-vokaalikartta, jossa vokaalit ovat luokiteltu artikulaatiopaikan ja ääntöväylän avoimuuden mukaan. Pisteellä erotetut vokaalit ovat pyöristämätön-pyöristetty-pareja. Kuvassa (b) on suomen kielen vokaalit F_2-F_1-formanttiavaruudessa, jossa voidaan huomata niiden asettuvan vokaalikartan muotoon. Vokaalit lausuttiin pitkinä (noin sekunnin pituisina), jonka vuoksi vokaaleissa kuten /e/ ja /u/ esiintyy liukumista. Formantit erotettiin äänisignaalista Praat-ohjelmalla ja kuvaaja muodostettiin R-ohjelmalla.

soinnittomiin ja soinnillisiin foneemeihin joista soinnilliset äänteet hyödyntävät äänihuulien tuottamaa lähdesignaalia ja soinnittomat äänteet muodostuvat äänihuulten levätessä henkitorven seinämillä. Konsonantit voivat olla soinnillisia tai soinnittomia ja osa konsonanteista muodostaa soinnillinen-soinniton foneemipareja, mutta kaikki vokaalit ovat soinnillisia. [24]

Soinnilliset foneemit voidaan tunnistaa niiden resonanssitaajuuksien suhteiden perusteella. Kukin soinnillinen foneemi voidaan asettaa moniulotteiseen formanttiavaruuteen, jossa tuntemattoman foneemin formanteista koostuva vektori F voidaan tunnistaa lähimmän foneemin sijainnin perusteella. Lähin foneemi löydetään formanttiavaruudesta kosinisimilaariudella

$$\text{sim}(F,\, t_k) = \frac{F^T \cdot t_k}{||F|| \cdot ||t_k||}, \tag{115}$$

jossa t_k ovat tunnettujen foneemien formanttivektorit. Koska ensimmäisellä formantilla f_0 ei ole artikulatorisia ominaisuuksia, se voidaan jättää huomiotta foneemeja tunnistaessa. Tällöin formanttiavaruus koostuu pelkästää yläsävelsarjasta $\sum_n f_{n+1}$ ja soinnillisten vokaalien tunnistaminen onnistuu melko tarkasti jo yläsävelsarjan

kahden alimman formantin suhteella [29].

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Kuvio 6. IPA-taulukko, jossa yleisimmät pulmoniset konsonantit luokitellaan ääntöpaikan ja ääntötavan mukaan. Kahden konsonantin esiintyminen samassa luokassa merkitsee niiden olevan soinniton-soinnillinen-pari. Kuva: [30]

Suomen konsonanteista soinnittomia ovat klusiilit (Plosives) /p/, /t/, /k/ ja /q/ sekä frikatiivit (Fricatives) /f/, /s/, /ʃ/ ja /h/. Frikatiivit voidaan jakaa soinnillisiin ja soinnittomiin soinnittoman osan keston avulla (Duration of unvoiced portion, DUP). Frikatiivien spektrejä karakterisoi ääntöväylässä puristetun ilmavirtauksen turbulenssin aiheuttama satunnainen kohina. Kohina painottuu suun etuosassa tuotetuilla frikatiiveilla korkeille taajuuksille ja suun takaosassa matalemmille taajuuksille, joten frikatiivien artikulaatiopaikkaa ja siten foneemia voidaan arvioida kolmen spektriominaisuuden perusteella. Sibilantit eli alveolaarit ja postalveolaarit voidaan tunnistaa normitetun spektrin maksimikulmakertoimen perusteella (Maximum normalized spectral slope, MNSS)

$$\mathrm{MNSS} = \frac{\mathrm{argmax}_i(u[k]_i - u[k]_{i-1})}{\sum_i u[k]_i}, \tag{116}$$

jossa indeksi i kertoo minkä taajuuden MFC- tai muu painokerroin on kyseessä.

Postalveolaarit voidaan edelleen eritellä kohinan spektrin massakeskipisteellä (Spectral center of gravity, SCG)

$$\mathrm{SCG} = \frac{\sum_{i:f<1{,}2\mathrm{kHz}} i \cdot u[k]_i}{\sum_{i:f<1{,}2\mathrm{kHz}} u[k]_i}. \tag{117}$$

ja alveolaareihin verrattuna matalataajuisten maksimien avulla. [31]

Labiodentaaliset ja glottaaliset frikatiivit puolestaan voidaan määrittää deduktiivisesti siten, että ne eivät sovi kumpaankaan joukkoon. Labiodentaalit ja glottaalit voidaan edelleen erottaa toisistaan glottaalisen kohinan matalan intensiteetin perusteella. Frikatiivit pystytään siis tunnistamaan sen perusteella, ovatko ne soinnillisia vai soinnittomia ja ovatko ne sibilantteja, postalveolaareja, labiodentaaleja vai glottaaleja.

Koska soinnittomat klusiilit eivät omaa äänihuulien tuottamaa lähdesignaalia, niiden tunnistamiseen ei voida käyttää formanttianalyysia. Kaikilla klusiileilla, niin soinnittomilla kuin soinnillisilla, on kuitenkin tunnistettava ääntämisaika (Voice onset time, VOT) eli klusiilin ja klusiilia seuraavan äänteen välinen tauko. [32]

Frikatiivien ja klusiilien tunnistamiskriteerit kuten VOT ja DUP vaihtelevat kielittäin, jonka vuoksi niille ei voida asettaa universaaleja standardeja. Myös soinnillisten foneemien ääntäminen vaihtelee suuresti puheen aikana ääntämisen pituuden, foneemiyhdistelmien ja ääntämistavan mukaan. Semanttisen fonetiikan lisäksi puheesta tekeekin ymmärrettävää ja tunnistettavaa lähinnä puheen pragmaattinen konteksti, eli lausuttujen äänteiden ja sanojen ympärillä olevat äänteet ja sanat [24]. Kontekstiominaisuuksien tunnistamiseen tarvitaan siis klassisten fyysisten ominaisuuksien lisäksi monitulkintaisuuteen ja todennäköisyyteen perustuvia malleja.

5 Puheentunnistus

Viimeisessä luvussa käydään läpi nykyaikaisen puheentunnistuksen (Automatic speech recognition, ASR) toimintaperiaatteita. Tämän tutkielman julkaisuaikaan tarkimmat automaattiset puheentunnistusmallit perustuvat neuroverkkoihin ja koneoppimiseen [33], joiden perusteet esitellään alaluvuissa 5.1 ja 5.2. Puheesta tekstiksi -puheentunnistusmalli edellyttää sekä kielimallia, joka on sanasekvenssien todennäköisyysjakauma, että akustista mallia, joka on foneemi- tai puhe-esitysten todennäköisyysjakauma. Muutama kielimalli ja huomiomekanismi esitellään alaluvussa 5.3, jonka jälkeen puheenkäsittelyosio päättyy akustisten mallien alalukuun 5.4 jossa käydään läpi muutamia akustisia malleja sekä esitellään suomenkielinen puheentunnistusmalli, wav2vec2-xlsr-1b-finnish-lm-v2 [34].

5.1 Neuroverkot

Neuroverkot ovat graafimalleja, jotka koostuvat neuroneista ja niitä yhdistävistä synapseista [35]. Kukin neuroni omaa summaajan, joka vastaanottaa synapsin painolla w varustetun syötteen x, jolloin neuroni k antaa lineaarisen aktivaation

$$u_k(x) = \sum_j^m w_{kj} x_j + b, \tag{118}$$

jossa b on neuronin vakiotermi. Jos signaalisyöte kulkee lineaarisesti yhteen suuntaan neuronista seuraavaan, neuroverkon sanotaan olevan eteensyöttävä neuroverkko (FNN) ja jos signaali syötetään takaisin johonkin verkon neuroniin, verkkoa sanotaan rekursiiviseksi neuroverkoksi (RNN). Kun neuroneissa käytetään tavanomaisen matriisitulon sijaan konvoluutiota, niin kukin neuroni käsittelee syötteen synapsin painomatriisin tai painomatriisien w_i avulla. Jos neuroverkko hyödyntää konvoluutiota yhdessä tai useammassa kerroksessa, sitä kutsutaan konvoluutioneuroverkoksi (CNN) ja sen konvoluutiota hyödyntävän neuronin k aktivaatio on

$$u_k(x) = \sum_j^m w_{kj} * x_j + b. \tag{119}$$

Konvoluutioneuroverkoissa käytetään usein myös ristikkäiskorrelaatiota. Ristikkäiskorrelaatio $\star$ ja konvoluutio $*$ linkittyvät suhteella

$$A * B = A \star \text{rot180}(B), \tag{120}$$

jossa rot180-operaatio vaihtaa matriisin B kaikki alkiot matriisin vastakkaisille puolille. Kun konvoluutioneuroverkossa käytetään ristikkäiskorrelaatiota, aktivaatio on

$$u_k(x) = \sum_j^m w_{kj} \star x_j + b \tag{121}$$

eli ainoa ero tavalliseen konvoluutioaktivaatioon on se, ettei painomatriisia tarvitse kääntää kuten konvoluutiossa. Ristikkäiskorrelaatio ja konvoluutio ovat molemmat similaarisuusmittoja ja korostavat syötteen x ominaisuuksia jotka omat samankaltaisia kuin painomatriisien ominaisuudet. Tämän vuoksi konvoluutioneuroverkkojen painomatriiseja kutsutaan myös suotimiksi. [35]

Koska usean lineaarisen aktivaation tuottavan neuronin verkko voidaan aina esittää yhden lineaarisen aktivaation tuottavan neuronin avulla, syvempien neuroverkkojen hyöty saadaan esittelemällä kullekin neuronille epälineaarinen aktivaatiofunktio ϕ, joita on listattu liitteessä I. Nyt neuronin k siirtofunktio on

$$y_k = \phi(u_k). \tag{122}$$

5.2 Koneoppiminen

Neuroverkkojen parametreja muokkaamalla niistä voidaan tehdä malleja, jotka antavat haluttuja vastauksia annetuilla syötteillä. Koska suurilla neuroverkoilla parametrien muokkaaminen käsin on kuitenkin vaivalloista tai mahdotonta, neuroverkot voivat eri algoritmeja seuraamalla optimoida itse sisäiset parametrinsa. Itsenäistä optimointiprosessia kutsutaan koneoppimiseksi ja sitä voidaan hyödyntää datan luokittelutehtävissä. [35]

Ennen luokittelua data tulee esikäsitellä eli muotoilla ja koodata numeeriseen muotoon. Jos datassa on kategorista dataa eli merkkiketjuja, yksinkertaisin tapa koodata ne numeeriseen muotoon on one-hot-esitys. Esityksessä kutakin datajoukon A merkkiketjua vastaa $n = \dim(A)$ -ulotteinen identiteettivektori, eli one-hot-esitys on kuvaus $W : A \rightarrow R^n$. Jos käsitellään datasekvenssejä jotka ovat keskenään

eripituisia, liian lyhyeeseen dataan tulee lisätä täytettä (Padding) eli siihen tulee lisätä dimensioita nolla-alkioilla jotta datavektorit olisivat samankokoisia.

Luokitellessa datapisteitä neuroverkon siirtofunktio y jakaa data-avaruuden eri alueisiin, joissa olevilla pisteillä on yhteisiä ominaisuuksia. Luokittelun oppiminen voi tapahtua monin eri tavoin. Ohjatussa oppimisessa luokiteltuja tuloksia verrataan valmiisiin ja oikeisiin luokitteluihin, kun taas ohjaamattomassa oppimisessa luokitteluille ei ole valmista vertailukohdetta ja neuroverkko yrittää löytää datasta itsenäisesti rakenteita joita luokitella. [35]

Itseohjautuvassa oppimisessa neuroverkko esikoulutetaan antamalla verkolle ensin tunnettua dataa, jonka jälkeen neuroverkko piilottaa itseltään datapisteitä eli naamioi osan datasta ja kouluttaa itseään vertaamalla omia tuloksiaan piilottamiinsa datapisteisiin. Alkukoulutuksen jälkeen itseohjautuvasti oppiva neuroverkko hienosäädetään joko ohjatusti tai ohjaamattomasti. [36]

Optimoinnissa mitataan neuroverkon antamien arvojen eroa haluttuihin arvoihin hukkafunktiolla (Loss function), joka halutaan minimoida. Yksinkertaisin hukkafunktio on neliöllinen hukkafunktio

$$\mathbb{L}_{PNS} = \sum_{i=1}^{n} (t_k - y_k)^2, \tag{123}$$

jossa t_k on haluttu arvo ja y_k on neuroverkon antama arvo. Lisää hukkafunktioita on listattu liitteessä J.

Vastavirta (Backpropagation) on tapa laskea hukkafunktion gradientti graafiverkon painojen suhteen. Nimensä mukaisesti derivaatat lasketaan verkon painojen suhteen vastavirtaan eli aloittaen viimeisestä synapsista

$$\frac{\partial \mathbb{L}}{\partial w_n} = \frac{\partial \mathbb{L}}{\partial \phi_n} \frac{\partial \phi_n}{\partial w_n} \tag{124}$$

ja edeten Leibnizin ketjusääntöä noudattaen kunkin synapsin w_k läpi

$$\frac{\partial \mathbb{L}}{\partial w_k} = \frac{\partial \mathbb{L}}{\partial \phi_n} \frac{\partial \phi_n}{\partial \phi_{n-1}} \frac{\partial \phi_{n-1}}{\partial \phi_{n-2}} \frac{\partial \phi_{n-2}}{\partial \phi_{n-3}} \cdot \ldots \cdot \frac{\partial \phi_{k+1}}{\partial \phi_k} \frac{\partial \phi_k}{\partial w_k} \tag{125}$$

kunnes saavutetaan ensimmäinen synapsi w_0. [37][38]

Gradienttilasku (Gradient descent) on optimointialgoritmi, jossa lasketaan hukkafunktion gradientti ja pyritään etenemään gradientinvastaiseen eli laskevaan suuntaan etsien hukkafunktion globaalia minimiä. Algoritmissa lasketaan yhden iteraation

aikana kullekin synapsin painolle w_k uusi paino w'_k

$$w'_k = w_k - r\sum_{i=1}^{n} \nabla_w \mathbb{L}, \tag{126}$$

jossa säädettävä parametri r on oppimisnopeus eli iterointiaskeleen koko. Tavallisessa gradienttilaskussa lasketaan kutakin uutta painoa varten koko hukkafunktiogradienttien joukon (Batch) summa, jonka johdosta algoritmi on sekä tarkka että laskennallisesti kallis. [39]

Gradienttilaskun hitaus ja laskennallinen kalleus voidaan välttää stokastisella gradienttilaskulla, jossa lasketaan kullekin synapsille uusi paino hyödyntäen vain yhtä gradienttia

$$w'_k = w_k - r\nabla_w \mathbb{L}. \tag{127}$$

Vaikka stokastisessa gradienttilaskussa ei sinällään ole satunnaisia arvoja, se saa nimensä stokastista käyrää muistuttavasta optimointireitistään yksittäisten ja siten epätarkkojen gradienttiarvojen johdosta. Oppimisnopeuden r suositaan olevan suuri joukkogradienttilaskussa ja pieni stokastisessa gradienttilaskussa. [39]

Yleinen oppimisalgoritmi voisi olla siis seuraavanlainen:
1. Signaali kuljetetaan ensin neuroverkon läpi
2. Lasketaan hukkafunktion $\mathbb{L}$ gradientti kunkin synapsin painon w_i suhteen vastavirran avulla.
3. Lasketaan kullekin painon arvolle w_i uusi arvo w'_i jonkin optimointialgoritmin, kuten stokastisen gradienttilaskun avulla
4. Toistetaan kunnes optimointialgoritmi saavuttaa optimitilan.

5.3 Kielimallit

Kielimallit ovat sanasekvenssien todennäköisyysjakaumia, joissa jokaisella sanajoukon S sanalla s on jokin todennäköisyysarvo $P(s_0,s_1,s_2,...,s_n)$ olla sanasekvenssin seuraava sana. Todennäköisyysarvojen määrittämistä varten kullekin luonnollisen kielen sanalle tulee antaa numeerinen vektoriesitys. [40]

Sanat voidaan koodata one-hot-esityksellä, jolloin jokainen sanaesitys on toisista sanaesityksestä riippumaton eli sanaesityksillä ei ole semanttisia (merkityksellisiä) tai syntaktisia (kieliopillisia) ominaisuuksia. Aluksi kaikki sanat ovat yhtä todennäköisiä, mutta sanavektoreille voidaan opettaa ominaisuusesityksiä kouluttamalla

niitä neuroverkossa tekstidatalla. Koneoppimisen aikana sanavektorit levittäytyvät useampaan ulottuvuuteen sanaominaisuusavaruudessa ja klusteroituvat koulutusdatan perusteella läheisten sanojen pariin jolloin sanaesityksien ominaisuuksia voidaan tutkia aritmeettisin operaatioin. Onnistuneilla sanaesityksillä on sekä semanttisia ominaisuuksia [41]

$$v(\text{"kuningas"}) - v(\text{"mies"}) = v(\text{"kuningatar"}) \tag{128}$$

että syntaktisia ominaisuuksia

$$v(\text{"ilmeisesti"}) - v(\text{"ilmeinen"}) + v(\text{"nopea"}) = v(\text{"nopeasti"}). \tag{129}$$

Kielimallia jossa sanoilla ei ole järjestystä kutsutaan sanasäkiksi (Bag of words, BOW), sillä sanasekvenssin seuraava sana valikoituu kuin säkillisestä satunnaisia sanoja. Sanasäkkimallissa sanavektoreiden ominaisuudet ovat sanojen määrät ja esiintymistiheydet koulutusdatassa. Jos neuroverkossa käytetään softmax-aktivointifunktiota sanaesitykset ovat jatkuvia ja kielimallia kutsutaan jatkuvaksi sanasäkiksi (CBOW). Sanasäkkimallien tarkkuus paranee jos valittavan sanan s_k ympärillä olevien sanojen vektorit summataan yhteen $\overline{v}(s_k) = ... + v(s_{k-1}) + v(s_{k+1}) + ...$ ja lasketaan sanan todennäköisyys vektoreiden kosinisimilaariuden avulla $P(s_k) = \text{sim}(v(s_k),\overline{v}(s_k))$. [41]

Kun kielimalli ottaa huomioon n-kappaletta sanaa s_k edeltäviä sanoja s_{k-1},s_{k-2},..., seuraavan sanan s_k todennäköisyys on ehdollinen $P(s_k|s_{k-1},...,s_{n-(k-1)})$ ja kielimallia kutsutaan n-gram-malliksi. 1-gram malli antaa sanan s_k todennäköisyydeksi

$$P(s_k|s_{k-1}) = \frac{P(s_k \cap s_{k-1})}{P(s_{k-1})}, \tag{130}$$

kun taas 2-gram malli antaa

$$P(s_k,|s_{k-1},s_{k-2}) = \frac{P(s_k \cap s_{k-1} \cap s_{k-2})}{P(s_{k-1} \cap s_{k-2})} \tag{131}$$

ja niin edelleen. [42]

n-gram-mallilla yritetään ymmärtää sanan kontekstia, mutta sekvenssi voi jumittua muutaman sanan kehään ja sanasekvenssien pidentyessä malli käy laskennallisesti kalliiksi. Kontekstin ymmärtämiseen vaaditaan myös usein sanan s_k jälkeisiä sanoja, jotka voivat muuttaa sanan merkitystä. Huomiomekanismi (Attention) kuvaa kysely-

vektorin Q (Query) ja avain-arvo (Key-Value) vektorit K ja V yhdeksi sanavektoriksi. Skaalatussa pistetulohuomiossa annetaan ensin kyselyvektorina sana $v(s_k)$, joka kerrotaan pistetulolla avainvektorien eli kaikkien sanavektorien $v(s_0)$,$v(s_1)$...,$v(s_{n-1})$ mukaanlukien itsensä kanssa antaen n skalaaria. Tämän jälkeen skalaarit skaalataan sanavektorien dimensioiden neliöjuurella $\sqrt{d_s}$ ja ne ajetaan liitteessä I esitellyn softmax-funktion läpi antaen painotetun jakauman. Jakauma toimii itsessään painona kun se kerrotaan pistetulolla arvovektorin kanssa eli jälleen kaikkien sanavektorien kanssa, jolloin huomiomekanismi lopulta palauttaa sanavektorit kontekstualisoituina. Huomiomekanismi voidaan siis kirjoittaa muodossa [5]

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_s}}\right)V. \tag{132}$$

Yhden huomiomekanismin kerran laskemisen sijaan voidaan saman huomion Q, K ja V -vektorit jakaa kukin h-määrään päitä joille lasketaan huomio rinnakkaisesti. Vektorit ovat nyt tavallisen huomion d_{model}-ulotteisten vektoreiden sijaan d_q, d_k ja d_v -ulotteisia ja kunkin pään vektoreilla on eri sanaesitykset. Rinnakkaisten huomiolaskujen jälkeen kaikki kontekstoidut sanavektorit kiinnitetään yhteen concat-funktiolla ja projisoidaan halutunmuotoiseksi. Mekanismia kutsutaan monipäiseksi huomioksi (Multi-head attention) ja se voidaan esittää muodossa

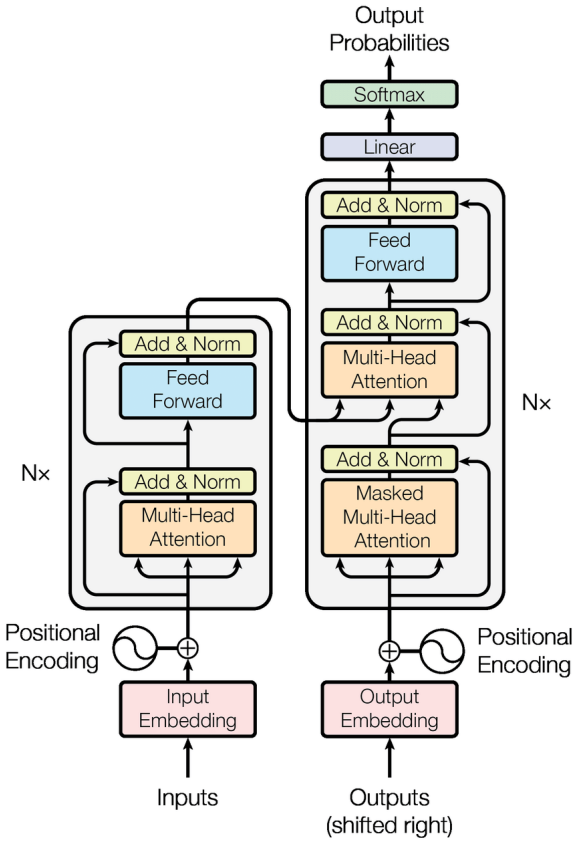
$$\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1,...,\text{head}_h)W^O \tag{133}$$

jossa lineaariprojektio on $W^O \in \mathbb{R}^{hd_V \times d_{\text{model}}}$ ja kukin head_i on

$$\text{head}_i = \text{Attention}(QW_i^Q,KW_i^K,VW_i^V) \tag{134}$$

joissa lineaariprojektiot ovat $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ ja $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$. Monipäinen huomio mahdollistaa samojen sanojen kontekstualisoimisen sanaominaisuusavaruuden eri aliavaruuksissa ja eri sijainneilla samanaikaisesti, tuottaen paremmin kontekstualisoituja sanaesityksiä. Naamioitu monipäinen huomio (Masked multi-head attention) naamioi kultakin sanalta niitä seuraavat sanat. [5]

Kuviossa 7 esitelty huomioon perustuva muuntajamalli (Transformer) kehitettiin kielenkääntämiseen, mutta on löytänyt käyttötarkoituksia useissa tekoälysovelluksissa. Muuntaja koostuu kooderiosasta ja dekooderiosasta sekä datan esi- ja jälkiprosessoinnista. Esiprosessoinnissa syötevektoreille annetaan haluttu muoto ja satunnaisia



Kuvio 7. Muuntajamallin arkkitehtuuri, jossa vasen osio on muuntajan kooderi ja oikea osa dekooderi. Kooderin ulostulo on dekooderin syöte. Kuva: [5]

alkuarvoja lineaariprojektiolla (Input embedding) ja sanavektorit paikkakoodataan (Positional encoding, PE). Paikkakoodaaminen muuntajamallissa on olennaista, sillä huomiomekanismiin nojaavalla kielimalleilla ei ole tietoa sanasekvenssin järjestyksestä kuten rekursiivisilla ja konvolutionaalisilla malleilla. Paikkakoodaukseen voidaan käyttää eritaajuuksisia sinifunktioita [5]

$$\mathrm{PE}_{(\mathrm{pos},2i)} = \sin\left(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}}\right), \tag{135}$$

$$\mathrm{PE}_{(\mathrm{pos},2i+1)} = \cos\left(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}}\right), \tag{136}$$

jossa pos on sanan paikka sekvenssissä ja i on sanan ulottuvuus. Näin sanavektoreille koodautuu paikkaominaisuus sanaominaisuusavaruudessa.

Muuntajan kooderi ottaa syötteekseen esikäsiteltyjä sanavektoreita $s_1,...,s_n$ ja ajaa ne monipäisen huomion ja eteenpäinsyöttävän neuroverkon läpi palauttaen kontekstualisoituja sanavektoreita $z_1,...,z_n$. Sekä monipäisen huomion että eteenpäin

syöttävän neuroverkon jälkeen sanavektorit ynnätään käsittelemättömien sanavektoreiden kanssa ja ne normitetaan. Kun koodereita on pinossa N määrä, vektorit käsitellään N kooderin kautta ennenkuin dekooderiosa palauttaa sanavektorit z_1,...,z_n. Kooderin ja dekooderin välissä vektorit esikäsitellään uudestaan muotoilun ja paikkakoodauksen avulla.

Dekooderi ottaa vastaan uudelleen esikäsitellyt sanavektorit z ja ajaa ne ensin naamioidun monipäisen huomion ja sitten tavallisen monipäisen huomion läpi ennen kuin ne jälleen asetetaan eteenpäinsyöttävään neuroverkkoon, joka palauttaa dekoodatut sanavektorit y_1,...,y_n. Sanavektorit jälleen normitetaan ja ynnätään kunkin vaiheen jälkeen kuvion 7 osoittamalla tavalla vanhojen vektoreiden kanssa. Kun dekoodereita on pinossa N määrä, vektorit käsitellään N dekooderin kautta ennenkuin dekooderiosa palauttaa sanavektorit y_1,...,y_n. Jälkikäsittelyssä vektorit muotoillaan lineaariprojektiolla halutun muotoiseksi ja syötetään softmax-funktioon, joka antaa kielimallin lopullisen todennäköisyysjakauman kullekin sanasekvenssille. [5]

Muuntajamallista on tehty lukuisia variantteja, jotka hyödyntävät muuntajaa tai sen osia. Jos mallissa käytetään vain enkooderipinoa, sitä kutsutaan BERT-malliksi (Bidirectional encoder representations from transformers) ja jos mallissa käytetään vain dekooderipinoa, sitä kutsutaan GPT-malliksi (Generative pre-trained transformers). BERT-malleja voidaan käyttää sanasekvenssien analysointiin, kun taas GPT-mallit kuten ChatGPT kykenevät tuottamaan hyvinkin luonnollisia sanasekvensssejä.

5.4 Akustiset mallit

Akustiset mallit ovat todennäköisyysjakaumia, joissa kullakin syötteenä annetun äänisignaalin ikkunalla on todennäköisin foneemi tai puhe-esitys z. Akustinen malli ottaa syötteekseen äänisignaalin x joka jaetaan T ikkunaan ja palauttaa n alkion sekvenssin $(z_1, z_2, ..., z_n)$ eli se on kuvaus $h : X \rightarrow Z$ kun $n \leq T$. Foneemien ominaisuusavaruus voi olla alaluvulussa 4 esitelty formanttiavaruus tai muihin puheominaisuuksiin perustuva ominaisuusavaruus.

Akustisen mallin vastaanottama äänisignaali voidaan esikäsitellä luvussa 3 esitellyin metodein. Kutakin ikkunaspektriä voidaan kohdella kuvana josta voidaan tunnistaa MFC-kertoimia kuvaominaisuuksina, tehden puheominaisuuksien tunnistustehtävästä kuvaominaisuuksien tunnistustehtävän. Kuvaominaisuuksia tunnistetaan tehokkaasti konvoluutioneuroverkoilla, joten saamme aikaan yksinkertaisen

akustisen mallin kouluttamalla konvoluutioneuroverkko tunnistamaan foneemeja ikkunaspektrien MFC-kerrointen ominaisuuksista. Tällöin konvoluutioneuroverkko antaa todennäköisimmän foneemin y_i n-määrälle syötteenä annettuja ikkunoita x_i, jolloin todennäköisin foneemisekvenssi on

$$p[x] = (y_1, y_2, ..., y_n). \tag{137}$$

Tätä luokittelumetodia kutsutaan ikkunoittaiseksi luokitteluksi (Framewise classification) ja sen lähtöavaruus on ikkunoitujen vektoreiden sekvenssijoukko $X = (\mathbb{R}^m)^*$ ja maaliavaruus foneemien L sekvenssijoukko $Z \in L^*$. Ikkunoittaisen luokittelun ongelmana on äänisignaalin ajallisuus eli luokittelu esittää kustakin ikkunasta tunnistetun foneemin uutena foneemina eikä ota huomioon foneemien pituuksia ikkunoiden suhteen, mikä aiheuttaa ongelmia oikeinkirjoituksessa. Tämä korjaantuu ajallisilla luokittelumetodeilla (Temporal classification), jotka käyttävät neuroverkon viimeisenä kerroksena softmax-funktiota jolloin verkko palauttaa jatkuvan todennäköisyysjakauman foneemien ja tyhjien ikkunoiden eli joukon $L' = L \cup \{\quad\}$ yli ja tunnistavat ikkunoiden foneemit ehdollisella todennäköisyydellä [43]

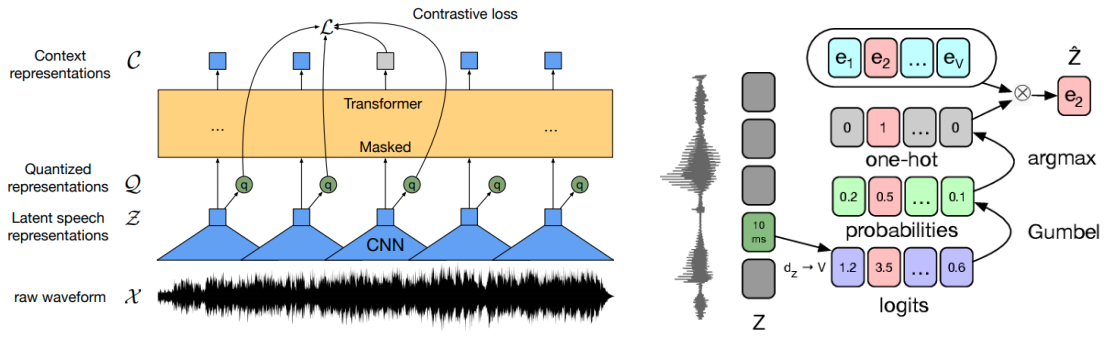
$$p(\pi|x) = \prod_{t=1}^{T} y_{\pi_t}. \tag{138}$$

Jos ajallisen luokittelumetodin konvoluutioneuroverkko on myös rekursiivinen, sitä sanotaan yhdistelmälliseksi ajalliseksi luokitteluksi (Connectionist temporal classification, CTC). Yhdistelmällisessä ajallisessa luokittelussa sekvenssistä poistetaan todennäköisimmin tyhjät ikkunat ja peräkkäiset todennäköiset foneemit eli muodostetaan uusi kuvaus $B : L'^T \rightarrow L^{\leq T}$ joka palauttaa uudet todennäköisimmät foneemit l_i ehdollisella todennäköisyydellä [43]

$$p(l|x) = \sum_{\pi \in B^{-1}(l)} p(\pi|x) = \sum_{\pi \in B^{-1}(l)} \prod_{t=1}^{T} y_{\pi_t}. \tag{139}$$

Yhdistelmällinen ajallinen luokittelu käyttää hukkafunktionaan nimitysvirhetasoa, joka vertaa sekvenssipareja $(x,f) \in S$ joukossa $S \subset \mathbb{D}_{X \times F}$ jossa f_i ovat mallifoneemit ja laskee niiden liitteessä J esitellyn normalisoidun Levenshtein-etäisyyden

$$\mathbb{L}_{LER}(h, S) = \frac{1}{N} \sum_{(x,f) \in S} \mathrm{ED}(h[x], f). \tag{140}$$



(a) Wav2vec2-mallin rakenne. Kuva: [33]

(b) Kvantitusmoduulin rakenne. Kuva: [44]

Kuvio 8. Raa'an äänisignaalin puhe-esityksiksi muuntava wav2vec2-malli koostuu kuvan (a) mukaisesti kolmesta osasta: konvoluutioneuroverkosta, kvantitusmoduulista ja naamioidusta muuntajaneuroverkosta. Kuvassa (b) esitelty kvantitusmoduuli ottaa konvoluutioneuroverkon tuottaman latentin puhe-esityksen ja yhdistää sen koodikirjan g alkioon e luoden kandidaattiesityksiä. Kvantitettujen esitysten ja muuntajamallin kontekstualisoitujen esitysten avulla voidaan käyttää kontrastiivista hukkafunktiota mallin kouluttamiseen.

Puhe-esityksien koodaamiseen on kehitetty myös konvoluutio- ja muuntajaneuroverkkoihin perustuva wav2vec2, jonka arkkitehtuuri esitellään kuviossa 8. Malli ottaa syötteekseen äänisignaalin $x \in X$, jonka se ajaa konvoluutioneuroverkon läpi muodostaen foneemiesitykset $z \in Z$. Tämän jälkeen wav2vec2-mallin kvantitusmoduulissa luodaan ensin koodikirjoja (Codebooks) g joissa on V satunnaista alkiota eli lineaariprojektiota $e \in \mathbb{R}^{(V \times d)/G}$ ja joita on G kappaletta. Foneemiesitykset z lineaariprojisoidaan logiteiksi $l \in \mathbb{R}^{G \times V}$ jolloin ne voidaan asettaa liitteessä I esiteltyyn Gumbel-softmax-funktioon joka palauttaa jatkuvan todennäköisyysjakauman. Gumbel-softmaxin jälkeen todennäköisyysjakauma diskretoidaan ottamalla argmax-funktiolla todennäköisimmät foneemiesitykset, jotka lopulta lineaariprojisoidaan koodikirjoihin e tuottaen kandidaattiesitykset q_i. Gumbel-softmax kertoo todennäköisyyden saada koodikirjan g alkio e_v

$$p_{g,v} = \frac{\exp\left((l_{g,v} - \log\left(-\log\left(u_v\right)\right)\right)/\tau\right)}{\sum_{j=1}^{K} \exp\left((l_{g,v} - \log\left(-\log\left(u_v\right)\right)\right)/\tau\right)}, \qquad p : \mathbb{R}^{G \times V} \rightarrow [0,1]^V. \tag{141}$$

Eteenpäin kulkiessa kvantitusmoduuli antaa diskreetit arvot q, mutta kvantitusmoduuli voidaan edelleen optimoida käänteisetenemisellä derivoituvan todennäköisyysjakauman ansiosta. [44]

Puhe-esitykset z viedään myös muuntajaneuroverkkoon, jossa ne kontekstualisoidaan alaluvussa 5.3 kuvaillulla tavalla. Ainoina eroina on huomiomekanismin naamioiminen ja paikkaesityksen vaihto konvoluutioneuroverkkoon, eli sinimuotoinen absoluuttinen paikkaesitys (135, 136) korvataan konvoluutioneuroverkolla joka koodaa puhe-esityksiin suhteellisen paikkaesityksen [45]. Muuntajaneuroverkko kuvaa puhe-esitykset z kontekstualisoiduiksi esityksiksi c.

Malli antaa siis tähän mennessä kandidaattiesitykset q ja kontekstualisoidut esitykset c. Wav2vec2 on itseohjautuvasti oppiva malli ja se esikouluttautuu hukkafunktion

$$\mathbb{L} = \mathbb{L}_m + \alpha \mathbb{L}_d \tag{142}$$

avulla, jossa $\mathbb{L}_m$ on kontrastiivinen hukkafunktio (231), $\mathbb{L}_d$ on diversiteettihukkafunktio (232) liitteen J mukaisesti ja α on säädettävä parametri.

$\mathbb{L}_m$ on hukkafunktio naamioidulle ajanhetkelle t jossa kontekstualisoidun esityksen c_t tulee tunnistaa kandidaattiesitys q_t, kandidaattiesitysjoukosta $\tilde{q} \approx Q_t$ jossa muut esitykset kuin q_t ovat naamioituja esityksiä muista ajahetkistä. Tällöin kontrastiivinen hukkafunktio on siis

$$\mathbb{L}_m = -\log \frac{\exp(\mathrm{sim}(c_t, q_t)/\kappa)}{\sum_{\tilde{q}\tilde{Q}_t} \exp(\mathrm{sim}(c_t, \tilde{q})/\kappa)}. \tag{143}$$

Diversiteettihukkafunktio $\mathbb{L}_d$ on mukana kasvattaakseen kandidaattiesityksien merkitystä hukkafunktiossa ja sen tehtävänä on maksimoida jatkuvan todennäköisyysjakauman $p_{g,v}$ keskiarvotettu entropia kunkin koodikirjan g ja koodikirja-alkion v suhteen

$$\mathbb{L}_d = \frac{1}{GV} \sum_{g=1}^{G} -S(p_g) = \frac{1}{GV} \sum_{g=1}^{G} \sum_{k=1}^{K} p_{g,v} \log\left(p_{g,v}\right). \tag{144}$$

Esikoulutuksen jälkeen wav2vec2-hienosäätö aloitetaan asettamalla satunnaisesti alustettu lineaarikerros kontekstiverkon päälle ja jaetaan puhe-esitykset sanastoon C. Hienosäätö tapahtuu varsinaisesti käyttämällä yhdistelmällisen ajallisen luokittelun ehdollista todennäköisyyttä (139) ja optimoimalla sen negatiivisten logaritmien summa minimiin koulutusjoukossa $\mathbb{D}$

$$\mathbb{L}_{CTC} = \sum_{x,l \in \mathbb{D}} -\log(p(l|x)) = \sum_{x,l \in \mathbb{D}} -\log(\sum_{\pi \in B^{-1}(l)} \prod_{t=1}^{T} y_{\pi_t}). \tag{145}$$

Koska wav2vec2 on kehitetty vain englannin kielellä, siitä on kehitetty myös moni-

kielinen variantti wav2vec2 XLS-R [46] joka on esikoulutettu 128 eri kielellä suomi mukaanlukien ja 436000 tunnilla monikielistä puhedataa. Mallin esikoulutus vastaa wav2vec2 esikoulutusta, mutta koulutuspuhedataa näytteistetään todennäköisyysjakauman $(p_l)_{l=1,\dots,L}$ mukaan jolla $p_l = a_l(n_l/N)$ ja jossa l merkkaa tiettyä kieltä, L on kielten määrä, n_l on on kielen l koulutusdata tunteina, N kokonaiskoulutusdata tunteina ja a_l on kunkin kielen kerroin jolla voidaan muuttaa tietyn kielen merkittävyyttä esikoulutuksen aikana. [47]

Puheentunnistusmallien tarkkuutta mitataan usein sanavirhetasolla (word error rate, WER) joka voidaan laskea naiivilla Levenshtein-etäisyydellä (237) liitteen J mukaisesti eli

$$\mathbb{L}_{WER} = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}, \tag{146}$$

jossa S on vaihdettujen sanojen määrä, D on poistettujen sanojen määrä, I on insertoitujen sanojen määrä, C on oikeiden sanojen määrä ja N on kaikkien käsiteltyjen sanojen määrä. Koska suomi on taipuva ja agglutinatiivinen kieli, on analyyttisillä kielillä kehitellyillä puheentunnistusjärjestelmillä usein verrattain korkea sanavirhetaso. Tämän takia suomen kielen puheentunnistuksessa otetaan huomioon myös kirjainvirhetaso (character error rate, CER) joka määritetään samalla tavalla kuin sanavirhetaso eli

$$\mathbb{L}_{CER} = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}, \tag{147}$$

jossa S on vaihdettujen kirjainten määrä, D on poistettujen kirjainten määrä, I on insertoitujen kirjainten määrä, C on oikeiden kirjainten määrä ja N on kaikkien käsiteltyjen kirjainten määrä.

Suomen kielelle XLS-R-malli on hienosäädetty Tanskasen ja Toivasen mallissa wav2vec2-xlsr-1b-finnish-lm-v2, joka on saavuttanut 275,6 tunnin suomenkielisellä hienosäätödatalla kolmelle testidatasetille (Common Voice 7.0, Common Voice 9.0, FLEURS) sanavirhetasot 9,73% (CV 7.0), 8,96% (CV 9.0) ja 14,89% (FLEURS) ja kirjainvirhetasot 1,65% (CV 7.0), 1,52% (CV 9.0) ja 6,06% (FLEURS). Lisäämällä mallin dekooderiin suomen kielellä koulutetun n-gram-pohjaisen kenLM-kielimallin [48] sanavirhetasot laskevat lukuihin 4,09% (CV 7.0), 3,72% (CV 9.0) ja 12,11% (FLEURS) ja kirjainvirhetasot lukuihin 0,88% (CV 7.0), 0,80% (CV 9.0) ja 5,65% (FLEURS). [34]

6 Päätäntö

Tässä tutkielmassa on luotu katsaus puheen fysikaalisiin perusteisiin ja automaattisen puheentunnistuksen periaatteisiin.

Tutkielman luvussa 2 annettiin puheelle akustinen pohja kuvaamalla pitkittäisiä aaltoja matemaattisesti ja yhdistämällä kuvaus fysikaaliseen maailmaan antamalla aalloille väliaineeksi kolmiulotteinen kaasu. Äänen todettiin olevan herätteen aiheuttama aalto, jota voidaan kuvata väliaineen hiukkasten siirtymällä, siirtymän aikaderivaatalla eli siirtymänopeudella tai siirtymän paikkaderivaatalla eli paineella. Ääni leviää pistemäisestä herätteestä pallomaisesti, mutta sitä voidaan kohdella tasoaaltona kaukana herätteestä ja eri geometrioissa, kuten putkissa. Koska ääniaallot heijastuvat impedanssieron rajapinnalla, erojen rajaamaan tilaan muodostuu seisovien aaltojen äänikenttä joka vaimenee ajassa kentän energiatiheystason aikaderivaatan mukaisesti.

Luvun 3 alussa esiteltiin yleinen metodi kaasussa etenevien aaltojen muuntamiseksi sähköisiksi diskreeteiksi signaaleiksi ja jälkimmäisellä puoliskolla tutkittiin kuinka signaaleja voidaan lyhytaikakäsitellä ja esittää spektrogrammeina. Ääniaalto voidaan siirtää sähköiseen muotoon mikrofonilla ja sähköinen signaali puolestaan voidaan diskretisoida A/D-muuntajalla digitaaliseen muotoon. Diskreettiä signaalia voidaan käsitellä tietokoneella, joka mahdollistaa signaalin reaaliaikaisen lyhytaikakäsittelyn ja Fourier-muuntamisen jolloin puheesta voidaan muodostaa puheominaisuuksia korostava Mel-taajuuksinen spektrogrammi.

Puheentuotantoa eli äänihuulien tuottamien ääniaaltojen modulointia ääntöväylässä tarkasteltiin luvussa 4 foneettisesti ja luvun lopuksi käytiin läpi yksittäisten foneemien laskennallisia tunnistamismetodeja. Puhetta voidaan tutkielman perusteella kutsua tavaksi koodata ääniaaltoihin informaatiota moduloimalla ääniaaltoja ääntöväylässä. Lähde-suodin-malli erottaa Bernoullin lain mukaisesti tuotetun lähdesignaalin ja ääntöväylän artikulatoristen elinten muodostaman suotimen. Ääniaaltoon koodatut foneemit voidaan tunnistaa niiden ylä-äänisarjojen ja artikulatoristen ominaisuuksien perusteella laskennallisesti formanttiavaruuden ja spektrogrammien avulla.

Luvun 5 aluksi luotiin pohjatiedot puheentunnistamista varten tutustumalla neuroverkkoihin ja koneoppimiseen, joiden pohjalta kyettiin käymään läpi neuroverkoilla muodostettuja kielimalleja ja akustisia malleja, joihin automaattinen puheentunnistus perustuu. Koska tietokone ei ymmärrä puhetta, automaattinen puheentunnistus perustuu puheen ja sanojen numeerisiin puhe- ja sanaesityksiin. Sanaesitykset ovat sanaominaisuusavaruudessa sijaitsevia vektoreita, jotka klusteroituvat avaruudessa sekä semanttisten että syntaktisten ominaisuuksiensa mukaisesti. Puhe-esitykset ovat samankaltaisesti puheominaisuusavaruudessa sijaitsevia vektoreita, joita voidaan muodostaa foneemiesityksillä tai abstraktimmin huomiomekanismilla. Foneemiesitykset perustuvat ikkunoiduista spektrogrammeista saataviin puheominaisuuksiin kuten Mel-taajuuksisiin kepstrikertoimiin, kun taas wav2vec2-mallin tuottamat puheesitykset muodostuvat puhe-esitysten kontekstista. Sanaesityksiä vastaavien sanojen sekvenssin todennäköisyysjakaumaa kutsutaan kielimalliksi ja puhe-esityksiä vastaavien foneemien tai sanojen sekvenssin todennäköisyysjakaumaa kutsutaan akustiseksi malliksi. Malleja voidaan optimoida neuroverkoilla erinäisiin tehtäviin koneoppimisen keinoin. Yhdistämällä akustinen malli ja kielimalli voidaan tuottaa puheesta tekstiksi -tyylinen automaattinen puheentunnistusmalli.

Tutkielman lopussa tutustuttiin suomenkieliseen puheentunnistusmalliin wav2vec2-xls-r-1b, joka koostuu puhe-esityksiä tuottavasta Wav2Vec2 XLS-R -mallista ja todennäköisimpiä sanasekvenssejä muodostavasta KenLM-kielimallista. Mallin tekijät loivat mallin tekoäly-yhteisö HuggingFacen tapahtumassa Robust Speech Event, jossa koulutettiin puheentunnistusmalleja eri kielille. Alaluvussa 5.4 esitellyt sanavirhetasot ja kirjainvirhetasot ovat mallin HuggingFace-sivulta ja ne ovat toisinnettavissa mallin ja datasettien ollessa julkisesti saatavia.

Akustiikka ja fonetiikka ovat vakiintuneita tieteitä, joita kehitetään edelleen tarkempien simulaatioiden ja mallien avulla. Automaattinen puheentunnistus on puolestaan jatkuvasti muuttuva ala, joka on edennyt harppauksin viimeaikaisten tekoälykehitysten seurauksena. Kehitys on ollut nopeinta kielimallien parissa, joskin myös akustiset mallit ovat hyötyneet kielimallien innovaatioista, kuten huomiomekanismista. Suomenkielinen automaattinen puheentunnistus on hyvällä tasolla, josta tämä kappale on puhumanani esimerkkinä. Puheentunnistumallin tunnistama puhe löytyy liitteestä K, jossa esiintyvät oikeinkirjoitusvirheet ovat korjattavissa yhdistämällä puheentunnistusmalliin kehittyneempi kielimalli.

Lähteet

[1] I. D. Rowland, T. N. Howe ym. *Vitruvius:'Ten books on architecture'*. Cambridge University Press, 2001.

[2] M. Mersenne. *Harmonie universelle: contenant la théorie et la pratique de la musique*. Paris: Sebastien Cramoisy, 1636, Viitattu: 17.8.2023. URL: https://archive.org/details/imslp-universelle-mersenne-marin/PMLP156089-MersenneM_HarmUniv_Pt1_01/mode/2up.

[3] I. Newton. *The Principia: mathematical principles of natural philosophy*. Univ of California Press, 1999.

[4] C. B. Boyer ja U. C. Merzbach. *A history of mathematics*. John Wiley & Sons, 2011.

[5] A. Vaswani ym. "Attention is all you need". *Advances in neural information processing systems* 30 (2017).

[6] S. T. Thornton ja J. B. Marion. *Classical dynamics of particles and systems*. Cengage Learning, 2021.

[7] P. Ceperley. *Three types of waves: traveling waves, standing waves, and rotating waves*. 2016, Viitattu: 17.8.2023. URL: http://resonanceswavesandfields.blogspot.com/2016/03/three-types-of-waves-traveling-waves.html.

[8] L. E. Reichl. *A modern course in statistical physics*. 1999.

[9] F. M. White. *Fluid mechanics*. New York, 1990.

[10] D. J. Griffiths ja D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

[11] P. Filippi ym. *Acoustics: basic physics, theory, and methods*. Elsevier, 1998.

[12] H. Kuttruff. *Room acoustics*. Crc Press, 2016.

[13] C. E. Shannon. "A mathematical theory of communication". *The Bell system technical journal* 27.3 (1948), s. 379–423.

[14] R. C. Weast. *CRC handbook of chemistry and physics.* CRC Press Inc., Boca Raton, FL. 1986.

[15] I. Müller. "The coldness, a universal function in thermoelastic bodies". *Archive for Rational Mechanics and Analysis* 41 (1971), s. 319–332.

[16] R. D. Knight. *Physics for scientists and engineers.* Pearson Education, Limited, 2022.

[17] S. K. Mitra. *Digital signal processing: a computer-based approach.* McGraw-Hill Higher Education, 2001.

[18] T. Bäckström ym. *Introduction to Speech Processing.* 2. painos. 2022, Viitattu: 17.8.2023. DOI: 10.5281/zenodo.6821775. URL: https://speechprocessingbook.aalto.fi.

[19] Banco. *Profile of a dynamic microphone.* 2005, Viitattu: 17.8.2023. URL: https://commons.wikimedia.org/wiki/File:Mic-dynamic.PNG.

[20] U. Bharathi. *How do Analog to Digital Converters (ADCs) work?* 2022, Viitattu: 17.8.2023. URL: https://www.circuitbread.com/ee-faq/how-do-analog-to-digital-converters-adcs-work.

[21] L. Rabiner ja R. Schafer. *Theory and applications of digital speech processing.* Prentice Hall Press, 2010.

[22] G. Fant. *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations.* 2. Walter de Gruyter, 1971.

[23] M.-c. Park. "Understanding the multi-mass model and sound generation of vocal fold oscillation". *AIP Advances* 9.10 (2019), s. 105002.

[24] K. Ojutkangas ym. *Johdatus kielitieteeseen.* WSOY oppimateriaalit, 2009.

[25] J. Koreman. "THE EFFECTS OF STRESS AND F 0 ON THE VOICE SOURCE". *Phonus* 1 (elokuu 2023), s. 105–120.

[26] I. Titze, T. Riede ja T. Mau. "Predicting achievable fundamental frequency ranges in vocalization across species". *PLoS computational biology* 12.6 (2016), e1004907.

[27] D. Houkema. "The coherent perception of speech within Cognitive Science". Väitöskirja. University of Groningen, 2001.

[28] M. Wieling, E. Margaretha ja J. Nerbonne. ”Inducing phonetic distances from dialect variation”. *Computational Linguistics in the Netherlands Journal* 1 (joulukuu 2011), s. 109–118.

[29] P. Ladefoged. ”Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech” (1967).

[30] S. R. M. Tarek. ”Analyzing the Stress Pattern of Bangla Songs through Metrical Phonology”. Väitöskirja. University of Dhaka, 2019.

[31] A. A. Ali, J. Van der Spiegel ja P. Mueller. ”An acoustic-phonetic feature-based system for the automatic recognition of fricative consonants”. Teoksessa: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*. Vol. 2. IEEE. 1998, s. 961–964.

[32] G. Korvel ja B. Kostek. ”Voiceless stop consonant modelling and synthesis framework based on MISO dynamic system”. *Archives of Acoustics* 42 (2017), s. 375–383.

[33] A. Baevski ym. ”wav2vec 2.0: A framework for self-supervised learning of speech representations”. *Advances in neural information processing systems* 33 (2020), s. 12449–12460.

[34] R. Tanskanen A & Toivanen. *Wav2vec2-xls-r-1b for Finnish ASR*. 2016, Viitattu: 17.8.2023. URL: `https://huggingface.co/Finnish-NLP/wav2vec2-xlsr-1b-finnish-lm-v2`.

[35] J. D. Kelleher. *Deep learning*. MIT press, 2019.

[36] A. Baevski, M. Auli ja A. Mohamed. ”Effectiveness of self-supervised pre-training for speech recognition”. *arXiv preprint arXiv:1911.03912* (2019).

[37] D. E. Rumelhart, G. E. Hinton ja R. J. Williams. ”Learning representations by back-propagating errors”. *Nature* 323.6088 (1986), s. 533–536.

[38] S. Linnainmaa. *Algoritmin kumulatiivinen pyöristysvirhe yksittäisten pyöristysvirheiden Taylor-kehitelmänä*. Pro Gradu. University of Helsinki, 1970.

[39] S. Ruder. ”An overview of gradient descent optimization algorithms”. *arXiv preprint arXiv:1609.04747* (2016).

[40] S. Bengio ja G. Heigold. ”Word Embeddings for Speech Recognition”. Teoksessa: *Proceedings of the 15th Conference of the International Speech Communication Association, Interspeech*. 2014.

[41] T. Mikolov ym. ”Efficient estimation of word representations in vector space”. *arXiv preprint arXiv:1301.3781* (2013).

[42] J. Jurafsky D & Martin. *Speech & language processing, 3rd ed.* 3. painos. Jan 7, 2023 draft, Viitattu: 17.8.2023. URL: `https://web.stanford.edu/~jurafsky/slp3/`.

[43] A. Graves ym. ”Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks”. Teoksessa: *Proceedings of the 23rd international conference on Machine learning*. 2006, s. 369–376.

[44] A. Baevski, S. Schneider ja M. Auli. ”vq-wav2vec: Self-supervised learning of discrete speech representations”. *arXiv preprint arXiv:1910.05453* (2019).

[45] A. Mohamed, D. Okhonko ja L. Zettlemoyer. ”Transformers with convolutional context for asr”. *arXiv preprint arXiv:1904.11660* (2019).

[46] A. Babu ym. ”XLS-R: Self-supervised cross-lingual speech representation learning at scale”. *arXiv preprint arXiv:2111.09296* (2021).

[47] Y. Getman. *End-to-End Low-Resource Automatic Speech Recognition for Second Language Learners*. Diplomityö. Aalto University, 2021.

[48] K. Heafield. ”KenLM: Faster and smaller language model queries”. Teoksessa: *Proceedings of the sixth workshop on statistical machine translation*. 2011, s. 187–197.

[49] D. Hendrycks ja K. Gimpel. ”Gaussian error linear units (gelus)”. *arXiv preprint arXiv:1606.08415* (2016).

[50] Y. Wu ja K. He. ”Group normalization”. Teoksessa: *Proceedings of the European conference on computer vision (ECCV)*. 2018, s. 3–19.

[51] J. L. Ba, J. R. Kiros ja G. E. Hinton. ”Layer normalization”. *arXiv preprint arXiv:1607.06450* (2016).

[52] E. Ukkonen. ”On approximate string matching”. Teoksessa: *International Conference on Fundamentals of Computation Theory*. Springer. 1983, s. 487–495.

[53] V. I. Levenshtein ym. ”Binary codes capable of correcting deletions, insertions, and reversals”. Teoksessa: *Soviet physics doklady*. Vol. 10. 8. Soviet Union. 1966, s. 707–710.

[54] T. von Neumann ym. ”On word error rate definitions and their efficient computation for multi-speaker speech recognition systems”. Teoksessa: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, s. 1–5.

[55] A. Ali ja S. Renals. ”Word error rate estimation for speech recognition: e-WER”. Teoksessa: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2018, s. 20–24.

[56] D. A. Fleisch. *A student's guide to vectors and tensors*. Cambridge University Press, 2011.

A Liike-energia ja potentiaalienergia

Olkoon järjestelmä lepotilassa eli vakaassa tasapainotilassa kun

$$q_j = q_{j0}, \quad \dot{q}_j = 0, \quad \ddot{q}_j = 0, \quad j = 1,2,3,...,n$$

ja valitaan $q_{j0} = 0$. Liike-energia on yleisissä koordinaateissa [6]

$$T = \frac{1}{2}\sum_{j,k}^{n} m_{jk}\dot{q}_j\dot{q}_k, \tag{148}$$

jossa m_{jk} on järjestelmän massamatriisi ja vapausasteiden määrät $j = 1,2,3,...,n$ ja $k = 1,2,3,..,n$ ovat sen sarake- ja rivilukunumerot. Potentiaalienergian U selvittämiseksi tulee tutkia Lagrangen yhtälöä (2), jonka paikkaderivaatta on tasapainopisteissä

$$\frac{\partial \mathcal{L}(q_{j0})}{\partial q_j} = \frac{\partial \mathcal{L}(0)}{\partial q_j} = \frac{\partial T(0)}{\partial q_j} - \frac{\partial U(0)}{\partial q_j} = 0, \quad j = 1,2,...,n \tag{149}$$

ja koska

$$\frac{\partial T(0)}{\partial q_k} = 0, \quad k = 1,2,...,n$$

voidaan todeta että

$$\frac{\partial U(0)}{\partial q_k} = 0, \quad k = 1,2,...,n.$$

Nyt potentiaalienergiaa voidaan approksimoida tasapainopisteen $U(q_{k0}) = 0$ lähellä Taylorin sarjalla:

$$U(q_1,q_2,...,q_n) = U_0 + \sum_{k}^{n} \frac{\partial U(0)}{\partial q_k} q_k + \frac{1}{2}\sum_{j,k}^{n} \frac{\partial^2 U(0)}{\partial q_j \partial q_k} q_j q_k + ... \tag{150}$$

Voidaan todeta että alkutermit $U_0 = U(q_{k0}) = 0$ ja $\sum_k^n \frac{\partial U(0)}{\partial q_k} = 0$. Potentiaalienergian tarkempi määrittely edellyttää fyysisiä rajoitteita: käsittelemme harmonista värähtelyä jossa esiintyy enintään toisen asteen termejä, joten sarjan laajennuksesta voidaan myös jättää huomiotta toisen asteen derivaattaa korkeammat termit. Tällöin

Taylorin sarja typistyy muotoon

$$U = \frac{1}{2} \sum_{j,k}^{n} \frac{\partial^2 U(0)}{\partial q_j \partial q_k} q_j q_k,$$

joka voidaan kirjoittaa muodossa

$$U = \frac{1}{2} \sum_{j,k}^{n} C_{jk} q_j q_k. \tag{151}$$

B Ominaisvektorit ja normaalikoordinaatit

Ominaisarvo-ongelmasta voidaan ratkaista myös ominaisvektorit A_j, jotka muodostavat ortonormaalin joukon. Sekulaarin yhtälön (10) s:lle ja r:lle juurelle voimme kirjoittaa yhtälön eri indeksein

$$\begin{cases} \omega_s^2 \sum_j^n m_{jk} A_{js} = \sum_j^n \kappa_{jk} A_{js} \\ \omega_r^2 \sum_k^n m_{jk} A_{kr} = \sum_k^n \kappa_{jk} A_{kr} \end{cases} . \tag{152}$$

Jos kerromme j-indeksin yhtälöä tekijällä A_{kr} ja k-indeksin yhtälöä tekijällä A_{js}, saamme yhtälöparin oikeanpuoleiset termit yhtäsuuriksi

$$\begin{cases} \omega_s^2 \sum_{j,k}^n m_{jk} A_{js} A_{kr} = \sum_{j,k}^n \kappa_{jk} A_{js} A_{kr} \\ \omega_r^2 \sum_{j,k}^n m_{jk} A_{js} A_{kr} = \sum_{j,k}^n \kappa_{jk} A_{js} A_{kr} \end{cases} \tag{153}$$

ja voimme kirjoittaa yhtälöparin muodossa

$$(\omega_s - \omega_r) \sum_{j,k}^{n} m_{jk} A_{js} A_{kr} = 0 \tag{154}$$

joka toteutuu kun $s = r$ jolloin $\omega_s - \omega_r = 0$. Tilanteessa $s \neq r$ termi ei välttämättä häviä, joten muutoin on siis oltava

$$\sum_{j,k}^{n} m_{jk} A_{js} A_{kr} = 0, \quad s \neq r. \tag{155}$$

Tilanteessa $s = r$ summan arvoa ei kuitenkaan tunneta. Jos sijoitamme ratkaisun reaalisen osan liike-energian yhtälöön saamme identtisen summan

$$T = \frac{1}{2} \sum_{j,k}^{n} m_{jk} \dot{q}_j \dot{q}_k$$

$$= \frac{1}{2} \sum_{j,k}^{n} m_{jk} \Big(\sum_s \omega_s A_{js} (\sin(\omega_s t - \delta_s)) \Big) \Big(\sum_r \omega_r A_{kr} (\sin(\omega_r t - \delta_r)) \Big)$$

eli kun $s = r$, voidaan liike-energia kirjoittaa muodossa

$$T = \frac{1}{2} \sum_{s,r}^{n} \omega_s \omega_r \sin^2 (\omega_r t - \delta) \sum_{j,k}^{n} m_{jk} A_{js} A_{ks} \tag{156}$$

jossa

$$\omega_s \omega_r \sin^2 (\omega_r t - \delta) \geq 0.$$

Koska liike-energia on aina positiivinen suure ja menee nollaan vain lepotilassa, voidaan todeta että

$$\sum_{j,k}^{n} m_{jk} A_{js} A_{js} \geq 0$$

joka voidaan normittaa määräämällä summa unitaariseksi

$$\sum_{j,k}^{n} m_{jk} A_{js} A_{js} = 1 \tag{157}$$

jolloin yhtälöiden (155) ja (157) perusteella nähdään summatermin muodostavan Kroneckerin deltan

$$\sum_{j,k}^{n} m_{jk} A_{js} A_{kr} = \delta_{sr}. \tag{158}$$

Tällöin A_{js} on s:nen ominaisvektorin j:nes alkio ja vektorit $\mathbf{A}_r$ voidaan kirjoittaa identiteettivektorien $\mathbf{e}$ avulla

$$\mathbf{A}_r = \sum_{j}^{n} A_{jr} \mathbf{e}_j, \tag{159}$$

jolloin voidaan todeta että vektorit $\mathbf{A}_r$ ovat keskenään ortogonaalisia ja normitettuja, eli ne muodostavat ortonormitetun joukon. Ominaisvektorien epäfyysisen normituksen myötä tulee yleisille koordinaateille määritellä uudet kompleksiset skaalausvakiot α_r ja β_r joilla

$$q(t) = \sum_{r} \alpha_r A_{jr} e^{i(\omega_r t - \delta_r)}, \quad q(t) = \sum_{r} \beta_r A_{jr} e^{i(\omega_r t)}. \tag{160}$$

Nyt voimme määritellä järjestelmälle normaalikoordinaatin

$$\eta_r(t) = \beta_r e^{i\omega_r t} \tag{161}$$

jolloin yleinen koordinaatti on

$$q_j(t) = \sum_{r} A_{js} \eta_r(t). \tag{162}$$

Normaalikoordinaatit ovat harmonista värähtelyä yhdellä taajuudella edustavia suureita, jotka ratkaisevat yhtälön (7) muotoa olevia yhtälöitä eli

$$\ddot{\eta} + \omega_r^2 \eta = 0. \tag{163}$$

C Kylmyystekijän johto mikrotilojen avulla

Jos järjestelmässä on kahden energian omaavia molekyylejä ja niiden kokonaisenergia on summa $E = E_1 + E_2$, järjestelmän mikrotilojen määrä voidaan esittää muodossa

$$\Omega = \Omega_1(E_1)\Omega_2(E_2) = \Omega_1(E_1)\Omega_2(E - E_1) \tag{164}$$

joka saavuttaa tasapainotilan kun

$$\frac{\partial \Omega}{\partial E_1} = \Omega_2(E_2)\frac{\partial \Omega_1(E_1)}{\partial E_1} + \Omega_1(E_1)\frac{\partial \Omega_2(E_2)}{\partial E_2}\frac{\mathrm{d}E_2}{\mathrm{d}E_1} = 0 \tag{165}$$

jossa

$$\frac{\mathrm{d}E_2}{\mathrm{d}E_1} = -1 \tag{166}$$

koska $E = E_1 + E_2$. Tällöin saadaan

$$\Omega_2(E_2)\frac{\partial \Omega_1}{\partial E_1} - \Omega_1(E_1)\frac{\partial \Omega_2}{\partial E_2} = \frac{\partial \Omega_1}{\partial E_1}\frac{1}{\Omega_1} - \frac{\partial \Omega_2}{\partial E_2}\frac{1}{\Omega_2} = 0 \tag{167}$$

eli

$$\frac{\partial \ln \Omega_1}{\partial E_1} = \frac{\partial \ln \Omega_2}{\partial E_2} \tag{168}$$

jonka yleistettyä muotoa kutsutaan kylmyystekijäksi [15]

$$\beta = \frac{\partial \ln \Omega}{\partial E}. \tag{169}$$

D Boltzmannin jakauman johto

Entropian maksimi voidaan etsiä optimoimalla Lagrangen kertoimien avulla, sillä järjestelmällä on kaksi rajoitetta: todennäköisyyksien tulee todennäköisyysteorian toisen aksiooman mukaan ynnäytyä yhteen

$$\sum_i p_i = 1 \tag{170}$$

ja energian tulee säilyä energian säilymislain mukaisesti eli energian odotusarvon tulee olla järjestelmän sisäenergia

$$U = \sum_i p_i E_i = \langle E \rangle. \tag{171}$$

Nyt siis Lagrangen funktio on

$$\mathcal{L} = S + \lambda_1(1 - \sum_i p_i) + \lambda_2(U - \sum_i p_i E_i) \tag{172}$$

joka Euler-Lagrangen yhtälössä derivoidaan aikariippumattomuudesta johtuen vain todennäköisyyden suhteen

$$\frac{\partial \mathcal{L}}{\partial p_i} = -k_B \sum_i \ln p_i - k_B - \lambda_1 - \lambda_2 E_i = 0 \tag{173}$$

josta saadaan maksimoivan jakauman muoto

$$p_i = e^{\frac{-k_B - \lambda_1 - \lambda_2 E_i}{k_B}} = a_1 e^{-a_2 E_i}. \tag{174}$$

Ensimmäinen a_1 voidaan ratkaista todennäköisyysrajoitteella

$$\sum_i p_i = a_1 \sum_i e^{-a_2 E_i} = 1 \iff a_1 = \frac{1}{\sum_i e^{-a_2 E_i}} \tag{175}$$

ja a_2 saadaan Gibbsin entropiasta

$$S = -k_B \sum_i p_i \ln p_i = -k_B \sum_i p_i (\ln a_1 - a_2 E_i)$$

$$= -k_B \ln a_1 + k_B a_2 \sum_i p_i E_i = -k_B \ln a_1 + k_B a_2 U$$

jolloin lämpötilan määritelmän (38) mukaan

$$\left(\frac{\partial S}{\partial U}\right) = k_B a_2 = \frac{1}{T} \iff a_2 = \frac{1}{k_B T} = \beta. \tag{176}$$

Sijoittamalla tulokset yhtälöön (174) saamme entropian maksimoivaksi jakaumaksi Boltzmannin jakauman

$$p_i = \frac{1}{\sum_i e^{-\beta E_i}} e^{\beta E_i} = \frac{1}{Z} e^{\beta E_i}, \tag{177}$$

jossa Z on jakauman partitiofunktio.

E Virtausmekaanisen liikemääräyhtälön johto

Jos virtaava suure on liikemäärä $p = m\mathbf{V}$, on Reynoldisn jatkuvuuslauseen yhtälö

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_{\mathrm{syst}} = \sum \mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathrm{CV}} \mathbf{V}\rho d\mathbb{V}\right) + \int_{\mathrm{CS}} \mathbf{V}\rho(V\cdot\mathbf{n})\mathrm{d}A \tag{178}$$

ja jos virtauspinnat ovat yksiulotteisia

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_{\mathrm{syst}} = \sum \mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathrm{CV}} \mathbf{V}\rho\mathrm{d}\mathbb{V}\right) + \sum(\dot{m}_i\mathbf{V}_i)_{out} - \sum(\dot{m}_i\mathbf{V}_i)_{in}. \tag{179}$$

Kun tilavuuselementti on hyvin pieni, tilavuusintegraalia voidaan approksimoida tavallisena derivaattana

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathrm{CV}}(\rho\mathbf{V})\mathrm{d}\mathbb{V}\right) \approx \frac{\mathrm{d}}{\mathrm{d}t}(\rho\mathbf{V})\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \tag{180}$$

jossa kunkin tilavuuselementtikuution sivua vastaa saapuva ja poistuva vuo

$$\begin{cases} \sum(\dot{m}_i\mathbf{V}_i)_{x,in} = \rho u\mathbf{V}\mathrm{d}y\mathrm{d}z, & \sum(\dot{m}_i\mathbf{V}_i)_{x,out} = (\rho u\mathbf{V} + \frac{\partial}{\partial x}(\rho u\mathbf{V})\mathrm{d}x)\mathrm{d}y\mathrm{d}z \\ \sum(\dot{m}_i\mathbf{V}_i)_{y,in} = \rho v\mathbf{V}\mathrm{d}x\mathrm{d}z, & \sum(\dot{m}_i\mathbf{V}_i)_{y,out} = (\rho v\mathbf{V} + \frac{\partial}{\partial y}(\rho v\mathbf{V})\mathrm{d}y)\mathrm{d}x\mathrm{d}z \\ \sum(\dot{m}_i\mathbf{V}_i)_{z,in} = \rho w\mathbf{V}\mathrm{d}x\mathrm{d}y, & \sum(\dot{m}_i\mathbf{V}_i)_{z,out} = (\rho w\mathbf{V} + \frac{\partial}{\partial z}(\rho w\mathbf{V})\mathrm{d}z)\mathrm{d}x\mathrm{d}y \end{cases} \tag{181}$$

joilla yhtälö (179) saadaan muotoon

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_{\mathrm{syst}} = \left(\frac{\partial}{\partial t}(\rho\mathbf{V}) + \frac{\partial}{\partial x}(\rho u\mathbf{V}) + \frac{\partial}{\partial y}(\rho v\mathbf{V}) + \frac{\partial}{\partial z}(\rho w\mathbf{V})\right)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z$$

$$= \left(\mathbf{V}\left(\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{V})\right) + \rho\left(\frac{\partial\mathbf{V}}{\partial x} + u\frac{\partial\mathbf{V}}{\partial x} + v\frac{\partial\mathbf{V}}{\partial y} + w\frac{\partial\mathbf{V}}{\partial z}\right)\right)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z. \tag{182}$$

Yhtälön ensimmäistä suluissa olevaa termiä kutsutaan jatkuvuusrelaatioksi, jonka tulee olla nolla massan säilymislain mukaisesti [6] [9]

$$\left(\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{V})\right) = 0, \tag{183}$$

jolloin yhtälö (182) voidaan kirjoittaa muodossa

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_{\mathrm{syst}} = \sum \mathbf{F} = \rho \frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z. \tag{184}$$

Kukin tilavuuselementti kokee painovoiman ja tilavuuselementin pinnalla jännityksen. Painovoiman differentiaali voidaan kirjoittaa muodossa

$$\mathrm{d}\mathbf{F}_G = -\rho\mathbf{g}\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z, \tag{185}$$

jossa $\mathbf{g} = g\hat{k}$ on painovoimakiihtyvyys.

Jännitys σ_{ij} ilmenee tilavuuselementtikuution sivujen pinnoilla ympäröivien elementtien paineena p ja nopeusgradienteista johtuvista viskositeettijännitteistä τ_{ij}. Jännitystensori voidaan kirjoittaa muodossa [9]

$$\sigma_{ij} = \begin{pmatrix} -P + \tau_{xx} & \tau_{yx} & \tau_{zx} \\ \tau_{xy} & -P + \tau_{yy} & \tau_{zy} \\ \tau_{xz} & \tau_{yz} & -P + \tau_{zz} \end{pmatrix}, \tag{186}$$

jolloin jännityksestä aiheutuvan voiman differentiaali on

$$\mathrm{d}\mathbf{F}_\sigma = \begin{cases} \left(\frac{\partial}{\partial x}\sigma_{xx} + \frac{\partial}{\partial y}\sigma_{yx} + \frac{\partial}{\partial z}\sigma_{zx}\right)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \\ \left(\frac{\partial}{\partial x}\sigma_{xy} + \frac{\partial}{\partial y}\sigma_{yy} + \frac{\partial}{\partial z}\sigma_{zy}\right)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \\ \left(\frac{\partial}{\partial x}\sigma_{xz} + \frac{\partial}{\partial y}\sigma_{yz} + \frac{\partial}{\partial z}\sigma_{zz}\right)\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z \end{cases} \tag{187}$$

josta saadaan

$$\left(\frac{\mathrm{d}\mathbf{F}}{\mathrm{d}\mathbb{V}}\right)_\sigma = \begin{cases} -\frac{\partial}{\partial x}P + \frac{\partial}{\partial x}\tau_{xx} + \frac{\partial}{\partial y}\tau_{yx} + \frac{\partial}{\partial z}\tau_{zx} \\ -\frac{\partial}{\partial y}P + \frac{\partial}{\partial x}\tau_{xy} + \frac{\partial}{\partial y}\tau_{yy} + \frac{\partial}{\partial z}\tau_{zy} \\ -\frac{\partial}{\partial x}P + \frac{\partial}{\partial x}\tau_{xz} + \frac{\partial}{\partial y}\tau_{yz} + \frac{\partial}{\partial z}\tau_{zz} \end{cases} \tag{188}$$

eli

$$\left(\frac{\mathrm{d}\mathbf{F}}{\mathrm{d}\mathbb{V}}\right)_\sigma = -\nabla P + \left(\frac{\mathrm{d}\mathbf{F}}{\mathrm{d}\mathbb{V}}\right)_\tau, \tag{189}$$

jossa

$$\left(\frac{\mathrm{d}\mathbf{F}}{\mathrm{d}\mathbb{V}}\right)_\tau = \left(\frac{\partial}{\partial x}\tau_{xx} + \frac{\partial}{\partial y}\tau_{yx} + \frac{\partial}{\partial z}\tau_{zx}\right)\hat{i} + \left(\frac{\partial}{\partial x}\tau_{xy} + \frac{\partial}{\partial y}\tau_{yy} + \frac{\partial}{\partial z}\tau_{zy}\right)\hat{j}$$

$$+\left(\frac{\partial}{\partial x}\tau_{xz}+\frac{\partial}{\partial y}\tau_{yz}+\frac{\partial}{\partial z}\tau_{zz}\right)\hat{k}=\nabla\cdot\vec{\tau}_{ij}. \tag{190}$$

Nyt voimme kirjoittaa virtaavan aineen infinitesimaalisen elementin differentiaalisen liikemääräyhtälön muodossa [9]

$$\rho\mathbf{g}-\nabla P+\nabla\cdot\vec{\tau}_{ij}=\rho\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}, \tag{191}$$

jossa ensimmäinen termi vastaa elementin kokema painovoimaa, toinen painetta, kolmas elementin viskositeettia ja yhtälön oikea puoli kertoo elementin tiheyden kerrottuna elementin kokonaiskiihtyvyydellä eli elementin kokeman kokonaisvoiman.

F Pallomaisen aaltoyhtälön separointi

Sijoittamalla (65) yhtälöön (64) ja jakamalla yhtälö tulolla $R(r) \cdot \Theta(\theta) \cdot \Phi(\phi) \cdot T(t)$, saadaan

$$\frac{1}{R}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{1}{\Theta}\frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right) + \frac{1}{\Phi}\frac{1}{r^2\sin^2(\theta)}\frac{\partial^2\Phi}{\partial\phi^2} = \frac{1}{T}\frac{1}{v^2}\frac{\partial^2 T}{\partial t^2} \tag{192}$$

jonka vasen puoli on aikariippumaton. Tällöin yhtälö voidaan kirjoittaa kahdessa toisistaan riippumattomassa osassa [7]

$$\begin{cases} \frac{1}{R}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{1}{\Theta}\frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right) + \frac{1}{\Phi}\frac{1}{r^2\sin^2(\theta)}\frac{\partial^2\Phi}{\partial\phi^2} = -k^2 \\ \frac{1}{v^2}\frac{\partial^2 T}{\partial t^2} = -k^2 T \end{cases} \tag{193}$$

joissa k on aaltoluku. Jälkimmäinen yhtälö tunnetaan Helmholtzin yhtälönä ja sen ratkaisut ovat aaltofunktion reaaliosan muotoa

$$T(t) = A\cos\omega t, \tag{194}$$

kun taas aikariippumattomasta aaltoyhtälöstä voidaan edelleen tehdä riippumaton muuttujan ϕ suhteen

$$-r^2\sin^2(\theta)\left(\frac{1}{R}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{1}{\Theta}\frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right) + \kappa^2\right) = \frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2} \tag{195}$$

josta saadaan jälleen kaksi toisistaan riippumatonta yhtälöä

$$\begin{cases} r^2\sin^2(\theta)\left(\frac{1}{R}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \frac{1}{\Theta}\frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right) + k^2\right) = m^2 \\ \frac{\partial^2\Phi}{\partial t^2} = -m^2\Phi \end{cases} . \tag{196}$$

Toisella yhtälöllä on ratkaisumuotonaan jälleen aaltofunktion reaaliosa

$$\Phi(\phi) = B\cos(m\phi) \tag{197}$$

ja ensimmäisestä yhtälöstä yritetään taas eristää muuttuja θ ja R toisistaan riippumattomiksi

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+r^2k^2=-\frac{1}{\Theta}\frac{1}{\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right)+\frac{m^2}{\sin^2(\theta)}, \tag{198}$$

joka onnistuu separaatiolla $l(l+1)$ [7] [10] antaen toisistaan riippumattomat yhtälöt

$$\begin{cases}\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+r^2k^2=l(l+1)\\ -\frac{1}{\Theta}\frac{1}{\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta}{\partial\theta}\right)+\frac{m^2}{\sin^2(\theta)}=l(l+1)\end{cases}. \tag{199}$$

Merkkaamalla $x=\cos(\theta)$ saadaan $\sin^2(\theta)=1-x^2$ ja $\mathrm{d}x=-\sin(\theta)\mathrm{d}\theta$ jolloin toinen yhtälö voidaan uudelleenjärjestellä ja kirjoittaa yleisen Legendren yhtälön muodossa

$$\frac{\partial}{\partial x}\left((1-x^2)\frac{\partial\Theta}{\partial x}\right)+\left(l(l+1)-\frac{m^2}{1-x^2}\right)\Theta=0, \tag{200}$$

jonka ratkaisut ovat Legendren liittofunktioita

$$\Theta(\theta)=P_\ell^m(\cos(\theta))=(1-\cos^2(\theta))^{m/2}\frac{\mathrm{d}^m}{\mathrm{d}\cos^m(\theta)}\left(P_\ell(\cos(\theta))\right). \tag{201}$$

Radiaalimuuttujan R yhtälö voidaan kirjoittaa Besselin yhtälön muodossa [7] [10]

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+2r\frac{\mathrm{d}R}{\mathrm{d}r}+(r^2k^2-l(l+1))R=0, \tag{202}$$

jolloin sen ratkaisut ovat Besselin pallofunktioita

$$R(r)=Cj_\ell(kr)=C\sqrt{\frac{\pi}{2kr}}J_{n+\frac{1}{2}}(kr), \tag{203}$$

jossa $J_n(x)$ on Besselin integraali

$$J_n(x)=\frac{1}{\pi}\int_0^\pi\cos(n\tau-x\sin\tau)\,\mathrm{d}\tau=\frac{1}{2\pi}\int_{-\pi}^\pi e^{i(n\tau-x\sin\tau)}\,\mathrm{d}\tau. \tag{204}$$

Pallovärähtelijän aaltofunktio on siis kokonaisuudessaan

$$\begin{aligned}\psi(r,\theta,\phi,t)&=R(r)\cdot\Theta(\theta)\cdot\Phi(\phi)\cdot T(t)=(ABC)j_\ell(kr)P_\ell^m(\cos(\theta))\cos(m\phi)\cos(\omega t)\\&=Dj_\ell(kr)Y_\ell^m(\theta,\phi)\cos(\omega t).\end{aligned} \tag{205}$$

G Bernoullin yhtälön johto

Jos virtauksessa on mitätön viskositeetti $\vec{\tau}_{ij} << 1$, voidaan liikemääräyhtälö (191) kirjoittaa Eulerin yhtälönä

$$\rho\mathbf{g} - \nabla P = \rho\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}, \tag{206}$$

josta näemme virtausnopeuden muutoksen aiheuttavan muutoksen väliaineen paineeseen. Hyödyntämällä yhtälöä (54) ja vektori-identiteettiä [56]

$$(\mathbf{V}\cdot\nabla)\mathbf{V} \equiv \nabla\left(\frac{1}{2}|\mathbf{V}|^2\right) + \zeta\times\mathbf{V} \tag{207}$$

voimme kirjoittaa Eulerin yhtälön muodossa

$$\left(\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} + \nabla\left(\frac{1}{2}|\mathbf{V}|^2\right) + \zeta\times\mathbf{V} - \mathbf{g} + \frac{1}{\rho}\nabla P\right)\cdot\mathrm{d}\mathbf{r} = 0 \tag{208}$$

josta kolmas termi häviää $(\zeta\times\mathbf{V})\cdot\mathrm{d}\mathbf{r} = 0$ virtauksen ollessa putken myötäistä. Tällöin

$$\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}\cdot\mathrm{d}\mathbf{r} + \mathrm{d}\left(\frac{1}{2}|\mathbf{V}|^2\right) + g\ \mathrm{d}z + \frac{\mathrm{d}P}{\rho} = 0 \tag{209}$$

joka on integroitavissa minkä tahansa kahden vuon myötäisen pisteen välillä. Kun virtaus tapahtuu pisteiden 1 ja 2 välillä, saadaan Bernoullin epävakaan ja kitkattoman virtauksen yhtälö

$$\int_1^2 \frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}\cdot\mathrm{d}s + \int_1^2\frac{\mathrm{d}P}{\rho} + \mathrm{d}\left(\frac{1}{2}(|\mathbf{V}_2|^2 - |\mathbf{V}_1|^2)\right) + g(z_2 - z_1) = 0 \tag{210}$$

ja jos väliaineen hiukkaset virtaavat hitaasti ($u < 0{,}3$ Ma), voidaan kokoonpuristuvaa väliainetta approksimoida kokoonpuristumattomana ja tiivistää edellinen yhtälö kirjoittamalla se kokoonpuristumattomana Bernoullin yhtälönä

$$P + \frac{1}{2}\rho|\mathbf{V}|^2 + \rho g z = C, \tag{211}$$

jossa ensimmäinen termi on painetermi, toinen termi on kineettinen energia, kolmas termi on potentiaalienergia ja C on vakio. Massan säilymisen (86) ja Bernoullin

lain (87) avulla näemme, että nopeuden kasvaessa putken läpileikkauspinta-alan pienentymisen johdosta väliaineen paine pienenee ja virtausnopeuden laskiessa pinta-alan kasvamisen johdosta väliaineen paine kasvaa.

H Impulssivaste ja huonevaste

Kun heräte on lyhyt impulssi hetkellä $t = 0$, se aiheuttaa impulssivasteen $g(t)$ jota kuvaa Fourierin muunnos [12]

$$g(t) = \int_{-\infty}^{\infty} p_\omega e^{i\omega t} \mathrm{d}\omega, \tag{212}$$

jonka ratkaisut ovat muotoa

$$g(t) = \begin{cases} 0 & \text{kun } t < 0 \\ \sum_n A_n e^{(-\beta_n t)} e^{i\omega_n t} & \text{kun } t \geq 0 \end{cases}. \tag{213}$$

Jos heräte taas on jatkuva ja loppuu hetkellä $t = 0$, se aiheuttaa huonevasteen eli kaiun

$$h(t) = \int_{-\infty}^{0} s(\tau) g(t - \tau) \mathrm{d}\tau = \sum_n A_n e^{-\beta_n t} (a_n \cos(\omega_n t) + b_n \sin(\omega_n t)) \tag{214}$$

aika-alueelle $t \geq 0$. Huonevasteen neliö on verrannollinen huoneen energiatiheyteen

$$E(t) \propto (h(t))^2 = \sum_m \sum_n C_m C_n e^{(-(\beta_m + \beta_n)t)} (\cos(\omega_m t) + (\cos(\omega_n t)), \tag{215}$$

jossa vakiot ovat koottu vakioihin $C_i = A_i \sqrt{a_i^2 + b_i^2}$. Energiatiheys voidaan keskiarvottaa

$$\langle E(t) \rangle = \sum_n C_n e^{-2\beta_n t} \tag{216}$$

ja kirjoittaa integraalina vaimennusjakauman $H(\beta)$ yli [12]

$$\langle E(t) \rangle = \int_0^{\infty} H(\beta) e^{-2\beta t} \mathrm{d}\beta, \tag{217}$$

jossa vaimennusjakauma riippuu niin herätteen laadusta kuin herätteen ja tarkkailijan sijainnista huoneessa.

I Aktivaatiofunktioita ja normitus

Yleisiä aktivaatiofunktioita ovat logistinen käyrä eli sigmoidifunktio σ

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}}, \qquad \sigma : \mathbb{R} \to [0,1] \tag{218}$$

ja sen nollakeskitetty sekä skaalattu ja täten käytännöllisempi variantti, hyperbolinen tangenttifunktio

$$\tanh(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{-x} + \mathrm{e}^{x}}, \qquad \tanh : \mathbb{R} \to [-1,1]. \tag{219}$$

Hierarkkisille neuroverkoille on todettu tehokkaaksi aktivaatiofunktioksi tasasuunnattu lineaarinen yksikkö eli ReLU (Rectified Linear Unit)

$$\mathrm{ReLU}(x) = \frac{|x| + x}{2} = \begin{cases} 0 & \text{kun } x \leq 0 \\ x & \text{kun } x > 0 \end{cases} \qquad \mathrm{ReLU} : \mathbb{R} \to [0,\infty] \tag{220}$$

ja sen sileä variantti GELU (Gaussian Error Linear Unit)

$$\mathrm{GELU}(x) = x \cdot \Phi(x) = x \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} -\frac{t^2}{2} \, \mathrm{d}t$$

$$\approx 0{,}5x \cdot \left(1 + \tanh\left(\sqrt{\frac{2}{\pi}}(x + 0{,}044715x^3)\right)\right), \qquad \mathrm{GELU} : \mathbb{R} \to [-0{,}17; \infty] \tag{221}$$

jossa Φ on kumulatiivinen normaalijakauma, jonka keskihajonta on $\sigma = 1$ ja odotusarvo on $\mu = 0$. [49]

Boltzmannin jakaumaa (177) voidaan käyttää aktivointifunktiona ja se tunnetaankin yleisessä muodossaan myös softmax-funktiona. Softmax ottaa syötteeksi diskreetin K-ulotteisen vektorin z ja palauttaa jatkuvan todennäköisyysjakauman

$$\mathrm{softmax}(z_i) = \frac{\mathrm{e}^{\beta z_i}}{\sum_{j=1}^{K} \mathrm{e}^{\beta z_j}} = \frac{\mathrm{e}^{\beta z_i}}{Z}, \qquad \mathrm{softmax} : \mathbb{R}^K \to [0,1]^K \tag{222}$$

jossa β on lämpötilakerroin. Softmaxin todennäköisyysjakauma on kuitenkin deterministinen ja sen todennäköisimmät arvot saadaan argmaxilla, joka palauttaa ne diskreetteinä. Argmaxin käyttö voidaan välttää softmax-funktion variantilla Gumbel-softmaxilla, jossa softmaxiin esitellään satunnaisia Gumbel-jakauman arvoja $u \in [0,1]$ ja funktio palauttaa jatkuvan todennäköisimpien arvojen jakauman. Gumbel-softmax edellyttää syötteen z muuntamisen logiteiksi $l \in \mathbb{R}^{M \times K}$. Tällöin jonkin tilan k todennäköisyys joukossa $m \in M$ on

$$p_{m,k} = \frac{\exp\left((l_k - \log\left(-\log\left(u_k\right)\right))/\tau\right)}{\sum_{j=1}^{K} \exp\left((l_j - \log\left(-\log\left(u_j\right)\right))/\tau\right)}, \qquad p : \mathbb{R}^{G \times K} \rightarrow [0,1]^K \tag{223}$$

jossa $\beta = 1/\tau$ on lämpötilakerroin.

Sigmoidi- ja softmax-funktiot ovat normitettuja todennäköisyysteorian toisen aksiooman (170) mukaisesti, jonka vuoksi niitä käytetään usein neuroverkkojen viimeisen kerroksen aktivointifunktioina. Neuroverkon arvoja voidaan normittaa myös muilla tavoilla, kuten joukkonormituksella (batch normalization) tai kerrosnormituksella (layer normalization). Joukkonormituksessa normitetaan kaikki kunkin kerroksen vastaanottamat syötteet asettamalla niiden keskiarvot ja varianssit vakioiksi. Normitus suoritetaan minijoukoilla B jotka ovat kokoa m, joiden keskiarvoksi ja varianssiksi asetetaan siis [50]

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{ja} \qquad \sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2. \tag{224}$$

Kerrokseen annetun d-ulotteisen syötteen, jokainen ulottuvuus $x = (x^{(1)},...,x^{(d)})$ normitetaan erikseen

$$\hat{x}_i^{(k)} = \frac{x_i^{(k)} - \mu_B^{(k)}}{\sqrt{\left(\sigma_B^{(k)}\right)^2 + \epsilon}}, \tag{225}$$

jossa $k \in [1,d]$ ja $i \in [1,m]$. $\mu_B^{(k)}$ ja $\sigma_B^{(k)}$ ovat ulottuvuuskohtaiset keskiarvo ja keskihajonta, kun taas ϵ on pieni vakio joka lisätään numeerisen stabiiliuden vuoksi. Jotta normitettu aktivaatio $\hat{x}^{(k)}$ kompensoisi normituksessa menetettyjä esityspiirteitään, se asetetaan uudelleen lineaarikuvaukseen

$$y_i^{(k)} = \gamma^{(k)}\hat{x}_i^{(k)} + \beta^{(k)}, \tag{226}$$

jossa optimoitavat parametrit $\gamma^{(k)}$ ja $\beta^{(k)}$ alustetaan satunnaisesti.

Kerrosnormitus toimii muuten samoin, mutta kullekin kerrokselle normitus tehdään ominaisuuksien j suuntaan eikä joukon i suuntaan [51]

$$\mu_B = \frac{1}{m}\sum_{j=1}^{m} x_{ij} \qquad \text{ja} \qquad \sigma_B^2 = \frac{1}{m}\sum_{j=1}^{m}(x_{ij} - \mu_B)^2, \tag{227}$$

sekä

$$\hat{x}_{ij}^{(k)} = \frac{x_{ij}^{(k)} - \mu_i^{(k)}}{\sqrt{\left(\sigma_B^{(k)}\right)^2 + \epsilon}} \tag{228}$$

jolloin

$$y_{ij}^{(k)} = \gamma^{(k)}\hat{x}_{ij}^{(k)} + \beta^{(k)}. \tag{229}$$

J Hukkafunktioita

Muita hukkafunktiota ovat muun muassa ristikkäisentropia

$$\mathbb{L}_S = -\sum_{i=1}^{n} t_k \log y_k, \tag{230}$$

ja kosinisimilaariutta hyödyntävä kontrastiivinen hukkafunktio

$$\mathbb{L}_m = -\log \frac{\exp(\text{sim}(y_k,t_k)/\kappa)}{\sum_q \exp(\text{sim}(y_q,t_q)/\kappa)}, \tag{231}$$

jossa $\text{sim}(p_k,t_k) = (p_k^T \cdot t_k)/(||_k|| \cdot ||t_k||)$.

Diversiteettihukkafunktio

$$\mathbb{L}_d = \frac{1}{GK} \sum_{g=1}^{G} -S(p_g) = \frac{1}{GK} \sum_{g=1}^{G} \sum_{k=1}^{K} p_{gk} \log (p_{gk}) \tag{232}$$

keskiarvoittaa softmax-jakaumien p_{gk} maksimoiman entropian kun $g \in G$ ja $k \in K$. Nimitysvirhetasohukkafunktio (label error rate, LER) vertaa sekvenssipareja $(x,z) \in Z$ joukossa $Z \subset \mathbb{D}_{X \times Z}$ jossa $h[x]$ ovat verkon antamia sekvenssejä ja z mallisekvenssejä, jolloin nimityshukkafunktio laskee niiden normalisoidun editointietäisyyden (editing distance, ED) [43]

$$\mathbb{L}_{LER}(h, Z) = \frac{1}{N} \sum_{(x,z) \in Z} \text{ED}(h[x],z). \tag{233}$$

Nimitysvirhetasohukkafunktiossa N on mallifoneemien lukumäärä ja editointietäisyys on Levenshtein-etäisyys [53] joka kertoo kuinka monta Levenshtein-operaatiota eli merkin insertiota (ins), poistamista (del) ja vaihtoa (sub) sekvenssiin $h[x]$ tulee minimissään tehdä jotta se vastaa mallisekvenssiä z. Levenshtein-etäisyys voidaan laskea Wagner-Fischer algoritmillä [52] [54] eli

$$\text{ED}_{i0} = \sum_{k=1}^{i} w_{\text{del}}(h[x_k]), \quad \text{kun } 1 \leq i \leq m \tag{234}$$

$$\mathrm{ED}_{0j} = \sum_{k=1}^{i} w_{\mathrm{ins}}(f_k), \quad \text{kun } 1 \leq i \leq n \tag{235}$$

$$\mathrm{ED}_{ij} = \begin{cases} d_{i-1,j-1} & \text{kun } a_i = b_j \\ \min \begin{cases} d_{i-1,j} + w_{\mathrm{del}}(a_i) \\ d_{i,j-1} + w_{\mathrm{ins}}(b_j) & \text{kun } a_i \neq b_j \\ d_{i-1,j-1} + w_{\mathrm{sub}}(a_i, b_j) \end{cases} \end{cases} \quad \text{kun } 1 \leq i \leq m, 1 \leq j \leq n, \tag{236}$$

tai naiivisti

$$\mathbb{L}_{LER} = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}, \tag{237}$$

jossa S on vaihdettujen nimitysten määrä, D on poistettujen nimitysten määrä, I on insertoitujen nimitysten määrä, C on oikeiden nimitysten määrä ja N on kaikkien käsiteltyjen nimitysten määrä.

K Tunnistettua puhetta

Tässä liitteessä on wav2vec2-xlsr-1b-finnish-lm-v2-puheentunnistusmallilla tunnistettua puhetta tutkielman kirjoittajan lausumana.

akustiikka ja fonetiikka ovat vakiintuneita tieteitä joita kehitetään edelleen tarkempien simulaatioiden ja mallien avulla automaattinen puheentunnistus on puolestaan jatkuvasti muuttuva ala joka on edennyt harppauksin viimeaikaisten tekoäly kehitysten seurauksena kehitys on ollut nopeinta kieli mallien parissa joskin myös akustiset mallit ovat hyötyneet kielimallien innovaatioista kuten huomio mekanismista suomenkielinen automaattinen puheentunnistus on hyvällä tasolla josta tämä kappale on puhumaan esimerkkinä puheen tunnistusmallin tunnistama puhe löytyy liitteestä k jossa esiintyvät oikein kirjoitusvirheet ovat korjattavissa yhdistämällä puheentunnistus malliin kehittyneempi kielimalli

Tunnistetussa puheessa on 6 yhdyssanavirhettä jotka ovat sanavirhetason kannalta kalliita, sillä jokainen yhdyssanavirhe merkitsee yhden sanan poistamista ja toisen sanan substituutiota. Tunnistetun puheen sanavirhetaso on yhtälön 146 mukaisesti $\mathbb{L}_{WER} = 17{,}10\%$ kun $S = 7$, $D = 6$ ja $C = 63$. Sanavirhetaso on siis korkea kielen agglutinatiivisuudesta johtuen, mutta yhtälö 147 puolestaan antaa hyvin matalan kirjainvirhetason $\mathbb{L}_{CER} = 1{,}14\%$ kun $D = 6$, $I = 2$ ja $C = 692$. Tunnistetun puheen pienestä mittakaavasta huolimatta virhetasot vastaavat kokoluokaltaan Tanskasen ja Toivasen ilmoittamia virhetasoja. [34]