

TUTKIMUSDATAN MÄÄRITTELYN JA AINEISTONHALLINNAN ONGELMAT IHMISTIETEISSÄ



Mikko Ojanen

*kirjastonhoitaja, Helsingin yliopiston kirjasto
tohtorikoulutettava, Musiikkitiede, Helsingin yliopisto
orcid.org/0000-0002-7833-9659
mikko.ojanen@helsinki.fi*

**HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI**

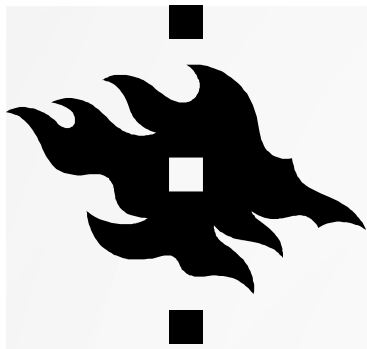


Informaatiotutkimuksen päivät 3.–4.11.2016
Digitoidut kokoelmat digitaalisten ihmistieteiden tutkimusaineistona



ESITYKSEN TEEMAT

- Tutkimusaineistojen ja -datan hallinnan ongelmat
 - Aineiston ja datan määrittelyä
 - Tutkijoiden usein pohtimia ongelmia aineistonhallintaan liittyen
- Esimerkki äänitedigitoinnista ja digitaalisten kuvantamismenetelmien hyödyntämisestä musiikkitieteellisessä tutkimuksessa



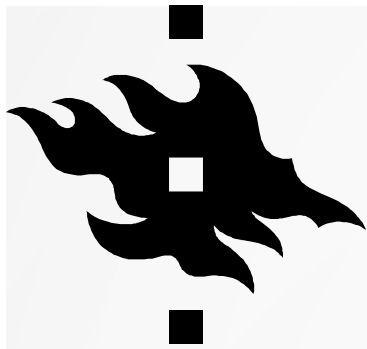
DIGITAALISET IHMISTIETEET

- Digitaaliset ihmistieteet; ks. esim. <http://www.aka.fi/digihum>
 - *Digitaaliset ihmistieteet -akatemiaohjelman alueeseen kuuluvat erilaiset uudet menetelmät ja tekniikat, joissa digitaalista teknologiaa ja modernin datatieteen menetelmiä hyödynnetään humanististen ja yhteiskuntatieteellisten aineistojen keruussa, hallinnassa ja analyysissä.*
- Painotus koneluettavissa, laajoissa aineistokokonaisuuksissa; ei tarkkaa määritelmää; erilaisia lähestymistapoja
 - Esim. *making things* ja *theorizing things* -jaottelu
- Samaistun etäisesti digihumanistiksi, vaikka omassa työssäni en (ainakaan toistaiseksi) hyödynnä laskennallisia menetelmiä
- Äänitedigitointi ja digitaalisten äänitteiden kuvantamismenetelmät tarjoavat havainnollisen pohjan analyysille; hyödyllisiä työkaluja perinteisestä musiikkianalyysistä poikkeavien kohteiden kuvailuun



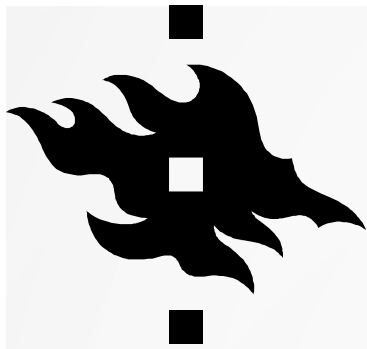
AINEISTON JA DATAN MÄÄRITELMIÄ

- **Aineiston ja datan määrittelyn ongelmat** hämmentävät tutkijoita
 - Englanninkielisistä fraaseista kopioitunut data-sana
 - RDM, research data management
 - DMP, data management plan(ing)
 - Big data (<http://bigdata.fi/big-data-maaritelma>)
 - Open data (<http://www.hri.fi/fi/mita-on-avoin-data/>; <https://www.avoindata.fi/fi/opas/jatko>)
 - My data (<http://urn.fi/URN:ISBN:978-952-243-418-0>) jne.
 - Miten määritellään esim. aineisto, data, dataset, metadata jne.?
 - Tutkimuksessani ei ole dataa!
- **Avoimuuden ja avaamisen vaatimukset** hämmentävät tutkijoita
 - Miten ihmeessä voin avata arkaluonteista tai tekijänoikeuden alaista materiaalia?
 - **HUOMAA!** Avoin tiede ei tähtää kaiken datan avaamiseen, vaan auttaa tutkijoita tiedostamaan aineistojen hallintaan (esim. sensitiivisyyteen) liittyvät yksityiskohdat
 - ”Avataan niin paljon kuin mahdollista suojataan niin paljon kuin tarpeen.” / ”As open as possible, as closed as necessary” (ks. esim. Jones <http://www.dcc.ac.uk/news/new-h2020-dmp-guidelines>)



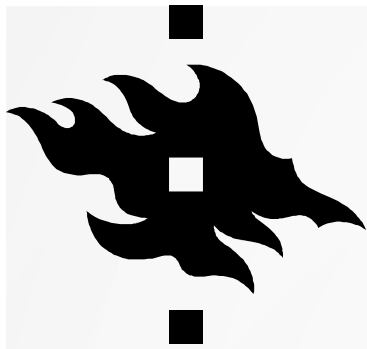
AINEISTON JA DATAN MÄÄRITELMIÄ

- Aineisto = Data
 - Arkikielessä määritelmään kiinnitetään harvoin huomiota
 - Käytetään usein toistensa synonyymeinä
 - Data voidaan ymmärtää yleiskäsitteenä, johon pitää erikseen liittää sitä selittävä määre; määrällinen, laadullinen, laskennallinen, tulkittu jne.
- Aineisto ≠ Data
 - Datan ajatellaan usein viittaavaan vain digitaaliseen aineistoon
 - Data ajatellaan usein olevan laskennallista, määrällistä, koneluettavaa jne.



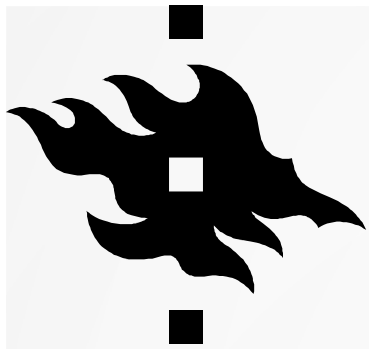
AINEISTON JA DATAN MÄÄRITELMIÄ

- Praktinen näkökulma datan määritelmään
 - *Data* voidaan ymmärtää havainnoiksi mistä tahansa *aineistosta*; havainnoiksi, joilla sinällään ei vielä ole esim. semanttista merkitystä; saa merkityksen tulkinnan tai analyysin kautta
- Data life-cycle
 - Raakamateriaalista dataksi, datasta tulkinnoiksi/analyysiksi, tulkinnoista ja analyyseistä tutkimusraportiksi
 - Yhden tason data voi olla toisen tason raaka dataa
 - Aineiston ja datan jatkojalostaminen ja kopioituminen
 - Erotetaan toisistaan tutkimuksen aikainen datan käsittely ja tutkimuksen jälkeiseen aikaan liittyvät toimenpiteet



AINEISTON JA DATAN MÄÄRITELMIÄ

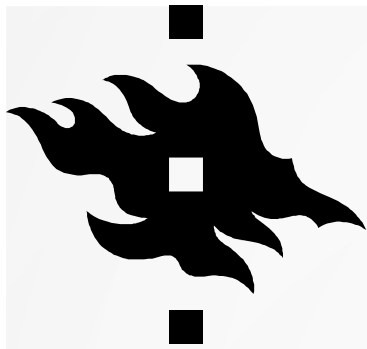
- Erilaisten aineistotyyppien/tasojen erottelu
 - Tutkimusta edeltävät aineistot ja data:
 - itse kerätty, yksilöllinen dokumenttien kokoelma
 - jo ennalta arkistoitu, luetteloitu ja kuvailtu tai täysin järjestämätön mutta asianmukaisesti säilytetty kokoelma
 - Tutkimuksen myötä tuotetut aineistot ja data:
 - tulkitut, kommentoidut tai koodatut aineistot (kvali ja/tai kvanti) ja raaka data/materiaali (esim. haastattelu äänitteellä)
 - taiteellinen teos tai prosessi ja siitä tehtävät huomiot – eroaako tämä edellisistä ja jos niin miten?
 - Fyysiset, digitoidut, digi-natiivit -jaottelu



ERKKI KURENNIEMEN SÄHKÖSOITTIMET

- Soittimia koskeva tutkimusaineisto:
 - Fyysiset soittimet
 - Electric-Sound-Machine (1964–)
 - Electric Quartet (1968)
 - Andromatic (1968)
 - DICO (1969)
 - DIMI-A (1970)
 - DIMI-O (1971)
 - DIMIX (1972)
 - DIMI-S (1972)
 - DIMI-T (1973)
 - DIMI-6000 (1973–75)
 - Valokuvat soittimista
 - Historialliset aikalaiskuvat; digitoidut
 - Tutkimuksen aikana kuvatut; diginatiivit
 - “Käyttöohjeet”, spesifikaatiot ja tekniset tiedot
 - Henkilöhaastattelut
 - Historialliset aikalaisdokumentit; digitoidut
 - Tutkimuksen aikana tuotetut; diginatiivit
 - Äänitiedostot sekä litteraatiot



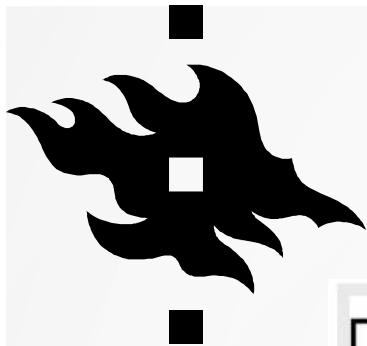


HELSINGIN YLIOPISTON ELEKTRONIMUSIIKKISTUDION ÄÄNITEARKISTO

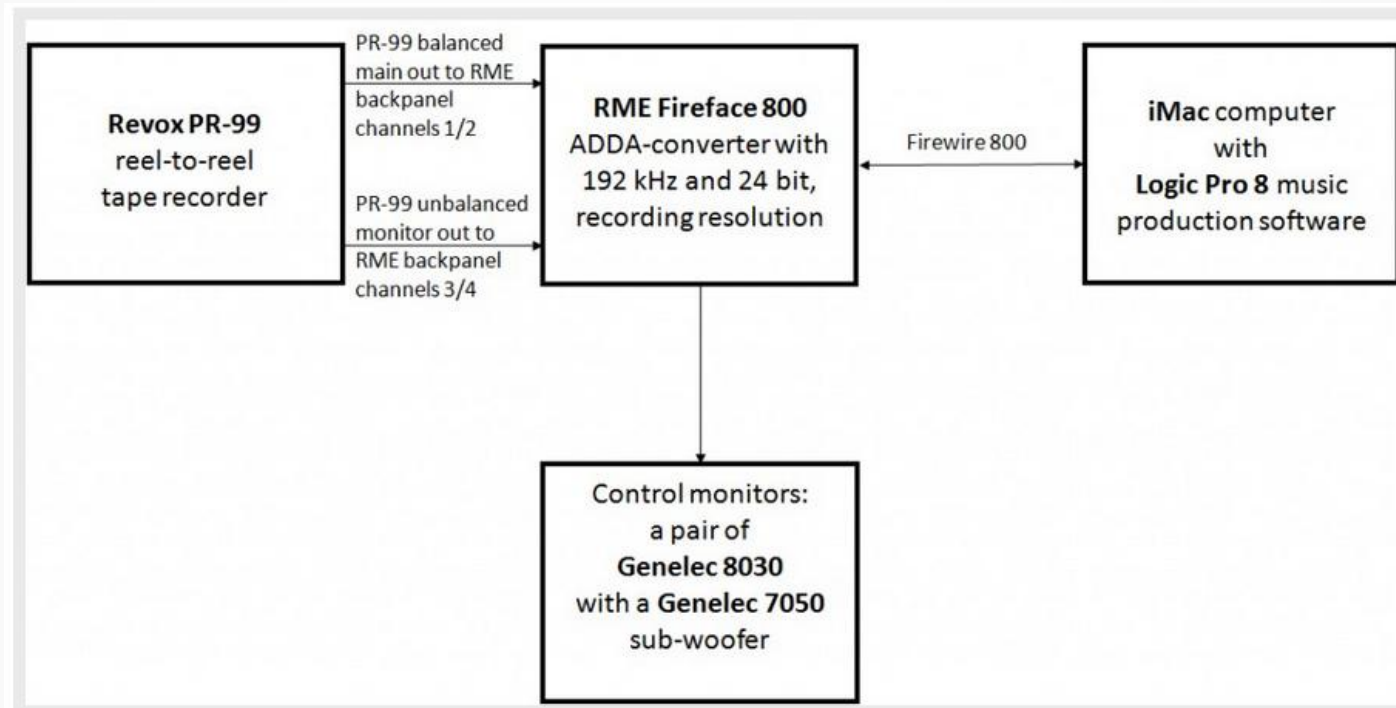
- Pääosin ¼”-kelanauhoja
- Noin 350–400 kelanauhaa (koteloa)
- Erkki Kurenniemen kokoelma (~10-15% koko kokoelmasta):
 - Musiikkiteoksia ja muita äänitteitä
 - Noin 30 itsenäistä teosta ja äänitystä
 - Elektroakustista nauhamusiikkia, nauhakollaaseja, performansseja ja improvisaatioita
 - Musiikkia ja äänisuunnittelua näyttelyihin, elokuvaiin mainoksiin, näytelmiin ja radiokuunnelmiin
 - Äänipäiväkirjat; noin 100 tuntia c-kasettiäänityksiä (1970–75) (Kansallisgallerian kokoelmassa)



Kuva: Mikko Ojanen

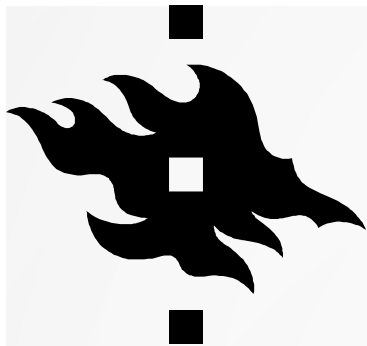


ÄÄNITEDIGITOINTI



Kuva: Mikko Ojanen

- Yksityiskohtainen kuvaus digitointiprosessista ks. <http://arpjournal.com/mastering-kurenniemis-rules-2012-the-role-of-the-audio-engineer-in-the-mastering-process/>



ÄÄNITEDIGITOINTI

TAPE/REEL DETAILS

Reference tones: Yes

* Date of recording: ? --Se

Date of recording (1st date):

Date of recording (2nd date):

Date of recording (3rd date):

* Tape/Reel type: --Se

If other, specify what?

* Center hub type: --Se

If other, specify what?

Tape type [brand & type e.g. AGFA PER 525]:

For tape manufacturers and types see

Tape includes...

... leader tape in the beginning

... in the middle

... leader tape in the end

Hours [00] Minutes [00]

Tape length:

* Tape width: --Se

Tape speed (if multiple, choose all applicable):

4.76 cm/s (1 7/8 ips) 38 cm/s (15 ips)

5.53 cm/s (3 3/4 ips) 57.15 cm/s (22.5 ips)

19.05 cm/s (7 1/2 ips) Other

If other, specify what?

Channel orientation (if multiple, choose all applicable):

Full track / Mono Stereo

2 track 4 track

Dual mono Other

If other, specify what?

* Tape direction: --Select--

Markings/Other details:

DIGITIZATION DETAILS

Digitization engineer:

File name:

File location:

Date of digitization:

* Sample rate:

* Bit rate:

* File format:

If other, specify what?

Digitization hardware: Height [cm]

Digitization software:

Container size:

Signal path:

* Date of recording: ? --Select--

Date of recording (1st date):

Other notes:

CONTAINER DETAILS

Tape content & details (supposed):

Tape type [brand & type e.g. AGFA PER 525]:

* Box material:

If other, specify what?

Archive ID:

Archive signum [location on the shelf]:

* Collection: --Select--

PHYSICAL ARCHIVE INFORMATION

AUDIO CONTENT DETAILS

Musical work Instrument test

Music (identified) Documentary recording

Music (unidentified) Interview

Live performance Other

Improvisation

If tape contains an identified musical work, submit cataloguing details for the musical work with a dedicated lomake via this link: <https://elomake.helsinki.fi/lomakkeet/71985/lomake.html>

If other, specify what?

Keywords ?

Other notes:

Kuva: Mikko Ojanen

- Tiedonkeruu ja luettelointi toistaiseksi e-lomakkeella Excelliin
- Kerätään ja luetteloidaan:
 - digitointiyksityiskohdat
 - kelan kotelon tiedot,
 - kelanauhan ja keskiön tiedot
 - soivan äänitteen tiedot
 - fyysisen arkiston tiedot
 - kokoelma ja sijainti



ÄÄNITEDIGITOINTI

- Digitaalisen äänitiedoston kuvaajat
- Spektrogrammi- ja äänekkyys kuvaajat mahdollistavat teosten ja äänitteiden eri versioiden havainnollisen vertailun

Kuvat:

<http://arpjournal.com/mastering-kurenniemis-rules-2012-the-role-of-the-audio-engineer-in-the-mastering-process/>

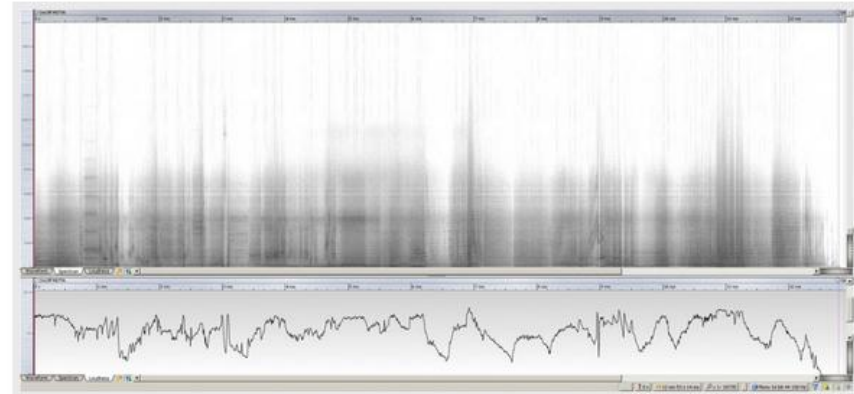


Figure 5a. The spectrogram and loudness contour of the 1992 version of *On-Off*.

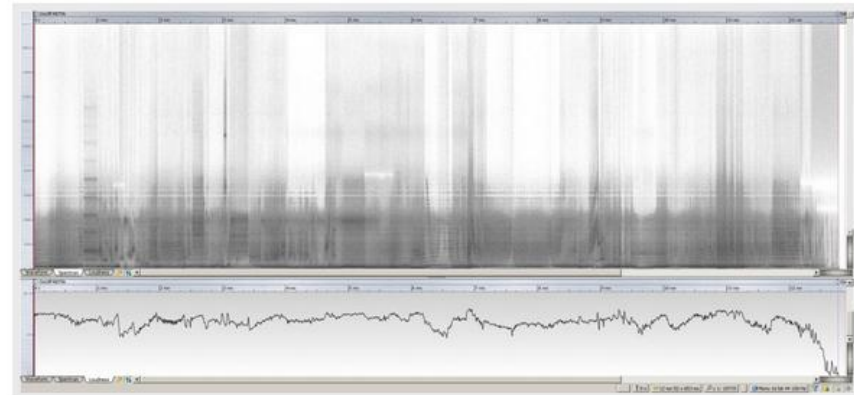


Figure 5b. The spectrogram and loudness contour of the 2002 version of *On-Off*.

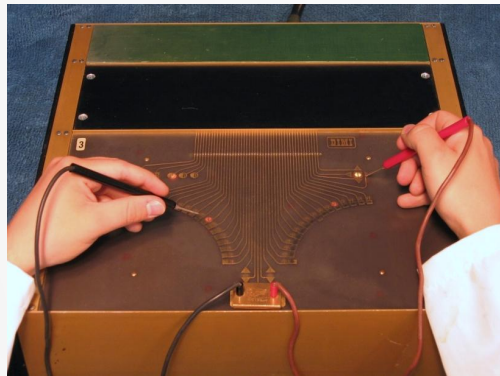
Kuva: Mikko Ojanen

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)

Kurenniemen sovitus Johann Sebastian Bachin a-molli-inventiosta (nro 13; BWV 784)

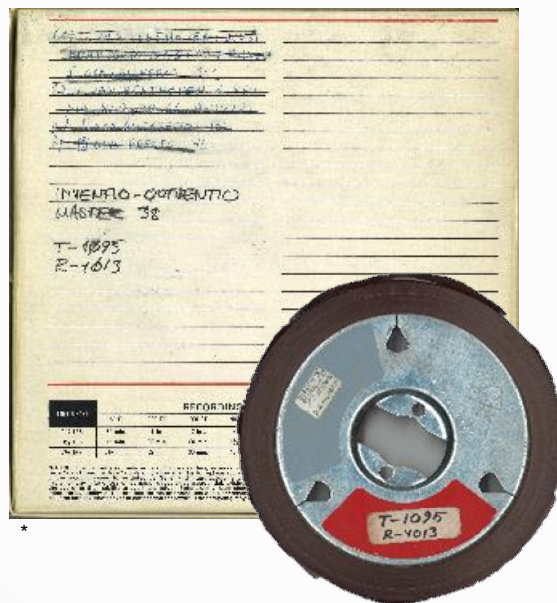
Julkaistu *Dimi is born* -vinyylisinglenä 1970 DIMI-A-soittimen markkinointitarkoituksessa yhdessä *Outvention* (yhteistyössä Ruohomäen kanssa) ja Jukka Ruohomäen *Mikä aika on* -teoksen kanssa.

PROSESSI: Kelanauhasta äänitteeksi, äänitteestä digitoiduksi tiedostoksi (192 kHz, 24 bit; .wav-tiedosto), korkean resoluution tiedostosta analyysiä ja kuuntelukopioita varten (44.1 kHz, 24 bit ja 320 kbps .mp3 -tiedostot), digitaalisesta äänitiedostosta (kommentoiduiksi) kuvaajiksi



**

DIMI-A-syntetisaattori (1970)



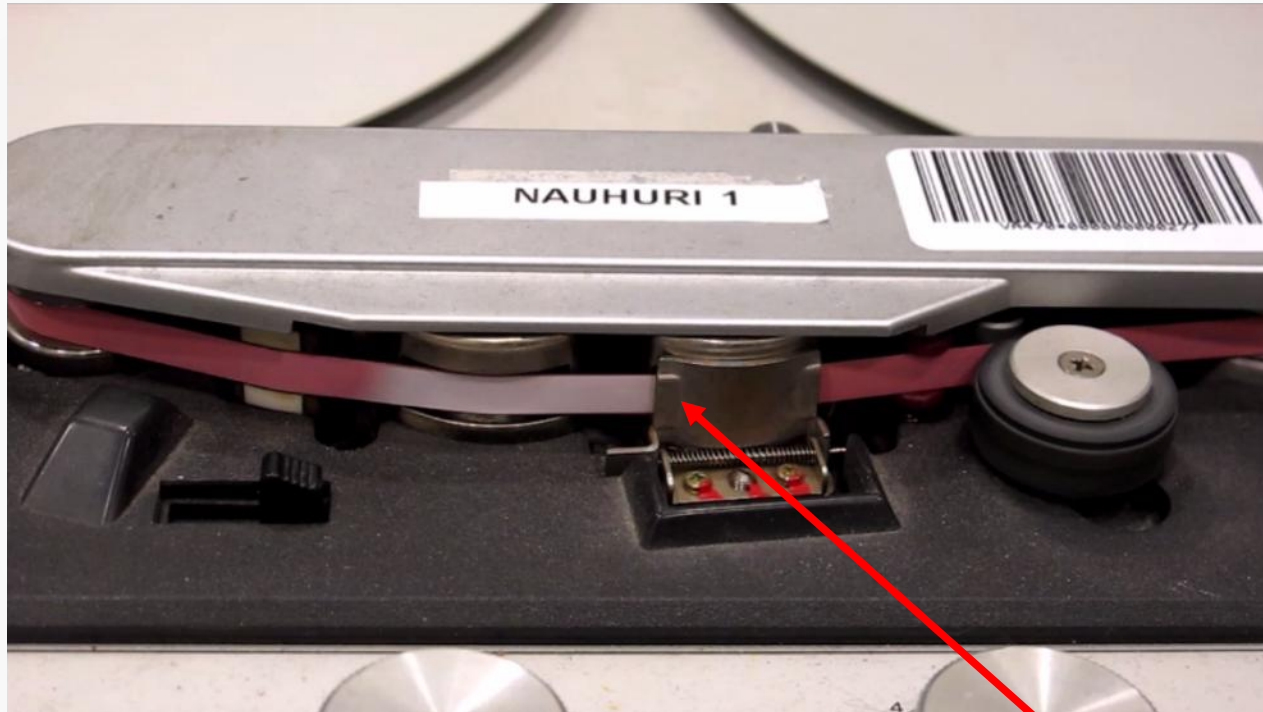
*

Invention alkuperäinen
masternauha (1970)



Dimi is born vinyylisingle-
julkaisu (1970)

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)

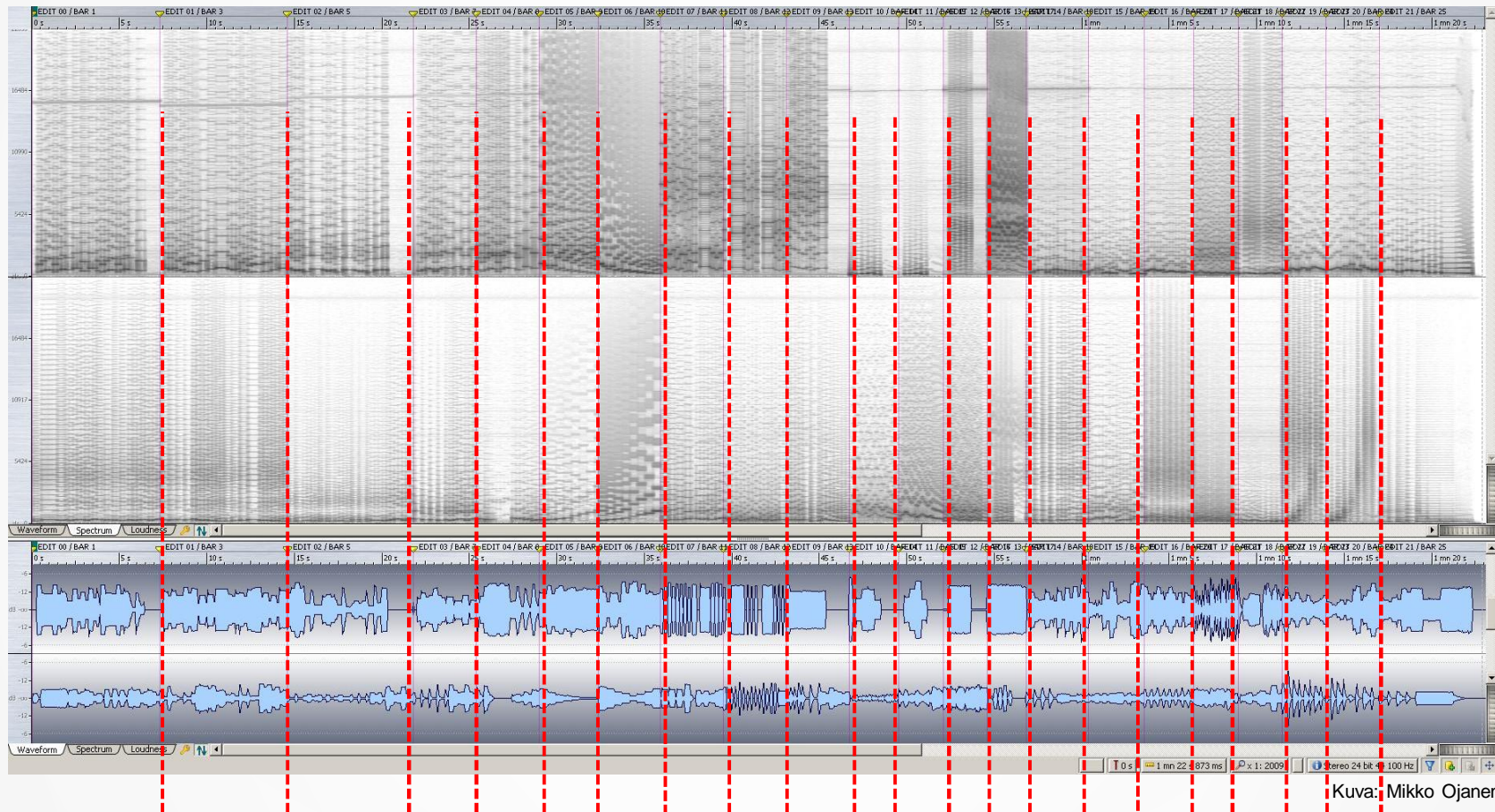


Kuva: Mikko Ojanen

Valkoinen teippi ohittaa
äänipään = nauhan
leikkauskohta

Inventio-analyysi; masternauhan leikkauskohdat

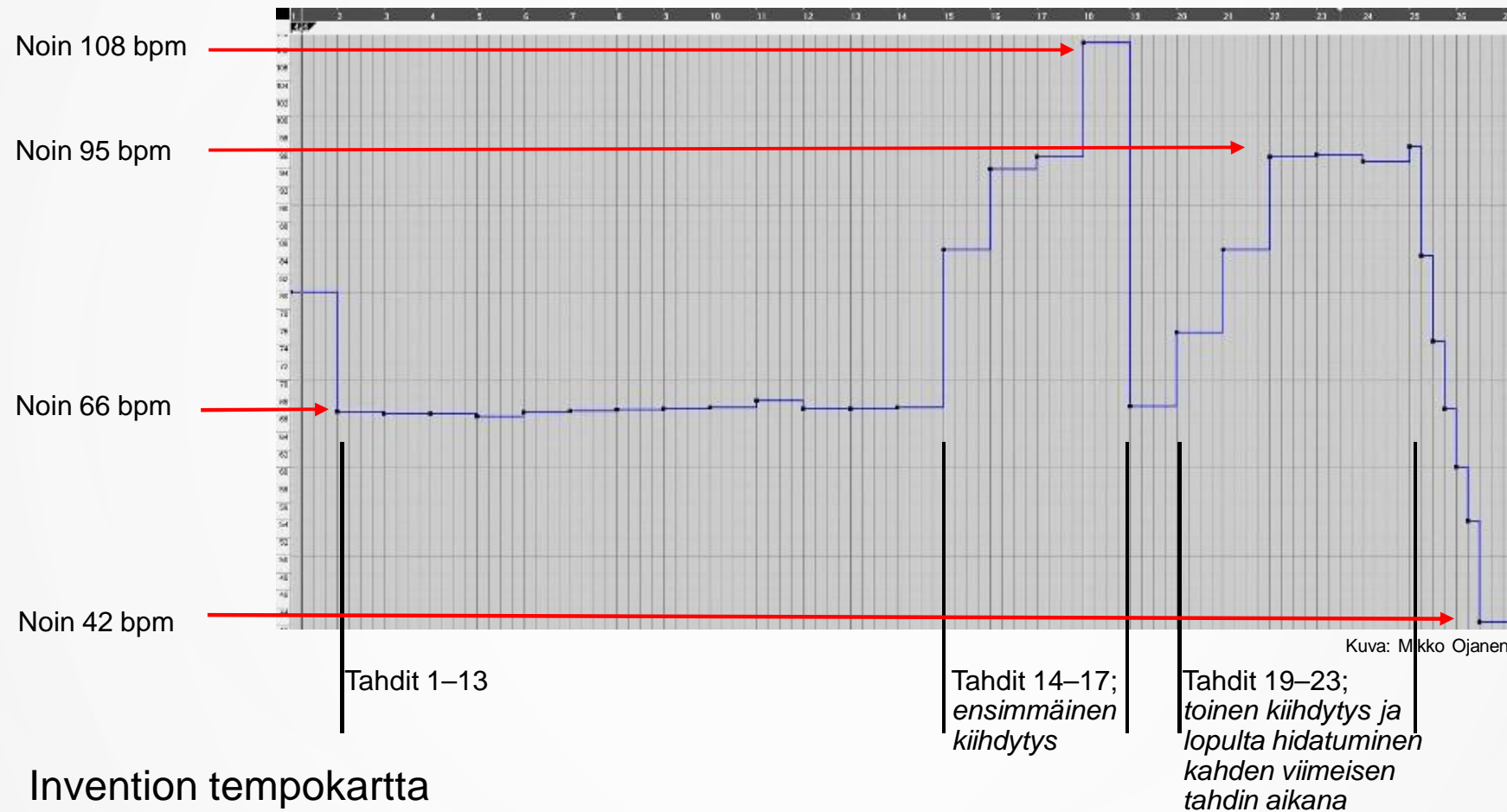
ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)



Invention sonogrammikuvaaja ja masternauhan leikkauskohdat

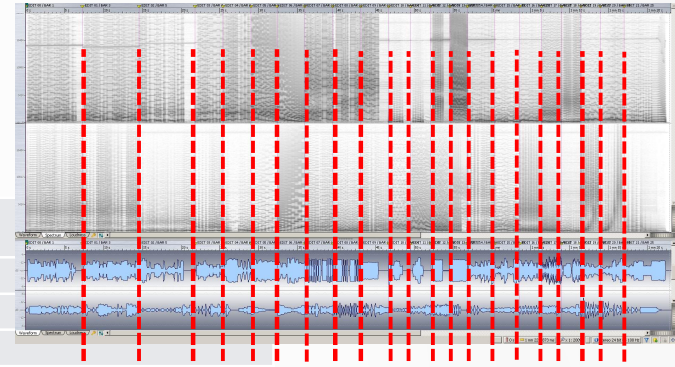
Leikkauskohdat näkyvät myös tekstuurin muutoksina sonogrammikuvaajassa

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)

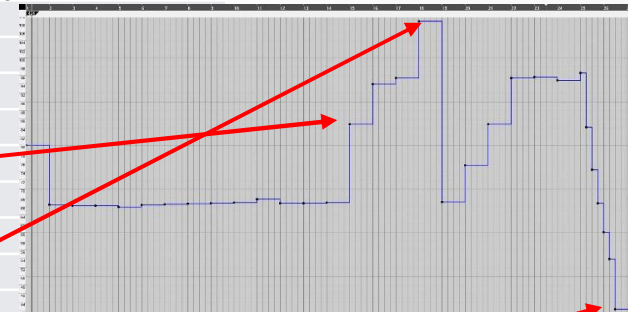


Invention tempokartta
Tuotettu Steinberg Cubase -ohjelmalla; midi-dataa

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)

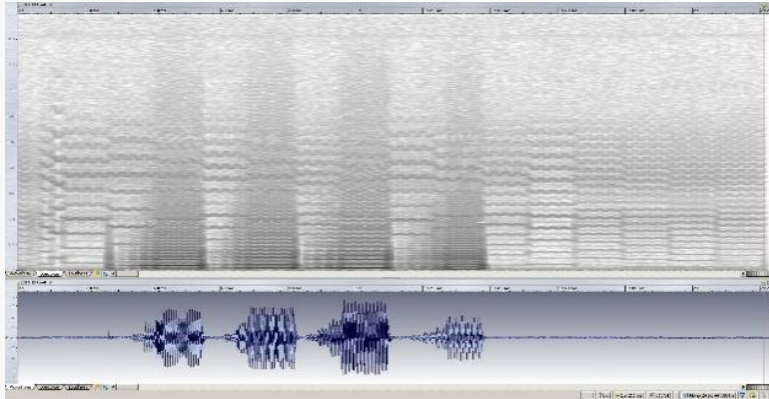


EDIT	SLICE	BAR	POSITION [mn:s:ms]	TEMPO	COMMENT
0	1	1	00:00:000	66	Inventio start; Tempo 66 bpm
1	2	3	00:07:316	66	[n/a]
2	3	5	00:14:583	66	[n/a]
3	4	7	00:21:810	66	Careless edit; unintentional snap on the left channel
4	5	8	00:25:414	66	[n/a]
5	6	9	00:29:011	66	Left and right part switch position in the middle; 16th note missing in the end
6	7	10	00:32:386	66	Left and right part switch position in the middle
7	8	11	00:35:935	66	[n/a]
8	9	12	00:39:531	66	[n/a]
9	10	13	00:43:121	66	[n/a]
10	11	14	00:46:712	85	Tempo change: 85 bpm; Left and right part re-arranged
11	12	15	00:49:547	95	Tempo change: 95 bpm; Left and right part re-arranged
12	13	16	00:52:079	95	Left and right part rearranged
13	14	17	00:54:606	108	Tempo change: 108 bpm; Left and right part re-arranged
14	15	18	00:56:817	66	Tempo change: 66 bpm
15	16	19	01:00:394	75	Tempo change: 75 bpm; Left and right part re-arranged (second and last 4 16th notes)
16	17	20	01:03:575	85	Tempo change: 85 bpm
17	18	21	01:06:402	95	Tempo change: 95 bpm (Possibly Ring Modulation in action)
18	19	22	01:08:926	95	[n/a]
19	20	23	01:11:445	95	[n/a]
20	21	24	01:13:966	95	Ritardando start: Tempo 95 bpm
21	22	25	01:16:992	45	Tempo at the 45 bpm



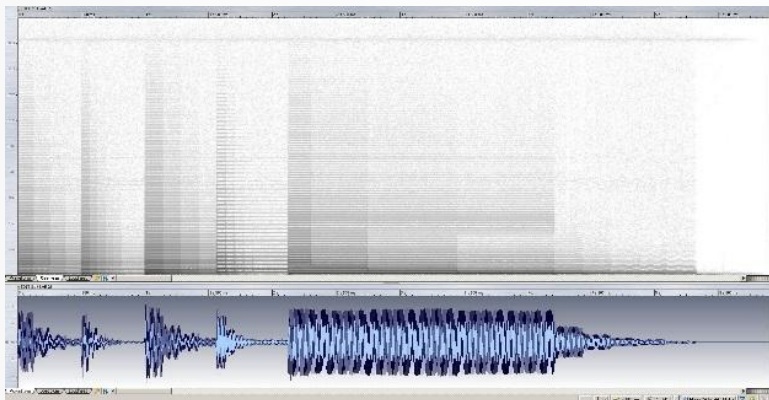
Kuvat: Mikko Ojanen

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)



Invention ohjelmointiesimerkki; tahti 17

Soittimessa ei ole erillistä verhoikäyrageneraattoria ja muun muassa yksittäisten nuottien dynamiikka pitää ohjelmoida soittimen muistiin askel kerrallaan.



Invention ohjelmointiesimerkki; tahti 25

Esimerkeissä tahti jaettu 1/64-osiin, jotta 1/8-osa nuoteille voidaan ohjelmoida nuotin sisäisiä dynamiikan vaihteluita.

1/64-osat luettavissa sonogrammikuvaajasta selkeästi.

Kuvat: Mikko Ojanen

ANALYYSIESIMERKKI: J. S. BACH INVENTIO A-MOLLIN SOVITUS DIMI-A:LLE (1970)

Masterkelanauhan ja vinyylituloksisun
(1970; julkaistu CD:llä 2002) vertailu

Kuvat:

<http://arpjournal.com/mastering-kurenniemis-rules-2012-the-role-of-the-audio-engineer-in-the-mastering-process/>

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

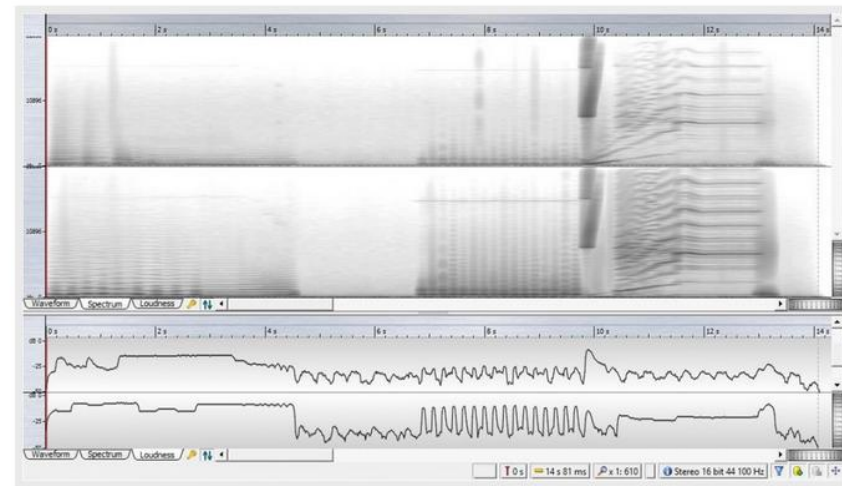


Figure 10a. The 2002 released CD version of *Inventio-Outventio* is copied from the original vinyl single – including audible snaps and leakage between the channels (14 sec. excerpt from the middle section of the work). The channel orientation also differs from the original master tape (see Figure 10b).

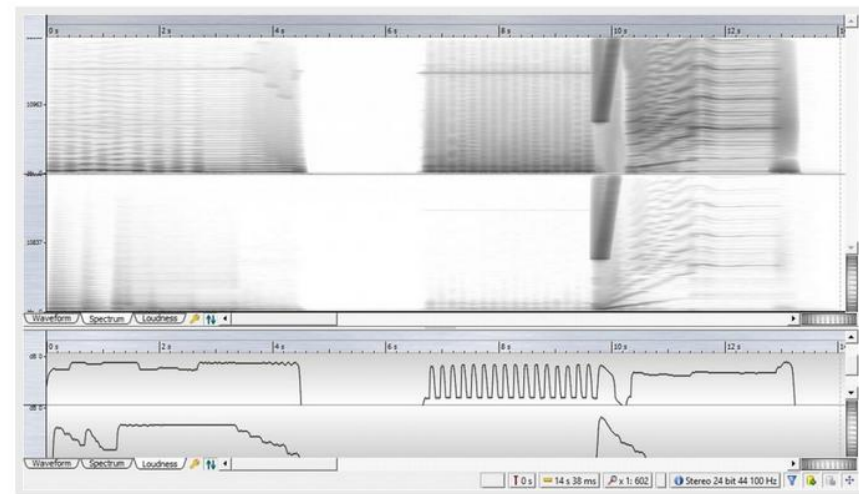


Figure 10b. The 2012 digitization of the *Inventio-Outventio* master tape (14 sec. excerpt from the middle section of the work).

Kuvat: Mikko Ojanen



KIITOS!

TUTKIMUSDATAN MÄÄRITTELYN JA AINEISTONHALLINNAN ONGELMAT IHMISTIEETEISSÄ



Mikko Ojanen

*kirjastonhoitaja, Helsingin yliopiston kirjasto
tohtorikoulutettava, Musiikkitiede, Helsingin yliopisto
orcid.org/0000-0002-7833-9659
mikko.ojanen@helsinki.fi*

**HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI**



Informaatiotutkimuksen päivät 3.–4.11.2016
Digitoidut kokoelmat digitaalisten ihmistieteiden tutkimusaineistona