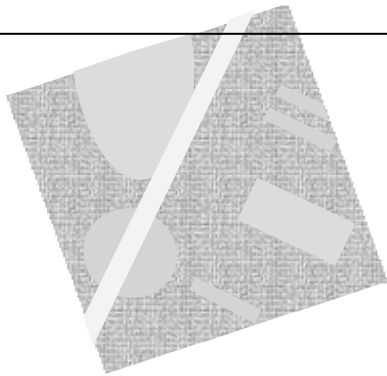


Luento 6

Tiedon esitysmuodot



- Lukujärjestelmät
- Luvut, merkit,
merkkijonot,
totuusarvot, oliot
- Kuvat, äänet, hajut(?)
- Ohjelmat

14.8.2000 Teemu Kerola, K2000 1

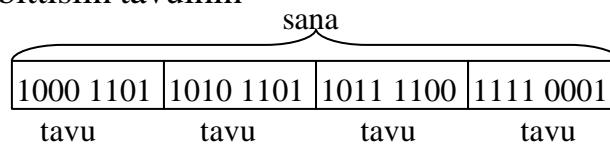
Tiedon tyypit

- **Kommunikointi ihmisen kanssa**
 - kuva, ääni, merkit, ...
- **Laitteiston sisäinen talletus**
 - kuvaformatit, ääniformatit, pakkausstandardit, ...
 - kokonaisluvut, liukuluvut
 - merkit, merkistöt
- **Suorittimen omana lajinaan ymmärtämät tyypit**
 - on olemassa konekäskyjä tälle tietotyypille
 - kokonaisluvut
 - liukuluvut (useimmat suorittimet nykyään)
 - totuusarvot (jotkut suorittimet)
 - merkit (jotkut suorittimet)

14.8.2000 Teemu Kerola, K2000 2

Tiedon esitys laitteistossa ⁽⁴⁾

- Kaikki tieto koneessa on binääribitteinä (0 tai 1)
 - binäärijärjestelmän numerot: 0, 1
 - helppo toteuttaa piireillä
 - helppo suunnitella logiikkaa Boolean algebran avulla
- Muisti jaettu tasapituisiin sanoihin (word)
 - sana = word = 32 bittiä (16 bittiä, 64 bittiä, ...)
- Usein sana on jaettu tasapituisiin (byte) 8-bittisiin tavuihin



14.8.2000

Teemu Kerola, K2000

3

Tiedon esitys laitteistossa ⁽⁴⁾

- Tietoa siirretään muistiväylää pitkin sanoina
 - joskus useampi kuin yksi sana kerrallaan (lohko)
- Prosessorin rekisterit ovat yleensä yhden tai kahden sanan mittaisia
 - 1 sana: kokonaisluku, pieni liukuluku
 - 1 sana: 1 merkki tai 4 merkkiä
 - 2 sanaa: pitkä kokonaisluku, iso liukuluku

14.8.2000

Teemu Kerola, K2000

4

Tiedon esitys ⁽⁷⁾

- Kysymys: miten esittää eri tyyppisiä tietoja?
- Vastaus: koodataan ne biteiksi
 - kaikki tieto on koneessa bitteinä
- Kaikelle käsitellylle tiedolle on omat koodausmenetelmänsä
 - kaikkia koodausmenetelmiä ei ole standardoitu
 - samalla tietotyypille voi olla useita koodausmenetelmiä
 - kokonaisluvut, liukuluvut, merkit, merkkijonot, kuvat, ...
 - ongelma: ymmärtävätkö koneet toisiaan?
 - tiedon esitysmuotoa voidaan joutua muuttamaan kun tietoa siirretään koneelta toiselle

14.8.2000

Teemu Kerola, K2000

5

Suorittimen ymmärtämä tieto ⁽⁹⁾

- Kaikki tieto koneessa on koodattuna biteiksi
- Muistissa voidaan esittää kaikki tieto sovitulla esitystavalla (koodauksella)
- Suoritin osaa tehdä operaatioita joillakin esitystavoilla koodatuille tiedoille

TTK-91: kokonaisluvut

 - kokonaisluvut ja liukuluvut (aina)
 - totuusarvot, merkit ja merkkijonot (joskus)
 - kuvat, äänet ja hajut (ei yleensä)
- Muiden tietojen käsittely tapahtuu ohjelmallisesti
 - esim. merkkejä voidaan käsitellä kokonaislukuoperaatioilla ja aliohjelmilla

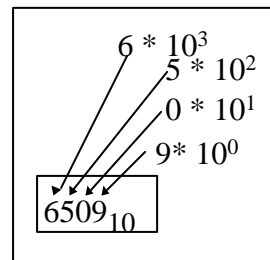
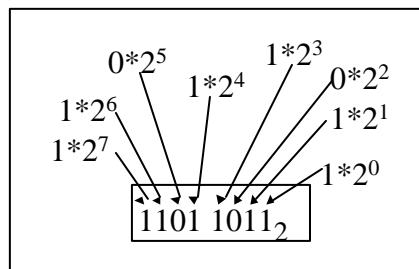
14.8.2000

Teemu Kerola, K2000

6

Binäärijärjestelmä (2)

- Kantaluku 2, numerot 0 ja 1
 - numeroiden painoarvot oikealta vasemmalle:
 $1=2^0, 2=2^1, 4=2^2, 8=2^3, 16=2^4, 32=2^5, \dots$
 - kymmenjärjestelmässä painoarvot ovat
 $1=10^0, 10=10^1, 100=10^2, 1000=10^3, \dots$

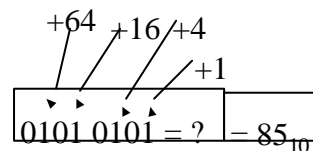
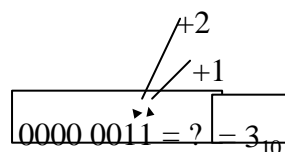
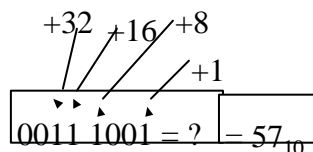


14.8.2000

Teemu Kerola, K2000

7

Binäärilukuesimerkkejä



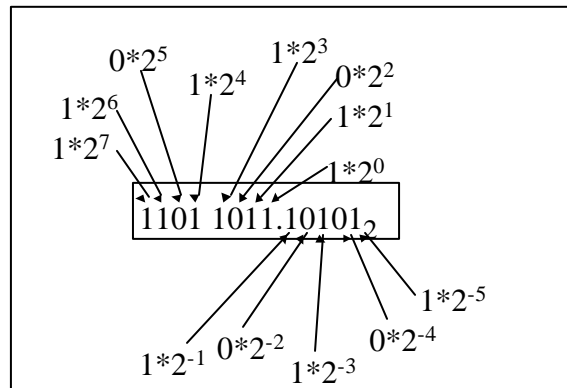
14.8.2000

Teemu Kerola, K2000

8

Binääripiste (2)

- Binääriluvuilla voi olla myös binääriosa (vrt. desimaaliosa)

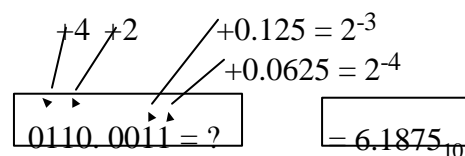
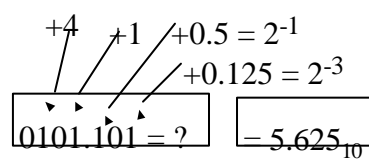


14.8.2000

Teemu Kerola, K2000

9

Binääripiste-esimerkkejä



14.8.2000

Teemu Kerola, K2000

10

Muunnokset lukujärjestelmien välillä ⁽⁵⁾

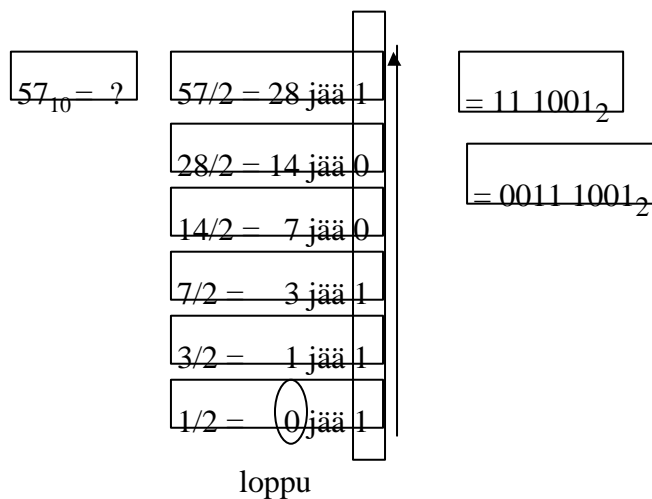
- 2-järjestelmä \Rightarrow 10-järjestelmä
 - esitettiin jo edellä
- 10-järjestelmä \Rightarrow 2-järjestelmä
 - kokonaisosa ja desimaaliosa erikseen
 - kokonaisosa:
 - jaa toistuvasti 2:lla, kunnes 0 jäljellä
 - ota jakojäännökset käännetyssä järjestyksessä

14.8.2000

Teemu Kerola, K2000

11

10-järj \Rightarrow 2-järj kokonaislukuesimerkki ⁽¹¹⁾



14.8.2000

Teemu Kerola, K2000

12

10-järj \Rightarrow 2-järj binääriosalle

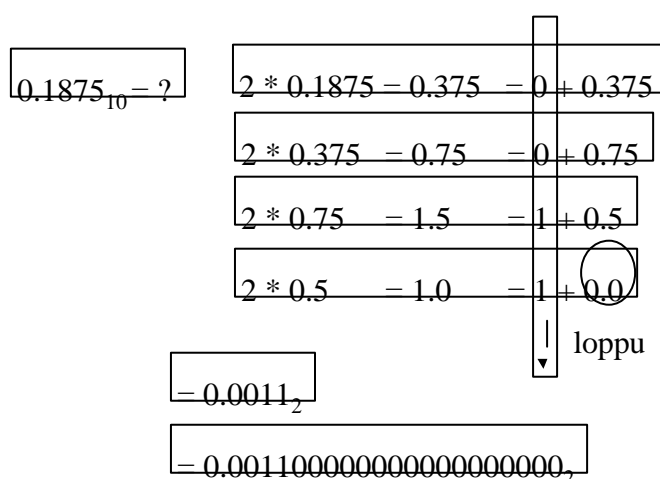
- Kerrotaan toistuvasti desimaaliluvun desimaaliosa 2:lla, kunnes
 - desimaaliosa = 0 (tarkka binääriesitys)
 - tarpeeksi numeroita haluttuun tarkkuuteen
- Tulos saadaan ottamalla saatujen desimaalilukujen kokonaisosat (0 tai 1) lasketussa järjestyksessä

14.8.2000

Teemu Kerola, K2000

13

10-järj \Rightarrow 2-järj binääriosaesimerkki



14.8.2000

Teemu Kerola, K2000

14

Heksadesimaaliesitys ⁽⁶⁾

- Binäärilukuja käyttö on tarpeellista mutta niitä on ikävä kirjoittaa
 - liikaa numeroita
- Kirjoitetaan ne 16-järjestelmässä eli heksadesimaalijärjestelmässä
- 4 bittiä vastaa aina yhtä 16-järjestelmän numeroa
- Yksi 16-järjestelmän numero vastaa aina 4 bittiä
- 16-järjestelmän numerot ovat:
0,1,2,3,4,5,6,7,8,9, A, B, C, D, E ja F

10	11	12	13	14	15
----	----	----	----	----	----

14.8.2000

Teemu Kerola, K2000

15

Heksadesimaaliesimerkkejä ⁽¹¹⁾

binääri:

0100	0111	1001	1010	1111
------	------	------	------	------

16-järj: 4 7 9 A F

= 479AF ₁₆

= 0004 79AF ₁₆ = 0x 479AF

16-järj:

120ADF ₁₆

1	2	0	A	D	F
---	---	---	---	---	---

binääri: 0001 0010 0000 1010 1110 1111

14.8.2000

Teemu Kerola, K2000

16

Oktaaliesimerkkejä ⁽¹³⁾

Numerot: 0, 1, 2, 3, 4, 5, 6, 7

binääri: 01 000 111 100 110 101 111

8-järj: 1 0 7 4 6 5 7 = 1074657₈

= 0001074657₈ = 01074657

8-järj: 120371₈ 1 2 0 3 7 1

binääri: 001 010 000 011 111 001

14.8.2000
Teemu Kerola, K2000
17

Big vs. Little Endian ⁽³⁾

- Miten monitavuiset arvot talletetaan?

0x1200:

Sanan osoite

0x1200 0x1201 0x1202 0x1203

talleta 0x11223344 ?? tavuosoitteet

Big-Endian: eniten merkitsevä tavu pienimpään osoitteeseen

→ 0x11 0x22 0x33 0x44

Little-Endian: vähiten merkitsevä tavu pienimpään osoitteeseen

→ 0x44 0x33 0x22 0x11

14.8.2000
Teemu Kerola, K2000
18

Big vs. Little Endian ⁽⁵⁾

- Monitavuisen tiedon (sana-) osoite on sama molemmissa tapauksissa
- Tavujen sisäinen järjestys vaihtelee
- Suorittimen suunnittelija päättää
 - Matematiikkapiirien tulee tietää miten luvut esitetty
 - Täytyy ottaa huomioon siirrettäessä tietoa verkon yli
- Power-PC: bi-endian - molemmat moodit käytössä
 - voidaan valita ohjelmakohtaisesti
 - etuoikeutetussa tilassa voidaan vielä valita erikseen
 - suoritin osaa laskea kummallakin tavalla talletetuilla luvuilla

TTK-91: big-endian

14.8.2000
Teemu Kerola, K2000
19

Kokonaislukujen esitysmuoto ⁽⁸⁾

Kaikki esitetty biteillä 0 ja 1
ei etumerkkejä
ei desimaalipistettä

- Etumerkittömät kokonaisluvut helppoja
- Positiiviset luvut helppoja
 - normaali binäärilukuesitys
- Negatiiviset luvut
 - etumerkkibitti erikseen
 - kahden komplementtiesitys

hexadesimaali esitys: $3 \cdot 16 + 9 \cdot 1 = 0x39$

binääri: $57 = 0011\ 1001$

negatiivinen: $-57 = 1100\ 0111$

14.8.2000
Teemu Kerola, K2000
20

Kahden ja yhden komplementti esitykset

- Kahden komplementti käytössä useimmiten

– Tilaa 8 bittiä?

- 7 bittiä datalle
- 1 bitti "etumerkille"

+2 = 0000 0010
+1 = 0000 0001
0 = 0000 0000
-1 = 1111 1111
-2 = 1111 1110

- Yhden komplementti:

ones complement: -0 = 1111 1111
-1 = 1111 1110

14.8.2000

Teemu Kerola, K2000

21

Liukuluvut

- Tietokoneessa ei ole realilukuja tai rationaalilukuja (matemaattiset käsitteet)
- Aina on rajallinen esityksen tarkkuus
 - lukuja π , $\sqrt{2}$, tai $1/3$ ei voi esittää tarkasti
 - luvut 1.000000000 ja luvut 1.000000001 ovat yhtäsuuria (joissakin esityksissä)
- Yleinen realilukuja vastaava esitysmuoto on liukukuesitysmuoto float, double, real
 - 32 bittiä, noin 7-8 desimaalinumeron tarkkuus
 - 64 bittiä, noin 16-17 desimaalinumeron tarkkuus

14.8.2000

Teemu Kerola, K2000

22

Liukulukujen esitys ⁽⁴⁾

$-0.000\ 000\ 000\ 123 = -1.23 * 10^{-10}$
--

$+0.123 = +1.23 * 10^{-1}$

$+123.0 = +1.23 * 10^2$

$+123\ 000\ 000\ 000\ 000 = +1.23 * 10^{14}$
--

“+”	“14”	“1.23”
-----	------	--------

sign	exponent	mantissa or significand
	(exponentti)	(mantissa)

14.8.2000
Teemu Kerola, K2000
23

IEEE 32-bit Floating Point Standard ⁽³⁾

IEEE Standard 754

“+”	“14”	“1.1875” = “1.0011”
-----	------	---------------------

sign	exponent	mantissa or significand
	(2:n potenssi)	

- Etumerkki
 - 1 bitti etumerkille, 1 ⇒ “-”, 0 ⇒ “+”
 - etumerkki bitti $S \Rightarrow$ etumerkin arvo = $(-1)^S$

14.8.2000
Teemu Kerola, K2000
24

IEEE 32-bit FP Standard

“+”	“15”	“1.1875” = “1.0011”
sign	exponent	mantissa or significand

- 8 bittiä eksponentille, lisättynä 127:llä (biased form)

exponent = 5	store	5+127 = 132 = 1000 0100
exponent = -1	store	-1+127 = 126 = 0111 1110
exponent = 0	store	0+127 = 127 = 0111 1111

– eksponentit 0 and 255 erikoistapauksia

- talletettu arvoalue: 1 - 254 ⇒ todellinen arvoalue: -126 - 127

14.8.2000
Teemu Kerola, K2000
25

IEEE 32-bit FP Standard (7)

“+”	“15”	“0.1875” = “0.0011”
sign	exponent	mantissa or significand

1/8 = 0.1250
1/16 = 0.0625
0.1875

- 23 bittiä mantissalle, siten että ...

0.0011	“15”
1.1000	“12”
1000	“12”

- 1) Binääripiste (.) on heti ensimmäisen bitin jälkeen
- 2) Mantissa on normalisoitu: vasemmanpuolimmainen bitti on 1
- 3) Vasemmanpuolimmaista (eniten merkitsevä) bittiä (1) ei talleteta (implied bit)

24 bitin mantissa!

14.8.2000
Teemu Kerola, K2000
26

IEEE 32-bit FP Values ⁽⁸⁾

$$23.0 = +10111.0 * 2^0 = +1.0111 * 2^4 = ?$$

$$4+127=131$$

0	1000 0011	011 1000 0000 0000 0000 0000
---	-----------	------------------------------

sign exponent mantissa or significand
 1 bit 8 bits 23 bits

$$1.0 = +1.0000 * 2^0 = ?$$

$$0+127 = 127$$

0	0111 1111	000 0000 0000 0000 0000 0000
---	-----------	------------------------------

sign exponent mantissa or significand
 1 bit 8 bits 23 bits

14.8.2000
Teemu Kerola, K2000
27

IEEE 32-bit FP Values ⁽⁶⁾

0	1000 0000	111 1000 0000 0000 0000 0000
---	-----------	------------------------------

sign exponent mantissa or significand
 1 bit 8 bits 23 bits

$X = ?$	$X = (-1)^0 * 1.1111 * 2^{(128-127)}$
	$= -1.1111_2 * 2$
	$= -(1 + 1/2 + 1/4 + 1/8 + 1/16) * 2$
	$= -(1 + 0.5 + 0.25 + 0.125 + 0.0625) * 2$
$= -1.9375 * 2$	$= -3.875$

14.8.2000
Teemu Kerola, K2000
28

Konekäskyjen esitysmuoto muistissa ⁽⁴⁾

- Konekohtainen, jokaisella omansa
- Käskyt ovat 1 tai useamman tavun mittaisia
 - SPARC, kaikki käskyt: 1 sana eli 4 tavua
 - PowerPC, kaikki käskyt: 1 sana eli 4 tavua
 - Pentium II: 1-16 tavua, paljon variaatioita
- Käskyillä on yksi tai useampi muoto, kussakin tietty määrä erilaisia kenttiä
 - opcode, Ri, Rj, Rk, osoitusmoodi
 - pitkä tai lyhyt vakio

TTK-91, kaikki käskyt: 1 sana, 1 muoto

14.8.2000

Teemu Kerola, K2000

29

TTK-91 konekäskyn rakenne

- Käskyn esitys bittitasolla on aina:

OPER käskykoodi 8 bit field	Rj 3 bit	M 2 bit	Ri 3 bit	ADDR osoiteosa 16 bit field
--------------------------------	-------------	------------	-------------	--------------------------------

31 24 21 19 16 15 0

Rj = käskyn ensimmäinen operandi

Ri = indeksirekisteri

M = muistinoutojen määrä toiseen operandiin
(ennen mahdollista muistiin talletusta)

00 eli 0 kpl, rekisteri tai välitön osoitus

01 eli 1 kpl, suora osoitus

10 eli 2 kpl, epäsuora osoitus

(11 eli 3 kpl, epäkelpo arvo → poikkeustilanne)

muistiosoite tai
(pienehkö) vakio

(addressing
mode)

14.8.2000

Teemu Kerola, K2000

30

Konekäslyn operandit ja tulos

- Tulos: rekisteri Rj
 - paitsi WRITE- tai PUSH-käskeyissä muistipaikan sisältö
- Ensimmäinen operandi: rekisteri Rj
- Toinen operandi
 - laske ensin arvo $R_i + ADDR$ ja käytä sitä sellaisenaan tai käytä sitä muistisoitteena

Kone-
kielen
tiedon
osoitus-
moodit

- arvo: $R_i + ADDR$
- muistipaikan $M[R_i + ADDR]$ sisältö
- muistipaikan $M[M[R_i + ADDR]]$ sisältö

jos $R_i = R_0$,
niin pelkkä ADDR

14.8.2000

Teemu Kerola, K2000

31

Merkit ⁽⁴⁾

- Yleensä 1 tavu per merkki
- ASCII, 7 bittiä/merkki (+ tark. bitti?)
- EBCDIC, 8 bittiä/merkki
- ISO/IEC 8859-15 ('Latin-9'),
 - 8-bittiä/merkki, 256 eri merkkiä käytössä
 - mukana myös ä, ö, š, €

'A' - 0x41, 'a' - 0x61, LF - 0x0A

Lisää tietoa: ks.

<http://www.tieke.fi/edisty/edis699/stand699.htm>

14.8.2000

Teemu Kerola, K2000

32

UCS ja Unicode ⁽⁵⁾

- UCS - Universal Character Set
- Samat merkistöt, eri standardit
- 2 tavua eli 16 bittiä per merkki
 - 65536 merkkiä käytössä oleville n. 200000 symbolille
- Kontrollimerkit
 - 0x0000-001F and 0x0080-009F
 - 0x007F = DELETE, 0x0020 = SPACE
- UCS:ssä myös 8-bittiset koodi ”rivit”
 - eri alueille tai tarkoituksiin (zone) omat 8-bittiset koodinsa?

14.8.2000

Teemu Kerola, K2000

33

UCS ja Unicode

- Merkit välillä 0x0000-00FF samassa järjestyksessä kuin Latin-9 merkistössä
 - 16-bittisen UCS:n ”rivi 00” = 8-bittinen Latin-9
- Myös muut aakkoset:
 - I-zone = Kanji (0x4E00-9FFF, 20992 merkkiä)
- Ei omia konekäskyjä, manipulointi aliohjelmilla

14.8.2000

Teemu Kerola, K2000

34

Merkkijonot

- Yleensä peräkkäin talletettu joukko tavuja
- Lisäksi tarvitaan jollain tavalla koodata merkkijonon pituus
 - laitetaan loppuun erikoismerkki
 - C-kieli: `'\0'` = 0x00
 - toteutetaan tietueena

20	"Ei yleensä nyt enää!"
----	------------------------

pituus merkkijono
 - ei omia konekäskyjä, manipulointi aliohjelmissa
 - kokonaisluku- ja bittimanipulointikäskyt

14.8.2000

Teemu Kerola, K2000

35

Totuusarvot ⁽⁴⁾

- Boolean TRUE ja FALSE
- Yleensä koodattu TRUE=1, FALSE=0
 - muttei aina!
- Usein Boolean arvo per sana
 - loput 31 bittiä nolliä
 - ohjelmointikielten Boolean muuttujat
- Joskus pakatussa muodossa 32/arvoa per sana
- Ei omia konekäskyjä, manipulointi aliohjelmissa
 - kokonaisluku ja bittimanipulointikäskyt

14.8.2000

Teemu Kerola, K2000

36

Taulukot ja tietueet

- Taulukot peräkkäisrakenteena
 - kuten esimerkit aikaisemmin
 - riveittäin tai sarakkeittain
 - ei omia konekäskyjä, manipulointi aliohjelmilla tai loopeilla (paitsi ns. vektorikoneet, joilla on omia konekäskyjä vektorioperaatioita varten)
- Tietueet peräkkäisrakenteena
 - osoite on jonkin osoitemuuttujan arvo
 - ei omia konekäskyjä, manipulointi aliohjelmilla tai kääntäjän generoimien vakiolisäysten avulla

14.8.2000

Teemu Kerola, K2000

37

Oliot

- Oliot
 - kuten tietueet, yleensä varattu kasasta (heap)
 - useat olion kentistä sisältävät vuorostaan osoitteen kasasta suoritusaikana varattuun toiseen olioon
 - metodit ovat aliohjelmien osoitteita
 - ei omia konekäskyjä, manipulointi aliohjelmilla

14.8.2000

Teemu Kerola, K2000

38

Kuvat

- **Monta kuvastandardia**
 - yleisyys, siirrettävyys, pakkaustiheys
 - näyttöä varten tarvittavan laskennan määrä
- **Kuvatiedoston alussa otsake kertoo talletusformaatin**
- **Viiva- ja vektorikuvat**
 - kuva koodattuna objekteina
 - ympyrä, monikulmio, käyrä, alueen väri
- **Rasterikuvat**
 - kuva koodattuna pisteinä
 - kunkin pisteen väri koodattu esim. 24 bitillä

14.8.2000

Teemu Kerola, K2000

39

Kuvat

- **Kuvat ovat yleensä pakattu mahdollisimman vähän tilaa vievää muotoon**
 - purkaminen vaatii laskentaa
- **GIF, JPEG, TIFF, BMP,**
- **Ei omia konekäskyjä, manipulointi aliohjelmilla**

14.8.2000

Teemu Kerola, K2000

40

Video kuva

- Vie hyvin paljon muistitilaa
- Talletus kuva kerrallaan, esim. 25 kuvaa/sek
 - 1 sekunti hyvälaatuista videokuvaa pakkaamattomassa muodossa 20 MB
- Talletus ”incrementaalisesti”
 - kun seuraava kuva poikkeaa edellisestä vain vähän
 - talleta vain muutokset edelliseen

14.8.2000

Teemu Kerola, K2000

41

Videostandardit

- MPEG (Moving Pictures Expert Group)
- AVI (Audio Visual Interleave)
- MOV, INDEO, FLI, GL, ...
- ei omia konekäskyjä, manipulointi aliohjelmilla (tai erikoisprosessoreilla, joissa erikoiskäskyjä)
 - grafiikkakortit

14.8.2000

Teemu Kerola, K2000

42

Äänet

- Syntetisoitu ääni tai täydellinen äänidata
- Syntetisoitu ääni
 - MIDI-käskyjä
 - Music Instrument Digital Interface
 - ”Soita nuotti N voimakkuudella V”
- ei omia konekäskyjä, manipulointi aliohjelmilla (tai erikoisprosessoreilla, joissa erikoiskäskyjä)
 - äänikortit

14.8.2000

Teemu Kerola, K2000

43

Ääni

- Täydellinen äänidata
 - Perustuu ääninäytteeseen 44.1 KHz taajuudella
 - 44100 näytettä / sek
 - koodataan esim. 16 bitillä (2 tavua): 88KB /sek
- Äänistandardit
 - MIDI, WAV, AU

14.8.2000

Teemu Kerola, K2000

44

Maku, haju, tunto ja muu data ⁽³⁾

- Tähtien kirkkaus, hajut, ks. HS artikkeli 5.5.2000
veneen tyyppi, tunteen palo,
- Sovelluskohtaisesti, ei vielä yleisiä standardeja
 - kokonaisluvut (diskreetti data)
 - liukuluvut (jatkuva data)
- Ei omia konekäskyjä, manipulointi omilla aliohjelmilla

14.8.2000

Teemu Kerola, K2000

45

-- Jakson 6 loppu --



14.8.2000

Teemu Kerola, K2000

46