

정보

Sangwook Lee
Duil Middle School



저작자를 밝히면 이용이 가능하지만, 영리 목적으로 이용할 수 없으며, 내용을 수정해서도 안 됩니다

II 데이터

1 데이터의 표현과 관리

2 데이터의 구조화와 해석

1 데이터의 표현과 관리

01 데이터의 디지털 표현

02 데이터의 수집과 관리

01 데이터의 디지털 표현 (45쪽)

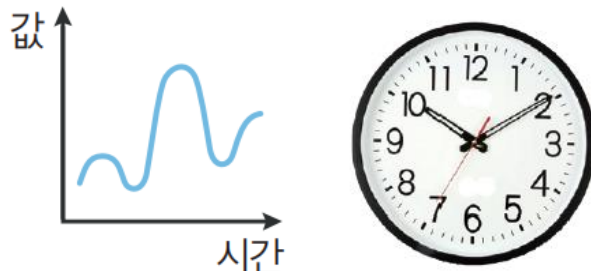
- 데이터

- 관찰, 측정, 실험, 조사 등으로 얻을 수 있는 값

- 아날로그 데이터 vs 디지털 데이터

- 아날로그 데이터: 연속적인 값

- 디지털 데이터: 불연속적인 값



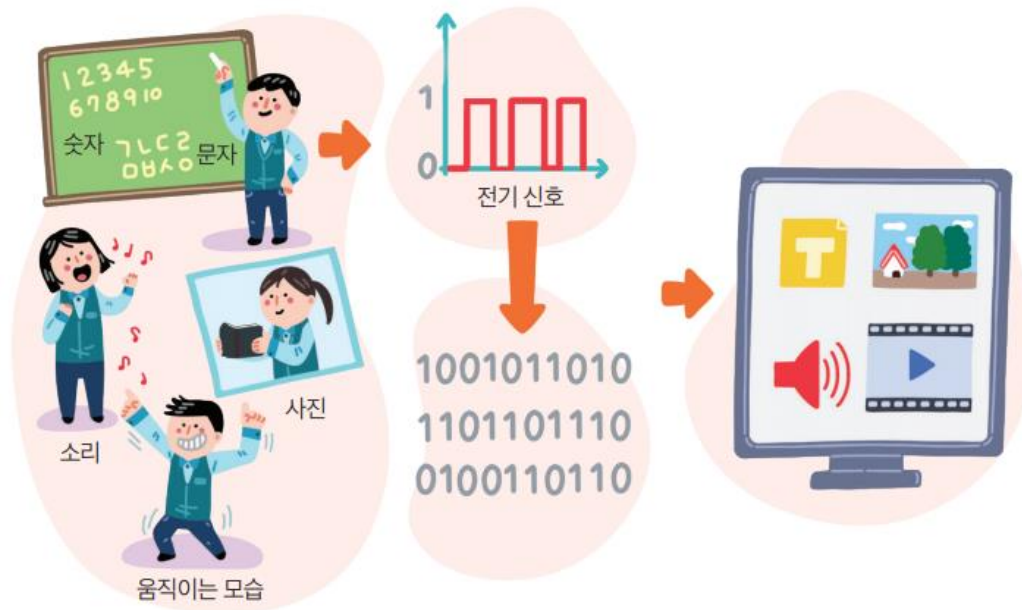
[아날로그 데이터]



[디지털 데이터]

01 데이터의 디지털 표현 (46쪽)

- 컴퓨터의 데이터 표현
 - 숫자 0과 1만을 사용
 - 불연속적으로 값을 표현



[실생활 데이터의 디지털 변환 과정]

01 데이터의 디지털 표현 (46쪽)

- 진법

- 숫자의 위치에 따라 자릿값을 다르게 부여하는 수의 표현 방법

- N진법으로 표현된 수의 자릿값

- 오른쪽 끝자리부터 **N**의 거듭제곱으로 증가

- 진수

- 특정 진법으로 표현된 수
(예: 2진법으로 표현된 수는 2진수)

01 데이터의 디지털 표현 (46쪽)

- 진법별로 사용하는 숫자(기호)의 개수와 종류
 - 2진법
 - 개수: 2개
 - 종류: 0, 1
 - 10진법
 - 개수: 10개
 - 종류: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
 - 16진법
 - 개수: 16개
 - 종류: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

N진수의 크기

- N진수의 크기를 구하는 방법
(= N진수를 10진수로 변환하는 방법)
 - ① 각 자리의 숫자와 자릿값을 곱함
(* 각 자리 숫자가 실제로 의미하는 값을 구하는 과정)
 - ② 곱한 결과를 모두 더함

(예) 16진수 2345의 각 자리 숫자가 실제로 의미하는 값 구하기

자릿값	16^3	16^2	16^1	16^0
각 자리 숫자	2	3	4	5
숫자의 실제 값	2×16^3	3×16^2	4×16	5×1

N진수의 크기

예제) 2진수 10101의 크기

$$\begin{array}{c} 10101 \\ \swarrow \quad \downarrow \quad \searrow \\ 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ = \\ 1 \times 16 + 0 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1 \\ = \\ 16 + 0 + 4 + 0 + 1 \\ = \\ 21 \end{array}$$

따라서
 $10101_{(2)} = 21_{(10)}$

N진수의 크기

예제) 16진수 AD4를 10진수로 변환

$$\begin{aligned} & \text{AD4} \\ & \quad \swarrow \quad \downarrow \quad \searrow \\ & A \times 16^2 + D \times 16^1 + 4 \times 16^0 \\ & \quad \quad \quad = \\ & 10 \times 256 + 13 \times 16 + 4 \times 1 \\ & \quad \quad \quad = \\ & 2560 + 208 + 4 \\ & \quad \quad \quad = \\ & 2772 \end{aligned}$$

〈16진수 문자 의미〉

A = 10

B = 11

C = 12

D = 13

E = 14

F = 15

따라서

$$\text{AD4}_{(16)} = 2772_{(10)}$$

디지털 데이터의 양을 나타내는 단위 (46쪽)

- 비트(bit)

- 디지털 데이터의 최소 단위
- 하나의 0 또는 1 (즉, 2진수 한 자리)
- 예: '011101'은 6비트

- 바이트(byte)

- 디지털 데이터의 기본 단위
- 여덟 개의 0 또는 1 (즉, 2진수 여덟 자리)
- 예: '10010011 11001011'은 2바이트(=16비트)

바이트는 데이터를 저장하는 기억 장치의 용량을 나타낼 때 사용하기도 함

기억 장치의 용량이 작게 표시되는 이유

제조업체와 컴퓨터가 용량을 계산하는 방식이 다르기 때문!

	제조업체	컴퓨터
1k	$10^3 = 1,000$	$2^{10} = 1024$
1M	$10^6 = 1,000,000$	$2^{20} = 1,048,576$
1G	$10^9 = 1,000,000,000$	$2^{30} = 1,073,741,824$
1T	$10^{12} = 1,000,000,000,000$	$2^{40} = 1,099,511,627,776$

예를 들면..

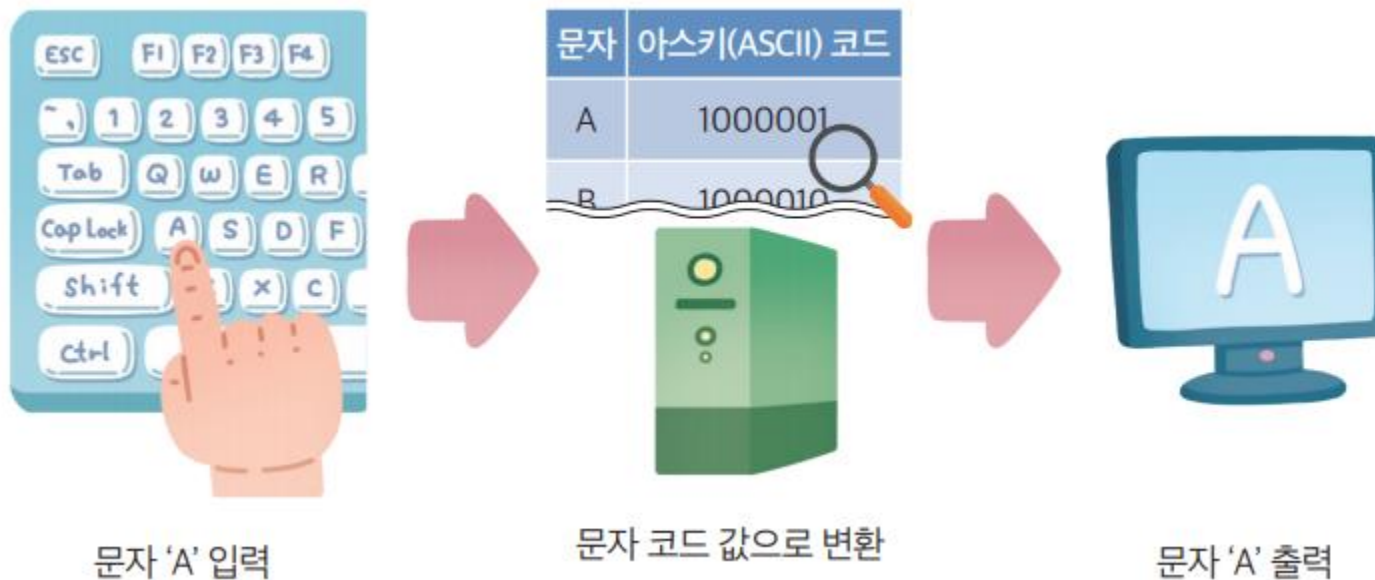
제조사는 1,000,000,000,000 바이트의 기억 장치를 만든 후
1 TB로 표시하고, 컴퓨터는 0.931 TB로 인식



01 데이터의 디지털 표현 (46쪽)

- 문자 데이터의 디지털 표현

- 문자별로 약속된 번호를 이진수의 형태로 나타낸 문자 코드를 사용하여 표현



[문자 데이터의 디지털 표현 예]

01 데이터의 디지털 표현 (46쪽)

- 문자 코드

- 문자별로 약속된 번호를 이진수의 형태로 나타낸 것

- 문자 코드 예

- 아스키(ASCII) 코드

- 미국 표준 협회에서 개발
 - 7비트를 사용하여 총128개의 문자를 표현

- 유니코드(Unicode)

- 세계의 모든 문자를 표현하기 위해 개발
 - 최대 32비트를 사용하여 문자를 표현

<아스키 코드 예>

A



65번 부여



이진수 1000001로 저장

01 데이터의 디지털 표현 (46쪽)

• 유니코드 한글 문자

가 AC00 (44032)
 각 AC01 (44033)
 각 AC02 (44034)
 갓 AC03 (44035)
 간 AC04 (44036)

11,172자

힛 D79F (55199)
 힐 D7A0 (55200)
 힐 D7A1 (55201)
 효 D7A2 (55202)
 흥 D7A3 (55203)

	AC0	AC1	AC2	AC3	AC4	AC5	AC6	AC7	AC8	AC9	ACA	ACB	ACC	ACD	ACE	ACF
0	가 AC00	감 AC10	갸 AC20	갸 AC30	갈 AC40	각 AC50	갸 AC60	거 AC70	검 AC80	겐 AC90	겉 ACA0	결 ACB0	결 ACC0	겉 ACD0	고 ACE0	곰 ACF0
1	각 AD01	갑 AC11	갸 AC21	갸 AC31	갸 AC41	갸 AC51	갸 AC61	걱 AC71	겁 AC81	갸 AC91	갸 ACA1	갸 ACB1	갸 ACC1	갸 ACD1	곡 ACE1	곱 ACF1
2	각 AD02	갸 AC12	갸 AC22	갸 AC32	갸 AC42	갸 AC52	갸 AC62	겉 AC72	겉 AC82	갸 AC92	갸 ACA2	갸 ACB2	갸 ACC2	갸 ACD2	곡 ACE2	곱 ACF2
3	갸 AD03	갸 AC13	갸 AC23	갸 AC33	갸 AC43	갸 AC53	갸 AC63	갸 AC73	갸 AC83	갸 AC93	갸 ACA3	갸 ACB3	갸 ACC3	갸 ACD3	곡 ACE3	곱 ACF3
4	간 AD04	갸 AC14	갸 AC24	갸 AC34	갸 AC44	갸 AC54	갸 AC64	갸 AC74	갸 AC84	갸 AC94	갸 ACA4	갸 ACB4	갸 ACC4	갸 ACD4	곡 ACE4	곱 ACF4

01 데이터의 디지털 표현 (47쪽)

- 이미지 데이터의 디지털 표현
 - 비트맵 방식
 - 벡터 방식

비트맵 방식

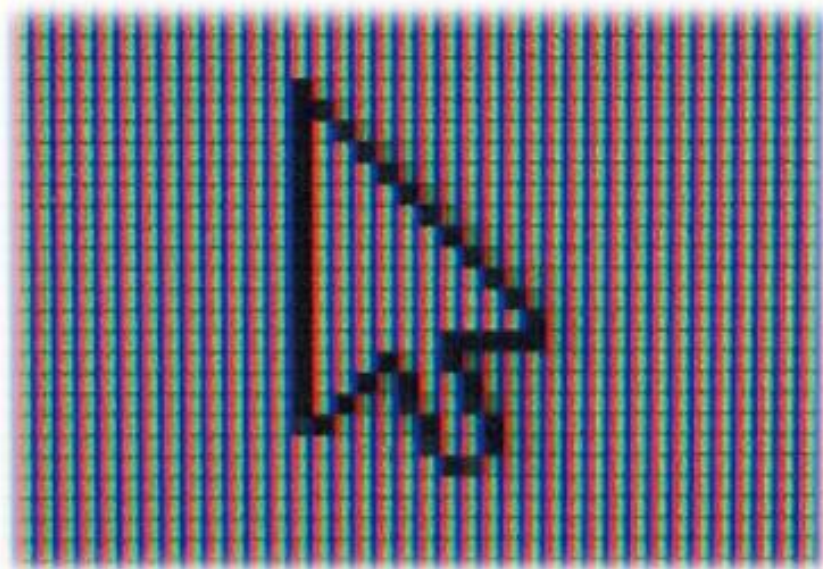


벡터 방식



01 데이터의 디지털 표현 (47쪽)

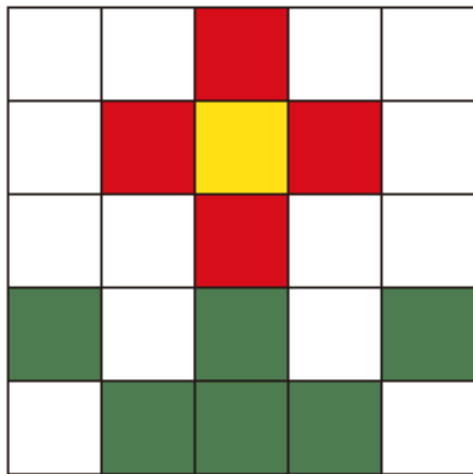
- 픽셀(pixel) = 화소(畫素)
 - 이미지나 화면을 구성하는 기본 단위
 - 한 픽셀은 여러 색상을 동시에 표현할 수 없음



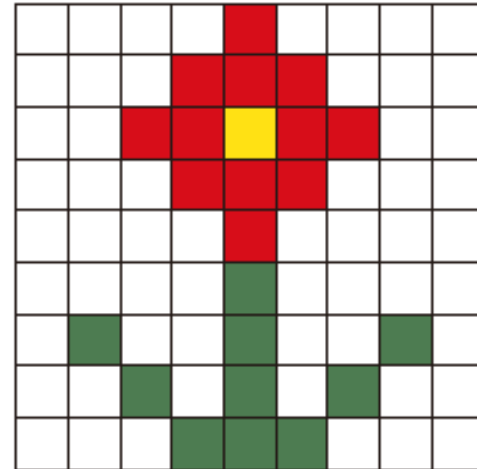
[이미지나 화면을 구성하는 픽셀]

01 데이터의 디지털 표현 (47쪽)

- 해상도(resolution)
 - 전체 픽셀 수 (또는 1인치당 픽셀 수)
 - 해상도가 높을수록 더 섬세한 표현이 가능



▲ 5픽셀 × 5픽셀 해상도

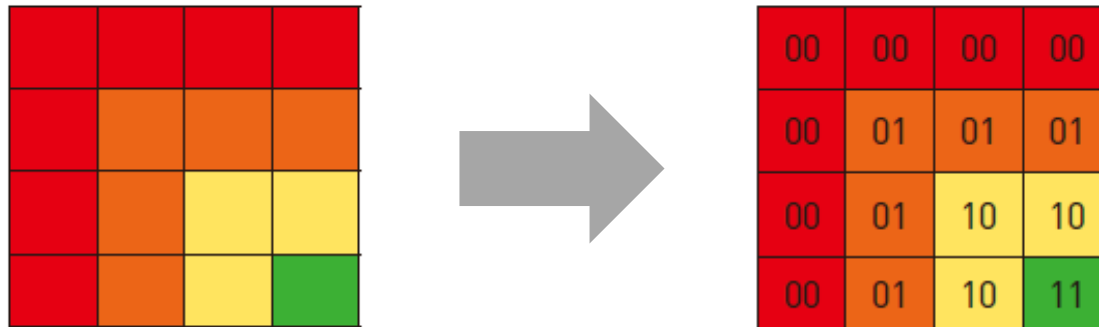


▲ 9픽셀 × 9픽셀 해상도

01 데이터의 디지털 표현 (47쪽)

- 비트맵(bitmap) 방식

- 픽셀의 색상 정보로 그림을 표현하는 방식
- 색상별로 번호를 정한 후 모든 픽셀의 색상 번호를 저장함

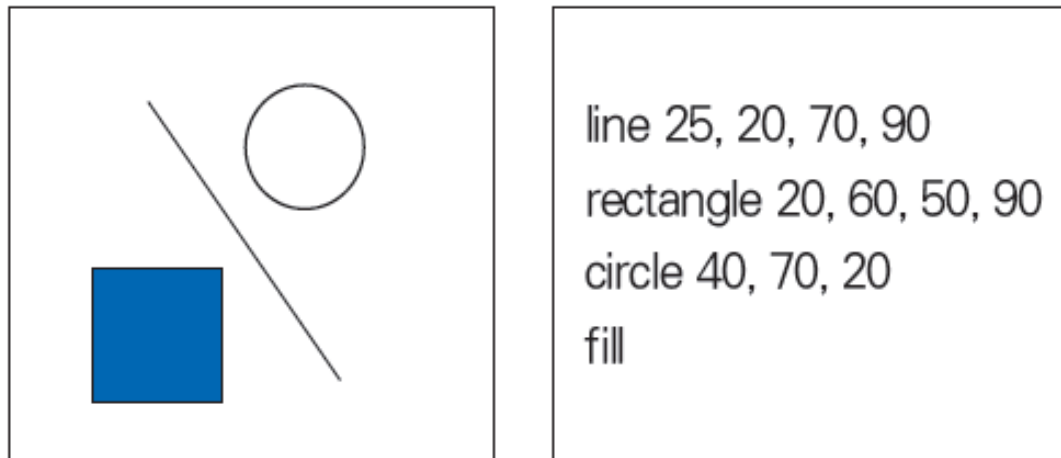


n 개의 비트로 색상을 표현하면, 최대 2^n 개의 색상을 표현할 수 있음
(픽셀당 사용하는 비트 수 많을 수록, 표현 가능한 색상 수도 많아짐)

01 데이터의 디지털 표현 (47쪽)

- 벡터(vector) 방식

- 도형 종류, 위치 등 개체의 모양 정보로 그림을 표현하는 방식
- 표현하려는 모든 개체를 그릴 수 있는 정보를 저장함



01 데이터의 디지털 표현 (47쪽)

- 비트맵 방식으로 만들어진 이미지의 특징

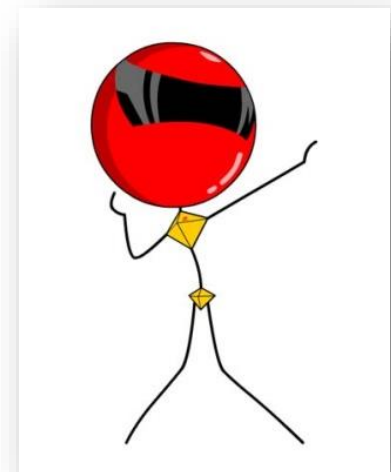
- 실세계의 복잡한 형상을 표현하기에 효율적이다
 - ☞ 해상도가 동일하면 내용이 복잡해지더라도 용량이 증가하지 않기 때문
- 이미지 속 개체별로 수정하기 어렵다
 - ☞ 개체의 픽셀 정보가 개체별로 저장되어 있지 않기 때문
- 확대 시 계단 현상이 발생한다



01 데이터의 디지털 표현 (47쪽)

- 벡터 방식으로 만들어진 이미지의 특징

- 실세계의 복잡한 형상을 표현하기에 비효율적이다
 - ☞ 개체 수가 많아져서 내용이 복잡해질수록 용량이 증가하기 때문
- 이미지 속 개체별로 수정하기 쉽다
 - ☞ 개체의 모양 정보가 개체별로 저장되어 있기 때문
- 확대 시 계단 현상이 발생하지 않는다

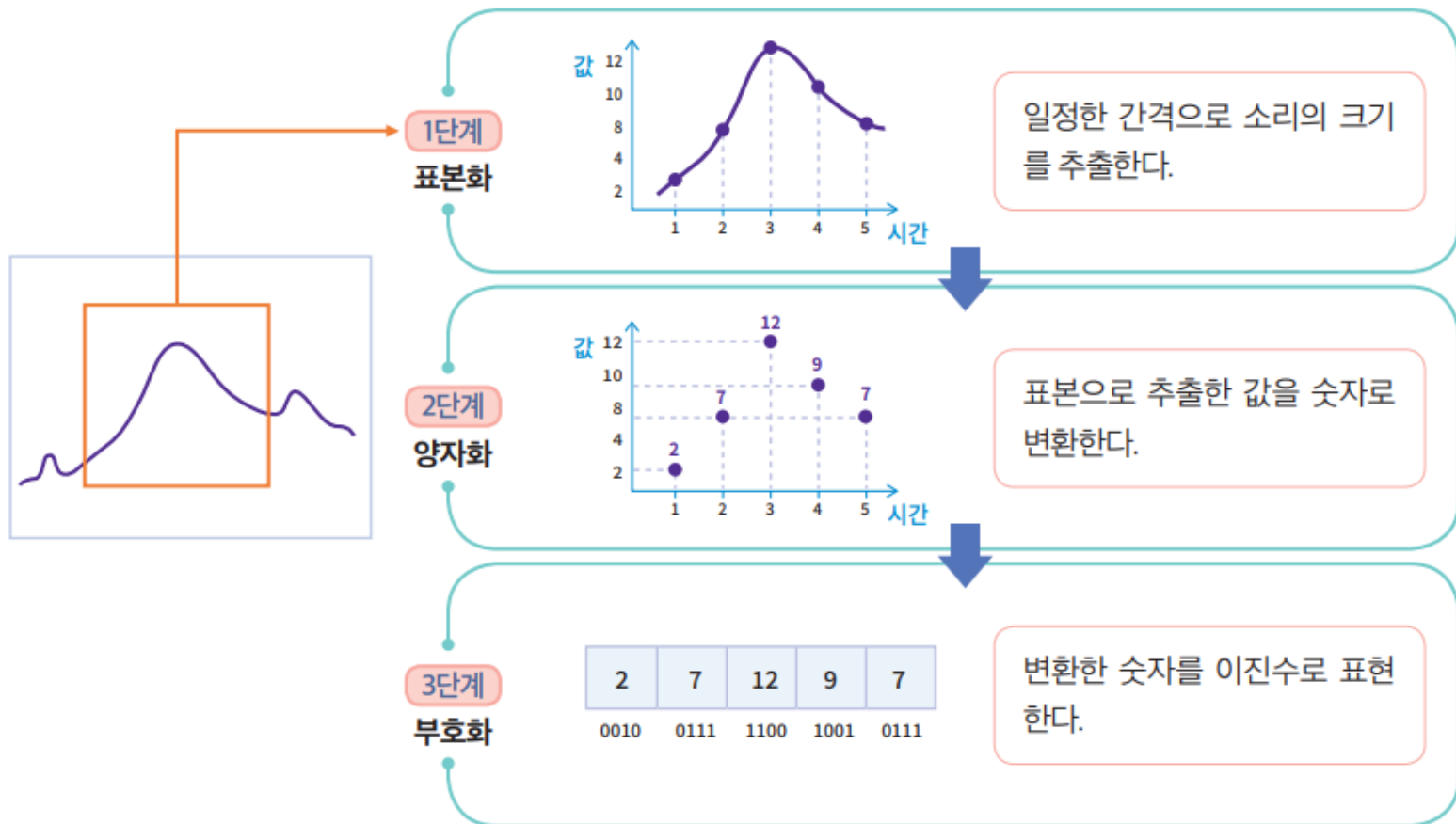


비트맵 이미지 vs 벡터 이미지

	비트맵 이미지	벡터 이미지
장점	복잡해지더라도 용량은 동일함	개체를 선택하고 수정하는 것이 쉬움 확대해도 화질이 유지됨
단점	개체를 선택하고 수정하는 것이 어려움 확대하면 화질이 저하됨	복잡해질수록 용량이 증가됨

01 데이터의 디지털 표현 (48쪽)

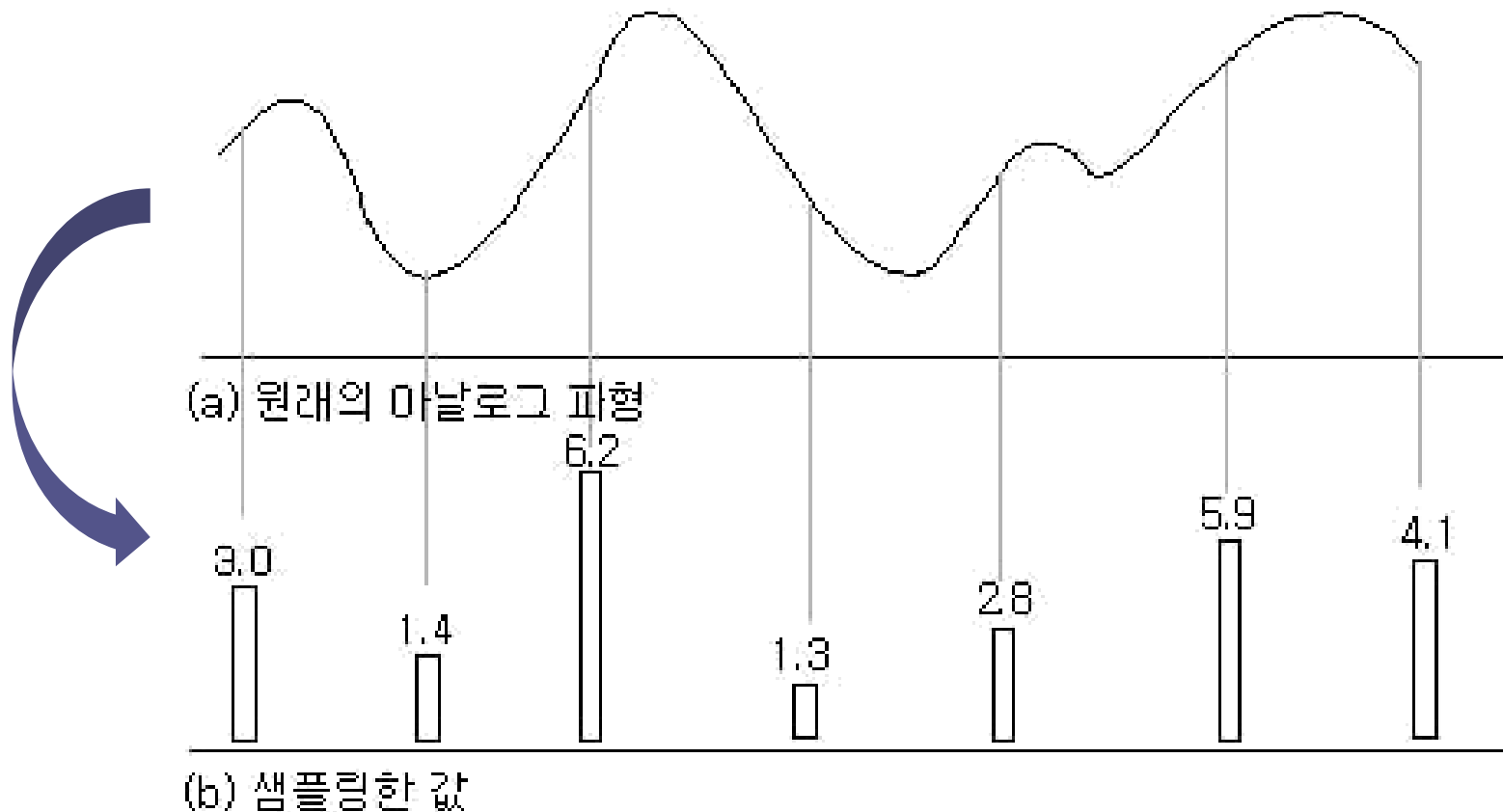
- 소리 데이터의 디지털 표현
 - 파형을 일정한 간격으로 나누고 그 값을 이진수로 바꿔서 저장



01 데이터의 디지털 표현 (48쪽)

- 표본화(sampling)

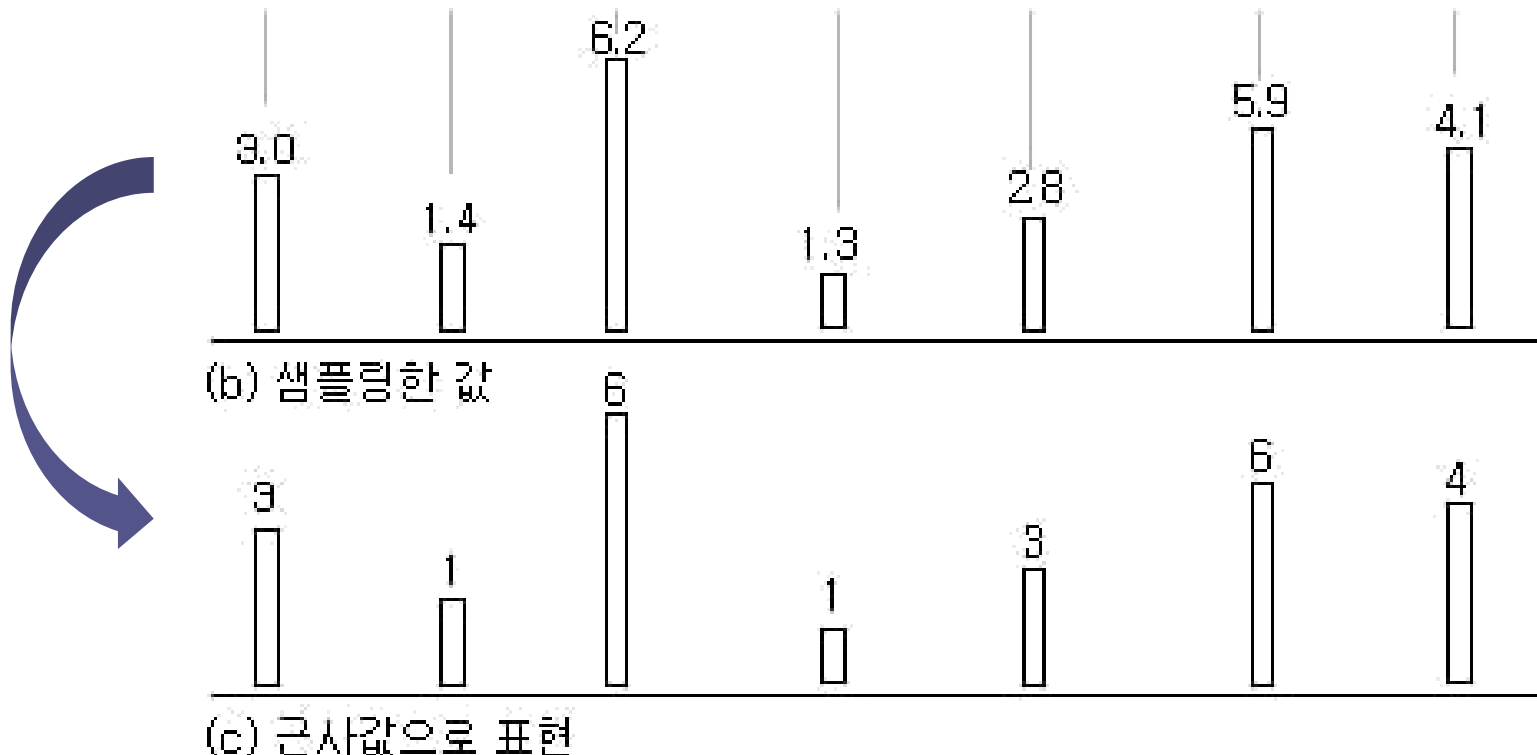
- 아날로그 신호의 진폭값을 일정한 간격으로 측정하는 것



01 데이터의 디지털 표현 (48쪽)

- 양자화(quantization)

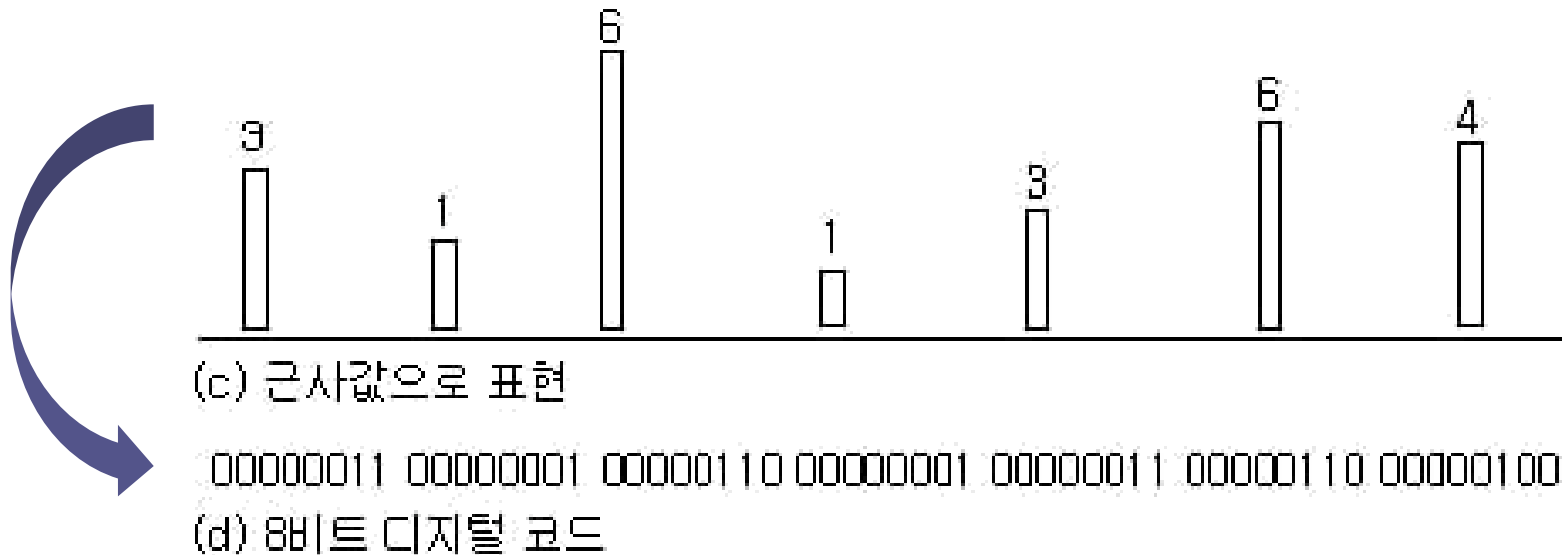
- 실숫값으로 측정된 진폭값을 가까운 정숫값으로 표현하는 것



01 데이터의 디지털 표현 (48쪽)

- 부호화(encoding)

- 정숫값으로 표현된값을 이진수로 변환하는 것



01 데이터의 디지털 표현 (48쪽)

- 초당 샘플 수

- 1초 동안 측정한 아날로그 신호의 진폭값 수
- 표본화 비율(sampling rate)이라고도 함

- 샘플당 비트 수

- 진폭값 하나를 저장할 때 사용하는 비트 수
- 양자화 비트 수(quantization bit depth)이라고도 함

소리 파일의 음질은

초당 샘플 수와 샘플당 비트 수에 의해 결정

01 데이터의 디지털 표현 (48쪽)

- 소리 파일의 크기(용량)

초당 샘플 수 × 샘플당 비트 수 × 시간(초) × 채널 수

Q. 계산 결과 값의 단위는?

A. 비트(bit)

오디오 파일의 채널 수란?

하나의 소리 파일 안에서
독립적으로 재생되는 소리 수

Q. 단위를 바이트(byte)로 바꾸려면?

A. 8로 나눔

채널 수가 많으면
보다 공간감(현장감) 있는
소리를 들을 수 있음

[문제] 1초에 44,100개의 샘플을 만드는 비율로 16비트 크기의 샘플을 4분 동안 생성하였다. 저장된 소리 파일 원본의 크기는 몇 바이트인가? (단, 소리 파일은 스테레오로 생성되었다.)

2채널

(풀이)

소리 파일 크기 =

초당 샘플 수 × 샘플당 비트 수 × 시간(초) × 채널 수

44,100 × 16 × 240 × 2

= 338,688,000 (비트)

☞ 338,688,000 ÷ 8 = 42,336,000 (바이트)

≒ 42M (바이트)



비트레이트

초당 샘플 수

* 비트레이트(bit rate) *
소리를 1초 동안 표현하는 데 사용되는 비트 수

[문제] 초당 샘플 수가 44,100, 샘플당 비트 수가 16, 채널 수가 2인 mp3 파일의 비트레이트는? (단, mp3 파일은 원본 wav 파일을 1/11로 압축하였다고 가정한다.)

(풀이)

비트레이트 =

초당 샘플 수 × 샘플당 비트 수 × 채널 수

44,100 × 16 × 2

= 1,411,200 (비트)

☞ 1,411,200 ÷ 11 = 128,290 (비트)

≒ 128k (비트)

1 데이터의 표현과 관리

01 데이터의 디지털 표현

02 데이터의 수집과 관리